III. INTERNATIONAL APPLIED STATISTICS CONFERENCE

(UYIK 2022)

In the name of Prof. Dr. Tahsin KESİCİ

PROCEEDINGS BOOK

22 JUNE - 24 JUNE 2022 SKOPJE, N. MACEDONIA

III. INTERNATIONAL
APPLIED STATISTICS CONFERENCE

Proceedings Book of the UYIK-2022

ISBN: 978-975-7328-85-8

HYBRID (ONLINE AND FACE TO FACE) - SKOPJE / N. MACEDONIA

Note: All administrative, academic and legal responsibilities of the departments belong to their authors

Proceedings Book of the III. International Applied Statistics Conference (UYIK-2022)

Publisher:
Tokat Gaziosmanpasa University
Editors:
Yalcin TAHTALI
E-Book Layout, Preparation and Composition:
Yalcin TAHTALI, Samet Hasan ABACI, Kaan KAPLAN, Furkan YILMAZ
All published articles were peer-reviewed by Scientific Committee
The organizers do not have any legal liability for to contents of the presentation texts
Organized by
Tokat Gaziosmanpasa University, Tokat, Turkey
International Balkan University, Skopje, N. Macedonia

COMMITTEES

HONORARY CHAIRMAN

Prof. Dr. Mustafa ŞENTOP, Speaker of the Grand National Assembly of Türkiye

Prof. Dr. Vahit KİRİŞÇİ, Republic of Türkiye Minister of Agriculture and Forestry

HONOURARY COMMITTEE

Prof. Dr. Bünyamin ŞAHİN, Rector of Tokat Gaziosmanpaşa University

Prof. Dr. Mehmet Dursun ERDEM, Rector of the International Balkan University

CONFERENCE CO-CHAIRMANS

Prof. Dr. İsa GÖKÇE, Tokat Gaziosmanpaşa University

Prof. Dr. Orhan KAVUNCU, Kastamonu University

Prof. Dr. Rüstem CANGİ, Tokat Gaziosmanpaşa University

Prof. Dr. Soner ÇANKAYA, Ondokuz Mayis University

Assoc. Prof. Dr. Yalçın TAHTALI, Tokat Gaziosmanpaşa University

Dr. Saniye DEMİR, Tokat Gaziosmanpaşa University

CONFERENCE SECRETARIAT

Assoc. Prof. Dr. Samet Hasan ABACI, Ondokuz Mayıs University

Dr. Nur İlkay ABACI, Ondokuz Mayıs University

Dr. Lütfi BAYYURT, Tokat Gaziosmanpaşa University

Furkan YILMAZ, Tokat Gaziosmanpaşa University

Burak FIRAT, Tokat Gaziosmanpaşa University

ORGANIZING COMMITTEE

Prof. Dr. Marti Jane ANDERSON, Massey University, New Zealand

Prof. Dr. Kaspar BIENEFELD, Humboldt University of Berlin, German

Prof. Dr. Aşir GENÇ, Necmettin Erbakan University, Türkiye

Prof. Dr. Khalid JAVED, University of Veterinary and Animal Sciences, Pakistan

Prof. Dr. Michael JORDAN, University of California, Berkeley, USA

Prof. Dr. Sıddık KESKİN, Yüzüncü Yıl University, Türkiye

Prof. Dr. Nurhan KESKİN, Yüzüncü Yıl Üniversity, Türkiye

Prof. Dr. Zahide KOCABAŞ, Ankara University, Türkiye

Prof. Dr. Masood H. SIDDIQUI, University of Lucknow, India

Prof. Dr. Mustafa ŞAHİN, K. Maraş Sutcu Imam University, Türkiye

Prof. Dr. Vjekoslav TANASKOVIK, Ss.Cyril and Methodius University, Macedonia

Prof. Dr. Arda YILDIRIM, Tokat Gaziosmanpaşa University, Türkiye

Assoc. Prof. Dr. Songül AKIN, Dicle University, Türkiye

Assoc. Prof. Dr. Francesca BERTOLINI, Technical University of Denmark, Denmark

Assoc. Prof. Dr. Şener BILALLI, International Balkan University, Macedonia

SCIENTIFIC COMMITTEE

BIOGRAPHY OF PROF. DR. TAHSİN KESİCİ

Tahsin Kesici was born on 01.03.1938 in Ermenek. He completed his primary and secondary education in this city. He graduated from Konya High School in 1956 and from Ankara University Faculty of Agriculture in 1960. With a scholarship given by Italy, he conducted research in this country for eight months between 1960-1961. He attended vocational courses after graduation at the University of Milan.

In June 1961, Kesici entered Ankara University, Faculty of Agriculture, Department of Animal Science as an assistant, and completed his doctorate in 1964. Between 1965 and 1967, he fulfilled his military service. After his military service, he started to work in the newly established Genetics and Statistics Department at the Faculty of Agriculture. Between 1969 and 1971, he did post-doctoral studies on his subject in Federal Germany for two years. He became an associate professor in 1971 and served as the Chair of Genetics and Statistics between 1972-1982. Later, he was the head of this unit, which was restructured as the Department of Biometrics-Genetics, for seventeen years.

Kesici, who became a professor in 1979, graduated from A.U. between 1980-1984. He served as the Deputy Dean of the Faculty of Agriculture, the Dean of the Faculty of Agriculture in 1984-1987, and the Vice Rector of Ankara University between 1987-1992. Between 1993 and 2002, he served as a member of the Ankara University Board of Directors for nine years. He served as the Founding President of the Informatics Department affiliated to the Rectorate and the Director of the School of Foreign Languages.

He was appointed as a member of the Council of Higher Education in July 2002. While he was still in this position, he was appointed as the Rector of TOBB University of Economics and Technology in September 2004. prof. Dr. Tahsin Kesici's foreign languages are German and Italian. He knows enough French and English to follow the literature in his field. He is married and has two children.

PREFACE

Dear Scientists and Dear Friends,

The third International Applied Statistics Conference, hosted by the International Balkan University and in cooperation with Tokat Gaziosmanpaşa University, was held in Skopje with the contribution and participation of a large academic group, both face-to-face and online.

As it is known, we held our first conference in the name of Prof. Dr. Yüksel Bek in October 2020 under pandemic conditions. The second conference, on the twenty-fifth anniversary of his death, we had the honor of holding our congress in memory of our late Prof Dr. Orhan Duzguneş who gave his life to statistics and genetics and had great efforts in the training of many teachers in our country. This year, we had our third conference was in the name of Prof. Dr. Tahsin KESİCİ.

As the congress organizing committee, we focus on two themes while organizing these congresses; investment in young people and fidelity to those who contribute to this science. On this occasion, while we organize our congresses in the name of our professors to keep their memories alive, we try to stand by the young researchers by organizing various courses with the efforts of valuable teachers within the congress.

For this purpose, 7 different course programs were organized within the scope of the congress, and approximately 175 trainees participated in these courses. In addition, 105 researchers from 7 different countries participated in the congress with approximately 133 oral presentations.

On the occasion of the conference, the honorary president of the conference and the President of the Turkish Grand National Assembly, Prof. Dr. Mustafa Şentop, for his material and moral contributions to the organization of this conference; especially, Tokat Gaziosmanpasa University rector Prof. Dr. Bünyamin Şahin and to the entire university administration, the rector of the International Balkan University, Prof. Dr. Mehmet Dursun Erdem, we want to thank.

We would also like to thank all our teachers and colleagues who contributed to the organization of the congress and courses.

The third of our congress, “3. International Applied Statistics Conference”, again with your active contribution and participation, our esteemed teacher, in the name of Prof. Dr. Tahsin KESİCİ, we are planning to hold it in Skopje in June 2022, hosted by the International Balkan University and with the cooperation of Tokat Gaziosmanpaşa University. We will be honored and happy to host you in Macedonia during one of the most beautiful times of the year.

Our next congress is “4. International Congress of Applied Statistics" We are planning to hold the in Sarajevo in April 2023, hosted by the International University of Sarajevo, in cooperation with Tokat Gaziosmanpasa University and International Balkan University, again with your active contribution and participation. We will be honored and happy to host you in Bosnia and Herzegovina in one of the most beautiful times of the year.

We hope that our congress will contribute to all applied statistics fields, be beneficial to young researchers and set an example for those who contribute to science.

Congress Organizing Committee

TABLE OF CONTENT

ORAL PRESENTATIONS

ORAL PRESENTATIONS CONTENTS

Long Non-coding RNA and Their Roles in Livestock

Gülşah Keklik[1*], **Bahri Devrim Özcan**[1], **Elif Dikkaya**[1]

[1] Çukurova Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, 01250, Adana, Turkey

*Corresponding author e-mail: gulsahkeklk@gmail.com

Abstract

The genes on DNA molecule are divisible into two groups, coding and non-coding genes. While coding genes are responsible for synthesizing a protein as the end product, in non-coding genes the end products are always RNA molecules. Transfer RNA (tRNA) and ribosomal RNA (rRNA) molecules are known as non-coding RNA (ncRNA) molecules from the 1950s, and the other ncRNA molecules such as small non-coding RNAs which less than 200 nucleotides in size and include both small interfering RNAs (siRNAs) and microRNAs (miRNAs), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), piwi-interacting RNA (piRNA), circular RNA (circRNA) are some of non-coding RNA molecules. Long non-coding RNA molecules (lncRNA) are transcripts that as longer than 200 nucleotide size. There are some databases related to lncRNAs, and information of certain species can be accessed through these databases. While sense lncRNAs are transcribed from the sense strand of DNA, and containing exons of protein-coding genes, antisense lncRNAs are transcribed from the antisense DNA strand. Long non-coding RNAs are common transcripts existing in livestock genomes, including pigs, poultry, cattle, and small ruminants. However, the roles of lncRNAs in livestock animals are insufficient because of the limited studies of lncRNAs in livestock production. In animal products, meat quality and lean percentage are vital economic traits closely related to adipose tissue deposition. In this review, we summarize the lncRNAs and their roles on livestock species.

Key words: *Long non-coding RNA, livestock, poultry, ruminants, pig, honeybee*

The Role and Importance of Microarray Applications in Farm Animals

Elif Dikkaya[1*], Gülşah Keklik[1], Bahri Devrim Özcan[1], Makbule Baylan[2]

[1] Çukurova University, Faculty of Agriculture, Animal Science, 01250, Adana, Turkey
[2] Çukurova University, Faculty of Fisheries, Basic Science, 01250, Adana, Turkey

*Corresponding author e-mail: elifdikkaya@gmail.com

Abstract

Over the past 11.000 years, humans have domesticated many animals, from reptiles to mammals. Since then, In particular, in the domestication of farm animals, it was aimed to obtain a more effective yield from animals by using not only observation but also science-based breeding methods. Developments in heredity and molecular genetics have revealed that the progress made with classical breeding methods can be made faster and more effectively by using molecular methods. One of these molecular methods is microarray studies using microchips. With microarray studies, the interactions of nucleic acids and/or proteins with some other molecules can be revealed at the genome, transcriptome, or proteome level. Thus, it is possible to rapidly identify the genes in animals that are resistant to diseases or to reveal the genetic basis of the characteristics that cause yield loss in farm animals under extreme conditions, to select the animal from the herd, and to perform breeding more effectively with animals that adapt to these conditions thanks to their genetic characteristics. In this review, the role and importance of microarray applications in farm animals were discussed.

Key words: *Microarray, Farm animals, Biotechnology, Molecular genetics*

Genetik Algoritmaların İşleyişi ve Genetik Algoritma Uygulamalarında Kullanılan Operatörler

Gülşah Keklik[1*], Elik Dikkaya[1], Bahri Devrim Özcan[1], Mervan Bayraktar[1]

[1] Çukurova University, Faculty of Agriculture, Department of Animal Science, 01250, Adana, Turkey

*Corresponding author e-mail: gulsahkeklk@gmail.com

Abstract

Associated with the genes are the constituents that determine characteristics such as appearance, personality, behaviour in living things, and units used to obtain information about these creatures, they are part of the chromosome that controls any characteristics of a living being. Firstly used by Holland, an engineer and biologist, in 1975, the Genetic Algorithm (GA) aims to intuitively achieve the best solution using genetic cipher logic in living systems. With this method, the best result is obtained, since the strong individual is most likely to survive. Genetic algorithms encode every point in a solution space with a binary set of bits called chromosomes, and each point has a conformity value. Therefore, it maintains a set of points instead of a single point. The genetic algorithm creates a new population using genetic operators such as crossovers and mutations. Genetic algorithms include the coding of solutions, the calculation of conformity, the implementation of proliferation, cross-examination and mutation operators

Key words: *Genetic algorithm, Genetic operators, Crossing, Mutation*

Neo-malthusian Population Theory in Europe

Fazıl Kayıkçı[1*], Arda Altay[2]

[1] Yıldız Teknik University, 34000, İstanbul, Turkey

*Corresponding author e-mail: fkayikci@hotmail.com

Abstract

This study aims at analyzing the determinants of the population in Europe according to air pollution levels, human development indexes and fundamental rights in EU member states. Articles related to Neo-Malthusian and other population theories were examined. In addition, the Balanced Growth Approach, Rosenstein-Roden Big Push Approach, Harrod-Domar Model, Classical and Neoclassical Growth Models, Endogenous Growth Models, Structural Analysis, New Institutionalists, Fei-Ranis Model, Demographic Transition Model and Low Level Equilibrium Trap Models were examined. Those with missing data from European countries were derived using the 'interpolate' method. Based on the largest data set obtained, the time interval of the analysis was determined as between 1997 and 2017. First, cross-section dependence and homogeneity tests were performed. According to the results obtained, it was concluded that there is a cross-section dependence and heterogeneity in the series. Secondly, first and second generation unit root tests were performed. According to the results obtained, it was seen that the series at different levels were stationary. Third, the cointegration test was performed. According to the result obtained, the existence of a cointegration relationship has been determined. Fourth, the short- and long-term coefficients were obtained with the pooled mean group estimator and the mean group estimator. Hausman test was used to test which of these estimators was significant and it was seen that the pooled mean group estimator was more significant. According to the results obtained, long-run coefficients were found to be significant. The error correction coefficient was negative and significant and converged to the equilibrium in the long run. The model used in the study is as follows: 〖lnfertil〗_(i,t)=β_0+β_1 〖lnhdin〗_(i,t)+β_2 〖lnfuri〗_(i,t)+β_3 〖lnimad〗_(i,t)+ε_(i,t) lnfertile: Births Per Woman (logarithmic) lnhdin: Human Development Index (logarithmic) Infuri: Fundamental Rights (logarithmic) lnimad: Impartial Management (logarithmic)

Key words: *Malthusian theory; Population; Panel estimation*

Ege Bölgesi ve Çevresine Ait Depremlerin Poisson Süreçleri Yardımıyla İncelenmesi

Gamze Özel[1*], Tuğba Eroğlu Azak[2], Senem Tekin[3]

[1] Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Ankara, Türkiye
[2] Milli Savunma Üniversitesi, Kara Harp Okulu, İnşaat Müh., 34334, Ankara, Türkiye
[3] Adıyaman Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Madencilik ve Maden Çıkarma Bölümü, 02040, Adıyaman, Türkiye

*Corresponding author e-mail: gamzeozl@hacettepe.edu.tr

Özet

Dünyada olduğu gibi Ülkemizde de depremler çok sayıda can ve mal kaybına neden olmaktadır. Türkiye dünyanın en aktif deprem kuşaklarından biri olan Alp-Himalaya deprem kuşağı üzerinde bulunmakta olup, yüzölçümünün %42'si birinci derece deprem kuşağı üzerinde yer almaktadır. Ege Bölgesi, aktif tektonik açısından önemli ve araştırılması gereken bir bölgedir. Batı Anadolu Açılma Rejiminin etkisi altında Gediz graben sisteminin batı ucunda bulunan bölge, geçmişten günümüze depremlerden kaynaklı çok fazla hasara maruz kalmıştır. Özellikle son yıllarda meydana gelen 20 Mart 2019 Acıpayam (Denizli) (Mw 5.5), 08 Ağustos 2019 Bozkurt (Denizli) (Mw 6.0), 26 Haziran 2020 Manisa (Mw 5.5), 30 Ekim 2020 (Mw 6.6) Seferihisar (İzmir) depremleri Ege Bölgesi'nin depremselliğinin araştırılması gerektiğini göstermektedir. Deprem araştırmalarında deprem frekansı diğer bir deyişle sıklığının tahmini için zamana bağlı çok sayıda stokastik model geliştirilmiştir. Bu stokastik modellerden en fazla kullanılanları Poisson ve Markov modelleridir Çalışmada öncelikle Ege Bölgesi'nde 1900-2021 yılları arasında meydana gelen öncü ve artçı depremler ana depremlerden ayrıştırılacaktır. Ayrıştırılan deprem veri kümesinden yararlanarak deprem sıklık ve büyüklükleri Poisson süreçlerinin özel durumları olan homojen Poisson süreci, birleşik Poisson süreci, bağımlı iki değişkenli birleşik Poisson süreci yardımıyla incelenecektir.

Key words: *Poisson süreci, Deprem tahmini, Olasılıksal model*

Ege Graben Sisteminde Meydana Gelen Depremlerin Büyüklük ve Derinliklerinin Markov Zincirleri ile Değerlendirilmesi

Gamze Özel[1*], Ceren Ünal[1], Tuğba Eroğlu Azak[2], Senem Tekin[3], Tolga Çan[4]

[1] Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06800, Ankara, Türkiye
[2] Milli Savunma Üniversitesi, Kara Harp Okulu, İnşaat Müh., 34334, Ankara, Türkiye
[3] Adıyaman Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Madencilik ve Maden Çıkarma Bölümü, 02040, Adıyaman, Türkiye
[4]Çukurova Üniversitesi, Mühendislik Fakültesi, Jeoloji Mühendisliği Bölümü, 01330, Adana, Türkiye

*Corresponding author e-mail: gamzeozl@hacettepe.edı.tr

Özet

Sismik tehlike değerlendirmesi için olasılıklı deprem oluşumu modelleri sıklıkla kullanılmaktadır. Bunlardan en önemlileri arasında Poisson modelleri ve Markov modelleri bulunmaktadır. Poisson modeli, orta sıklıktaki depremlerle karakterize edilen bölgelere uygulanırken, Markov modelleri, nadir depremlerin bulunduğu bölgelerdeki olay dizilerini daha iyi tanımlamaktadır. Markov modellerinin deprem oluşumu analizi için uygunluğu, bir deprem meydana geldiğinde fay altında toplanan enerjilerin boşaltılması gerçeğiyle açıklanabilir. Bu durum belirli bir bölgedeki depremin zaman ve büyüklüğünün bölgedeki bir önceki depreme bağlı olduğunu göstermektedir. Bu çalışmada öncelikle Ege Bölgesi'nde 1900-2021 yılları arasında meydana depremler neotektonik bölgelere göre sınıflandırılacaktır. Her bölgeye ait deprem büyüklük ve derinlik sınıfları elde edilerek depremlerin büyüklük ve derinliklerinde zamana bağlı değişimler Markov zincirleri yardımıyla analiz edilecektir.

Key words: *Markov zinciri, Deprem büyüklük tahmini, Geçiş olasılığı, Olasılıksal*

On Jumps Models and Tampered Innovations in Volatility Models and Applications

Oyebimpe Adeniji[1*]

[1] Ibadan University, Independent National Electoral Commission, 200132, Ibadan, Nigeria

*Corresponding author e-mail: emmanuel14444real@yahoo.com

Abstract

Generalised Autoregressive Conditional Heteroskedasticity (GARCH) models have been used to model non-constant variances in financial time series models. Previous works have assumed error innovations of GARCH models of order (p,q) as: Normal, Student-t and Generalised Error Distribution (GED), but these distributions failed to capture conditional volatility series adequately, leading to low forecast performance. This study is therefore aimed at developing variants of GARCH(p,q), Asymmetric Power ARCH (APARCH(p,q)) models, Exponential GARCH EGARCH(p,q) model and comparison with Jumps GARCH models such as Generalized Autoregressive Score (GAS) , the Exponential GAS (EGAS) and the Asymmetric Exponential GAS (AEGAS)) with asymmetric error innovations for improved forecast estimates. The two error innovations considered were the Generalised Length Biased Scaled-t (GLBST) and Generalised Beta Skewed-t (GBST) distributions, obtained by remodifying Fisher Concept of Weighted Distribution and McDonald Generalised Beta Function, respectively, in the Student-t distribution. The properties of the proposed distributions were investigated. The proposed innovations were imposed on GARCH(1,1), EGARCH(1,1) APARCH(1,1) models to obtain GARCH-GLBST(1,1) and APARCH-GLBST(1,1), EGARCH-GLBST(1,1) models, respectively. Similarly, GARCH-GBST(1,1), EGARCH -GBST(1,1), APARCH-GBST(1,1) models were also obtained by incorporating proposed innovations into GARCH(1,1), EGARCH(1,1) APARCH(1,1) models. Data from the Central Bank of Nigeria All Share Index (ASL) were used to illustrate the models. The proposed models were compared with jumps and classical models. The performance of the proposed models over the existing ones were investigated using the Log-likelihood function, Root Mean Square Error (RMSE), Adjusted Mean Absolute Percentage Error (AMAPE) and Akaike Information Criterion (AIC). Out of the 18 models in consideration, EGARCH-GLBST(1,1) was the best, followed by APARCH-GLBST(1,1) and EGAS models, in terms of the AIC values (7.856,7.988 and 9.984).The forecast evaluation criteria (RMSE, AMAPE), EGARCH-GLBST(1,1) model also ranked best (RMSE =0.281, AMAPE = 0.280), followed by APARCH-GLBST(1,1) model (RMSE = 0.291, AMAPE = 0.290) and EGAS model (RMSE = 0.309, AMAPE = 0.301). The least performing in terms of forecasts was the GARCH(1,1)-Normal model. The proposed volatility models with error innovations outperformed existing models in terms of fitness of conditional volatility and forecasts. The proposed models will be good alternatives for volatility modelling of symmetric and asymmetric stock returns.

Key words: *GARCH models, Generalised length biased Scaled distribution, Root mean square error, Jumps Models*

INTRODUCTION

Volatility models are dynamic models that address unequal variances in financial time series, the first and formal volatility model is the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle Robert (1982). The history of ARCH is a very short one but its literature has grown in a spectacular fashion. Engle's Original ARCH model and it various generalization have been applied to numerous economic and financial data series of many countries. The concept of ARCH might be only a decade old, but its roots go far into the past, possibly as far as Ba0chelier (190), who was the first to conduct a rigorous study of the behaviour of speculative prices. There was then a period of long silence. Mandelbrot (1963, 1967) revived the interest in the time series properties of asset prices with his theory that random variables with an infinites population variance are indispensable for a workable description of price changes. His observations, such as unconditional distributions have thick tails, variance change over time and large(small) changes tend to be follow by large(small) changes of either sign are stylized facts for many economic and financial variables. Empirical evidence against the assumption of normality in stock return has been ever since the pioneering articles by Mandelbrot (1963), Fama (1965), and Clark (1973) which they argued that price changes can be characterized by a stable Paretian distribution with a characteristic exponent less than two, thus exhibiting fat tails and an infinite variance. Volatility clustering and leptokurtosis are commonly observed in financial time seies (Mandelbrot, 1963). Another phenomenon often encountered is the so called "leverage effect" (black 1976) which occur when stock price change are negatively correlated with changes in volatility. Such studies is scared in Nigeria Stock Exchange Market and observations of this type in financial time series have led to the use of a wide range of varying variance models to estimate and predict volatility.

In his seminal paper, Engle (1982) proposed to model time-varying conditional variance with Autoregressive Conditional Heteroskedasticity (ARCH) processes using lagged disturbances; Empirical evidence based on his work showed that a high ARCH order is needed to capture the dynamic behaviour of conditional variance. The Generalised ARCH (GARCH) model of Bollerslev (1986) fulfils this requirement as it is based on an infinite ARCH specification which reduces the number of estimated parameters from infinity to two. Both the ARCH and GARCH models capture volatility clustering and Leptokurtosis, but as their distribution is symmetric. They fail to model the leverage effect. To address this problem, many nonlinear extensions of GARCH have been proposed, such as the Exponential GARCH (EGARCH) model by Nelson (1991), the so-called GJR model by Glosten et al (1993) and the Asymmetric Power ARCH (APARCH) model by Ding et al (1993).

Another problem encountered when using GARCH models is that they do not always fully embrace the thick tails property of high frequency financial times series. To overcome this drawback Bollerslev (1987), Baille and Bollerslev (1987) and Beine et al (2002) have used the Student's t- distribution. Similarly to capture skewness Liu and Brorsen (1995) used an asymmetric stable density. To model both skewness and kurtosis Fernandez and Steel (1998) used the skewed Student's t-distribution which was later extended to the GARCH framework by Lambert and Laurent (2000, 2001). To improve the fit of GARCH and EGARCH models into international markets, Harris et all (2004) used the skewed generalised Student's t-distribution to capture the skewness and leverage effects of daily returns.

The Beta probalility distribution missed with the student- t distribution and the resulting mixed- distribution applied to the GARCH model, with little modification to obtain the volatility model that is robust in modelling jumps. The Oil and stock markets stress of 1987 and 2008- 2009, respectively are good examples of jumps in volatility series (see Bates, 2000, Pan, 2002). Eraker, Johnannes and Polson (2003) apply continuous time stochastic volatility models with jumps components in returns and volatility of S&P500 and Nasdaq stocks indices ad observe significant evidence of jumps components, both in the volatility and in the returns. Generalized Autoregressive Score (GAS) , the Exponential GAS (EGAS) and the Asymmetric Exponential GAS (AEGAS) are new classes of volatility models that simultaneously account for jumps and asymmetry.

These jumps in ASI were experience as a result of influence of news, politics and global crisis on Nigeria economy. This project seek to estimate volatility in the Nigeria Stock Market along with forecasting performance of GARCH and new classes of volatility models that simultaneously account for jumps and asymmetry together with different density functions and recommending the most robust model for financial analysts and portfolio managers in the finance market. These jumps in ASI were experience as a result of influence of news, politics and global crisis on Nigeria economy.

DATA SOURCE: A daily data of the All Share Index (ASL) from the period January 3, 2000 to December 22 , 2017 were obtained from CBN statistical bulletin 2018

MATERIAL AND METHODS

Material

A daily data of the All Share Index (ASL) from the period January 3, 2000 to December 22 , 2017 were obtained from CBN statistical bulletin 2018.

Methods

We define as a (t x1) vector of assests log- returns up to time t that is

$$r_t = \varepsilon_t = \sigma_t z_t \qquad (1)$$

Where follows a particular probability distribution, and is the square root of the conditional variance. The mean equation of the model ie the deterministic aspect of the series follows Autoregressive Model, AR(p),

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \qquad (2)$$

The Standard GARCH (p,q) model by Bollerslev in (1986) is given as :

$$\sigma_t^2 = \alpha_0 + \sum_{t=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{t=i}^{p} \beta_i \sigma_{t-i}^2 \qquad (3)$$

Where> 0, ≥0 (for i=1, ---, q) , ≥0 (for j=1, …, p) is sufficient for conditional variance to be positive and stationarity. To capture asymmetry observed in series, a new class of ARCH model was introduced: The asymmetric power ARCH (APARCH) model by Ding et'al (1993), the exponential GARCH (EGARCH) model by Nelson (1991), the GJR by Glosten et al. (1993). This model can generate many model when the parameters are relaxed and is expressed as:

$$\sigma_t^\delta = \alpha_0 + \sum_{i=1}^{q} \alpha_i \left(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i} \right)^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta \qquad (4)$$

The parameter permit us to catch the asymmetric effects. The conditional standard deviation process and the asymmetric absolute residuals in the model were imposed in term of a Box- transformation. The well-known Leverage effect is the asymmetric response of volatility to negative and positive shocks.

Harvey (2013) developed three sets of volatility models that take into account robust capturing occasional changes in financial series known as jumps, he considered the EGARCH and AEGARCH types of the Beta-GARCH models each with two distributions assumptions applied. The Beta- EGARCH model specified without the leverage effect is given:

$$\log\sigma_t^2 = \omega + \alpha\mu_{t+1} + \varphi_1 log\sigma_{t-1}^2 \qquad (5)$$

Introducing the leverage effect we have the Beta-AEGARCH model (ie EGAS):

$$\log\sigma_t^2 = \omega + \alpha\mu_{t+1} + \gamma I_{t-1}\varphi_1 + log\sigma_{t-1}^2 \qquad (6)$$

Where I_(t-1)=sgn(〖-z〗 _t)(μ_t+1)when student-t distribution is considered and I_(t-1)=sgn(〖-z〗 _t)(μ_t+1) for skewed the skewed student-t distribution.Again,combing the same student-t with EGAS model leads to Beta-t AEGARCH ie AEGAS model.

The generalized beta distribution of the first kind was introduced by McDonald (1984), with link function

$$g(y) = \frac{c}{\beta(a,b)}\left[F(y)\right]^{ac-1}\left[1-F(y)^c\right]^{b-1} f(y) \qquad (7)$$

Where a, b, c are the shape parameter, f(y) is the probability function of student –t distribution, F(y) is the incomplete beta function and g(y) is the link function of Generalized Beta Skew-t distribution. The log-likelihood for estimation is:

$$l = LogL = n\log c - n\log B(a,b) + n\log\Big)\frac{v+1}{2} - n\log\Big)\frac{v}{2} - \frac{n}{2}\log\left[\pi(v-2)\sigma_t^2\right]$$

$$+(ac-1)\sum_{t=2}^{n} LogI + (b-1)\sum_{t=2}^{n}\log(1-I^c) - \left(\frac{v+1}{2}\right)\sum_{t=2}^{n}\log\left[1+\frac{\varepsilon_t^2}{\sigma_t^2(v-2)}\right] \qquad (8)$$

$\mu_w = E\left(w(y)\right) < \infty$ Then the random variables of Y_w having pdf $f_w(y) = \dfrac{w(y)f(y)}{E\left(w(y)\right)}, a < y < b$ Where

$E\left(w(y)\right) = \int_{-\infty}^{\infty} w(y)f(y)dy$, $-\infty < y < \infty$ is said to have weighted distribution length biased distribution is derived when the weighted function depend on the length of units of interest (i.e. w(y) = y) . The pdf of a length biased random variable is defined as:

The log-likelihood of equation (4) when the pdf is student-t is obtained as

$$L = \log\prod_{i=1}^{n} g(y) = n\log\Big)\frac{v+1}{2} - n\log\mu - n\log\Big)\frac{v}{2} - \frac{n}{2}\log\left[\pi(v-2)\sigma^2\right] + \sum_{i=1}^{n}\log y_i - \left(\frac{v+1}{2}\right)\sum_{i=1}^{n}\log\left[1+\frac{(y-\mu)^2}{(v-2)\sigma^2}\right] \qquad (10)$$

These two newly distributions will be incorporated into conventional and Jumps GARCH models. In the literature the most recent error innovation used along with volatility models are Normal, Student-t and GED. Below are parameter estimations of the three innovation: see Yaya *et al*, (2013) ,for Normal distribution ,the Log-likelihood is

$$\boldsymbol{l_t} = -\frac{\mathbf{1}}{\mathbf{2}}\left[Nlog(2\pi) + \sum_{i=1}^{N}\frac{\varepsilon_t^2}{\sigma_t^2} + \sum_{i=1}^{N} log\sigma_t^2\right] \qquad (11)$$

$\boldsymbol{\varepsilon_t = z_t\sigma_t}$Where $z_t = \frac{\varepsilon_t}{\sigma_t}$

Equation (9) is the log-likelihood $\boldsymbol{l_t}$of Normal , N is the sample sizes of the series, ε_t is the white noise ,$z_t is$ sequence of identical independent random variables and σ_t^2 $is\ the$ conditional variance.

The Log-likelihood for Student-t distribution is

$$l_t = -\frac{1}{2}\left[N\log\left(\frac{\pi(v-2)\gamma\left(\frac{v}{2}\right)^2}{\gamma\left(\frac{v+1}{2}\right)^2}\right) + \sum_{i=1}^{N}\log\sigma_t^2 + (v-1)\sum_{i=1}^{N} log\left(1+\frac{\epsilon_t^2}{\sigma_t^2(v-1)}\right)\right] \qquad (12)$$

In the estimation in equation (10) v is the degree of freedom and $\gamma(.)$ Is the gamma function, for GED it is

$$l_t = -\frac{1}{2}\left[N\log\left(\frac{\gamma(v^{-1})^3}{\gamma(3v^{-1})\,\gamma\left(\frac{v}{2}\right)^2}\right) + \sum_{i=1}^{N}\log\sigma_t^2 + (v-1)\sum_{i=1}^{N} log\left(\frac{\gamma(3v^{-1})\epsilon_t^2}{\sigma_t^2(v^{-1})}\right)\right] \qquad (13)$$

RESULTS

To obtain a stationary series, we use the returns R_t=100(log(Y_t)- 〖log(Y〗 _(t-1))) where is the closing value of index at month t. The sample statistics for the returns are exhibited in table 1. For NSE index (sample January 2000 to September 2018). The time plots which is the first step to examine hidden characteristic reveals non-stationarity, patterns and clustered volatility.

Table 1. Descriptive statistics for returns

Index	Min	Median	Mean	Max	Skewness	Kurtosis	Shapiro-Wilk Test
NSE	1.00	96.50	96.40	191.00	-0.00763	-3.2260	0.9538

The skewness is negatively skewed and also exist negative kurtosis which indicate anomalous distribution. Shapiro-Wilk test indicate non normality. Out of the 18 models in consideration, EGARCH-GLBST(1,1) was the best, followed by APARCH-GLBST(1,1) and EGAS models, in terms of the AIC values (7.856,7.988 and 9.984).The forecast evaluation criteria (RMSE, AMAPE), EGARCH-GLBST(1,1) model also ranked best (RMSE =0.281, AMAPE = 0.280), followed by APARCH-GLBST(1,1) model (RMSE = 0.291, AMAPE = 0.290) and EGAS model (RMSE = 0.309, AMAPE = 0.301). The least performing in terms of forecasts was the GARCH(1,1)-Normal model.

DISCUSSION AND CONCLUSION

Mixture innovations of GARCH models best explained Nigeria Stock volatilities. The forecasting we obtain are evaluated using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) predicting 24 steps ahead. The forecasting is reported by ranking the different models with respect to RMSE and MAPE for NSE index .The proposed volatility models with mixture error innovations outperformed conventional models in terms of fitness of conditional volatility and forecasts. The proposed models will be good alternatives for volatility modelling of symmetric and asymmetric stock returns.

References

Beine M, Laurent S, Lecourt C, 2002. Accounting for conditional leptokurtosis and closing days effects in FIGARCH models of daily exchange rates, Applied Financial Economics, 12, 589–601.

Black F, 1976. Studies of stock market volatility changes, Proceedings of the American Statistical Association, Business and Economic Statistics Section, 177–81.

Bollerslev T, 1986. Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31, 307–27.

Lambert P, Laurent S, 2000. Modelling skewness dynamics in series of financial data, Discussion Paper, Institut de Statistique, Louvain-la-Neuve

Shittu OI, Adepoju KA, Adeniji EO, 2014. On Beta Skew – t distribution in modeling Stock Returns in Nigeria, International Journal of Modern Mathematical Sciences, Florida, USA 11(2): 94-102

Harvey A, 2013. Dynamic Models for volatility and Heavy tails: with applications to financial and economic time series. Cambridge, University Press, London.

Harvey A, 2014. On the Formulation of ARCH in Mean Models. Proceedings of 15th Oxmetrics Users Conference. Cass Business School, City University, London.

Harvey AC, Luati A, 2014. Filtering with heavy tails. Journal of the American Statistical Association, page forthcoming.

Comparison of Beta Mean, Beta Model and Generalized Biparabolic Distributions Regression on Milk Traits

Ebru Ergüneş Berkin[1*], **Hasan Önder**[2]

[1] Ondokuz Mayıs University, 55020, Samsun, Turkey
[2] Islahiye Directorate Of District Agriculture And Forestry

*Corresponding author e-mail: ebru_berkin01@hotmail.com

Abstract

Percentage data are the case of taking values in the standard unit range (0,1) such as percentages, ratios, fractions and proportions. Analysis of percentage data has a general problem for researchers to work with quantitative data. Many studies have been conducted in recent years on statistical modeling for continuous percentage data. Commonly used regression models such as linear or nonlinear regression models are not suitable for percentage continuous variables. One of the proposed methods is based on the beta distribution by Ferrari and Cribari-Neto (2004), which is useful for modeling continuous variables that take values (0, 1). In beta regression, the parameterization distribution is modeled as beta with the help of mean and precision parameters. Here, the mean linked with the response, as in generalized linear models (GLMs), by means of a link function and a linear predictor. Besides, the precision parameter can be linked to another set of regressors via a second link function, which then results in a variable dispersion model. Many link functions such as probit, log, cauchit, log-log can be used for this process. Estimation is performed with maximum likelihood via link one of these link functions. In this study it was aimed to show the usage of beta mean, beta model and generalized biparabolic distributions regression methods using animal science data. Results showed that generalized biparabolic distributions method had the best fitting according to Log-Likelihood value for regression of milk protein percentage and milk yield on milk fat percentage.

Key words: *Regression, Percentage Data, Milk Traits, Beta Regression*

A New Flexible Discrete Distribution with Application to Zero-Inflated Regression Analysis

Yunus Akdoğan[1*]

[1] Selcuk University, Department of Statistics, Faculty Science, 42075, Konya, Turkey

*Corresponding author e-mail: yakdoğan.selcuk.edu.tr

Abstract

In this paper, alpha power transformation for continuous distributions is adapted to discrete distributions. A new family for discrete distributions called discrete alpha power transformation is proposed. The discrete alpha power transformation exponential distribution is studied in detail. Several distributional properties of introduced distribution including moments, survival and hazard rate functions, mode, and quantile function are discussed. The statistical inference on the model parameters is studied by maximum likelihood, moments, and least-squares estimation methods. A simulation study is performed to observe the performance of bias and mean square errors of these estimates. Three bootstrap methods are considered for constructing confidence intervals for the distribution parameters. As an application of the discrete alpha power transformation-exponential distribution, a new zero-inflated count regression model is proposed to be an alternative model for zero-inflated Poisson, zero-inflated geometric and zero-inflated negative binomial regression models. Two examples with real data are provided to illustrate the applicability of introduced distribution and count regression analysis.

***Key words**: Bootstrap confidence intervals, ratio test method, discrete distributions, count regression, zero-inflated model*

A New Lifetime Distribution Based on the Transmuted First Two Lower Records

Tenzile Erbayram[1*], **Ümmügülsüm Yıldırım**[1], **Yunus Akdoğan**[1]

[1] Selcuk University, Department of Statistics, Faculty Science, 42075, Konya, Turkey

*Corresponding author e-mail: tenzile.erbayram@selcuk.edu.tr

Abstract

This article introduces a new lifetime distribution by merging the first two lower records based on exponential distribution and discusses the different features of the distribution. Statistical inferences about the distribution parameters are discussed with three estimation methods, namely maximum likelihood, least squares, and weighted least squares. Monte Carlo simulation study is performed to evaluate of these estimators based on mean square errors estimation, mean absolute deviation, and mean relative errors of estimation for a sample of different sizes. A distribution simulation analysis based on real data is provided to demonstrate the adaptability of the proposed model

.

Key words: *Lower record value, Lifetime distribution, Monte Carlo Simulation, Estimation*

Stress Strength Reliability for Discrete Natural Lindley Distribution

Ümmügülsüm Yıldırım[1*], Yunus Akdoğan[2], Tenzile Erbayram[3]

[1] Selcuk University, Department of Statistics, Faculty Science, 42075, Konya, Turkey

*Corresponding author e-mail: acargulsum466@gmail.com

Abstract

The probability of estimating stress-strength plays an important role in reliability analysis. If stress is in the form of a number of shocks which is applied on a product and strength is in terms of the number of shocks that the product can withstand, then the stress and strength are discrete random variables. In the case of the demand-supply system of a production process, the number of items demanded or supplied is also discrete in nature. In this work, stress-strength reliability are discussed for Discrete natural Lindley Distribution. Two estimation methods such as maximum likelihood and proportion are obtained for stress-strength reliability. Simulations study is also performed to compare the estimates.

Key words: *Stress-strength reliability, Discrete natural Lindley distribution, Estimation, Monte Carlo simulation.*

Behaviour of LER for Some Policy Modification Based on Some Claim Size Distributions

İsmail Kınacı[1*], İsmail Hakkı Kınalıoğlu[1], Fahreddin Kalkan[1], Demet Sezer[1]

[1] Selcuk University, Department of Statistics, Faculty Science, 42075, Konya, Turkey

*Corresponding author e-mail: ikinaci@selcuk.edu.tr

Abstract

Premium is one of the most important concepts for insurance policies. Premium is a factor that directly affects both the profitability of companies and the sale of policies. As the premium for the policies increases, the number of sales of the policy will decrease, which is undesirable. As the number of policies decreases for companies, the uncertainty increases. For this reason, some modifications such as insurance, exemption and limit are frequently used on the policies in order to reduce the premium. Thanks to these modifications, there will be a decrease in both the policy premiums and the claims arising from the policy. Loss elimination ratio (LER) is a value that is frequently used to show how much policy modifications reduce claims. In this study, a simulation study was conducted in which policy modifications were compared in terms of LER, considering different distributions for damage sizes, and it was tried to determine which modification was more suitable for which damage distribution.

Key words: *Loss Elimination Ratio, Policy Modification, Premium, Claim Size Distribution.*

The Poisson Epanechnikov-Exponential Distribution with a New Count and Zero-Inflated Regression Analyses

Kadir Karakaya[1*]

[1] Selcuk University, Department of Statistics, Faculty Science, 42075, Konya, Turkey

*Corresponding author e-mail: kkarakaya@selcuk.edu.tr

Abstract

In this study, a new discrete distribution is introduced combining Poisson and Epanechnikov-exponential distributions. This new model is called as Poisson-Epanechnikov-exponential distribution. Some distributional properties of the new model, such as moments, skewness, kurtosis, moment generating function, etc. are obtained. Some estimators, such as maximum likelihood, moment and proportion are examined to estimate the model parameter. A new count regression and zero-inflated count regression models are also introduced based on the new distribution. The Monte Carlo simulation study is carried out to compare the estimators for different sample sizes and parameter settings. The usability of the new model is demonstrated by various practical data analyses.

Key words: *Poisson distribution, Epanechnikov-exponential distribution, Count regression model, Zero-inflated regression model, Real data analysis*

Some Simulation for Ruin Probability Based on Different Ruin Scenarios

Fahreddin Kalkan[1*], **İsmail Hakkı Kınalıoğlu**[1], **Demet Sezer**[1], **İsmail Kınacı**[1]

[1] Selcuk University, Department of Statistics, Faculty Science, 42075, Konya, Turkey

*Corresponding author e-mail: fahreddin.kalkan@selcuk.edu.tr

Abstract

Ruin is one of the most important issues for insurance companies. Ruin is briefly defined as a company's capital falling below zero. However, companies may not ruin as soon as their capital turns negative. Various ruin scenarios can be considered, such as when the capital of the companies is negative k times in a row, when it goes negative for the kth time, or when the time between two consecutive negatives is below a prefixed constant. In addition, premium incomes, claim frequencies and claim sizes are also factors affecting ruin. Calculating the probability of ruin in most cases is quite complex. For this reason, simulation is frequently used in calculating the probability of ruin. In this study, ruin simulations were carried out based on various ruin scenarios, considering distributions with different characteristics for claim size.

Key words: *Ruin, Claim Size, Claim Frequency, Risk Process.*

Topic Classification and Sentiment Analysis on Consumer Reviews on Shopping Sites

İsmail Hakkı Kınalıoğlu[1*], İsmail Kınacı[1], Demet Sezer[1], Fahreddin Kalkan[1]

[1] Selcuk University, Department of Statistics, Faculty Science, 42075, Konya, Turkey

*Corresponding author e-mail: ismailhakkikinali@gmail.com

Abstract

Technology has caused various paradigm shifts in many areas of human life. Access methods for products and services is one of them. Nowadays, we can easily reach any product or service we need through the internet. Especially during the pandemic process, the fact that people spend most of their time at home has led to a very high increase in the amount of online shopping. This provides consumers with benefits such as time savings, competitive prices, abundance of options, ease of return and exchange, etc. However, purchasing a product that is not physically visible can sometimes carry a risk. At this point, the most important guide of the users is the comments made by other consumers about the products. These comments contain information about various aspects of the product or service. While some comments provide information about the general features of the product, others may include information on the shipping process, seller, packaging, etc. The correct categorization of these comments will enable users of shopping sites to have an opinion more easily. In this study, sentiment analysis based on text mining was carried out on user experiences in shopping sites that are frequently used by Turkish consumers. User comments were collected with the developed web scraping tools. Afterwards, these interpretations, which were passed through a linguistic filter, were classified according to the determined themes. Machine learning algorithms were used for the classification process and the success of these algorithms was compared.

Key words: *Classification, Machine Learning, Sentiment Analysis, Text Mining, Topic Classification.*

A New Family of Distributions: Compound Transmuted Family of Distributions Properties and Applications

Coşkun Kuş[1*], Kadir Karakaya[1], Caner Tanış[2], Yunus Akdoğan[1], Sümeyra Sert[1], Fahreddin Kalkan[1]

[1] Selçuk University, 42075, Konya, Turkey
[2] Çankırı Karatekin University, 18100, Çankırı, Turkey

*Corresponding author e-mail: kkarakaya@selcuk.edu.tr

Abstract

Many lifetime distributions have been introduced in the last 10 years. These distributions do not fit the data collected in practice. In some cases, there may be problems on estimating the parameters of the proposed distributions. In this study, a flexible lifetime distribution is proposed, and some distributional properties are studied. Several estimation methods such as maximum likelihood, least squares, weighted least squares, Cramer-von Mises and Anderson-Darling type are discussed and compared according to their mean squared error criterias. Two numerical examples are also provided. The first example relates to modeling the distribution of of mechanical component failure times and the second example deals with modeling the distribution of survival times of patients given chemotherapy treatment. The proposed model is superior to other current models in these examples.

Key words: *Lifetime Distribution, Estimation, Monte Carlo Simulation*

Automatically Image Classification Based on a New CNN Architecture

Öznur Özaltın[1*], Murat Köklü[2], Aynur Yonar[3], Özgür Yeniay[1]

[1] Hacettepe University, Institute of Science, Department of Statistics, 6800, Ankara, Turkey
[2] Selcuk University, Faculty of Technology, Department of Computer Engineering, 42031, Konya, Turkey
[3] Selcuk University, Faculty of Science, Department of Statistics, 42031, Konya, Turkey

*Corresponding author e-mail: oznurozaltin@hacettepe.edu.tr

Abstract

These days, deep learning algorithms are used in many fields of our lives, especially image recognition. Convolutional neural networks (CNN), which are one of the most effective algorithms for image recognition, classification, or prediction, are preferred. Thanks to CNN, real-world problems can be solved easily in many fields such as biomedical, bio information, agriculture, livestock, and economy. In this study, we purpose that a CNN is created for classifying three real datasets. Primarily, we create a new CNN, called OzNet, having many convolutional layers. Next, we utilize widely preferred two datasets: CIFAR-10 and MNIST for measuring the performance of OzNet. As a conclusion of the classification from CIFAR-10 and MNIST datasets, the high success accuracy rates are acquired at 85.4% and 100%, respectively. The following study, we investigate a Pistachio image dataset which is having binary classes. Experimental results show that OzNet architecture obtained an accuracy of over 95%.

Key words: *Classification, CNN, Deep Learning, OzNet Architecture.*

INTRODUCTION

In recent years advancements, hardware, and data sets perform it practical to train deep neural networks (DNN) by using many hidden layers (Bankman, Yang, Moons, Verhelst, & Murmann, 2018; Bengio & LeCun, 2007; Krizhevsky, Sutskever, & Hinton, 2012a). DNN obtains state-of-the-art success on duties, for instance, image recognition and speech recognition (Bankman et al., 2018). Besides, Convolutional neural networks (CNN) which is one of the deep learning algorithms are widely used for two-dimensional (images) datasets (Indolia, Goswami, Mishra, & Asopa, 2018).

In this study, we develop a new CNN named as OzNet and evaluate it using two famous datasets: CIFAR-10 and MNIST (Mixed National Institute of Standards and Technology). MNIST and CIFAR-10 are accepted as the most cited the datasets in research as a test area for new opinions (Koonce, 2021). While the MNIST dataset consists of 28x28 grayscale images which are handcrafted digits of the numbers 0–9, the CIFAR-10 dataset composes of 32x32x3 real-world problems images (airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck). Whether it's a real-world problem isn't the only difference between the CIFAR-10 dataset and MNIST. The CIFAR-10 dataset has also RGB (red, green, blue) color channels. The color turns out to not be as complex an issue as considered by CNN (Koonce, 2021). MNIST dataset cannot be evaluated for overfitting when a new algorithm is created. In addition, the CIFAR-10 dataset should be also investigated to eliminate this issue. Many researchers have used these popular datasets in their studies (Baldominos, Saez, & Isasi, 2019; Darlow, Crowley, Antoniou, & Storkey, 2018; Doon, Rawat, & Gautam, 2018; Kayed, Anter, & Mohamed, 2020; Krizhevsky & Hinton, 2010; Schott, Rauber, Bethge, & Brendel, 2018; Thakkar, Tewary, & Chakraborty, 2018; Xiao, Rasul, & Vollgraf, 2017; Yang, Bankman, Moons, Verhelst, & Murmann, 2018).

Additionally, we examine a different real-world problem dataset: Pistachio to assess OzNet performance. Pistachio includes unsaturated fatty acids (Singh et al., 2022). For this reason, it is commonly utilized in the nutrition of people and many food industries(Singh et al., 2022). Pistachios quality is related to reflecting in the cost to the consumer. If the Pistachios are quality, the price is expensive. Thus, economy, export, and marketing are correlated with the quality of pistachios (Singh et al., 2022). In line with this consideration, practical determining the quality of pistachios in other words good classification is very crucial for producers. When the algorithm is created effectively, the producers find an answer to the question "How is the product's situation: good or not?". For this purpose, many researchers present some algorithms to classify the Pistachio dataset (Bonifazi, Capobianco, Gasbarrone, & Serranti, 2021; Cetin, Pearson, & Tewfik, 2004; Omid, Firouz, Nouri-Ahmadabadi, & Mohtasebi, 2017; Özkan, Köklü, & Saraçoğlu, 2021; Singh et al., 2022; Vidyarthi, Tiwari, Singh, & Xiao, 2020).

MATERIAL AND METHODS

Material

In this study, a new Convolutional Neural Network (CNN) which is called OzNet is presented to classify three different image datasets. As well as two widely used datasets which are MNIST and CIFAR-10, we also classify the Pistachio dataset using OzNet. Therefore, we evaluate to performance of OzNet through different datasets with binary and multinomial classes.

Methods

MNIST Dataset

The MNIST (Mixed National Institute of Standards and Technology) dataset consists of handwritten digits. It was presented in 1998 by Y. LeCun et al. Although the dataset includes 70,000 images, we utilize 10,000 images that are selected randomly from the main dataset in this study. The digits possess 10 classes and are labeled from 0 to 9. Essentially, each image has a size of 28x28. However, we resize to 227x227 .jpg format all images according to OzNet initial size. Figure 1 exhibits examples of the dataset.

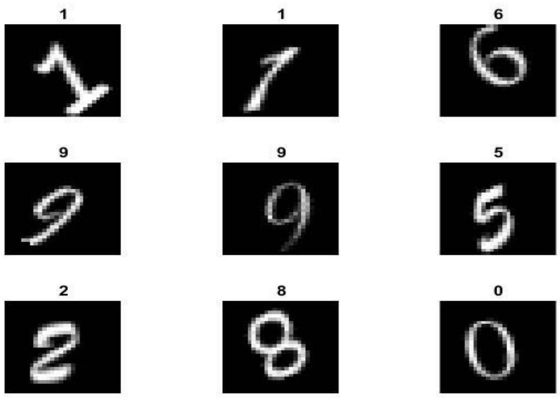


Figure 1. Examples of the MNIST dataset.

CIFAR-10 Dataset

CIFAR-10 is including 10 classes which are airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck (CIFAR-10). Notably, the dataset utilizes to detect image recognition(Chauhan, Ghanshala, & Joshi, 2018). In total, the dataset contains 60,000 images with 32x32x3 dimensions. Hence, there are 6000 images in each class. According to split the dataset test-train set of 20%-80%, 50,000 images are used for training, and 10,000 images are used for testing. In this study, we also utilize 50,000 images to train and 10,000 images to test, randomly. Figure 2 shows instances of the dataset.

Figure 2. Randomly selected instances of the CIFAR-10 dataset.

Pistachio Dataset

Pistachio Dataset consists of 2148 images that have binary classes: Kirmizi and Siirt pistachios. Here, 1232 images are labeled as Kirmizi, and 946 images are labeled as Siirt. The real dataset was created from original images. Additionally, each image size is 600x600x3 RGB (Red-Green-Blue) format (Özkan et al., 2021; Singh et al., 2022). Figure 3 shows the flowchart of this study. Figure 4 demonstrates randomly chosen samples of the Pistachio dataset. In this study, we declared that the dataset is our main dataset and we also resized the 227x227x3 .jpg format for each image in order to analyze the performance of OzNet.

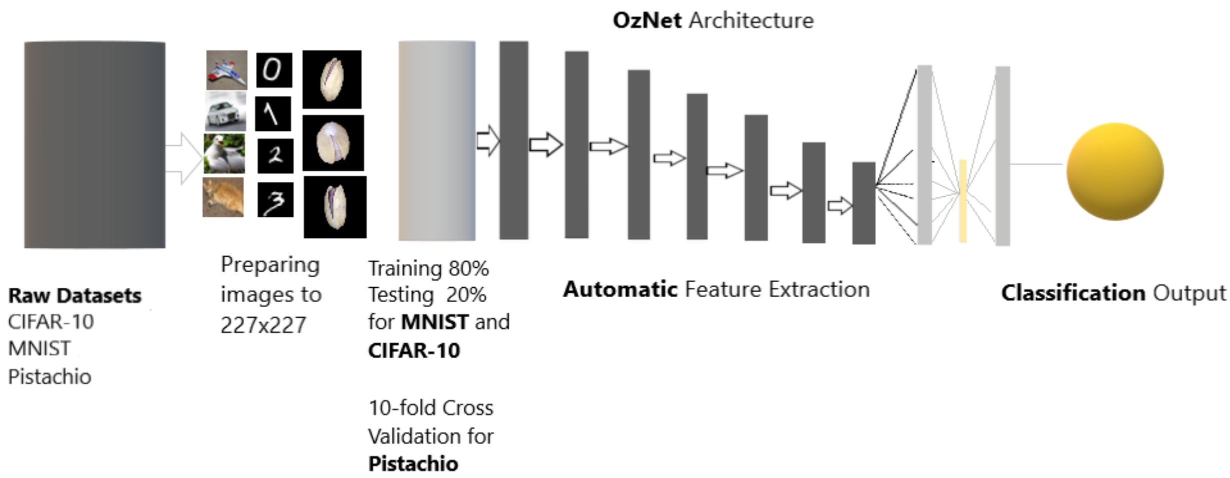


Figure 3. Flowchart of this study.

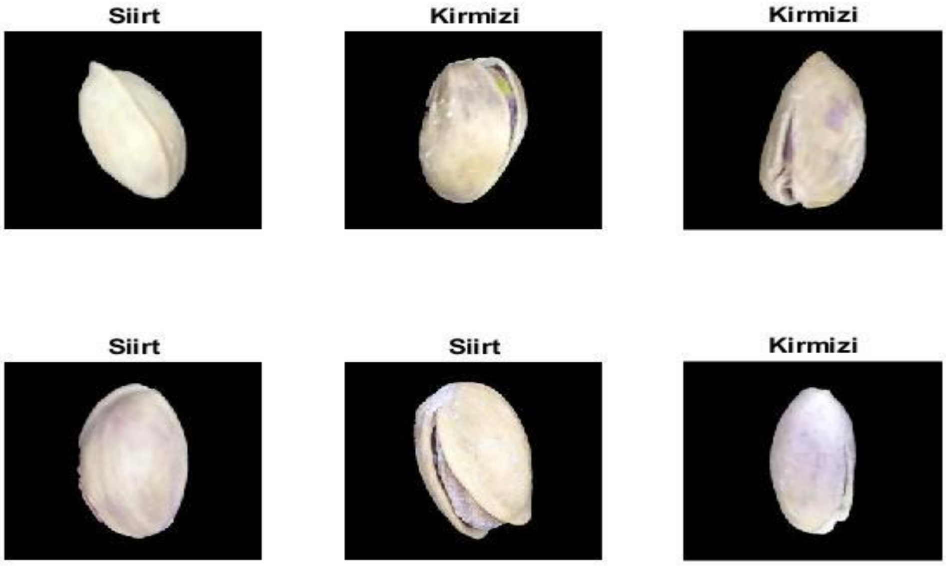


Figure 4. Randomly chosen samples of the Pistachio dataset.

OzNet: A new CNNs architecture

Convolutional Neural Networks (CNN) is a perfect deep learning algorithm able to tackle huge parameters(Chauhan et al., 2018). Over the limited history of automatic image recognition, rising the role of learning appears to have always made better the general yield of recognition systems(Yann LeCun, Bottou, Bengio, & Haffner, 1998). Therefore, hand-crafted feature extractions have been removed thanks to CNN (Yann LeCun et al., 1998). According to this property, experts' opinions do not need almost, and CNN architectures are developed by the day. In recent years, many architectures are presented to the literature by researchers. For example, AlexNet (Krizhevsky, Sutskever, & Hinton, 2012b), DenseNet201(Huang, Liu, Van Der Maaten, & Weinberger, 2017), GoogleNet (Szegedy et al., 2015), MobileNet (Howard et al., 2017), NASNetMobile (Zoph, Vasudevan, Shlens, & Le, 2018), ResNet-50 (He, Zhang, Ren, & Sun, 2016), SqueezeNet (Iandola et al., 2016), and VGG-16 (Simonyan & Zisserman, 2014) are some of the architectures.

In this study, we introduce a new deeper CNN architecture that can be automatically extracted features from images and classified the images. This architecture which had 34 layers and was named as OzNet (Özaltın & Yeniay, 2021), was designed with specific parameters. In Table 1, the parameters of OzNet are shown. In this study, OzNet architecture is based on 7 main block layouts. Each block consists of a convolution layer, a maximum pooling layer, a batch normalization layer, and a ReLU activation layer. The next layers are respectively two fully connected layers, a dropout layer, a SoftMax layer, and a classification layer. As for the answer to the question: How is it different from other CNN architectures? Here, we utilized different convolution layers, filter numbers, filter sizes, and fully connected layers. In general, filter sizes and the numbers of filters are determined regularly by rising. However, we do not use this procedure. When Table 1 is attentively investigated, it is seen that OzNet filter sizes are starting with 64, then the filter sizes in the next two layers are 128. Next, it is reaching 256. The following three layers of filter sizes are 128 again. The same is true for filter numbers. While many CNNs use 3x3 as filter numbers for convolutional layers, the situation is different in OzNet. Therefore, we can declare that OzNet is the original architecture.

Table 1. Parameters of OzNet

Name of layer	Parameters				Output size
	Size of Filter	**Number of Filters**	**Stride**	**Padding**	
Input	-	-	-	-	227x227x3
Conv-1	64	5x5	1	1	225x225x64
MaxPool-1	3x3	2	0	3x3	112x112x64
Conv-2	128	3x3	1	1	112x112x128
Conv-2	128	3x3	1	1	112x112x128
MaxPool-2		3x3	2	0	55x55x128
Conv-3	128	13x13	1	0	55x55x128
MaxPool-3		3x3	2	0	27x27x128
Conv-4	256	7x7	1	1	27x27x256
MaxPool-4		2x2	2	0	13x13x256
Conv-5	128	3x3	1	1	13x13x128
MaxPool-5		3x3	2	0	6x6x128
Conv-6	128	3x3	1	1	6x6x128
MaxPool-6		3x3	2	0	3x3x128
Conv-7	128	3x3	1	1	3x3x128
MaxPool-7	-	2x2	2	0	1x1x128
FC-8	4096				1x1x4096
Drop-8	50%				
FC-9	10 (number of class)				1x1x10
SoftMax					1x1x10
Output	Cross entropy				

Cross-Validation

It is very important to get reliable results from algorithms including black-box such as deep learning and machine learning algorithms. For reliable results, k-fold cross-validation is generally preferred by researchers (Lopez-del Rio, Nonell-Canals, Vidal, & Perera-Lluna, 2019; Saber, Sakr, Abo-Seida, Keshk, & Chen, 2021; Subasi, 2012). Another common problem is overfitting. In order to overcome this problem, k-fold cross-validation is used (Saber et al., 2021).

Cross-validation is a method in which the dataset is divided randomly with the determined number of folds and considering one of the sub-folds as the test fold, it trains the frame with left behind folds (Koklu & Ozkan, 2020). This process is recapped up to k folds and is tested in the frame (Arlot & Celisse, 2010). In this study, we determine k as 10 for trustful classification results.

Performance metrics

In this study, we measure OzNet with performance metrics: accuracy, sensitivity, specificity, precision, F1-Score, and G-mean (Rajinikanth, Joseph Raj, Thanaraj, & Naik, 2020; Sharifrazi et al., 2021; Singh et al., 2022). Here, Table 2 shows the performance metrics' equations

Table 2. Performance metrics' equations

Performance Metrics	Equations
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$
F1-Score	$(2\times TP)/(2\times TP+FP+FN)$
G-Mean	$\sqrt{Sensitivity\times Specificity}$
Precision	$TP/(TP+FP)$
Sensitivity	$TP/(TP+FN)$
Specificity	$TN/(TN+FP)$

TP : True Positive, FP : False Positive, TN : True Negative, and FN : False Negative

Receiver operating characteristic (ROC) curve

Sensitivity and specificity, which are based on the number of true positive (TP) and the number of true negatives (TN), respectively, form the fundamental metrics of the performance of tests (Park, Goo, & Jo, 2004). When the results of a test fall into one of two specified classes then the test possesses only one couple of sensitivity and specificity values (Park et al., 2004). However, making a decision is both hard and unsuitable in many categories of situations (Park et al., 2004). Therefore, a single couple of sensitivity and specificity values is not enough to determine the performance of a test, and the receiver operating characteristic (ROC) curve should be investigated in many categories' situations. This curve's x-axis shows the false positive rate and the y-axis shows the true positive rate. In general, the area under the curve (AUC) is also calculated to define if a certain condition is regarding test data. While the AUC value is approximately 1, the algorithm has a successful classification performance (Singh et al., 2022; Taspinar, Cinar, & Koklu, 2021). In this study, we also investigated ROC curves and computed AUC values for each dataset.

RESULTS

In this study, we classify three different image datasets by using OzNet in a MATLAB environment with Intel Core i7-7500U CPU, NVIDIA GeForce GTX 950M,16 GB RAM, 64-bit Operating System.
When a new CNN architecture is proposed, its performance is evaluated via famous datasets such as MNIST and CIFAR-10. Essentially, we proposed a new CNN architecture and wanted to see how it is performance on famous datasets. These datasets possess 10 different classes and include numerous images. Besides, we also use the Pistachio dataset to measure OzNet's performance on a real-world problem. This real dataset has two classes.

In this study, we apply no pre-processing method and resized images to 227x227 for input layer expectation. Also, we do not use data augmentation because of having enough images for each class. Moreover, 10-fold cross-validation is utilized while classifying the Pistachio dataset. Table 3 shows classification results for each experiment by using OzNet.

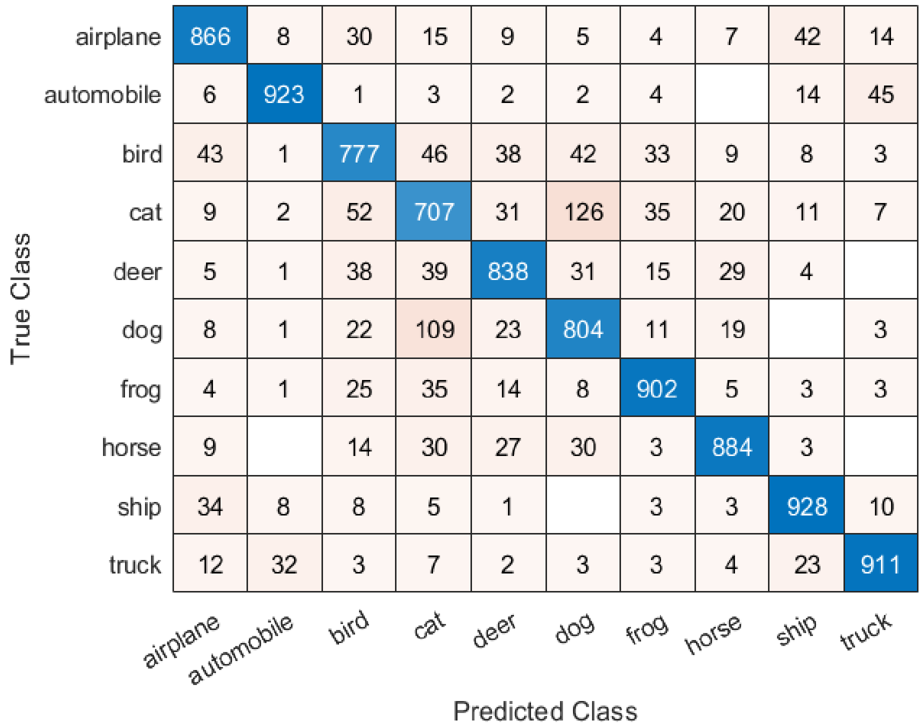
Table 3. Performance metrics of OzNet through different datasets.

Dataset	Performance Metrics						AUC
	Sensitivity	Specificity	Precision	F1-Score	G-Mean	Accuracy	
Pistachio	0.9570	0.9585	0.9688	0.9628	0.9577	0.9576	0.9931
CIFAR-10	0.8540	0.9838	0.8544	0.8541	0.9166	0.8540	0.9918
MNIST	1.00	1.00	1.00	1.00	1.00	1.00	1.00

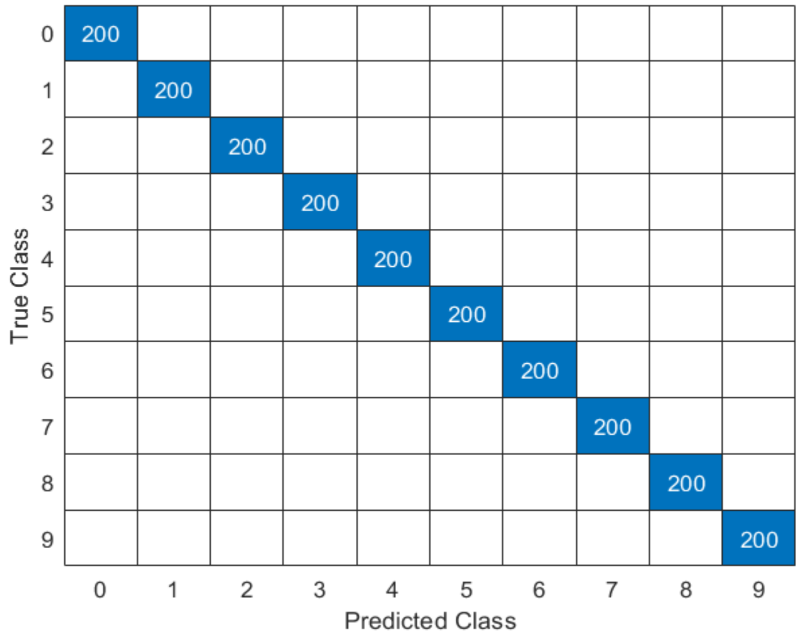
According to Table 3, the overall accuracies are 95.76 %, 85.4 %, and 100% for Pistachio, CIFAR-10, and MNIST datasets respectively. When we analyze other performance metrics for the Pistachio dataset, we view that the minimum metric (sensitivity) rate is 95.7%. Considering these performance results, we can express that OzNet has high successful classification output.

When we examine other performance metrics for the CIFAR-10 dataset, we observe that the maximum metric (specificity) rate is 98.38% and the minimum metric rate is 85.4%. However, OzNet's performance here is lower than other datasets. There could be several reasons for this situation. The first reason, the dataset has numerous images. The second reason, the images are not clear and have tiny in size. The last reason, the images are resized to 227x227. Although all negative situations, the performance of OzNet is very good.

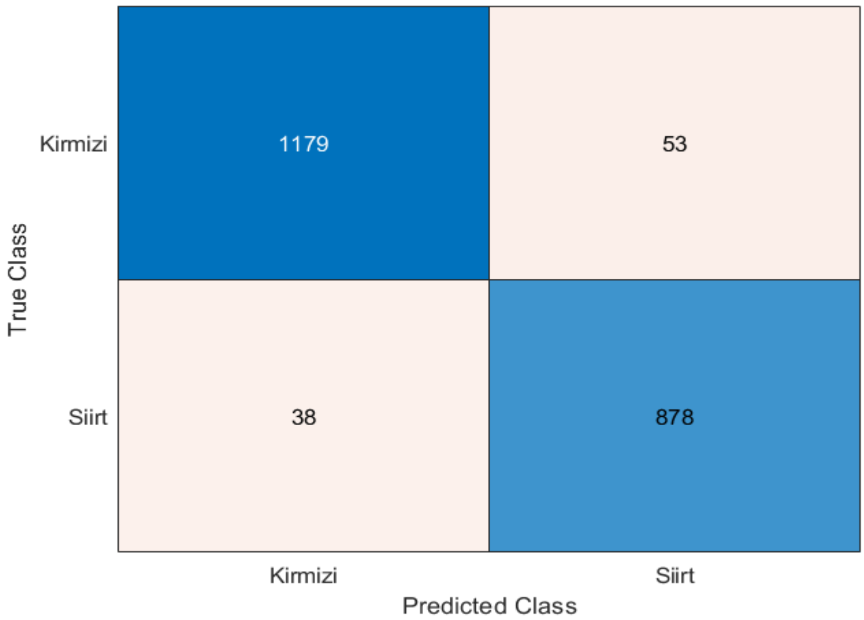
When we investigate other performance metrics for the MNIST dataset, we observe that all metric rates are 100%. As a consequence, we can state that OzNet performance is fairly well on these datasets. In addition, the confusion matrix is given for each dataset in Figure 5 and ROC Curves are exhibited in Figure 6.



a. CIFAR-10



b. MNIST



c. Pistachio

Figure 5. Confusion matrix for each dataset.

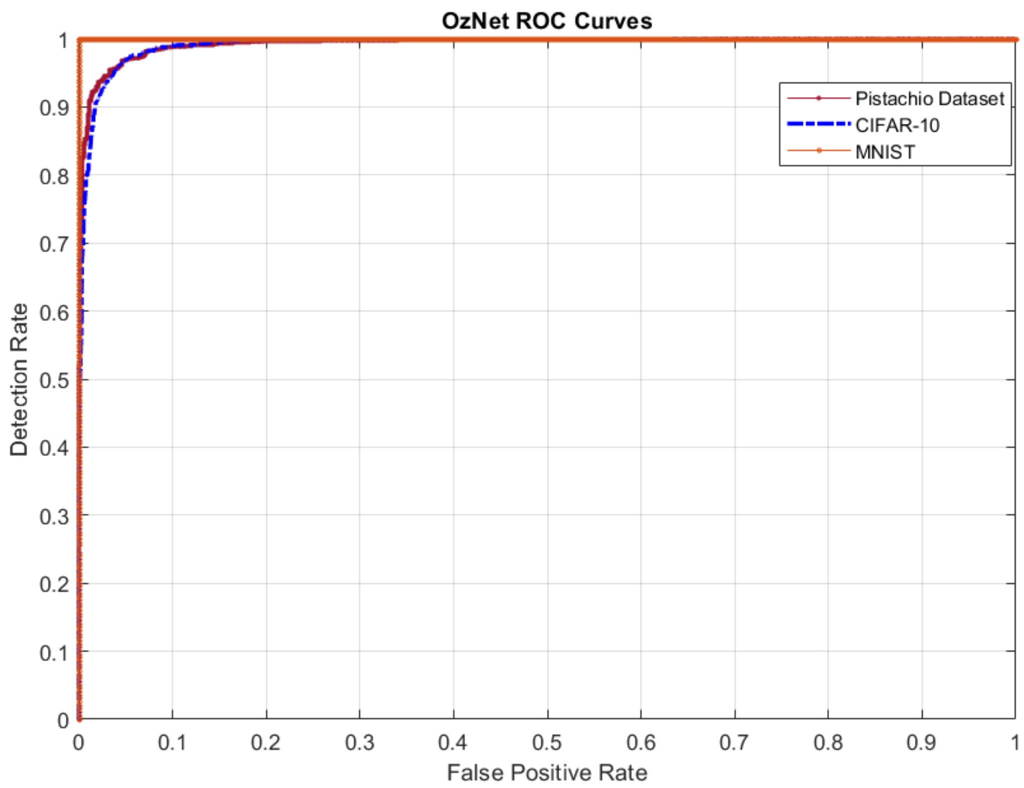


Figure 6. ROC Curves for three different datasets.

DISCUSSION AND CONCLUSION

Nowadays, deep learning algorithms are used popularly to solve classification problems. In this study, we also provide a deep learning algorithm to solve three different classification problems. This algorithm is one of the Convolutional Neural Networks architectures and we design that it has different parameters and layers, as detailed in Table 1. Additionally, we call it as OzNet architecture. As it is known, when a new architecture is designed, it is tested on frequently used datasets. Hence, we provide two widely used datasets: CIFAR-10 and MNIST. When the CIFAR-10 dataset is classified by using OzNet, we obtain an accuracy of 85.4%. Here, there are some limitations of the study. For instance, the dataset includes 60,000 tiny (32x32) images and we could not use 10-fold cross-validation owing to an insufficient GPU issue. Additionally, we resize all images to 227x227, and image qualifications decreased. Considering all these conditions, OzNet acquires very well results on the CIFAR-10 dataset. When the MNIST dataset is classified by using OzNet, we achieve an accuracy of 100%. The classification results on these datasets are highly successful, detailed in Table 3. Based on these good results, we consider a real-world problem of how OzNet is classification result. Therefore, we encounter the original dataset which is Pistachio. This dataset which is collected from Turkey has binary classes: Kirmizi and Siirt. When the Pistachio dataset is classified by using OzNet, we achieve an accuracy of 95.76%. According to all these results, we can express that OzNet is a very successful classifier.

In the next study, we will improve the classification results, particularly, for the CIFAR-10 and Pistachio datasets via created hybrid algorithms by using OzNet.

References

Arlot S, Celisse A, 2010. A survey of cross-validation procedures for model selection. Statistics surveys, 4, 40-79.

Baldominos A, Saez Y, Isasi P, 2019. A survey of handwritten character recognition with mnist and emnist. Applied Sciences, 9(15), 3169.

Bankman D, Yang L, Moons B, Verhelst M, Murmann B, 2018. An Always-On 3.8$\mu $ J/86% CIFAR-10 mixed-signal binary CNN processor with all memory on chip in 28-nm CMOS. IEEE Journal of Solid-State Circuits, 54(1), 158-172.

Bengio Y, LeCun Y 2007. Scaling learning algorithms towards AI. Large-scale kernel machines, 34(5), 1-41.

Bonifazi G, Capobianco G, Gasbarrone R, Serranti S, 2021. Contaminant detection in pistachio nuts by different classification methods applied to short-wave infrared hyperspectral images. Food Control, 130, 108202.

Cetin AE, Pearson T, Tewfik A, 2004. Classification of closed-and open-shell pistachio nuts using voice-recognition technology. Transactions of the ASAE, 47(2), 659.

Chauhan R, Ghanshala KK, Joshi R, 2018. Convolutional neural network (CNN) for image detection and recognition. Paper presented at the 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC).

CIFAR-10. Retrieved from https://www.cs.toronto.edu/~kriz/cifar.html

Darlow LN, Crowley EJ, Antoniou A, Storkey AJ, 2018. Cinic-10 is not imagenet or cifar-10. arXiv preprint arXiv:1810.03505.

Doon R, Rawat TK, Gautam S, 2018. Cifar-10 classification using deep convolutional neural network. Paper presented at the 2018 IEEE Punecon.

He K, Zhang X, Ren S, Sun J, 2016. Deep residual learning for image recognition. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Adam H, 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, 2017. Densely connected convolutional networks. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K, 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360.

Indolia S, Goswami AK, Mishra SP, Asopa P, 2018. Conceptual understanding of convolutional neural network-a deep learning approach. Procedia computer science, 132, 679-688.

Kayed, M., Anter, A., & Mohamed, H. (2020). Classification of garments from fashion MNIST dataset using CNN LeNet-5 architecture. Paper presented at the 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE).

Koklu M, Ozkan IA, 2020. Multiclass classification of dry beans using computer vision and machine learning techniques. Computers and Electronics in Agriculture, 174, 105507. doi:https://doi.org/10.1016/j.compag.2020.105507

Koonce B, 2021. Convolutional Neural Networks with Swift for Tensorflow: Apress Berkeley, CA.

Krizhevsky A, Hinton G, 2010. Convolutional deep belief networks on cifar-10. Unpublished manuscript, 40(7), 1-9.

Krizhevsky A, Sutskever I, Hinton GE, 2012a. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Krizhevsky A, Sutskever I, Hinton GE, 2012b. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

LeCun Y, Bottou L, Bengio Y, Haffner P, 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

LeCun Y, Cortes C, Burges CJC. The MNIST Database of handwritten digits. Retrieved from http://yann.lecun.com/exdb/mnist/

Lopez-del Rio A, Nonell-Canals A, Vidal D, Perera-Lluna A, 2019. Evaluation of cross-validation strategies in sequence-based binding prediction using deep learning. Journal of chemical information and modeling, 59(4), 1645-1657.

Omid M, Firouz MS, Nouri-Ahmadabadi H, Mohtasebi SS, 2017. Classification of peeled pistachio kernels using computer vision and color features. Engineering in agriculture, environment and food, 10(4), 259-265.

Özaltın Ö, Yeniay Ö, 2021. Ecg classification performing feature extraction automatically using a hybrid cnn-svm algorithm. Paper presented at the 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA).

Özkan İA, Köklü M, Saraçoğlu R, 2021. Classification of Pistachio Species Using Improved k-NN Classifier. health, 2, 3.

Park SH, Goo JM, Jo CH, 2004. Receiver operating characteristic (ROC) curve: practical review for radiologists. Korean journal of radiology, 5(1), 11-18.

Rajinikanth V, Joseph Raj AN, Thanaraj KP, Naik GR, 2020. A customized VGG19 network with concatenation of deep and handcrafted features for brain tumor detection. Applied Sciences, 10(10), 3429.

Saber A, Sakr M, Abo-Seida OM, Keshk A, Chen H, 2021. A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. IEEE Access, 9, 71194-71209.

Schott L, Rauber J, Bethge M, Brendel W, 2018. Towards the first adversarially robust neural network model on MNIST. arXiv preprint arXiv:1805.09190.

Sharifrazi D, Alizadehsani R, Roshanzamir M, Joloudari JH, Shoeibi A, Jafari M, Khozeimeh F, 2021. Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. Biomedical Signal Processing and Control, 68, 102622.

Simonyan K, Zisserman A, 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Singh D, Taspinar YS, Kursun R, Cinar I, Koklu M, Ozkan IA, Lee H-N, 2022. Classification and Analysis of Pistachio Species with Pre-Trained Deep Learning Models. Electronics, 11(7), 981.

Subasi A, 2012. Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines. Computers in Biology and Medicine, 42(8), 806-815.

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Rabinovich A, 2015. Going deeper with convolutions. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Taspinar YS, Cinar I, Koklu M, 2021. Classification by a stacking model using CNN features for COVID-19 infection diagnosis. Journal of X-ray science and technology (Preprint), 1-16.

Thakkar V, Tewary S, Chakraborty C, 2018. Batch Normalization in Convolutional Neural Networks—A comparative study with CIFAR-10 data. Paper presented at the 2018 fifth international conference on emerging applications of information technology (EAIT).

Vidyarthi SK, Tiwari R, Singh SK, Xiao HW, 2020. Prediction of size and mass of pistachio kernels using random Forest machine learning. Journal of Food Process Engineering, 43(9), e13473.

Xiao H, Rasul K, Vollgraf R, 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747.

Yang L, Bankman D, Moons B, Verhelst M, Murmann B, 2018. Bit error tolerance of a CIFAR-10 binarized convolutional neural network processor. Paper presented at the 2018 IEEE International Symposium on Circuits and Systems (ISCAS).

Zoph B, Vasudevan V, Shlens J, Le QV, 2018. Learning transferable architectures for scalable image recognition. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Acknowledgment

This study is motivated by O.Ozaltin's Ph.D. thesis.

Conflict of Interest

The authors have declared that there is no conflict of interest.

Author Contributions

Oznur Ozaltin considered this study and analyzed the datasets. Dr. Murat Koklu checked all results and supervised formal analysis, writing, and resources for the study. Aynur Yonar also created figures, checked all experimental results, and approved the manuscript. Ozgur Yeniay supervised the study and confirmed the final manuscript.

Classification of Breast Cancer Diagnosis with Machine Learning Algorithms

Aynur Yonar[1*], Harun Yonar[2], Öznur Özaltın[3]

[1] Selcuk University, Faculty of Science, Department of Statistics, 42031, Konya, Turkey
[2] Selcuk University, Faculty of Veterinary, Department of Biostatistics, 42031, Konya, Turkey
[3] Hacettepe University, Institute of Science, Department of Statistics, 6800, Ankara, Turkey

*Corresponding author e-mail: aynursahin@selcuk.edu.tr

Abstract

Breast cancer poses a high threat to women's health worldwide and causes the death of thousands of women every year. Therefore, the classification of benign and malignant breast cancer is vital. In this study, breast cancer classifications were first obtained with various Machine Learning (ML) techniques using features computed from digitized images of fine-needle aspiration of breast masses. Then, classifications were made again with only selected features from the genetic algorithm-based feature selection method. It was concluded that the accuracy rates of the ML techniques with the subset of features from the genetic algorithm were generally higher than those of baseline ML for the classification of breast cancer.

Key words: *Machine Learning, Classification, Genetic Algorithm, Breast Cancer*

The Complexity and Variation of Effect Size in Meta-Analyses

Harun Yonar[1*], **Aynur Yonar**[2]

[1] Selcuk University, Faculty of Veterinary, Department of Biostatistics, 42031, Konya, Turkey
[2] Selcuk University, Faculty of Science, Department of Statistics, 42031, Konya, Turkey

*Corresponding author e-mail: hyonar@selcuk.edu.tr

Abstract

A meta-analysis is an approach used to synthesize the results of individual studies through statistical methods. This method makes the literature knowledge about the researched subject measurable. It provides eliminates uncertainty by bringing together studies with conflicting results. The effectiveness of the studies examined in the meta-analysis can be revealed, and even studies with small effect sizes are brought together to provide a chance to achieve a large effect. In this respect, it contributes to creating different research designs in clinical studies. The meta-analysis is built on the concept of effect size, which directly impacts the results of heterogeneity and publication bias. Examining the effect size at different steps leads to some complexities. In this study, the individual effect size effect on overall effect size is examined through scenarios developed on the factors of the number of samples and the number of studies to examine these complexities. The results of different combinations are discussed comparatively. A strategic design is proposed to control the study design with comparative results, determination of the need for subgroup analyses, homogeneity, and publication bias reviews.

Key words: *Meta-Analyses, Effect Size, Heterogeneity, Publication Bias.*

Türkiye'de Farklı Finansal Okuryazarlık Seviyesindeki Yatırımcıların Kripto Paraya Karşı Tutumları

Deniz Varer[1], Öykü Yücel[1*], Zafer Yılmaz[1]

[1] TED Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, 06420, Ankara, Türkiye

*Corresponding author e-mail: oyku.yucel@tedu.edu.tr

Özet

Kripto para piyasası internet kullanımının yaygınlaşmasıyla hızlanan finansal entegrasyon süreciyle birlikte ülkemizde kendi dinamikleri çerçevesinde gelişmekte olan bir piyasadır. Yaygın bilinen Bitcoin, Ethereum gibi farklı birçok çeşidi olan kripto paralar yatırımcılarına kullanım alanlarının genişliği ve işlem maliyetlerinin düşük olması gibi avantajlar sağlamaktadır. Bu özellikleriyle geleceğin ekonomisi olma potansiyeli taşıyan kripto para piyasasındaki yatırımcıların tutumlarının demografik ve finansal okuryazarlık bilgileri çerçevesine incelenmesi yeni gelişmekte olan bu alana ışık tutması açısından önemlidir. Finansal okuryazarlık bilgisi finansal göstergeleri doğru değerlendirerek finansal piyasalarda planlama ve yatırım yapabilme yeteneğidir. Bu çalışmanın amacı farklı finansal okuryazarlık seviyesine sahip kişilerin kripto para farkındalık düzeylerinin araştırılması ve kripto para yatırımına olan tutumlarının anket çalışması yöntemiyle yatırım, güvenilirlik ve yenilik faktörleri bağlamında incelenmesidir. Anket çalışması Google forms aracılığıyla uygulanmıştır. Bu çalışmada olasılıklı olmayan örneklem tekniği uygulanmıştır. Örneklem sayısı %95 güven aralığında 0.05 standart sapmayla hesaplanmıştır. 384 katılımcıya ulaşılmıştır. Betimsel istatistik analizleri ve ANOVA testi SPSS programı aracılığıyla gerçekleştirilmiştir. Analiz sonucunda elde edilen bulgular doğrultusunda ülkemizdeki yatırımcıların kripto para piyasasına karşı tutumlarının farklı finansal okur yazarlık seviyelerinde anlamlı farklılıklar gösterip göstermediği değerlendirilmiştir. Güvenilirlik faktörü için finansal okuryazarlık seviyesinden bağımsız olarak kripto para piyasası için yüksek risk algısı olduğu görülmüştür. Bütün katılımcıların Bitcoin kripto parasını bildikleri anlaşılmıştır. Orta ve yüksek düzeyde finansal okuryazarlık seviyesine sahip katılımcıların kripto para piyasasına daha fazla yatırım yaptıkları anlaşılmıştır. Yenilik faktörü çerçevesinde kripto paraların gelecekte yaygın olarak kullanılacağının düşünüldüğü ve finans piyasasına yenilik getirdikleri algısı ortaya çıkmıştır. Yatırım faktörü kapsamında katılımcıların kripto para piyasalarının getiri ve risk düzeyini borsa işlemlerine benzer algıladıkları görülmüştür. Finansal okuryazarlık seviyesinden bağımsız olarak genel itibariyle kripto para piyasası yatırım işlemlerinin karmaşık olduğu düşünülmektedir. Yüksek finansal okuryazarlık seviyesindeki katılımcıların yüksek risk ve operasyonel zorluklar algısı nedeniyle kripto para piyasalarına yatırım yapmaktan kaçındıkları bulgusu doğrultusunda menkul kıymet piyasalarında olduğu gibi yönetici otoriteler tarafından uygulanacak düzenleyici kurallar ve uygulamarın geliştirilmesiyle kripto para piyasasının işlem

Key words: *Bitcoin, Kripto para, Finansal okuryazarlık*

Factors Affecting Consumer Purchasing Behavior of Environmentally Friendly Car

Hazal Vural[1], Gizem Çelik[1*], Zafer Yılmaz[1]

[1] TED University, Faculty of Economics and Administrative Sciences, Department of Business Administration, 06420, Ankara, Turkey

*Corresponding author e-mail: gizem.celik@tedu.edu.tr

Abstract

A green vehicle or clean vehicle is a vehicle that has less destructive impacts on the environment than other conventional internal combustion engine vehicles which use gasoline or diesel. The transport sector plays a critical role in ensuring sustainable environmental growth. With the global climate action taken against global warming and climate change, increasing pollution caused by the transportation sector due to CO2 emissions raises the importance of green vehicle use all around the world. Electric vehicles are one of the most widespread eco-friendly types of green vehicles. At the scope of the 2030 Agenda for sustainable development by the United Nations and global action taken through climate change, most municipalities and car manufacturing companies focus on electric vehicles. Also, this action affects the purchasing behavior of consumers with the consciousness of the environment and sustainability issues and there is an increasing demand for electric vehicles. This study aims to investigate the factors affecting consumer purchasing behavior of the private car and the environmental consciousness of consumers about energy consumption and carbon dioxide emission and examine the consumer purchasing behavior of electric cars by comparing the environmental effects of traditional vehicles and electric vehicles. Data is collected by conducting an online survey study through Google forms in Turkey. Within the scope of the research, data were collected using convenience sampling method from 384 participants. The survey data is analyzed through the SPSS-23 program and worked with a 95% confidence level. Then factors affecting consumer purchasing behavior toward environmentally friendly cars are measured through descriptive analyses and ANOVA test. As a result of the analysis made; it is observed that although the environmental consciousness and sustainability issues are not the considered enough in the purchasing process of private car, there is a high awareness about destructive impacts of transportation sector on the environment. In line with the findings, it is evaluated that both factors related to environmental friendly car such as price, design and attractiveness and brand related factor such as environmental consciousness and sustainability actions have impact on purchasing behavior of consumers in Turkey toward environmentally friendly cars.

Key words: *Consumer Behaviour, Purchasing Decision Process, Environmental Consumer Purchasing Behaviour, Electric Vehicle*

A Statistical Analysis of the System Parameters of a Fully Automated Parking System

Nurhan Dudaklı[1], Adil Baykaşoğlu[2*]

[1] Dokuz Eylül University, Engineering Faculty, Department of Industrial Engineering, 35397, İzmir, Turkey

*Corresponding author e-mail: nurhan.dudakli@deu.edu.tr

Abstract

An increase in the number of personal and commercial vehicles related to the rapidly increasing population all over the world, reveals parking lot problems, particularly in highly populated cities. At this point, especially in the last two decades, Fully Automated Parking Systems (FAPS) offer more efficient and eco-friendly solutions by utilizing new generation technology that employs automatic robots for parking. Although FAPS provide many advantages for the public and citizens, some operational issues should be improved. There are a limited number of studies that deal with planning, scheduling, and control of FSPS's operations. Therefore, academic research is in demand to increase the operational efficiency of FAPS. But, first of all, the system parameters should be determined and collected to develop proper methods and systems. Customer behaviour/profile is the most important factor that pressures on FAPSs since the FAPSs are in direct connection with the end-customers. However, collected raw data is not suitable for direct use. There is a need for additional analyses to produce proper data for the system. With this motivation, we collected real data from a FAPS located in Izmir/Alsancak to be used for this study. The provided data is utilised to extract statistical information such as the number of customers arriving in the FAPS, the arrival and departure patterns of the customers, and how long they parked their cars (parking durations) etc. This paper provides very important information about parameters that can be considered as inputs of any kind of FAPS model.

Key words: *Full Automated Parking System, Customers' parking behaviour, Goodness-of-fit Test, Arena software*

INTRODUCTION

FAPSs are state-of-the-art vehicle storage systems where robots perform lift up-set down, rotate, transport and allocation tasks. Therefore, they can be also termed robotic parking systems or robotic valet systems. There are different building shapes, transport systems and several possible slot (cell) layouts that can be arranged to maximize capacity and minimize retrieval time in FAPSs. Even though it is possible to provide several classifications based on the characteristic of the system and equipment used in the system, a classification mainly can be made based on the carriers that are used to transport and allocate the vehicles such as lifts, AGV, cranes, etc. (GIVT, 2022; Westfalia, 2022; Fata Automation, 2020; Automated Robotic Parking, 2016).

FAPSs are special cases of warehouses that are using automated storage/retrieval systems (AS/RS). For this reason, intrinsically FAPSs are similar to fully automated warehouse systems in terms of design characteristics. In both systems, the entities to be stored for a certain time are taken from an input-output (entry-exit) point and allocated to a predetermined location by automated carriers. The retrieve operation is also carried out by automated carriers and ended at the input-output point after transportation. All operations are performed without the need for an operator in a fully automated manner (Roodbergen & Vis, 2009).

However, it can be expressed that the most important point where FAPSs and warehouses are inherently separated from each other is the control of entities and events in these systems. For example, there is complete or partial control over how many products/materials when will arrive in the warehouse, how long they will be kept on the floor or the shelves when the products will be picked and shipped (Bartholdi & Hackman, 2011; Manzini 2012; Ertek, 2012). However, it is not possible to talk about the same situation in FAPSs. There is no chance to decide how many vehicles when will arrive in the system, how long each vehicle will be parked and when they will depart. At this point, we can summarize the most important contribution of the paper is managing this uncertainty in the most ideal way and providing an adequate level of control on the system parameters.

In addition to being an important source of uncertainty, the customer structure/profile is also one of the most important factors that pressure on FAPSs. Unlike warehouses, FAPSs are in direct connection with the end-customers. This situation makes it impossible to compensate for some mistakes in terms of customer satisfaction. The customer's request must be responded to properly and quickly as soon as it is realized. Moreover, with the fact that each vehicle in the system corresponds to a customer, and considering the nature of the parking process with the density within the cities, it is necessary to serve a lot of customers in a very short time. In the case of this dynamic structure and the uncertainty, it is very difficult to perform operations effectively and efficiently. With this motivation, real data that is obtained from a FAPS located in Izmir/Alsancak enables the extraction of statistical information such as the number of customers arriving in the FAPS, and the arrival and departure patterns of these customers, and how long they parked their cars (parking durations). This paper provides some important information and inputs for all kinds of models aiming to propose and analyze new planning methods and approaches for FAPSs.

MATERIAL AND METHODS

New methods and algorithms that address operational planning and control problems, which improve the performance and operational efficiency FAPSs are in high demand. The nature of the problem is depicted in Figure 1 (Dudaklı and Baykasoğlu, 2020). As it is emphasized by Dudaklı and Baykasoğlu (2020), the service times and system performance are primarily affected by operational decisions. However, the findings also show that operational problems and system performance are greatlyinfluenced by the structure of the building, the established transport system, and customer profile.

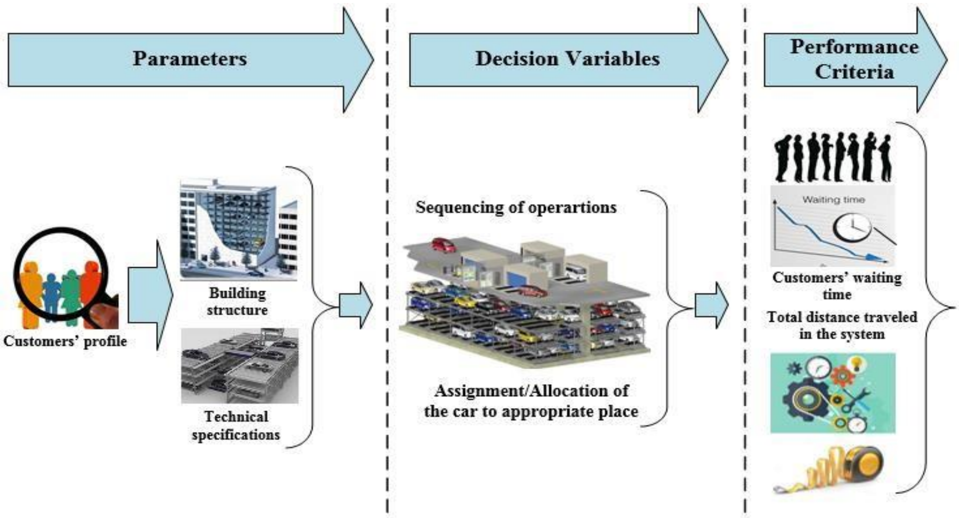


Figure 1. The general structure of the problem

The decision-maker has control over building structure and technical specification in the design phase of the system. However, there are uncontrollable parameters and thus should be considered when both designing and operating the system. One of these parameters is the customer profile, which is related to customer density and customer behaviours such as the arrival/departure patterns, and parking duration.

Considering the importance of the system parameters, it becomes more critical to take into account their deterministic and stochastic natures when providing inputs to the analytical, mathematical and simulation model of the system. Therefore, the following section and subsections provide detailed information about the system parameters of FAPSs.

System Parameters of FAPSS

It is possible to examine the parameters related to FAPSs in two categories as deterministic and stochastic system parameters. For the first group of parameters, the related values are (mostly) certain, and the acquisition of the data is relatively easy. However, the second group of parameters is difficult to collect and the collected raw data is not suitable for direct use. There is a need for additional analyses to produce proper data for the system. Hence, the parameters of the FAPS considered in the experimental study are examined in two subsections as follows.

Deterministic System Parameters

The aforementioned physical and technical features of the system are the members of the first group parameters. The size and layout of the slots and building, the speed of the transporters and other mechanical equipment, and the capacity of the system (the number of the slots) are parameters that should be identified properly. These kinds of parameters have deterministic values that are decided at the design and installation phase. The other design assumptions related to the system considered for theexperimental study can be summarized as follows. The FAPS has:

- ✓ Rectangular design
- ✓ Multi-storey structure
- ✓ Single entrance/ exit point
- ✓ Single aisle
- ✓ Single depth slot layout
- ✓ Single transporter (robot) moving vertically on the z-axis and horizontally on the x-axis
- ✓ Conveyors move horizontally on the y-axis.

Stochastic System Parameters

In addition to the physical and technical data, data including required information about the customer should also be provided before investigating system performance. The customer structure is already discussed as the main source of uncertainty and the highly dynamic nature of the system. Therefore, a detailed customer analysis should be carried out when designing the system and setting technical parameters. An accurate analysis is a key factor that leads to the efficient design and operating system.As a result of the analysis, important parameters like the daily number of customers, arrival and departure pattern of customers, and the parking duration of customers should be mined from raw customers' data.

The performance of FAPS highly depends on customers' behaviour. All parameters in question are not deterministic and have stochastic nature. The number of customers, arrival and departure times parking durations of the customers take probabilistic values within a day (Guo, et al., 2016). Therefore, to construct reliable models of the system and ultimately to give credible results, actual data about the customer's behaviour should be obtained and analysed. With this motivation, real data is provided froma FAPS located in Izmir/Alsancak. The provided data consists of customer records belonging to the first3 months (January, February, March) of 2017 and the following statistical information is extracted fromthese records.

The number of customers

1. The arrival and departure patterns of these customers, and
2. How long do they park their cars (parking durations)

Primarily, these records are filtered and cleaned to ensure proper data sets which enable to carry out overall analyses and consequently statistical analyses. Input Analyser of Arena® Simulation Software isutilized to determine the relevant statistical distributions of the customer profile. The software applies different goodness-of-fit tests like Chi-Square and Kolmogorov-Smirnov to the given data file and determines the distribution that will best fit to data. All applicable distributions are fitted to the data and then the distributions are sorted based on their corresponding square errors (Arena®, Chung, 2003; Gingu, & Zapciu, 2015). In this way, the most suitable distribution functions for the case of Izmir are discovered. The details of all analyses are clarified in the following subsections.

Analysis of Customer Numbers

During the time period (3 months) when the data was recorded, the system was used by more than 30,000 customers. Daily customer arrivals (demand) in the examined data varied from 92 to 578. This great variability leads to the investigation of the cause of this situation and to find out if variability depends on some variables. The first analysis carried out on charts showed that there is no great difference between months in terms of the number of customers. However, the same data when visualized depending on the day of the week indicates that the number of customers differentiates according to thedays of the week. Moreover, a similar week pattern is observed within the three months, except for verylimited cases. The next step is statistically analysing the number of customers varied on days of the weekto determine probability distribution functions that belong to each day.

By the way, it is proved whether there are meaningful statistical differences between the number of daily arrivals. With the help of the Input Analyser of Arena® software, it is possible and easy to determine the best distribution function that represents the data set belonging to each day. The software applies thegoodness of fit tests for all possible distributions and suggests the best-fit distribution function of the data based on square error. The detail of best-fit functions is introduced in Table 1.

Table 1. Distribution functions of customer numbers of the FAPS

	Distributions' Parameters	Distribution Function	[3]SSE
Monday	Triangular $f(x; a, b, c)$	TRIA(294, 383, 419)	0.056548
Tuesday	[1]Beta $f(x; \alpha, \beta)$	139 + 301 * BETA(1.25, 0.413)	0.057014
Wednesday	[1]Beta $f(x; \alpha, \beta)$	299 + 142 * BETA(0.884, 0.563)	0.044574
Thursday	Triangular $f(x; a, b, c)$	TRIA(299, 427, 482)	0.016156
Friday	[1]Beta $f(x; \alpha, \beta)$	282 + 253 * BETA(1.09, 0.414)	0.044131
Saturday	Triangular $f(x; a, b, c)$	TRIA(387, 559, 578)	0.083642
Sunday	[2]Normal $f(x; \mu, \sigma)$	NORM(177, 49.9)	0.045688

[1] $Beta\ distribution: 0 \leq x \leq 1\ and\ \alpha, \beta > 0$
[2] $Normal\ distribution: \mu = mean, \sigma = standart\ deviation$
[3] $SSE: Sum\ of\ Square\ Error$ = {fi − f(xi)}²

The distribution functions are produced by using 12 samples (12 weeks' data) of each day to represent the variance of customer numbers depending on the days of the week. Based on the given data, the number of customer arrivals fits the triangular distribution (Monday, Thursday, Saturday) and the beta distribution (Tuesday, Wednesday, Friday) for six days of the week, and only Sunday fits the normal distribution. Parameters of all distribution functions are provided by the software. The statistical test (Kolmogorov-Smirnov goodness-of-fit) results indicate a good fit for all cases according to acceptable SSE values. SSE is defined as the sum of {fi - f(xi)}² for overall histogram intervals and primarily represents the quality of a curve fit (Arena®). Thus, these distribution functions could be utilized in thesimulation model to generate different customer numbers for ensuring the validity of the experimental study and its results.

Analysis of Arrival and Departure Pattern of Customers

As stated before, customer arrivals and departures are not balanced in a day due to rush hours. Based onthe commute to work, mostly, the number of parking requests is higher than retrieval requests in early the morning, and vice versa in towards evening. This situation is valid for the downtowns where the carparking problem is an important issue. Of course, there are exceptions to this behaviour due to residentssettled near the FAPS and who park their cars after work. Considering all these situations, the density of requests within a day should be analysed carefully.

In order to statically discover underlying patterns and trends, the relevant data set is also analysed in Arena® Input Analyser. However, a reminder is needed before starting statistical analyses about the arrival and departure pattern of the customer. Unlike the queueing theory, the inter-arrival times are notconsidered for the case of FAPSs since it is not possible to observe a homogeneous customer distributionspread all the day, consequently a single inter-arrival pattern for a day. Therefore, to be more practicable,the moment of the day is considered instead of inter-arrival times of customers and the number of requests that arrive at those moments are analysed in this study. The detail of best-fit functions is presented in Table 2.

Table 2. Distribution functions of arrival patterns within the days

	Distributions' Parameters	Distribution Function	[3]SSE
Monday	[1]Gamma $f(x; k, \theta)$	2 + GAMM(11.5, 1.02)	0.003987
Tuesday	[2]Beta $f(x; \alpha, \beta)$	7 + 16 * BETA(2.13, 2.43)	0.004665
Wednesday	[1]Gamma $f(x; k, \theta)$	GAMM(13.7, 1.04)	0.004272
Thursday	[2]Beta $f(x; \alpha, \beta)$	24 * BETA(6.07, 4.11)	0.006749
Friday	Triangular $f(x; a, b, c)$	TRIA(7, 11.3, 24)	0.004850
Saturday	[1]Gamma $f(x; k, \theta)$	-0.001 + GAMM(2.28, 6.69)	0.008231
Sunday	Triangular $f(x; a, b, c)$	TRIA(-0.001, 19, 24)	0.030448

[1] $Gamma\ distribution\colon\ k, \theta > 0\ \ where\ k\ shape, \theta\ scale\ parameters\ and\ k * \theta = \mu$ (mean)
[2] $Beta\ distribution$: $0 \leq x \leq 1\ and\ \alpha, \beta > 0$ where α, β are shape parameters
[3] $SSE\colon Sum\ of\ Square\ Error = \{\mathrm{fi}\ - \mathrm{f(xi)}\}^2$

It can be seen from Table 2 that all distribution functions have a good-fit according to SSE values. As explained above, the parameters and the predicted variables of the distribution functions are related to continuous moments between (00:00-24:00) of the day. Also, another remarkable point is that in additionto the same distributions having different parameters, there is a wide variety of arrival distribution patterns like Gamma, Beta, and Triangular distribution patterns. This is another convincing result that proves that the customer attributes should be reflected according to the days of the week. The same analysis is conducted for the data that contains departure information of the customers. The detail of best-fit functions is introduced in Table 3.

Table 3. Distribution functions of departure patterns of the days

	Distributions' Parameters	Distribution Function	[3]SSE
Monday	[1]Beta $f(x; \alpha, \beta)$	-0.001 + 24 * BETA(4.58, 2.29)	0.005597
Tuesday	[2]Normal $f(x; \mu, \sigma)$	NORM(16.4, 4.41)	0.003739
Wednesday	[2]Normal $f(x; \mu, \sigma)$	NORM(16.1, 4.41)	0.015429
Thursday	[2]Normal $f(x; \mu, \sigma)$	NORM(16.1, 4.43)	0.006464
Friday	[2]Normal $f(x; \mu, \sigma)$	NORM(15.5, 5.88)	0.011663
Saturday	[1]Beta $f(x; \alpha, \beta)$	-0.001 + 24 * BETA(0.983, 0.789)	0.015944
Sunday	Triangular $f(x; a, b, c)$	TRIA(-0.001, 19, 24)	0.030448

[1] $Beta\ distribution$: $0 \leq x \leq 1$ and $\alpha, \beta > 0$ where α, β are shape parameters
[2] $Normal\ distribution$: $\mu = mean, \sigma = standart\ deviation$
[3] SSE: $Sum\ of\ Square\ Error = \{fi - f(xi)\}^2$

Analysis of Parking Durations

Parking duration refers to how long the customers leave their cars in the system. On one hand, this parameter is the key factor that shapes the departure pattern of the customer. On the other hand, this attribute highly affects the efficiency of allocation decisions and consequently the operational efficiencyof the system. Although there are a few cases where durations are more than 20 hours, it is seen that thevast majority of the customers park their cars for less than 10 hours. Moreover, most of these customersdo not exceed 5 hours of parking duration. When these patterns are analysed statistically based on daysof the week the following probability distribution functions are obtained. The distribution functions aregiven with their parameter values in Table 4.

As it can be seen in Table 4, there are two distribution functions, Lognormal and Gamma, that can represent parking duration patterns for all days of the week with quite low SSE values. The mean of thesamples (μ) and standard deviation (σ) are the parameters of Lognormal whereas a shape (k) and scale parameters (θ), where $* \theta = \mu$, are used in Erlang. As a practical interpretation, the discovered functionsshow that the mean parking duration and the variance of these durations are varied by day.

Table 4. Distribution functions of parking durations of customers depending on days of the week

	Distributions' Parameters	**Distribution Function**	**[3]SSE**
Monday	[1]Lognormal $f(x; \mu, \sigma)$	LOGN(2.96, 3.26)	0.004684
Tuesday	[1]Lognormal $f(x; \mu, \sigma)$	LOGN(2.79, 2.75)	0.001334
Wednesday	[2]Gamma $f(x; k, \theta)$	GAMM(1.52, 1.69)	0.013259
Thursday	[1]Lognormal $f(x; \mu, \sigma)$	LOGN(2.6, 2.97)	0.000498
Friday	[1]Lognormal $f(x; \mu, \sigma)$	LOGN(2.91, 2.58)	0.003551
Saturday	[2]Gamma $f(x; k, \theta)$	GAMM(1.77, 1.92)	0.011073
Sunday	[1]Lognormal $f(x; \mu, \sigma)$	LOGN(3.28, 3.09)	0.004846

[1] $Lognormal$: $0 \leq x$ and $\mu = mean, \sigma = standart\ deviation$
[2] $Gamma\ distribution$: $k, \theta > 0$ $where\ k\ shape, \theta\ scale\ parameters\ and\ k * \theta = \mu$ (mean)
[3] SSE: $Sum\ of\ Square\ Error = \{fi - f(xi)\}^2$

The real data set containing customer information was attained and statistically analysed to extract the stochastic customer behaviours. All these efforts to analyse customer behaviour statistically are carriedout to provide the reliable inputs for the simulation model that would be used for the investigation of FAPSs. Thus, in the next section, hypothetic FAPS is presented and simulated to verification of customerbehaviours that are statistically defined.

RESULTS

A simulation model is constituted to validate the statistical models of customer behaviour in this section. The developed simulation model and its important components are represented in Figure 2 Based on the problem structure, on one hand, physical and technical features of the considered FAPS and detail of the customer profile constitute inputs of the simulation model.

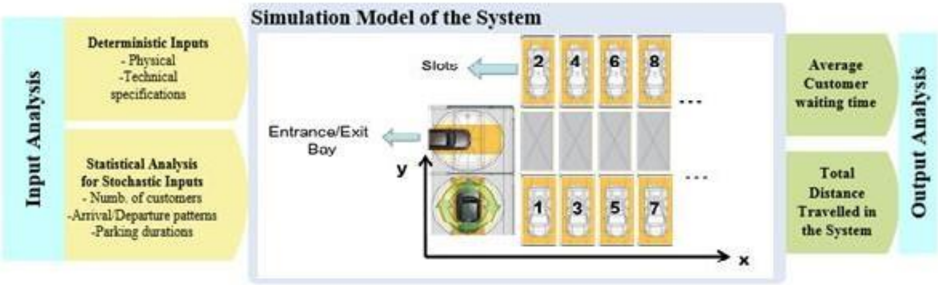


Figure 2. The developed simulation model and its components

The floor structure represented in Figure 2 might help to comprehend the system's physical-technical details. The horizontal moves carried out on each floor are considered on the (x-y) axis whereas the vertical moves between floors are considered on the z-axis. In this case, the transporter is responsible for respectively vertical and horizontal transportation on the z-axis and x-axis. Once the transportation on the x-axis is ended, the conveyors execute the horizontal transport on the y-axis. Ignoring the acceleration and deceleration, it is assumed that the speed of the transporter and conveyors on the corresponding axis is constant and the same as 0.5 meters in a second. Finally, the capacity of the hypothetic tower type FAPS is 100 slots, namely 100 cars. The considered FAPS consists of 10 floors and each floor consists of 10 slots.

The simulation is probably the most convenient tool for analysing dynamic and stochastic systems like FAPSs. However, at this point, there is an important decision about the simulation modelling approach. There is no doubt that discrete event simulation is suitable for FAPSs as variables change in discrete times and by discrete steps. In addition, there are two simulation modelling approaches depending on the characteristics of the real system and the purpose of the simulation study. These approaches are terminating or non-terminating simulation modelling. Terminating systems have initial starting conditions, a natural terminating event. On the other hand, in general, non-terminating systems can either have a terminating event but keep entities in the system between time periods, and not have a terminating event and run continuously (Chung, 2003). The other features of the simulation model developed in this study are summarized as follows:

- The non-terminating simulation model of the system is developed.
- The capacity of the considered system and the density of the customers are balanced. That means customer arrivals generally do not exceed the capacity at any time.
- The customers waiting for parking in the queue do not leave the system or they are not rejected unless the capacity is full.
- Stochastic parameters of the model, which are related to customers, are varied depending on the days of the week.

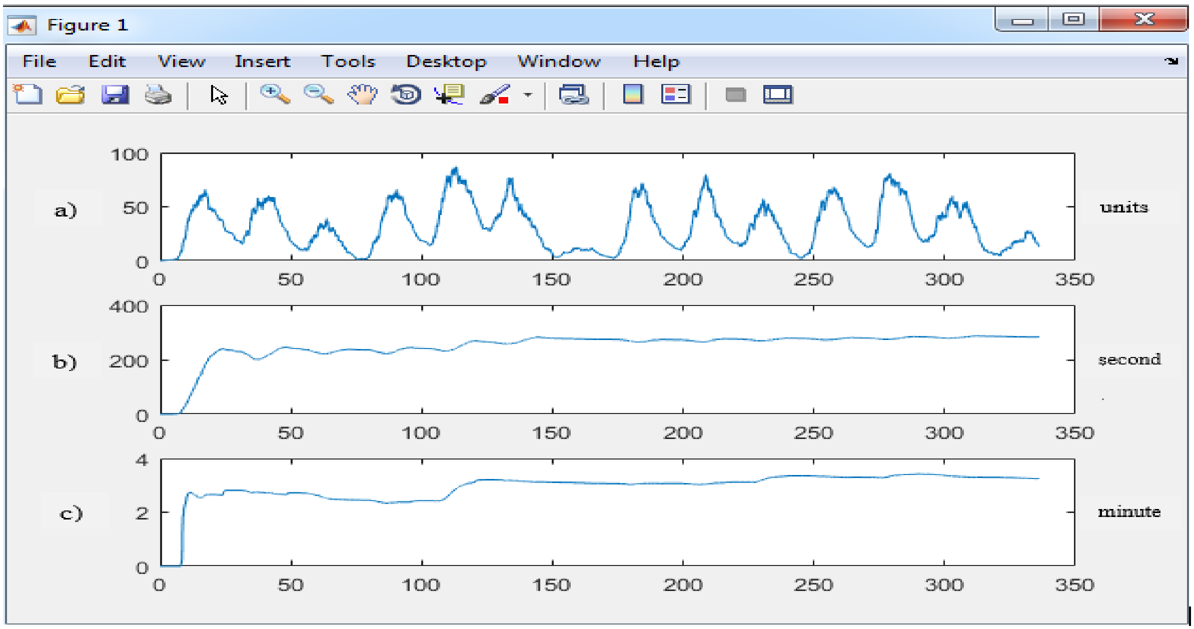


Figure 3. An overview of the system at the end of the 14 days simulation run: a) the number of cars inthe system, b) the average time spent per operation, c) the average customer waiting time per operation

Based on these facts and assumptions, the simulation model is developed in MATLAB. After that, the model is run for 2 weeks (14 days) time period, to check the pattern of system parameters and verify the simulation model. There is an assumption about the number of customers in our experimental study. The number of the customers that will arrive in our hypothetic system is assumed as half of the real system because the system considered in this study has the half capacity of Alsancak FAPS. Except forthat situation, all other customer parameters are used exactly as discovered.

The state variables that denote the number of cars in the system (a); the average time spent per operation(b); and the average customer waiting time per operation (c) is depicted hourly (0-350) through the simulation run in Figure 3. In addition to the number of cars at a certain time in the system, Figure 3 (a)also indicates arrival and departure patterns. As stated before, customer arrivals and departures are not balanced in a day due to rush hours. Based on the commute to work, mostly, the number of parking requests is higher than retrieval requests in early the morning, and vice versa in towards evening. In addition to the pattern that depends on work hours, a different parking and retrieval pattern arises on Friday and Saturday nights because customers' behaviour is changed at the weekend. Parking requests are increased after 19:00 and some of the retrievals requests sag to after midnight between 00:00-04:00.On the other side, Sunday has a quite different pattern from the arrival and departure patterns observedon other days. In conclusion, the relevant graph (Figure 3) shows that the customer behaviours are successfully reflected and modelled with help of the distribution functions discovered.

DISCUSSION AND CONCLUSION

This paper aanalyses the system parameters, the structure of the building, the established transport system, and customer profile, which greatly influence the performance and efficiency of FAPSs. Considering the importance of the system parameters, it becomes more critical to take into account theirdeterministic and stochastic natures. The decision-maker has control over the building structure and technical specifications in the design phase of the system. However, there is another uncontrollable parameter and thus should be considered when both designing and operating the system. This parameter

is the customer profile, which is related to the number of customers arriving in the FAPS, the arrival and departure patterns of these customers, and how long they parked their cars (parking durations).

At this point, as an important part of the study, customer structure behaviours are statistically analyzed.Real data is obtained from a FAPS located in Izmir/Alsancak. Primarily, these records are filtered and cleaned to ensure proper data sets, which enable to carry out of overall analyses and consequently statistical analyses. Input Analyser of Arena® Simulation Software is utilized to determine the relevant statistical distributions of the customer profile. The software applies different goodness-of-fit tests likeChi-Square and Kolmogorov-Smirnov to the given data file and determines the distribution that will best fit to data. The number of customers varied on days of the week, the arrival and the departure patternof the customer within the day, and parking duration refer to how long the customers leave their cars inthe system are analysed statistically to reveal proper distribution functions. The revealed distribution functions are embedded in a simulation model to ensure the validity and reliability of the results.

The real data set containing customer information was attained and statistically analysed to extract the stochastic customer behaviours. All these efforts to analyse customer behaviour statistically are carriedout to provide the reliable inputs for the simulation model that would be used for the investigation of FAPSs. Finally, it can be said that the interest in FAPSs is increasing worldwide, but the fact that the academic/scientific studies on this subject are limited thus there is still a need for work in this area.

References

Arena®, Access address: https://www.arenasimulation.com/search/eyJyZXN1bHRfcGFnZSI6InNlYXJjaCIsImtleXdvcmRzIjoiaW5wdXQgYW5hbHlzaXMifQ, Date of access: 06.05.2020

Automated Robotic Parking, 2016. Parking Equipment Types, Access address: http://www.automatedroboticparking.com/parking-equipment-types/ Date of Access: 20.04.2022.

Bartholdi JJ, Hackman ST, 2011. Warehouse & Distribution Science: Release 0.94. SupplyChain and Logistics Institute.

Chung CA, 2003. Simulation modeling handbook: a practical approach. CRC press.

Dudakli N, Baykasoğlu A, 2020. A research on fully automated parking systems and operational planning and control problems. Journal of the Faculty of Engineering and Architecture of Gazi University, 35(4), 2239-2254.

Ertek G, 2012. “Depolama Sistemleri (Warehousing Systems)”, Uluslar arası Lojistik, Anadolu Üniversitesi Yayınları, Açıköğretim Fakültesi Yayını No: 1593. Eds. Bülent Çatay and Gürkan Öztürk

Fata Automation, 2020. System Types. Access address: http://automatedparking.com/system-types/ Date of access: 20.04.2022.

Gıngu EI, Zapcıu M, 2015. Input Modellıng Usıng Statıstıcal Dıstrıbutıons and Arena Software. Annals: Series On Engineering Sciences (Academy of Romanian Scientists), 7(1), 63-70.

GIVT, 2022. Parking Access address: https://www.givt.de/index.php/en/mechan-automat-parksysteme-allg-en, Date of Access:20.04.2022.

Guo W, Zhang Y, Xu M, Zhang Z, Li L, 2016. Parking spaces repurchase strategy design viasimulation optimization. Journal of Intelligent Transportation Systems, 20(3), 255-269.

Manzini R, 2012. Warehousing in the global supply chain: Advanced models, tools and applicationsfor storage systems. Springer Science & Business Media.

Roodbergen KJ, Vis IF, 2009. A survey of literature on automated storage and retrieval systems. European journal of operational research, 194(2), 343-362.

Westfalia 2022. Parking systems, Access address: https://www.westfaliaparking.com/parking-systems/system-types/, Date of access: 20.04.2022.

Determination of Factors Affecting Bank Profits: Comparison of Multilinear Regression and Gradient Boosting Regression Methods

Ömer Faruk Rençber[1*], Abdurrahman Coşkuner[1]

[1] Gaziantep University, Economics and Business Faculty, Business Adm. 27150, Gaziantep, Turkey

*Corresponding author e-mail: dr.ofrencber@gmail.com

Abstract

Banks have a very important place in the financial system. Like traditional businesses, the main purpose of their activities in banks is to make a profit in a sustainable way. The management of financial or human resources in the banking sector is among the factors that directly affect bank profitability. However, which financial ratio affects profitability and at what level has always been a subject of curiosity in the literature. The aim of this study; The aim of this study is to examine the financial ratios that affect bank profitability in comparison with classical and data mining techniques. Accordingly, multiple linear regression analysis and gradient increasing regression tree methods were used in the study. Financially, profitability levels of banks; were analyzed with return on assets and return on equity ratios. The factors whose effects on profitability were investigated are as follows; consists of loans received/total assets, capital adequacy ratio, net interest income after provisioning/total assets, consumer loans/total loans, liquid assets/current liabilities, interest income/total income ratios. The study covers 26 banks operating in Turkey and the years between 2010 and 2020. As a result of the study, it has been concluded that the Gradient Enhancing Regression Tree method is more successful than the Multiple Linear Regression analysis method and the most important factors contributing to profitability are the capital adequacy ratio and the net interest income/total assets ratio after provisioning.

Key words: *Bank Profitability, Return on Assets Ratio, Return On Equity Ratio, Multiple Linear Regression Analysis, Gradient Boosting Regression*

A New Bivariate Copula and Its Properties

Rihab H.A. Abusaif[1], Buğra Saraçoğlu[2], Name Surname[3]

[1] Selçuk University, Institute of Natural and Applied Sciences, Department of Statistics, Phd Student, Konya, Turkey
[2] Selçuk University, Faculty of Science, Department of Statistics, Konya, Turkey

*Corresponding author e-mail: bugrasarac@selcuk.edu.tr

Abstract

Copulas are used to model the dependence between variables in a bivariate distribution and to obtaine new bivariate distributions. In this study, it is obtained a new bivariate copula by using arbitrary function in Ruschendorf method. We have proved that this new copula satisfies all properties needed in a copula. Furthermore, we have studied some properties such as the Kendall's τ, Spearman's ρ, tail dependence and stability for this new copula.

Key words: *Bivariate copula, Ruschendorf method, Stability, Tail dependence.*

Exploratory Data Analysis with Unsupervised Learning Algorithms for a Clinical Data Set

Harun Yonar[1*], Sadık Ibrahım Omar[1]

[1] Selcuk University, Faculty of Veterinary, Department of Biostatistics, 42031, Konya, Turkey

*Corresponding author e-mail: hyonar@selcuk.edu.tr

Abstract

Exploratory Data Analysis in data mining is an approach to analyzing data sets and obtaining information with visual methods. In the model-building stage, transforming the multidimensional structure into a more functional form, revealing unpredictable relationships, and recognizing the data are the prominent features of exploratory data analysis. In this sense, unsupervised learning algorithms include exploratory approaches that work on data and help solve complexity. This study investigates data exploratory with unsupervised learning algorithms in a complex clinical data set. For this purpose, unsupervised learning algorithms' clustering, association, and dimension reduction tasks are conducted as a preliminary for multivariate statistical methods.

Key words: *Data Mining, Exploratory Data Analysis, Unsupervised Learning Algorithms, Dimension Reduction, Clinical Data*

Estimation of the Factors Effecting Lactation Milk Yield of Holstein Cattle by the Adaptive Neuro-Fuzzy Inference System

Aycan Mutlu Yağanoğlu[1*]

[1] Ataturk University, Agriculture Faculty, Department of Animal Science, 25240 Erzurum, Turkey

*Corresponding author e-mail: myagan@atauni.edu.tr

Abstract

The aim of this study is to estimate lactation milk yield of Holstein cattles using first calving age, lactation period and service period with (ANFIS) adaptive neuro-fuzzy inference system. The input variables for the system in the study were first calving age, lactation period and service period. The output variable from the system was lactation milk yield. Predicted values which obtained from the ANFIS were compared with the observed values. 27 rule functions are used to obtain the best model and 1000 epochs are used to estimate the accuracy of the training and testing error. The relations between the output and input variables are showed with 3D graphics. R^2, RMSE and MAPE evaluation criterias were used to check the accuracy of the system's estimations. As a result, ANFIS estimates of lactation milk yield were quite close to the observed values and a positive correlation was found between them. The results showed that ANFIS can be successfully applied for estimate the lactation milk yield.

Key words: *Anfis, Holstein, Milk yield, Lactation period*

ROC Analysis for the Development of Diagnostic Tests in Medical Researches: An Application in SPSS

Adnan Ünalan[1*]

[1] Niğde Ömer Halisdemir University, Faculty of Medicine, Department of Biostatistics, 51240, Niğde, Turkey

*Corresponding author e-mail: unalanadnan@gmail.com

Abstract

ROC (Receiver Operating Characteristics) analysis was first used during 2nd World War for the analysis of radar signals before it was employed in signal detection theory. Then ROC analysis has been also used as a useful tool in the development of diagnostic test to new emerging diseases or if the current diagnostic tests are difficult or expensive to apply, especially in field studies. It is well known that diagnostic tests play an important role especially in the diagnosis of diseases for dichotomous outcome (positive/negative test results) in in medical research. ROC curve is a graphical plot that illustrate sensitivity (true positive rate: the ability of a test to correctly identify patients with a disease) on the Y axis and 1-specificity (specificity is true negative rate: the ability of a test to correctly identify people without the disease) on the X axis which are obtained from ROC analysis for different cut-off values. AUC (Area Under the Curve) is one of the mostly widely used metrics for model that area (probability of the curve) under the ROC curve and X-axis on the graph. When AUC approached to 1, then the classifier is able to perfectly distinguish between all the positive and the negative class points correctly. In this study, an SPSS application of ROC analysis was introduced in the development or use of diagnosis test for an appropriate data set related to medical research and the points to be considered in the interpretation of the analysis stages and results are emphasized.

Key words: *Medical researches, Diagnostic tests, ROC Analysis, AUC, SPSS Application*

Approximation Results for Adjustment Coefficient of Nonlinear Lundberg Risk Model

Tahir Khaniyev[1*], Zulfiye Hanalioglu[2], Basak Gever[3]

[1] TOBB University of Economics and Technology, Engineering Faculty, Department of Industrial Engineering, 06560, Ankara, Turkey
[2] Karabuk University, Faculty of Business, Department of Actuarial Sciences, 78050, Karabuk, Turkey
[3] University of Turkish Aeronautical and Association, Engineering Faculty, Engineering Faculty, 06790, Ankara, Turkey

*Corresponding author e-mail: tahirkhaniyev@etu.edu.tr

Abstract

Insurance firms are essential to the functioning of modern economies and states. In actuarial science, the examination of risk and ruin problems in insurance companies is critical. The premium income function $p(t)$ is commonly modeled as a linear function in the literature. However, because an insurance company's premium income does not always increase linearly, the linear model is not always accurate. This is especially true for insurance companies that are overburdened with policyholders. As a result, it's plausible to suppose that premium income is depicted as a function with a decreasing rate of growth over time, even if the function is monotonically rising. For this reason, in this study, a more realistic non-linear Cramér-Lundberg risk model is mathematically constructed and investigated, which is given as $V(t)=u+c\sum_{i=1}^{v(t)} g(W_i)-S(t)$. Here, $g(t)$ is monotonically increasing non-linear function whose rate of growth decreases with time; $W_i, i=1,2,3\ldots$ are positive-valued independent and identically distributed random variables describing inter-arrival times of claims; u – the initial capital of the company; c – the premium rate; $S(t)=\sum_{i=1}^{v(t)} X_i$ is the renewal-reward process that expresses the insurer's capital outflow; $v(t)$ – the process counting the total number of claims. Moreover, $X_i, i=1,2,3\ldots$ are independent and identically distributed random variables denoting the amount of payment for i^{th} claim, for $i=1,2,3,\ldots$. In other words, $V(t)$ expresses insurance company's capital balance at any time t. The main purpose of this study is to investigate ruin probability of the non-linear risk model $V(t)$. For this aim, Lundberg type upper bound is obtained for the ruin probability of this non-linear risk model, when the income function is defined as $p(t)=c\sqrt{t}$. The mathematical development of the Cramér-Lundberg model under the assumption that the premium income function is a non-linear function is the study's main contribution. Another contribution is the discovery of a Lundberg type upper bound for the ruin probability of this non-linear risk process. The upper bound is calculated using the generalized Lundberg adjustment coefficient (r). It is generally recommended that the coefficient r be obtained from an integral equation. However, calculating the coefficient r from this integral equation is a difficult problem. The fastest computers sometimes require longer times to calculate the coefficient r. Therefore, in this study, using numerical methods for solving the integral equations, the relation between the adjustment coefficient r and the coefficient ρ is obtained. Here, $\rho=\frac{\text{The average income in a cycle}}{\text{The average expense in a cycle}}$.

Key words: *Cramer-Lundberg risk model, Ruin probability, Non-linear risk model, Lundberg adjustment coeffcient.*

Impact of IFRS9 Implementation on Credit Risk of the Banking Sector in Kosova

Agnesa Krasniqi[1], **Skender Ahmeti**[2*], **Muhamet Aliu**[3]

[1] University Of Prishtina, Faculty Of Economy

*Corresponding author e-mail: agnesakrasniqi@live.com

Abstract

The banking industry is estimated to be one of the most profitable sectors in Kosovo. Kosovo banks serve the citizens and businesses of Kosovo by offering a wide range of financial services. Kosovo banks make it possible for businesses to start and expand their activities, increase employment and provide better services to their customers. Their products and services include bank accounts, loans, local and international payments, bank cards, bank guarantees, letters of credit, e-banking As of January 2020, commercial banks in Kosovo must report to the Central Bank of Kosovo (CBK) according to the new standards International Financial Reporting Standards 9 (IFRS 9) approved by the International Accounting Standards Board (IASB). Their implementation by banks has been a challenge in itself and has brought changes in the financial statements mainly in loan loss provisions (LLP), where the way of their calculation has changed in general. The purpose of this research is to investigate the impact of IFRS 9 implementation on Credit Risk. The research was done through multivariate regression, where as a dependent variable we have Loan Loss Provision (LLP) and as independent variables, we have Loans, Non-Performing Loans (NPL,) and IFRS 9 as a dummy variable. Based on the linear regression model, it is found that only three independent variables have a significant impact, in the dependent variable.

Key words: *Commercial banks, Credit risk, LLP, IFRS 9*

A Study on the Use of the Relationship Between Somatic Cell Count (SCC) and Some Milk Parameters in the Diagnosis of Mastitis

Serap Göncü[1], Nazan Koluman[1*]

[1] Çukurova University, Faculty of Agriculture, Department of Animal Science, 01330, Adana, Turkey

*Corresponding author e-mail: nazankoluman@gmail.com

Abstract

Abstract Relationships between various features in animal husbandry are used in the development of various technological facilities. In particular, the diagnosis of latent diseases such as mastitis can be predicted by changes in milk components and properties. Mastitis causes significant economic losses to businesses with loss of productivity, treatment and drug costs. For this reason, estimations that can accurately and easily diagnose mastitis are being studied intensively. Although there are many methods for the early diagnosis of mastitis, there is a need for an earlier, easier and more accurate estimation. For this reason, in this was carried out in order to identify the relations between Somatic cell count (SCC), Electrical conductivity of milk, initial milking temperature, milk pH and milk density in mastitic cows. The animal material of the study was mastitic Holstein Friesian cows in the Çukurova University Research Farm in which similar age and lactation number. SCC, electrical conductivity (EC), milk density, pH of milk and milk temperature parameters which changing with mastitis in milk were measured. Somatic cell counts higher than 200,000 / ml on the basis of quarter were accepted as indicator of subclinical mastitis. The obtained data were subjected by homogeneity test before analysis and it was determined that parameters except the temperature were not normal distribution in healthy group and SCC was not normal distribution in mastitis group. Analyze of variance, Regression and correlation analyses via SPSS package program were performed by applying logarithmic transformation of SCC data which is not normal distribution.

Key words: *Relationship, Somatic Cell Count (SCC), Milk Parameters, Diagnosis Of Mastitis*

Modeling Oil Demand in Turkey with Artificial Neural Networks, Xgboost and Time Series Analysis

Fatih Cemrek[1*], Özge Demir[2], Name Surname[3]

[1] Eskişehir Osmangazi Üniversitesi, 02640, Eskişehir, Turkey

*Corresponding author e-mail: zdemir94@gmail.com

Özet

The energy sector is currently facing increasing challenges in demand, efficiency, the analytical deficiency required for optimum management and changing patterns of supply and demand. In the last months of 2019, the global demand for oil has dropped in huge after the Covid19 outbreaks in China have become a worldwide epidemic. During the process of the effects of the pandemic on the markets, the demand for global crude oil in April decreased by 20 million V/g compared to the same period of the previous year, while the year -general year general year, the global oil demand decreased by 10 million V/g in 2020. calculated. (Turkey Petrolleri A.O., 2020). Turkey's demand for the effect of petroleum in the economy for modeling gasoline price, number of vehicles, oil imports, CPI, dollar, GDP data was discussed. Networks, XGBOOST Regression and Time Series Analysis Models are intended to compare with accuracy and accuracy. In this study, the machine learning methods to be used first and the prediction models will be created with Box-Gencins methods. It is formed by the modeling of the electrical bond between artificial neural networks and biological nerve cells and these cells between each other in the computer environment. Artificial neural networks are used in areas such as determination, classification, estimation, control, data association, data filtering, interpretation. It is not stated that it is very effective in complex problems in the literature. In the study, a multi -layered backward -spread algorithm will be created. XGBOOost (Extreme Gradient Boosting) is a high -performance state optimized by various regulations of the Gradient Boosting algorithm.Software and hardware optimization techniques have been applied to achieve superior results by using less resources. Decision tree-based algorithms are shown as the best. The other method is Box-Jenkins method. Average or moving average models are applied with average or authorization. With the development of computer facilities, the use of this method has become easier and widespread. The Box-Gencins method depends on the principle that each series of time is a function of past values and can be explained with them. The series, which is very successful in short -term estimation, is an important assumption of this method, which is a cut and static series consisting of observation values obtained at equal time intervals. Models where Box-Gencins method is applied; It has been developed depending on the assumption that time series related to time, events and time series related to these events. Box-Jenkins models are examined in three groups and these; Linear stable stocus models are non -static linear stochastic models and seasonal models. Box-Gencins Model Stages; Determining the appropriate model consists of estimated, control and prediction of the model. In the study, new modeling is desired to get out of classical methods. It is important to enrich data science with machine learning methods and to compare them with classical methods.

Key words: *Artificial Neural Networks, Time Series Analysis, XGBOOST Regression*

A New Bounded Distribution as an Alternative to the Beta and Kumaraswamy Distributions: Properties and Applications

Şule Sağlam[1*], Kadir Karakaya[2]

[1] Selçuk University, Department of Statistics, 42130, Konya, Turkey

*Corresponding author e-mail: sulesaglam75@gmail.com

Abstract

In this study, a new distribution is proposed as an alternative to the beta and Kumaraswamy distributions. Some mathematical properties of the new distribution such as moments, skewness, kurtosis, stochastic ordering etc. are studied. Six methods of estimation are obtained to estimate the unknown parameters of the new distribution. The simulated bias and mean squared error criteria are used to assess the performance of the estimation methods. As an alternative to the beta and Kumaraswamy regression models, we introduced a novel regression model based on the new distribution. The maximum likelihood method is studied to estimate regression parameters, and the performance of this method is evaluated via Monte Carlo simulation. In addition, the usability of the new distribution and regression model are demonstrated with several real data analyses.

Key words: *Unit distribution, NET distribution, maximum likelihood estimation, Monte Carlo simulation, regression model*

Life Expectancy Prediction Based on Animal Protein Consumption with Generalized Estimating Equations

Furkan Çağrı Beşoluk[1*], Harun Yonar[1], Mehmet Emin Tekin[1], Aynur Yonar[2]

[1] Selcuk University, Faculty of Veterinary, Department of Biostatistics, 42250, Konya, Turkey
[2] Selcuk University, Faculty of Science, Department of Statistics, 42250, Konya, Turkey

*Corresponding author e-mail: Furkan.besoluk@selcuk.edu.tr

Abstract

In this study, life expectancy was modeled based on animal protein consumption and health indicators with the Generalized Estimating Equations approach. The dataset used consists of 16 years of data from 69 countries. Different working correlation matrix structures and link functions were evaluated according to information criteria, and models were obtained for countries' OECD memberships, EU memberships, and income classifications. Animal protein consumption was determined as an effective factor in life expectancy predictions of countries, and the relations of these variables with health indicators such as obesity and cancer were examined. As a result, the direct effect of animal protein consumption on life expectancy and its indirect relationship with health indicators were revealed.

Key words: *Generalized Estimating Equations, Life Expectancy, Animal Protein Consumption, Health Indicators.*

Organik Tarım Yapan Meyve Üreticilerinin Tarım Sigortasına Yaklaşımları: Gaziantep ve Adıyaman İlleri Örneği

Sibel Ölmez Cangi[1*], **Hakan Karadağ**[2]

[1] Tokat Gaziosmanpaşa Üniversitesi, Turhal Uygulamalı Bilimler Fakültesi, Bankacılık ve Sigortacılık Bölümü, 60300, Tokat, Turkey
[2] Tokat Gaziosmanpaşa Üniversitesi, Ziraat Fakültesi, Bahçe Bitkileri Bölümü, 60240, Tokat, Turkey

*Corresponding author e-mail: sibel.olmez@gop.edu.tr

Özet

Organik tarım son yıllarda hızlı gelişen önemli sektörlerden birisidir. Günümüz dünyasının en büyük sorunları olarak sayılan; küresel ısınma, iklim değişikliği, sera gazı emisyonu, beslenme kaynaklı insan sağlığındaki bozulmalar ve tarımsal kaynaklı çevre kirliliği gibi durumlar için birçok otorite ve bilim insanı tarafından öne sürülen çözüm yollarından biri de organik tarım yapılmasıdır. Tarım, bir ülkenin bekası ve kalkınması için olmazsa olmaz bir sektör olarak kabul edilmektedir. Bununla birlikte tarım sektöründeki en büyük risk doğal afetlerin sıklıkla yaşanması ve bu durumun tarımsal üretimde verimliliği düşürmesidir. Bu riskler olduğunda, büyük ölçüde ürünlere hasar vermekte ve üretici gelirleri azalmaktadır. Gelişmiş ülkelere bakıldığında, tarımla uğraşan kesimin bu sorununu çözmek için tarım sigortası sisteminin uygulandığı görülmektedir. Bu araştırmada, Adıyaman ve Gaziantep illerinde organik tarım yapan üreticilerin tarım sigortasına yaklaşımları anket yoluyla belirlenmeye çalışılmıştır. Bu amaçla, Gaziantep ilinde 79, Adıyaman ilinde ise 84 adet organik tarım üreticisi ile yüz yüze anket yapılmıştır. Çalışmada, tarım sigortası yaptırıp yaptırmama durumunu etkileyen değişkenler lojistik regresyon analizi yapılarak incelenmiştir. Analiz sonuçlarına göre; Adıyaman yöresindeki üreticilerin, tarım sigortası yaptırma durumunun Gaziantep yöresindeki üreticilere oranla 0,501 kat; erkeklerin tarım sigortası yaptırmasının kadınlara oranla 2,172 kat olduğu sonucuna ulaşılmıştır. Yaş grubu 25-31, 32-38 ve 46 üstü olanların tarım sigortası yaptırması durumunun 39-45 olanlara oranla 0,432 kat fazla olduğu görülmüştür. Öğrenim durumu meslek yüksekokul olanların tarım sigortası yaptırmasının ilkokul, lise ve üniversite mezunlarına oranla 1,402 kat fazla olduğu sonucuna ulaşılmıştır. Gelir durumu tarım ve hayvancılık olanların tarım sigortası yaptırmasının diğer gelir kaynağı olanlar ya da sadece tarımla geçinenlere oranla 1,566 kat fazla olduğu görülmüştür.

Key words: *TARSİM, Lojistik Regresyon, Antep Fıstığı*

Entelektüel Sermayenin Emeklilik ve Hayat Sigorta Şirketlerinde Karlılığa Etkisi

Aylin Mercan Balkan[1], Sibel Ölmez Cangi[1*]

[1] Tokat Gaziosmanpaşa Üniversitesi, Turhal Uygulamalı Bilimler Fakültesi, Bankacılık ve Sigortacılık Bölümü, 60300, Turhal-Tokat, Turkey

*Corresponding author e-mail: sibel.olmez@gop.edu.tr

Özet

Bilgi çağıyla birlikte gelişen ve değişen işletmelerde şirketlerin değerini belirleyen unsurların sadece maddi varlıklar olmadığını, entelektüel sermaye olarak tanımlanan, insan kaynakları, çalışan bilgi birikimi, yapısal sermaye ve müşteri sermayesi gibi maddi olmayan varlıkların da son derece önemli olduklarını gösteren çeşitli çalışmalar yapılmıştır. Entelektüel sermayenin en önemli bileşenini, özellikle şirket performansındaki pozitif etkileriyle öne çıkan insan sermayesi oluşturmaktadır. İnsan sermayesi çalışanların sahip olduğu bilgi ve yaratıcılık gibi soyut değerleri şirketleri için faydaya dönüştürebilmesidir. Finansal hizmetler sektöründe önemli bir yere sahip olan sigortacılık sektörünün uzman bilgi ve becerisi gerektirdiği, bu nitelikten yoksun şirketlerin ise sürdürülebilirlikleri konusunda sorunlar yaşayabilecekleri söylenebilmektedir. Yapısı gereği oldukça karmaşık ve teknik işleyişi olan sigortacılık sektöründe de entelektüel sermayenin şirket karlılığına etkisinin olumlu yönde olabileceği düşünülmektedir. Bu çalışmada entelektüel sermayenin hayat ve emeklilik şirketlerinde karlılığa olan etkisinin araştırılması amaçlanmaktadır. Çalışmada 2017-2021 yılları arasında Türkiye'de faaliyette bulunan 15 hayat ve emeklilik sigorta şirketinin entelektüel sermaye katsayısı (VAIC), insan sermayesi etkinlik katsayısı, yapısal sermaye etkinlik katsayısı ve kullanılan sermaye etkinlik katsayısı VAIC modeli kullanılarak hesaplanmıştır. Sonrasında entelektüel sermaye katsayısı (VAIC) bileşenleri ile (VAIC) aktif karlılık (ROA) ile ilişkisi regresyon analizi ile test edilmiştir.

Key words: *Sigortacılık, VAIC, ROA*

Estimation of Lactation Milk Yield of Jersey Cattle by Isotonic Regression Analysis

Cem Tırınk[1*]

[1] Iğdır University, Faculty of Agriculture, Department of Animal Science, 76000, Iğdır, Turkey

*Corresponding author e-mail: cem.tirink@gmail.com

Abstract

This study used isotonic regression analysis to classify nonlinear regression models used to describe the lactation milk yield in relation to different quality fit evaluators. The best model selection was made by using the data obtained from Jersey cattle. Five nonlinear models measured by the fit quality determination coefficient, Akaike information criterion, Bayesian information criterion, mean quadratic estimation error, and estimated coefficient of determination were used. As a result of the different nonlinear models used to predict the lactation milk yield of Jersey cattle and the isotonic regression analysis applied to these models, considering the MSE (8.559) and R2 (0.878) values and the logistic model turned out to be the most appropriate model. As a result, it was revealed that the percentage of predictability and goodness of fit of the models increased significantly with isotonic regression analysis.

Key words: *Lactation Milk Yield, Lactation Period, Service Period, Nonlinear Models, Isotonic Regression*

Assessing Several Strategies for Estimating Average Treatment Effects in Nonparametric Model

Hülya Koçyiğit[1*]

[1] Karamanoğlu Mehmetbey University, 70100, Karaman, Turkey

*Corresponding author e-mail: hk20902@gmail.com

Abstract

Observational studies are frequently used to explore the connection between treatments and outcomes in many fields. Nevertheless, unless statistical methods are applied to account for the unbalance of confounders among treatment groups, they do not enable inferences about causal connections to be reached. The propensity score is a measure used to minimize imbalances in the distribution of confounding variables between treatment and control groups in studies evaluating the exposure effects on the outcome. Recent work has suggested that matching weighting (MW) outperforms inverse probability weighting (IPW) in linear settings. Researchers, on the other hand, require guidance in determining the best analytic method for any particular circumstance to estimate treatment effect based on the use generalized boosted model. To address this gap, the objectives of this study are threefold:(1) to present step-by-step instructions for researchers who aim learning IPW and MW for various treatments and outcomes, (2) to examine the two techniques (i.e., IPW and MW) in terms of unnecessary covariates, the complexity of the interaction between treatment or outcome and covariates, and residual variance, bias, and standard mean difference (SMD) in outcome and treatment, and (3) to propose the use of GBM to estimate the appropriate weights. This paper conducted a Monte Carlo simulation to provide clarity to analysts when determining between the two methods in GBM under various scenarios of outcomes or treatment based on the different covariates. Besides, this study describes the conditions under which one approach should be assumed to outperform the other, but we make no blanket assumption about whether one method is usually preferable to the other.

Key words: *Matching Weighting, Balancing Weight, Observational Study, Variable Selection, GBM*

Investigation of the Association Between Some Biochemical Properties in Patients with Metastatic Lung Cancer Receiving Palliative Radiotherapy

Feryal Karaca[1*], Selahattin Menteş[1], Sıddık Keskin[2]

[1] Adana City Training and Research Hospital, Department of Radiation Oncology, Adana, Turkey
[2] Van Yüzüncü Yıl University, Faculty of Medicine, Department of Biostatistics, Van, Turkey

*Corresponding author e-mail: feryalkaraca@gmail.com

Abstract

Introduction: Lung cancer is one of the most common malignant tumors. It is considered to be most important health problem for both sexes in developed countries. Cure is not possible most challenge in patients with Stage IV, these patients receive palliative radiotherapy (RT), chemotherapy and supportive care. In this study, it was aimed to determine the relationship between some biochemical properties in patients with metastatic lung cancer who received palliative RT. Methods: Forty-five patients who admitted to Adana City Training and Research Hospital for palliative RT between July 2016 and January 2021 were included in the study. Date were collected from Patients include age, Body mass index (BMI), gender, type of carcinoma (Squamous, Adeno and other), Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), Calcium (Ca), Sodium (Na), Potassium (K), Hematocrit value (HCT), Creatinine, Lymphocyte, Neutrophil, Platelet and Lactate dehydrogenase (LDH) values were measured. Pearson correlation coefficients were calculated for these features and Categorical Principal Component Analysis was performed. Results: All of the patients were Stage IV and the mean age was 63.84 ± 9.46 years. 10 (22.22%) were female and 35 (77.77%) were male. Nine patients have small cell lung cancer pathology. Thirty six patients had Non-small cell lung cancer, 14 (38.88%) patients had adenocarcinoma and 22 (61.11%) had squamous cell carcinoma. The result of the Categorical Principal Components Analysis demonstrated that 33.75% of the variation among the 15 variables was explained by the first dimension, 24.65% was explained by the second dimension, and the total explainable variance rate was found to be 58.40% with the two dimensions, BMI was found to be highly positively correlated with Lymphocyte, K , Na, Creatinine and LDH. In addition, HTC and Age were highly correlated, carcinoma type was also positively correlated with ALT and AST. Conclusions: The result of our study, associations between some biochemical features were determined in metastatic lung cancer patients who received palliative RT, and it was observed that categorical principal component analysis could be used in terms of examing the relationships between variables and reducing dimensions in high-dimensional data. RT is an effective treatment in terms of presence symptomatic palliation, symptoms such as hemoptysis, bronchiolar obstruction, cough, shortness of breath, and chest pain.

Key words: *Non-Small Cell Lung Cancer, Radiotherapy, Size Reduction, Correlation*

Sum-Lindley Poisson Model: Properties and Applications to New Count Regression

Çağrı Arsan[1*], Kadir Karakaya[1]

[1] Selçuk University, Department of Statistics, 42130, Konya, Turkey

*Corresponding author e-mail: arsan_13@hotmail.com

Abstract

In this study, a new discrete distribution is introduced by combining the sum-Lindley proposed by Chesneau et al. (2020) and the Poisson distribution. Some distributional properties of the new model, such as moments, expected value, variance, probability generating function, moment generating function, and deviation index are examined. For the estimation of the unknown model parameter, maximum likelihood, proportion, and moment methods are considered. A Monte Carlo simulation is conducted to observe the performances of these estmators. As an alternative to the Poisson and negative binomial regression models, we introduced a new count regression model based on the new distribution. The usability of the new distribution and regression model are demonstrated by some real data analyses.

Key words: *Sum-Lindley distribution, Poisson distribution, maximum likelihood estimation, Monte Carlo simulation, count-regression, Poisson regression*

Associations of CAST Gene Polymorphisms with Growth and Meat Quality Traits in Small Ruminants

Yasemin Gedik[1*], **Habiba Mohshina**[1]

[1] Eskisehir Osmangazi University, Faculty of Agriculture, Department of Animal Science, 26160, Eskisehir, Turkey

*Corresponding author e-mail: ygedik@ogu.edu.tr

Abstract

Identification of single nucleotide polymorphisms (SNPs) in the genes involved in the major biochemical pathways has great importance to improve farm animal performances. Calpastatin (CAST) is a highly abundant gene encoding a non-lysosomal protein which is a component of a proteolytic enzyme complex called the calpain/calpastatin system (CCS). Other components of CCS are a group of Ca2+ -dependent proteolytic enzyme calpains (CAPNs) inhibited by calpastatin (CAST). CAST restricts the proteolysis activity of calpains by preventing their activation and catalytic functions. The goat and sheep CAST genes consist of 34 and 29 exons that are found in the 7 and 5 chromosomes, respectively. Polymorphisms in different exonic and intronic regions of CAST have been identified using PCR-RFLP, PCR-SSCP, and nucleotide sequencing. Identified variants were connected with a variety of production traits such as birth weight, pre-weaning growth rate, final marketing weight, and other carcass characteristics. Besides these, CAST polymorphisms were also related to meat quality traits such as meat color, cooking loss, and tenderness. This study aimed to discuss the various CAST polymorphisms evidencing significant positive effects on different desirable traits in the sheep and goats and to reveal the most suitable variants that could be used widely in the selection program.

Key words: *Calpastatin, Goats, Polymorphisms, Sheep*

Estimation of Misclasification Rates Attribuable to Measurement Error in Non-Normal Case

Sümeyra Sert[1*], Coşkun Kuş[1]

[1] Selçuk University, Science Faculty, Department of Statistics, 42250, Konya, Turkey
[2] University, Faculty or Vocational school name, Department, Postal Code, City, Country
[3] University, Faculty or Vocational school name, Department, Postal Code, City, Country

*Corresponding author e-mail: sumeyra.sert@selcuk.edu.tr

Abstract

In measurement system analysis (MSA), the estimation of misclassification rates attributable to measurement error is generally discussed through normal distribution theory. However, there are too many situations where the data comes from a non-normal distribution. In this study, we discussed statistical inference on misclassification rates under non-normal distributed units. The likelihood ratio and bootstrap confidence intervals are constructed and a simulation study is performed to compare these intervals. A numerical example is provided to illustrate the discussed procedures.

Key words: *Measurement Error, Misclassification risk, Likelihood ratio, Bootstrap, Confidence intervals*

Ranking Çanakkale Districts in Terms of Pasture Quality with Multi-Criteria Decision Making Methods

Zeynep Gökkuş[1*], Sevil Şentürk[2], Fırat Alatürk[3]

[1] Kastamonu University, Vocational School, Department of Computer Technology, Kastamonu, Turkey
[2] Eskişehir Technical University, Faculty of Science, Department of Statistics, Eskişehir, Turkey
[3] Çanakkale Onsekiz Mart University, Faculty of Agriculture, Department of Field Crops, Çanakkale, Turkey

*Corresponding author e-mail: z.gokkus@gmail.com

Abstract

Based on the Project report of Determination of Pasture Availability and Range of Pasture Status Classes carried out by the Management of Doğu Anadolu Tarımsal Araştırma Enstitüsü, eleven districts of Çanakkale are ordered in terms of rangeland quality. For this reason, four different rankings were obtained by AHP, TOPSIS, VIKOR and WASPAS methods. According to the rankings made by the AHP, VIKOR and WASPAS methods, while the district with the highest pasture quality was Biga, the highest pasture quality according to the TOPSIS method was found to be the Center district.

Key words: *Ranking in agricultural experiments, AHP, TOPSIS, VIKOR, WASPAS.*

Ayçiçeği Fiyatları Üzerinde Etkileri Olan Faktörlerin ARDL Modeli ile Analizi

Kaan Kaplan[1*], Halil Kızılaslan[1]

[1] Tokat Gaziosmanpaşa Üniversitesi, Ziraat Fakültesi, Tarım Ekonomisi Bölümü, 60250, Tokat, Turkey

*Corresponding author e-mail: kaan.kaplan@gop.edu.tr

Özet

Bu çalışmada ayçiçeği fiyatlarına etkisi olan ayçiçeği üretim miktarı (ton), ayçiçeği ekili alan (ha), ithalat ve ihracat miktarları değişkenleri baz alınarak ayçiçeği fiyatları üzerindeki etkilerin ortaya konulması amaçlanmıştır. Bu amaçla 1980-2020 yıllarına ait ayçiçeği fiyatı (TL/kg), ayçiçeği üretim miktarı (ton), ayçiçeği ekili alan (ha) ve ihracat-ithalat miktarları (ton) veri olarak alınmıştır. Zaman serilerinde gerekli olan birim kök sınaması yapılmış ve ayçiçeği fiyatları ile ithalat miktarı değişkenlerinin düzeyde, ekili alan, üretim miktarı ve ihracat değişkenlerinin birinci düzeyde durağan olduğu tespit edilmiştir. Farklı seviyelerdeki durağanlıklara yönelik analiz imkanı veren ARDL modelinin çalışma için uygun model olduğuna karar verilmiş ve uygulanmıştır. Kısa dönem ARDL modeli için en uygun gecikme uzunlukları ARDL (1, 3, 0, 1, 1) olarak belirlenmiştir. Ayrıca seriler arasında otokorelasyona ve değişen varyansa rastlanılmamıştır. Uzun dönem ARDL modelinin uygun gecikme uzunluğu ise kısa dönem ARDL ile aynı şekildedir. Yapılmış olan sınır testine göre f istatistik değeri 9.5020 bulunmuştur ve bu değere göre %1 düzeyinde seriler arasında bir eşbütünleşme söz konusudur. Son olarak hata düzeltme modelinde yer alan CointEq değerine göre kısa dönemde serilerin karşılaşacağı bir şok uzun dönemde 1 yıl 4 aylık bir sürede normal hale gelmektedir.

Key words: *Ayçiçeği Fiyatı, Ayçiçeği, İhracat, İthalat, ARDL*

GİRİŞ

Günümüzde ayçiçeği büyük ölçüde yemeklik yağ gereksinimini karşılamak amacıyla kullanılmaktadır. Bazı ülkelerde yağlık ayçiçeğinin yanı sıra çerezlik çeşitlerin üretimi de yapılmaktadır. (Lofgren, 1978). Bu çeşitlerin taneleri besin maddelerince zengin olup, tuz, tereyağı ve bal ile karıştırılarak şekerleme yapımında kullanılmakta, sebze, balık, et ve salata üzerine çeşni olarak ilave edilmekte ve çerez olarak tüketilmektedir (Millete, 1974).

Ayçiçeği, soya, kolza (kanola) ve yerfıstığı ile beraber dünyada en fazla yağ elde edilen tek yıllık dört bitkiden biridir (Sencer ve ark., 1991). 20'den fazla kullanım alanına sahip olan ayçiçeği, aynı zamanda enerji ihtiyacının karşılanması amacıyla etanol ve biyodizel üretiminde de kullanılmaktadır (Pimentel ve Patzek, 2005).

Dünyada 2018/2019 pazarlama yılında toplam 600 milyon ton yağlı tohum üretimi gerçekleştirilmiş ve bu üretimi %8.4'ünü ayçiçeği tohumu oluşturmuştur. 2018/2019 pazarlama yılında toplam 25.8 milyon ha ayçiçeği ekimi yapılmıştır (USDA, 2020.). Ayçiçeği tarımı dünyada en fazla Ukrayna, Rusya, Arjantin ve Türkiye'de yapılmaktadır. Ayrıca dünyada Ayçiçek yağı ihracatı son dönemlerde artış göstermektedir. 2018/2019 pazarlama yılı itibariyle dünya bitkisel yağ ihracatının %13.4'ünü, ithalatın ise %11.5'ini ayçiçek yağı oluşturmaktadır.

Dünya ayçiçeği ve ayçiçek yağı fiyatları, yağlı tohum piyasalarındaki fiyat hareketleri, yerel ve uluslararası piyasalarındaki arz talep durumu, ülkelerin dış ticaret politikaları ve petrol fiyatlarından büyük ölçüde etkilenmektedir. 2011 yılında petrol fiyatlarının artmasıyla birlikte ayçiçeği tohumu ve ayçiçek yağı fiyatları oldukça yüksek seviyelere ulaşmıştır.

Türkiye'de en fazla üretilen yağlı tohum olması, bitkisel yağ üretiminde %80-85 civarındaki payı ve yüksek yağ içeriği (%40) nedeniyle en önemli yağlı tohum bitkisi olduğu söylenebilir. Ayrıca bitkisel yağ açığının azaltılabilmesi için öncelikli olarak üretiminin artırılması gereken yağlı tohumlu bitkidir. İklim ve toprak özellikleri dikkate alındığında yağlı tohumlu bitkilerin üretimi bakımından Türkiye'de büyük bir potansiyel mevcut olmasına rağmen yeterli ekim alanı ve üretim artışı sağlanamamıştır. Türkiye'de ayçiçeği üretiminin yeterli olmamasının nedenleri arasında üretim maliyetlerinin yüksek olması, dış pazar fiyatlarıyla rekabet edilememesi, birim alandan getirisinin düşük olması vb. nedenler söylenebilir (TEPGE, 2020).

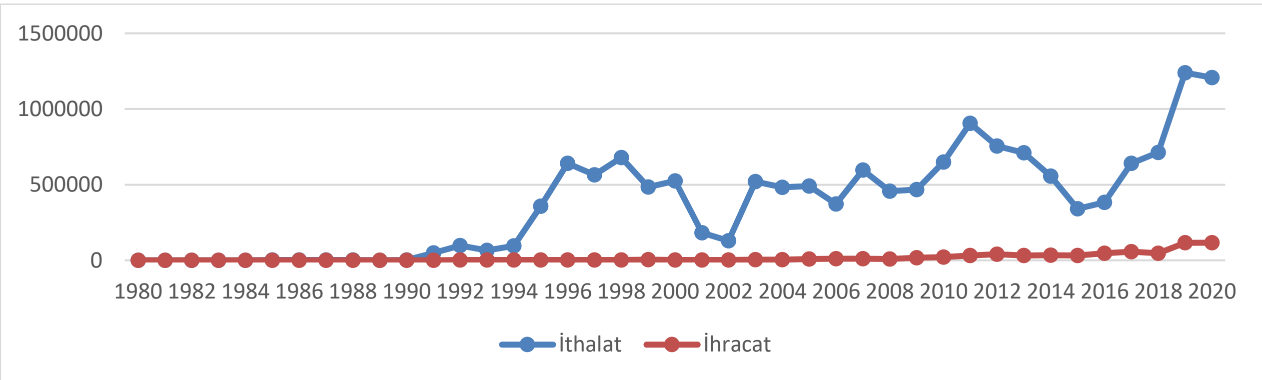
Tablo 1. Türkiye'de yıllara göre ayçiçeği üretim alanı (da) ve üretim miktarı (ton)

Yıllar	**1980**	**1990**	**2000**	**2005**	**2010**	**2015**	**2020**
Üretim Alanı (da)	575 000	714 599	542 000	566 000	641 343	685 174	728 853
Üretim Miktarı (Ton)	750 000	860 000	800 000	975 000	1 320 000	1 680 700	2 067 004

TÜİK (2022)'den elde edilen veriler ile belirli yıllardaki ayçiçeği üretim alanı (da) ve üretim miktarı Tablo 1'de verilmiştir. 1980'den 2020 yılına doğru üretim alanında yıllar itibariyle bir dalgalanma söz konusudur. Üretim miktarında ise dalgalanmalar olmakta birlikte bir artış eğilimi söz konusudur. 1980'de 750 000 ton olan üretim miktarı 2020 yılında 2 067 004 tona yükselmiştir. Üretim alanı ise 1980'de 575 000 dekar iken 2020 yılında 728 853 dekardır.

Türkiye'de ayçiçeği fiyatları Trakyabirlik tarafından belirlenen düzeylerde gerçekleştirilmekte ve özel firmalar bu fiyatlar seviyesinde alım yapmaktadır. Türkiye'nin ayçiçeğinde net ithalatçı olması nedeniyle yurtiçi ayçiçeği tohumu ve yağı fiyatları dünya fiyatlarındaki değişimlerden direkt etkilenmektedir ve ithal fiyatlarının artmasıyla Türkiye yurtiçi fiyatları da artış eğilimine geçmektedir (TEPGE, 2020).

Türkiye'de yağlık ayçiçeğinin, üretimde sürdürülebilirliğin sağlanması ve arz açığının giderilmesi amacıyla desteklenmesi önem arz etmektedir. Bu kapsamda, yağlık ayçiçeği uzun yıllardır üretimini teşvik etmek ve bitkisel yağ açığını giderebilmek amacıyla devlet desteklemesi kapsamındadır. Tarımda başlatılan yapısal reformlar ve destekleme yöntemlerindeki değişiklikler kapsamında; hem birliklerin dünya fiyatından ürün almasını sağlamak hem üreticiye yüksek fiyata vermek amacıyla 1999 yılında yağlık ayçiçeğine uygulanan prim sistemi halen devam etmektedir (TEPGE, 2020).



Şekil 1. Türkiye'de yıllar itibariyle ayçiçeğinde ithalat ve ihracat miktarları (ton)

Yukarıda bahsedildiği gibi dünya ve Türkiye'de ayçiçeği fiyatları üzerinde etkili olan birçok faktör yer almaktadır. Ancak Türkiye'de iç pazarda fiyatlar üzerinde etkili olduğu düşünülen değişkenler modele eklenmiştir. Bu sebeple çalışmada amaç ayçiçeği fiyatları üzerinde etkili olduğu düşünülen değişkenlerin varsa etki düzeylerinin ortaya çıkarılması ve sonuçların yorumlanmasıdır.

MATERYAL VE METOD

Materyal

Çalışmada ayçiçeği ürün fiyatı ile üretim miktarı, ekili alan, ithalat ve ihracat değerleri arasındaki ilişkiyi incelemek ve ürün fiyatını etkileyen değişkenlerin belirlenmesi amacıyla 2003 bazlı fiyat endeksi dikkate alınarak 1980-2020 dönemlerine ait veri seti kullanılmıştır. Zaman serisi analizlerinde tarımsal istatistiklerde geçmişe yönelik verilere farklı değişkenlerde ulaşılması zor olduğu için bu dönem tercih edilmiştir. Veriler Türkiye İstatistik Kurumu (TÜİK), Tarım ve Orman Bakanlığı (TOB) ve Food and Agricultural Organization (FAO) elde edilmiştir. Enflasyon etkisini barındıran cari fiyatlar ile elde edilen sonuçlar çoğu zaman yanlış değerlendirmelere sebep olabilmektedir. Bu sebeple daha sağlıklı sonuçlar elde edebilmek adına ayçiçeği ürün fiyatları üretici fiyat indeksi (ÜFE=2003) dikkate alınarak reel fiyata dönüştürülmüştür.

Metod

Birim kök testlerine yönelim zaman serilerinin genellikle durağan olmadığı gerçeği ile gerçekleşmiştir. Dickey ve Fuller (1981) tarafından geliştirilen Genişletilmiş Dickey-Fuller (ADF) testi ile Phillips ve Perron (1988) tarafından geliştirilen Phillips-Perron (PP) tesleri ile serilerin hangi türden trende sahip oldukları belirlenmektedir.

Granger ve Newbold (1974) tarafından belirtildiği gibi, zaman serilerinin çoğunlukla ortak yönde eğilim içermesi, seriler arasında sahte ilişkilerin ortaya çıkmasına neden olmaktadır. Gerçekte var olmayan bu sahte regresyon yüksek R2, düşük Durbin-Watson değerleri ve yüksek t istatistikleri ile geçersiz istatistiki sonuçlara neden olmaktadır. Tüm bu yanlışlıkları gidermek için, serilerin durağanlıktan kurtulması yani farklarının alınması gerekmektedir. Ancak bu durumda kısa dönem bilgileri elde edilmekte fakat uzun dönem bilgisinin kaybolmaktadır.

Granger (1981) ile Engle ve Granger (1987), uzun dönem bilgisinin kaybolması durumundan kaynaklanarak hem sahte regresyon sorununu aşmak hem uzun dönem katsayılarını kullanabilmek için kointegrasyon yöntemini geliştirmişlerdir. Ancak geliştirilen bu yöntemde bir bağımlı ve bir bağımsız değişken bulunmaktadır. Johansen (1980), Stock ve Watson (1988) ve Johansen ve Juselius (1990) kointegrasyonun hesaplanmasına yönelik Johansen Eşbütünleşme Yöntemini geliştirmişlerdir ve böyle iki ya da daha fazla sayıda değişken arasındaki uzun dönem ilişkisini tespit edebilmişlerdir.

Ancak bu yöntemlerin uygulanabilmesi için serilerin aynı seviyede durağan olmaları gerekmektedir. Değişkenlerin farklı seviyelerde durağan olması durumunda Peseran ve Shin (1997) tarafından ortaya konulan ARDL (Autoregressive Distrubuted Lag) Yöntemi kullanılmaktadır. Bu test ayrıca Sınır (Bound) testi olarak bilinmektedir ve değişkenlerin hem düzey hem de birinci farkta durağan olmaları durumunda kullanılabilmektedir. Engle ve Granger'ın (1987) ortaya koyduğu yönteme benzer şekilde iki aşamada uygulanmaktadır.

Bu yöntemin çeşitli avantajları literatürde bahsedilmektedir. Diğer yöntemlere nazaran bu yöntemin en önemli avantajı kullanılacak serilerin düzeyde veya birinci farkında durağan olmasının sorun teşkil etmemesidir. Yöntemin bir diğer avantajı ise eşbütünleşme testlerine göre hata düzeltme modelinde istatistiksel olarak daha güvenilir sonuçlar vermesidir. ARDL modeli; aynı anda modele dahil edilen seriler arasındaki hem kısa hem uzun dönem ilişkilerine yönelik bilgi vermektedir (Belen ve Karamelikli, 2016: Akel ve Gazel, 2014). Kullanılmış olan ARDL modeli (1) numaralı denklemde gösterilmiştir.

$$\mathrm{Y_t} = \alpha_0 + \alpha_1 \mathrm{t} + \sum_{i=1}^{p} \alpha_{2i}\ \mathrm{Y}_{t-i} + \sum_{i=0}^{q} \alpha_{3i} X_{t-i} + \mu$$

Ayrıca kısa dönem katsayıları tahmin etmek amacıyla yapılan hata düzeltme modeli ise (2) numaralı denklemde gösterilmiştir.

$$\Delta \mathrm{Y}_t = \alpha_0 + \alpha_1 \mathrm{t} + \alpha_1 \mathrm{EC}_{t-1} + \sum_{i=1}^{p} \alpha_{3i}\, \Delta \mathrm{Y}_{t-i} + \sum_{i=0}^{p} \alpha_{4i}\, \Delta \mathrm{X}_{t-i} + \mu_t$$

Çalışmada yer alan tüm serilerin öncelikle doğal logaritmaları alınmış ve ardından serilerin durağan olup olmadığının kontrolü için Genişletilmiş Dickey-Fuller (ADF) ve Phillips-Perron (PP) birim kök testlerinden yararlanılmıştır. Sonrasında nohut ekili alan bağımlı değişken alınarak diğer değişkenlerin ekili alan üzerindeki etkilerini araştırmak amacıyla ARDL yöntemi kullanılmıştır

SONUÇLAR

Serilerin durağanlıklarının test edildiği birim kök sınaması Tablo 2'de verilmiştir. Buna göre düzeyde ADF birim kök testinde ithalat miktarı (ton), PP birim kök sınamasında ise hem ithalat hem de fiyat değişkenleri diğer değişkenler birim kök içermektedir. Ancak birinci farkları alındığında tüm değişkenlerin birim kök içermediği sonucuna ulaşılmıştır. Serilerin farklı düzeylerde durağan olması I(0) ve I(1) araştırmada ARDL modelinin kullanılması gerekliliğini doğurmuştur.

Tablo 2. ADF ve PP birim kök sınaması

		Değişkenler	ADF	PP		*Değişkenler*	ADF	PP
Düzey	**Sabit**	**Ekili Alan (da)**	-1.5904	-1.7343	*Birinci Farklar*	**Ekili Alan (da)**	-6.2701***	-6.2612***
		Üretim Miktarı (ton)	-0.7019	-0.7019		**Üretim Miktarı (ton)**	-7.9487***	-7.9486***
		Ürün Fiyatı (TL/kg)	-1.4047	-3.2102**		**Ürün Fiyatı (TL/kg)**	-7.3382***	-13.5707***
		İthalat (Ton)	-4.3752***	-8.5191***		**İthalat (Ton)**	-4.0727***	-3.8865***
		İhracat (Ton)	-1.9106	-2.2299		**İhracat (Ton)**	-7.7663***	-7.8027***
	Sabit+Trend	**Ekili Alan (da)**	-1.7615	-1.9944		**Ekili Alan (da)**	-3.1108	-6.1408***
		Üretim Miktarı (ton)	-1.5455	-2.0828		**Üretim Miktarı (ton)**	-7.8401***	-7.8401***
		Ürün Fiyatı (TL/kg)	-1.9590	-4.8400***		**Ürün Fiyatı (TL/kg)**	-7.2052***	-12.8513***
		İthalat (Ton)	-3.3017*	-4.6111***		**İthalat (Ton)**	-4.5374***	-4.3345***
		İhracat (Ton)	-2.5267	-2.4007		**İhracat (Ton)**	-7.8707***	-8.1413***

*,**,*** sırasıyla %10, %5 ve %1 düzeylerinde anlamlılığı ifade etmektedir.

Eşbütünleşme testine ait sonuçlara Tablo 3'te yer verilmiştir. Eşbütünleşme ilişkisinin oluşabilmesi için Peseran tarafından belirlenen üst kritik değerlerinden F istatistik değerinin yüksek olması gerekmektedir. Tabloda yer alan sonuçlara göre, F istatistik değeri 9.502000 olarak bulunmuştur ve serilerin arasında eşbütünleşme ilişkisinin varlığı kabul edilebilir. Sonuç olarak ARDL modeli kurulması uygundur ve seriler arasında uzun ve kısa dönemde ilişkilerin belirlenebilir.

Tablo 3. Eşbütünleşme testi sonuçları (F Bound Test)

Test istatistiği	**Değeri**	**K**
F- istatistiği	**9.502000**	4
	Kritik Sınır Değerleri	
Anlamlılık Düzeyi	**Alt sınır**	**Üst Sınır**
%10	2.20	3.09
%5	2.56	3.49
%2.5	2.88	3.87
%1	3.29	4.37

ARDL modelinin optimum düzeyindeki gecikme uzunluğunun tespit edilmesinde AIC (Akaike Information Criterion) kullanılmıştır ve ayrıca SC (Schwarz Criterion) ve HQ (Hannan-Quinn Criterion) ilave olarak verilmiştir. Maksimum gecikmenin uzunluğu modelde yer alan değişkenlerin yıllık veri seti olduğu için 4 alınmış ve uygun gecikme uzunluğu ile tahminleme yapılmıştır. Ayrıca belirlenen gecikme uzunluğunun uygunluğuna yönelik otokorelasyonun ve değişen varyans testleri ile bu testlerin ortak sonucunda ARDL (1,3,0,1,1) şeklindeki modeli Tablo 4'de verilmiştir. Uzun dönem tahmin sonuçlarına göre değişkenlerin farklı ölçütlerde bulunmasından kaynaklı düzeltilmiş R2 oranı 0.667518 bulunmuştur. Değişkenlerin yıllar itibariyle farklarının alındığı gecikmeler anlamlı p değerlerini vermektedir.

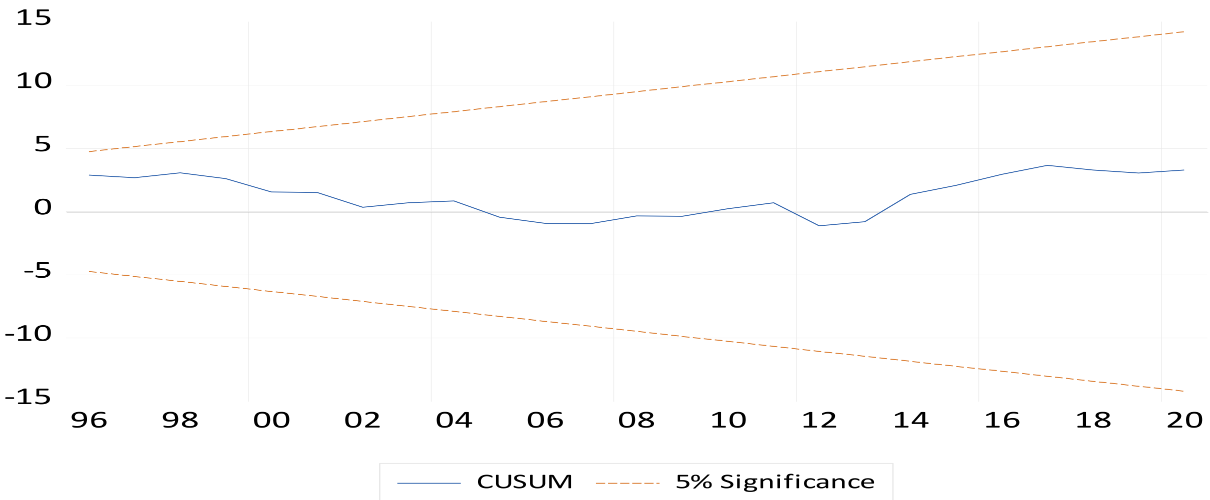
Tablo 4. ARDL (1, 3, 0, 1, 1) modeli ile uzun dönem tahmini

Değişken	**Katsayı**	**Std. Hata**	**t-istatistiği**	**P. value**
Ürün Fiyatı (-1)	0.273481	0.132361	2.066176	**0.0493**
İthalat	-0.040403	0.113863	-0.354835	0.7257
İthalat (-1)	-0.200287	0.164573	-1.217007	0.2350
İthalat (-2)	0.377404	0.154641	2.440521	**0.0221**
İthalat (-3)	-0.220827	0.086608	-2.549733	**0.0173**
Ekili Alan	-4.031850	1.391462	-2.897563	**0.0077**
İhracat	0.121537	0.141370	0.859710	0.3981
İhracat (-1)	0.370873	0.126404	2.934027	**0.0071**
Üretim Miktarı	0.869278	0.676966	1.284080	0.2109
Üretim Miktarı (-1)	-0.769432	0.550329	-1.398130	**0.1744**
C	1.432064	0.644754	2.221100	0.0356
R^2				0.762513
Adjusted R^2				0.667518
Prob (F-statistic)				**0.000012**
Akaike info criterion				1.017898
Schwarz criterion				1.501751
Hannan-Quinn criterion				1.186776

Tablo 5'de ARDL (1, 3, 0, 1, 1) modelinin uygunluğuna dair yapılmış olan sınamalara yer verilmiştir. Modelin anlamlılığına dair Breusch-Godfrey Otokorelasyon testi, ARCH Değişen varyans testi ve Jarque-Bera Normallik testine yer verilmiştir. Tabloya göre olasılık değeri 0.9398 bulunmuştur ve modelde otokorelasyon tespit edilmemiştir. Ayrıca değişen varyans testi olasılık değeri 0.3787 olarak bulunmuş ve modelde değişen varyans tespit edilmemiştir. Elde edilen bu sonuçlar uygulanan modelin çalışma için uygun bir model olduğu sonucunu doğurmaktır. Ayrıca Şekil 2'de yer alan CUSUM test sonucu ise, değişkenlere ait değerlerin %5 değer bandı içerisinde yer aldığı ve uzun dönem katsayılarının istikrarlı olduğunu göstermektedir.

Tablo 5. Modelin geçerliliği için yapılmış olan sınamalar

Breusch-Godfrey (LM) Otokorelasyon Testi		
Gecikme Uzunluğu	**LM-Test İstatistiği**	**Olasılık Değeri**
(1, 3, 0, 1, 1)	0.132284	0.9398
Heteroskedasticity ARCH Değişen Varyans Testi		
Gecikme Uzunluğu	**F-Test İstatistiği**	**Olasılık Değeri**
(1, 3, 0, 1, 1)	1.066071	0.3787
Normallik Testi		
Gecikme Uzunluğu	**Jarque-Bera**	**Olasılık Değeri**
(1, 3, 0, 1, 1)	0.398307	0.819424



Şekil 2. CUSUM test sonucu

ARDL (1, 3, 0, 1, 1) modelinde uzun dönem katsayıları Tablo 5'te verilmiştir. Test sonuçlarına göre ekili alan düzeyi ve ithalat miktarı %1 düzeyinde negatif yönde anlamlı bulunmuştur. Bu sonuca göre ekili alan ve ithalatın ayçiçeği fiyatları üzerinde negatif yönde bir etkisi olduğundan söz edilebilir.

Tablo 6. ARDL (1, 3, 0, 1, 1) modelinden elde edilen uzun dönem katsayıları

Değişkenler	Katsayı	Std. Hata	t-istatistiği	P. value
İthalat	-0.115774	0.053099	-2.180325	0.0389
Ekili Alan	-5.549545	2.208325	-2.513012	**0.0188**
İhracat	0.677766	0.353302	1.918389	0.0666
Üretim Miktarı	0.137431	1.475486	0.093143	0.9265
C	1.971130	0.708669	2.781454	0.0101
EC=Ürün Fiyatı – (-0.1158*İthalat-5.5495*Ekili Alan + 0.6778*İhracat + 0.1374* Üretim Miktarı + 1.9711)				

Modelde kullanılan değişkenler arasında bir kısa dönemdeki ilişkisini tespit etmek amacı ile ARDL modeline bağlı hata düzeltme yönteminden faydalanılmıştır. ARDL (1, 3, 0, 1, 1) modeline bağlı hata düzeltme modelinin sonuçları Tablo 6'da verilmiştir. Hatanın düzeltilmesine yönelik katsayının negatif bir işarete sahip olması ve istatistiksel olarak anlamlı olması durumunda denge düzeyinden sapma durumunda tekrardan denge düzeyine yönelik bir hareket olduğu sonucu çıkmaktadır (Bozkurt, 2007). Elde edilen sonuçlara göre istatistiksel olarak anlamlı olan CointEq değeri -0.726519 olarak bulunmuştur ve negatif bir değer taşımaktadır. Bu sonuca göre kısa dönemde denge halinden bir sapma olması durumunda tekrardan dengeye yönelik bir hareketin varlığından söz edilebilir. Kısa dönem sonuçlarına göre kısa dönemde meydana gelen bir değişme uzun döneme 1.4 aylık bir sürede etki ettiği söylenebilir.

Tablo 6. ARDL (1, 3, 0, 1, 1) modelinden elde edilen uzun dönem katsayıları

Değişken	Katsayı	Std. Hata	t-istatistiği	P. value
İthalat	-0.040403	0.081092	-0.498231	0.6227
İthalat (-1)	-0.156578	0.075861	-2.063999	0.0495
İthalat (-2)	0.220827	0.068878	3.206076	**0.0037**
İhracat	0.121537	0.070356	1.727449	0.0964
Üretim Miktarı	0.869278	0.265027	3.279960	**0.0031**
CointEq(-1)*	**-0.726519**	0.087836	-8.271300	**0.0000**
R^2				0.748352
Adjusted R^2				0.706411
Akaike info criterion				0.740120
Schwarz criterion				1.004040
Hannan-Quinn criterion				0.832235

SONUÇ VE TARTIŞMA

Seçilmiş olan ayçiçeği ürün fiyatı (TL), ayçiçeği üretim miktarı (ton), ayçiçeği ekili alan (da), ayçiçeği ithalat ve ihracat miktarı (ton) değişkenleri ile uygun model seçilerek ayçiçeği ürün fiyatı (TL) üzerinde diğer değişkenlerin etkisi araştırılmıştır. Birim kök sınaması yapılarak değişkenlerin farkı düzeylerde durağanlık içerdikleri belirlenmiş ve farklı düzeyler için en uygun model olan ARDL modeli uygulanmıştır. Yıllık veri seti kullanıldığı için uygun gecikme uzunluğu 4 olarak alınmış ve ARDL (1, 3, 0, 1, 1) modeli belirlenmiştir. Uzun dönemde test sonuçlarına göre ekili alan düzeyi ve ithalat miktarı %1 düzeyinde negatif yönde anlamlı bulunmuştur. Değişkenler arasında kısa dönem ilişkisini test etmek amacı ile ARDL modeline bağlı hata düzeltme yönteminden faydalanılmıştır. ARDL (3,4,1,4) modeline bağlı hata düzeltme modelinin sonuçlarına göre istatistiksel olarak anlamlı olan CointEq değeri negatif bir değer taşımaktadır. Bu sonuca göre kısa dönemde denge halinden bir sapma olması durumunda tekrardan dengeye yönelik bir hareketin varlığından söz edilebilir. Kısa dönem sonuçlarına göre kısa dönemde meydana gelen bir değişme uzun döneme 1.4 aylık bir sürede etki ettiği söylenebilir.

Ayçiçeğinde Türkiye'de kendine yeterlilik oranı 2000 yılında %70.1 iken 2020 yılında %62.5'e gerilemiştir. Yıllar itibariyle kendine yeterlilik oranı ayçiçeğinde %100 olmamakla birlikte bu durum dışa bağımlılığı arttırmaktadır. Net ithalatçı bir konumdan kurtulabilmek adına öncelikli olarak üretiminin artırılması gereken yağlı tohumlu bitkidir.

Ayçiçek yağı arzında ithalatın payı dikkate alındığında, son dönemde yaşanan ayçiçeği yurt içi tüketici cari fiyat artışında Türk Lirasının kaydedilen değer kaybının da etkili olduğu düşünülmektedir. Genel olarak Türkiye'de ayçiçeği yağı fiyatı dünya birim fiyatının üzerinde seyretmektedir. Dünyada döviz kurlarındaki yükseliş ve petrol fiyatlarının arttığı yıllarda, Türkiye'de yağlı tohum ve yağ fiyatları artmaktadır. Türkiye'nin net ayçiçeği ithalatçısı olması nedeniyle, Türkiye ayçiçeği fiyatları dünya fiyatları ve petrol fiyatlarından direkt olarak etkilenmektedir. Ayrıca fiyatlara kullanılan değişkenler dışında rekolte, iklim şartları, tüketim, petrol fiyatları, parite, ekonomik veriler ve devlet politikaları da etki etmektedir.

Kaynaklar

Belen M, Karamelikli H, 2016. Türkiye'de Hisse Senedi Getirileri ile Döviz Kuru Arasındaki İlişkinin İncelenmesi: ARDL Yaklaşımı. İstanbul Üniversitesi İşletme Fakültesi Dergisi, 4 (12): 55-72.

Dickey DA, Fuller WA, 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. Econometrica, 49 (4): 1057-1072.

Engle R, Granger CWJ, 1987. Co-integration and Error Correction: Representation, Estimation and Testing. Econometrica. 55 (2): 251-276.

FAO, 2022. Erişim Linki: www.fao.org; Erişim tarihi: 21.05.2022

Granger C, Newbold P, 1974. Spurious Regressions with Stationary Series. Applied Economics, 33: 899-904.

Granger CWJ, 1981. Some Properties of Time Series Data and Their Use in Econometric Model Specification. Journal of Econometrics, 16: 121-130.

Johansen S, 1988. Statistical Analysis of Cointegration Vector. Journal of Economic Dynamics and Control, 12: 231-254.

Johansen S, Juselius K, 1990. Maximum Likelihood Estimation and Inference on Cointegration-with Applications to the Demand for Money. Oxford Bulletin of Economics and Statistic, 52 (2): 169-210.

Lofgren JR, 1978. Sunflower for Confectionery Food, Bird Food and Pet Food. In J.F. Carter (ed) Sunflower Technology and Production. ASA. SCSA and SSSA Monograph.

Millete RA, 1974. Seed from the Sunflower. North Dakota State University Fargo Cir. HE-120.3p.

Perron P, 1997. Further Evidence on Breaking Trend Functions in Macroeconomic Variables. Journal of Econometrics, 80 (2): 355-385.

Pesaran HM, Shin Y, 1997. An Autoregressive Distrubuted Lag Modelling Approach to Cointegration Analysis. DAE Working Paper, No:9514, Department of Applied Economics, University of Cambridge.

Pimentel D, Patzek TW, 2005. Ethanol Production Using Corn, Switchgrass and Wood; Biodiesel Production Using Soybean and Sunflower, Natural Resources Research, 14 (1): 65-76.

Sencar Ö, Gökmen S, Yıldırım A, Kandemir N, 1991. Tarla Bitkiler Üretimi. Cumhuriyet Üniversitesi Ziraat Fakültesi Yayınları Yayın No:11, s. 206-215, Tokat.

Stock J, Watson MW, 1988. Testing for Common Trends. Journal of the American Statistical Association, 83: 1097-1107.

TEPGE, 2020. Tarımsal Ekonomi ve Politika Geliştirme Enstitüsü, Durum ve Tahmin, Ayçiçeği 2020. ISBN: 978-605-7599-46-9.

TUİK, 2022. Erişim Linki: www.tuik.gov.tr; Erişim tarihi: 21.05.2022

USDA, 2022. Erişim Linki: www.fas.usda.gov/psdonline/; Erişim tarihi: 11.06.2022

Çıkar Çatışması

Yazarlar çıkar çatışması olmadığını beyan ederler.

Yazar Katkısı

Yazarlar çalışmaya eşit oranda katkı yapmışlardır.

The Effects of Different Production Sytems on Wellfare of Dairy Goats

Ahmet Mert[1], Serap Öncü[1], Nazan Koluman[1*]

[1] Cukurova University, Faculty of Agriculture, Department of Anmal Sci.,01330, Adana-Turkiye

*Corresponding author e-mail: nazankoluman@gmail.com

Abstract

The aim of this study was to develop a wellfare systematic that could reveal the diferences between traditional and industrial dairy goat farming in Adana city of eastern Mediterranean region of Turkey. Thus, some data regarding to relationship between management systems and quality of products was determined in 9 dairy goat farms and almost 2000 dairy goats. An index was developed for objective evaluation of the animals. During the study dirtness index was improved for cleansing of udder, legs, back, horn and face. The dairy performances such as lactation milk yield and leghts of the does were recorded as well. At the end of the study, it was concluded that animal population and surface cleaning of barn had positive effects on milk quantity. The improved index could be used for easure wellfare of the dairy animals.

Key words: *Dairy Goat, Intensive, Extensive Sytems, Milk Quality, Dirtness Index*

An Application of Kernel Regression in Health Studies

Özge Pasin[1], **Senem Gönenç**[2*]

[1] Bezmialem University, 34000, İstanbul, Turkey
[2] Atatürk University, Science Faculty, Department of Statistics, 25030, Erzurum, Turkey

*Corresponding author e-mail: senemgonenc@atauni.edu.tr

Abstract

Regression models are the most frequently used statistical models in the field of health. These models are used to explain the causal relationship between dependent and independent variables. There are many different regression models in the literature. According to the number of dependent and independent variables, the way they are obtained, and the conditions under which the assumption conditions of the models are met, the regression model that is appropriate for the data set should be selected. The aim of this study is to introduce the theoretical features of the Kernel Regression Model, which is determined to have little application in the field of health, to explain in which cases it should be used, and to compare the results with Classical Linear Regression Analysis by making an application in the field of health. Kernel Regression Analysis is a non-parametric model. In cases where the assumptions of parametric regression analysis are not provided, non-parametric Kernel Regression Analysis can be used as an alternative. Kernel Regression Analysis does not make assumptions about whether the conditional expected function is linear, quadratic or exponential. The weights in the Kernel are defined by Epanechnikov, Biweight, Triangle, Gaussian, Parzen, etc. Kernel Functions. April June 23- June 16 2021, COVID-19 data in Turkey were used in the study. The number of deaths was taken as the dependent variable, confirmed, recovered, number of people who received at least one vaccine dose, and event cancellation status were taken as independent variables. Kernel Regression Analysis was used because parametric regression analysis assumptions were not provided. Epanechnikov Kernel Density Function and Li-Racine Density Function was used for calculations. Linear Regression Analysis was used to compare model predictions and residuals. Analyzes were performed with STATA 16. As a result of the study, the effect of the confirmed variable on the number of deaths was not statistically significant (p=0.724), while the effects of the recovered, vaccine and cancel event variables were found to be statistically significant in Linear Regression Analysis (p<0.001, p=0.001, p=0.001). In the Kernel Regression Analysis, the effect of the confirmed variable was not found significant (p=0.965), while the effects of the other variables were found to be significant (p<0.001; p=0.006; p=0.008). It is observed that the Kernel Function produces closer predictions. While the mean value of the absolute value differences of the observed death numbers with the estimates obtained as a result of the Kernel Regression Analysis was obtained as 186.79, the average value was obtained as 332.81 in the Linear Regression. The explanatory coefficient of Kernel Model was 0.995, the Linear Model was 0.989. Consequently, appropriate regression analysis models should be selected depending on the data structure. In cases where the assumptions of parametric regression analyzes are not met, a non-parametric method Kernel Regression Analysis can be used. We suggest that the use of Kernel Regression Analysis should be widespread and the usage areas of this method in the field of health should be increased.

Key words: *Regression, Non Parametric, Assumption, Kernel*

Application of Binary and Multinominal Logistic Regression in SPSS

Esra Yavuz[1*], **Şeyma Koç**[2], **Demet Çanga**[3], **Tolga Tolun**[4]

[1] Şırnak University, Cizre Vocational School, Department of Accounting and Tax, 73200, Şırnak, Turkey
[2] Kahramanmaras Sütçü Imam University, Faculty of Agriculture, Department of Animal Science, 46100, Kahramanmaraş, Turkey
[3] Osmaniye Korkut Ata University, Department of Chemistry and Chemical processing, Bahçe, Osmaniye, Turkey

*Corresponding author e-mail: yavuz7346@gmail.com

Abstract

Logistic regression measures the effects of independent variables on dependent variables as probabilities and aims to determine these risk factors as probabilities. Logistic regression is known as a method that calculates the relationship between the dependent variable and the independent variables in binary and multiple stages. The aim of this study is to examine binary and multinominal logistic regression analysis, which can describe the relationship between binary and multinominal dependent variables and independent variables consisting of discrete variables. For this purpose, it is to show the comparison of these two methods by analyzing the application of binary and multinominal regression analysis in SPSS. As a result, it is important to provide information on the use of binary and multiple logistic regression analysis in this study, to show the deficiencies in the applications and to make suggestions for future studies.

Key words: *Binary, Mutinominal, Logistics, Regression*

Addressing Multicollinearity Via Modified Shewhart Chart in Monitoring Com-Poisson Profiles

Ulduz Mammadova[1*], **M. Revan Özkale**[1]

[1] Çukurova University, Science Faculty, Department of Statistics, 01330, Adana, Turkey

*Corresponding author e-mail: ulduzozel@gmail.com

Abstract

Generalized linear profiles with count response are typically monitored with methods developed for Poisson profiles. However, not all counts have the equal mean and variance. In 1962 Conway and Maxwell introduced the Conway-Maxwell-Poisson (COM-Poisson) distribution to solve this problem. COM-Poisson distribution is the two-parameter generalization of Poisson distribution that overcomes limitations of Poisson distribution and handles count data with under-, equi-, and overdispersion. Multicollinearity is a serious concern that might arise during the monitoring of COM-Poisson profiles. We propose to address the problem in the monitoring process with modified residual-based Shewhart charts, which are ShewhartPCR and Shewhartr-k charts. In order to construct the proposed Shewhart charts, we defined the iterative principal component regression (PCR) and r-k class estimators for COM-Poisson model and obtained the PCR and r-k deviance residuals. These residuals are used as monitored observations to identify abnormalities in the process. The performance of proposed control charts is analyzed and evaluated via simulation and case studies in terms of run-length-based measurements.

Key words: *COM-Poisson distribution, Profile monitoring, Shewhart control chart, Principal component analysis, Ridge estimation.*

Cusum and Ewma Charts for Identifying Anomalies in Com-Poisson Profiles under Multicollinearity

Ulduz Mammadova[1*], **M. Revan Özkale**[1]

[1] Çukurova University, Science Faculty, Department of Statistics, 01330, Adana, Turkey

*Corresponding author e-mail: ulduzozel@gmail.com

Abstract

The real-life data sets from modern production processes may require addressing several problems simultaneously. We aim to detect shifts in generalized linear profiles while addressing the issues that arise from intercorrelated predictors known as multicollinearity and dispersed count responses. We propose utilizing modified CUSUM and EWMA control charts based on principal component regression (PCR) and r-k deviance residuals to detect shifts in the process. These deviance residuals are derived from the fitted Conway-Maxwell-Poisson (COM-Poisson) profile. While COM-Poisson distribution facilitates the analysis of under-, equi-, and overdispersed count data, PCR and r-k class estimators reduce the adverse effects of multicollinearity. The performance of the proposed control charts is illustrated through a simulation study and the analysis of the semiconductor manufacturing data set. Results of the CUSUM and EWMA control charts based on the deviance and ridge-deviance residuals are compared in terms of average and standard deviation of run length.

Key words: *COM-Poisson distribution, Profile monitoring, CUSUM control chart, EWMA control chart, Principal component analysis.*

Factor Analysis for Construct Validity: An Applied Study

Şeyma Koç[1*], Esra Yavuz[2]

[1] Kahramanmaraş Sütçü Imam University, Institute of Science and Technology, Department of Animal Science, Kahramanmaraş, Turkey
[2] Şırnak University, Cizre Vocational School, Department of Accounting and Tax, 73200, Cizre, Şırnak, Turkey

*Corresponding author e-mail: kockutlu1@gmail.com

Abstract

In this article, the 19-item information dimension of the scale, originally named "Sustainability Consciousness Questionnaire", was developed by Michalos, Creech, Swayze, Kahlke, Buckler & Rempel (2012) and updated by Gericke, Pauw, Berlung & Olsson (2018), whose factor structure was previously revealed. (latent variable) Construct validity was tested by applying it to a new data set of 307 people. IBM SPSS and AMOS statistical package programs were used in the analysis of the data. According to the results obtained in the study; It was found that the information dimension of the scale was gathered under three factors as in the original and it could explain 61.72% of the total variance. As a result of the reliability analysis, it was determined that the scale had a high level of reliability according to the Cronbach's-α coefficient (Cronbach's-α =0.923). In the confirmatory factor analysis, the improvement in the goodness of fit coefficients was examined by controlling the modification indices. In addition, the path coefficients of how much the 19 items belonging to the first level of the scale predict the latent variable are all significant. Among the items, it was found that the 13th and 14th items had the most effect ($\beta1= 0.828$, $p<0.01$).

Key words: *Goodness of Fit Indices, Structural Equation Modeling, Factor Analysis.*

INTRODUCTION

In the analyzes made in the fields of social sciences, educational sciences, psychology, biology and medicine, the relationship between measurable (observed) variables such as age, income, score, weight, pressure, etc., with non-directly measurable (latent) variables such as attitude, behavior, knowledge, anxiety, personality, intelligence. is frequently needed. It is also very important how confidently the items used in scale studies measure the variable to be measured. While looking at the relationship between several variables in classical methods, Structural equation modeling (SEM) detects the effects of latent and observed variables on each other simultaneously. Although it is a statistical method in which the results of factor analysis and regression analysis can be interpreted simultaneously, it is preferred by researchers even though it contains difficult techniques. Confirmatory factor analysis and exploratory factor analysis are a general and broad family of analyzes including multiple regression analysis and correlation analysis (Harrington, 2009).

Confirmatory Factor Analysis (CFA), which is the most widely used in the literature, is an extension of Explanatory Factor Analysis (EFA). It was first developed by Karl Jöreskog in 1967 in order to evaluate the construct validity of the data (Jöreskog, 1967). While EFA is trying to provide a determination function, to obtain information for forming hypotheses, CFA is used to test whether there is a sufficient level of relationship between these determined factors, which variables are related to which factors, whether the factors are independent from each other, and whether the factors are sufficient to explain the model (Özdamar, 2008). 2014). In summary, factor analysis as a whole (EFA or CFA) focuses on how and to what extent the observed variables are related to the underlying latent factors (Bryne, 2009).

The researcher may want to determine both the relationships between the items and the margins of error, the factor structure that will be formed with the help of these items, and the relationship between these factors, at the same time, with the survey study prepared in the Likert scale type. While DFA produces answers to all these questions, it also checks the prerequisites for these analyzes. Before starting CFA, it is necessary to analyze the normality of the data, outlier and outlier control, and missing data analysis. Whether the questions in the scale form a whole in a homogeneous structure is determined by the Cronbach Alpha Coefficient (Cronbach, 1951). If the questions in the form are reliable, this coefficient will be close to 1.

In this study, the validity of the Sustainability Knowledge dimension, which is one of the 3 different dimensions of the Sustainability Awareness Scale, was subjected to SEM processes with a new data set. Outlier analysis, item analysis, explanatory factor analysis, reliability analysis and confirmatory factor analysis, which should be used in the test of construct validity, were tried to be demonstrated in practice.

MATERIAL AND METHODS

Material

In this study, Sustainability Consciousness Questionnaire, originally called "Sustainability Consciousness Questionnaire", was developed by Michalos, Creech, Swayze, Kahlke, Buckler, and Rempel (2012) and updated by Gericke, Pauw, Berlung, and Olsson (2018), whose factor structure was revealed by using explanatory factor analysis. Consciousness Scale was used. The original scale consists of 50 items and 3 dimensions (Sustainability Knowledge, Sustainability Attitude and Sustainability Behavior) and is in the 5-point Likert type. The knowledge dimension includes 19 items, and the Attitude and Behavior dimensions include 14 and 17 items, respectively. Each dimension contains items belonging to environmental, social and economic components. The scale developed for primary and secondary school students was updated in the following years and applied to teachers as well. The descriptive mean of the scale's Sustainability Awareness dimension is M= 5.197 (SD= 1.112).

Methods

The Collection of the Data

The data of this study were drawn randomly from a survey of 1943 people applied through google forms between 17.3.2021-1.6.2021. There is no missing value in the data. After the extreme value (outlier) analysis, the number of data decreased to 297. The universe of the study is 7.,8.,9.,10. class students. Since the findings obtained in the study were not tried to be generalized to any universe, the universe-sample relationship was not entered and was carried out in a theoretical framework. There are different software used in the literature to perform SEM analysis. In this study, AMOS Version 23, which is preferred by researchers, was used. The exploratory factor analysis and the control of the prerequisites for these analyzes were made in SPSS Version 26 software.

Statistical Analysis

In this study, exploratory factor analysis was performed with SPSS Version 26 and a new data set on the information dimension consisting of the first 19 items of the Sustainability Awareness Scale, and then Confirmatory Factor Analysis was performed with AMOS Version 23. There is no missing value in the data. The normality of the data and the extreme values were checked. Normality assumption was checked with skewness and kurtosis values. In the exploratory factor analysis, 19 items belonging to the latent variable were grouped under 3 factors. The factorization structure is compatible with the original source. With item analysis in the Amos program, it is possible to determine the fit of the model as a whole with more than one model goodness coefficient. In addition, the path coefficients that give the latent variable explanation coefficient of each item can also be looked at. If acceptable coefficient ranges are not reached when the program is first run, evidence of mismatch is captured by modification indices (change indices), which can be conceptualized as a degree of freedom χ^2 statistic (Jöreskog & Sörbom, 1993).

Modification indices detect highly correlated error terms. By assigning covariance between these terms, it is tried to reach acceptable goodness of fit coefficients. In addition to the modification process, it is possible to remove items from the analysis for acceptable model fit. However, this move should not be made regardless of the significance of the path coefficients, the factor load of the items, and the item validity coefficients in the Item-Total statistics table. In Confirmatory Factor Analysis, the $\chi 2$ fit test, that is, the CMIN/DF value, is expected to be between 2 and 5 (Kelloway, 1998). Since the value of CMIN, i.e. $\chi 2$ (N-1) is equal to the value of (N-1)Fmin (N Sample volume F is the minimum fit function for any calculation methods used such as ML (Maximum Likelihood), GLS (Generalized Least Squares) or ULS (Unweighted Least Squares)) (Erkorkmaz et al. ., 2012. Due to the estimated number of parameters and sample size used in its formula, it can take values up to very high levels (Kline, 2011).

Researchers have taken these limitations into account and suggested more practical and utilitarian fit indices (Gerbing & Anderson, 1993; Hu & Bentler, 1995; Marsh, Balla & McDonald, 1988; Tanaka, 1993; Bryne, 1993). Other than the χ^2 chi-square test (CMIN/DF), the ones that are frequently used in the literature are; Root Mean Square Error of Approximation (RMSEA); Comparative Fit Index (CFI, Comparative Fit Index); Goodness of Fit Index (GFI, Goodness of Fit Index); Root Mean Errors (SRMR); It can be listed as scaled fit index (NFI; Normed Fit Index). In this study, the theoretical formulas of these indexes are not mentioned. Threshold values of these indices in various sources are given in Table 1.

Table 1. Threshold values of goodness of fit coefficient according to different sources

Reference Comment	Critical Value	Comment
Hooper, Coughlan and Müllen, 2008	GFI≥0.95	Perfect fit
Kline, 2011	0.95≥GFI≥90	Good fit
Anderson and Gerbing, 1984	0.85≥GFI≥0.90	Acceptable
Hu and Bentler, 1999	NFI≥0.95	Perfect fit
Kline, 2011	0.95≥NFI≥90	Good fit
Hooper, Coughlan and Müllen, 2008	0.80≥NFI≥0.90	Acceptable
Hooper, Coughlan and Müllen, 2008	0.03≥RMSEA	Perfect fit
Jöreskog and Sörborn, 1993	0.05≥RMSEA≥0.03	Good fit
Jöreskog and Sörborn, 1993	0.08≥RMSEA≥0.05	Acceptable
Tabachnick and Fidell, 2011	SMRM≈0	Perfect fit
Kline, 2011	0.05≥SMRM	Good fit
Anderson and Gerbing, 1984	0.08≥SMRM>0.05	Acceptable

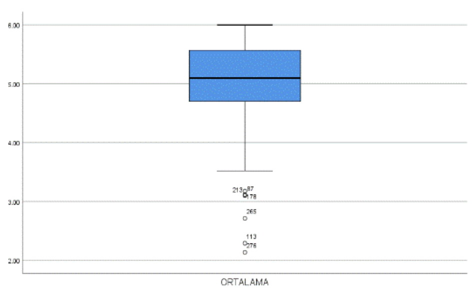
RESULTS

Demographic information of the participants participating in the study is given in Table 2.

Table 2. Demographic Information

	Frequency	Percentage
Gender		
Girl	146	49.1
Male	151	50.8
Class		
7. Class	62	20.9
8. Class	79	26.6
9. Class	78	26.3
10. Class	78	26.3
Mother Education Status		
Illiterate	10	3.3
Primary school	107	36.0
Middle School	51	17.2
High school	90	30.3
University	39	13.1
Father Education Status		
Illiterate	7	2.0
Primary school	69	23.2
Middle School	45	15.1
High school	112	37.7
University	65	21.9
Do You Have Own Room		
Yes	230	77.4
No	67	22.5
Do you have a computer at home		
Yes	217	73.1
No	80	26.9

Since the likert scale of the expressions to be analyzed affects the normality test, the skewness and kurtosis coefficients, which are the indicators of the normal distribution of each expression, were examined. In practice, it is seen that various coefficients regarding skewness and kurtosis values are ±1.5 (Tabachnick & Fidell, 2001) according to some authors and ±2 (George & Mallery, 2010) according to some authors. ±1.5 coefficient, which is a close value, is taken as basis.



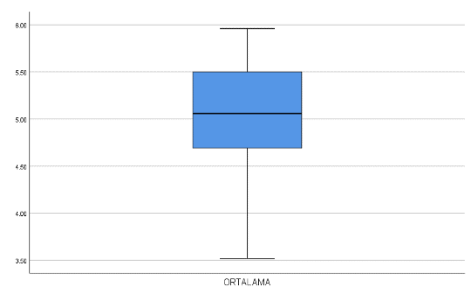
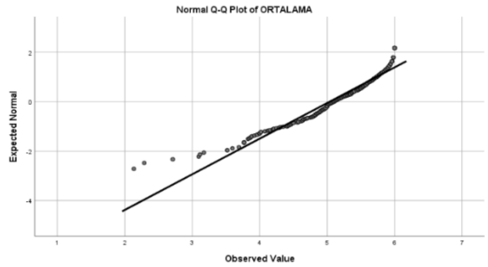


Figure 1. Boxplot for outlier (outlier) values check **Figure 2.** Boxplot when outliers are extracted



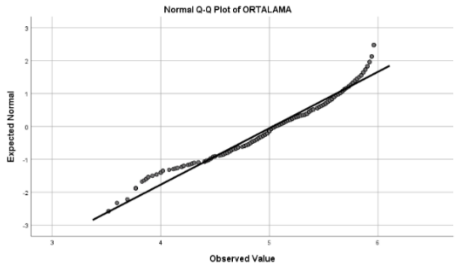


Figure 3. q-q plot for outlier (outlier) values control **Figure 4.** q-q plot when outliers are removed

The answers of some participants were outside the normal distribution values of ±1.5 at the confidence interval of 0.05. The data entry number of the participants who were not included in the distribution is shown in Figure 1. A total of 10 participants were excluded from the analysis by deleting them, as this would negatively affect the analyzes and factor dimensions to be made within the scope of factorization. After examining the contradictions between the answers, finding the values outside the answer range, and cleaning the data with extreme values, skewness=-0.559; std.err=0.141 and kurtosis=-0.313; It was found to be std.err=0.282. After clearing out the outliers, Boxplot and qq plot are given in Figure 2 and Figure 4, respectively.

Before the exploratory factor analysis, the reliability analysis of the item was also done. The Cronbach's Alpha Coefficient for the reliability analysis is 0.923. In addition, item-total (Item-Total) statistics were also examined. The results are presented in Table3. According to the Cronbach's Alpha if Item Deleted column, which shows the change in Cronbach's Alpha value in case of item deletion, it is seen that the coefficient will decrease if any item is deleted. According to the analysis output, it can be read that the coefficient will be preserved if only the M3 and M6 items are deleted. Similarly, if item M5 is deleted, it seems that the Cronbach's Alpha value will decrease from 0.923 to 0.918. This situation provides guidance to the researcher regarding the decision of item deletion while checking the model goodness coefficients in the confirmatory factor analysis, which is the next step.

Table 3. Item-Total Statistics

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
M1	93.34	157.779	.443	.922
M2	93.37	157.057	.491	.921
M3	93.61	155.273	.437	.923
M4	93.28	155.649	.567	.920
M5	93.57	150.928	.639	.918
M6	93.72	154.378	.458	.923
M7	93.60	151.504	.635	.918
M8	93.37	153.742	.536	.921
M9	93.39	152.571	.650	.918
M10	93.49	152.582	.589	.919
M11	93.47	151.825	.648	.918
M13	93.51	149.035	.724	.916
M14	93.56	147.571	.725	.916
M16	93.74	147.867	.643	.918
M17	93.73	149.947	.625	.919
M18	93.70	149.725	.599	.919
M19	93.60	149.782	.585	.920

In order to continue the factor analysis, the data must come from a multiple normal distribution. In order to detect this, the Bartlett sphericity test is performed. KMO and Bartlett tests were performed to determine whether the data set was suitable for factorization. The analysis output giving the test statistics values is given in Table 4.

Table 4. KMO and Bartlett' s Tests

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.913
Bartlett's Test of Sphericity	Approx. Chi-Square	3115.301
	df	171
	Sig.	.000

The KMO value, which indicates the suitability of the data for factor analysis, is required to be greater than 0.7 (Bartlett, 1950). After determining the suitability of the data for factorization, exploratory factor analysis was performed. The output of common variance values is given in Table 5.

Tablo5. Ortak Varyans (Communalities)

Communalities		
	Initial	Extraction
M1	1.000	.626
M2	1.000	.643
M3	1.000	.546
M4	1.000	.637
M5	1.000	.570
M6	1.000	.428
M7	1.000	.520
M8	1.000	.755
M9	1.000	.701
M10	1.000	.533
M11	1.000	.513
M12	1.000	.625
M13	1.000	.705
M14	1.000	.713
M15	1.000	.633
M16	1.000	.724
M17	1.000	.619
M18	1.000	.558
M19	1.000	.554

When the Common variance (Communality) table of the Exploratory Factor Analysis is examined, there is no value with a common variance value below 0.30. In this case, it is another clue that item deletion should not be preferred for this data set.

In exploratory factor analysis, there are 3 components with eigenvalues greater than one. This means that the 19 items used in the scale will be grouped under 3 factors. In the original of the scale, the items in this dimension were distributed over 3 factors. The percentage of explanation of the total variance of these factors is 61,072%. The graph showing the total variance and factorization structure explained in the program output table Table 5 is given in Table 6 and Figure 5, respectively.

Table 5. Explained total variance table

Components	Eigenvalues			Total Variance Explained		
	Total	Percent of variance	Percent of Shaped Variance	Total	Percent of variance	Percent of Shaped Variance
1	8.124	42.759	42.759	5.276	27.768	27.768
2	2.374	12.496	55.256	3.827	20.144	47.912
3	1.105	5.816	61.072	2.500	13.159	61.072
4	.929	4.890	65.962			
5	.766	4.033	69.995			
6	.650	3.423	73.418			
7	.605	3.187	76.605			
8	.542	2.854	79.459			
9	.511	2.692	82.151			
10	.496	2.609	84.761			
11	.484	2.547	87.308			
12	.434	2.284	89.592			
13	.382	2.010	91.602			
14	.341	1.797	93.399			
15	.305	1.607	95.006			
16	.283	1.489	96.495			
17	.257	1.355	97.849			
18	.224	1.180	99.029			
19	.185	.971	100.000			

As seen in Figure 5, a sharp return is observed after the first third eigenvalue. This is a sign that the scale will be three-dimensional.

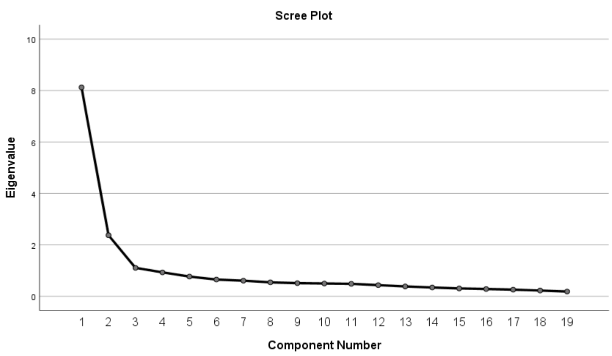


Figure 5. Scree plot showing the factorization structure

The rotated factor matrix giving the factor structure of the items is shown in Table 6.

Table 6. Rotated factor matrix

	Components		
	1	2	3
M16	.830		
M13	.785		
M14	.777		
M17	.768		
M15	.736		
M18	.724		
M12	.665		.371
M19	.600		.441
M11	.555		.383
M1		.785	
M2		.775	
M4		.759	
M3		.729	
M6		.616	
M5		.614	.320
M7	.391	.555	
M8			.818
M9	.362		.722
M10	.351		.596

When Table 6 is examined, it is seen that some M5, M7, M9, M10, M11, M12, M19 items are included in two factors. Items that create loads in more than one factor are called overlapping items, and if the fraction between two loads is less than 0.1, it is recommended to exclude the item from the analysis (Pallant, 2002). Since the load difference between the overlapping items in the table was greater than 0.1, the next step was continued without deleting the item.

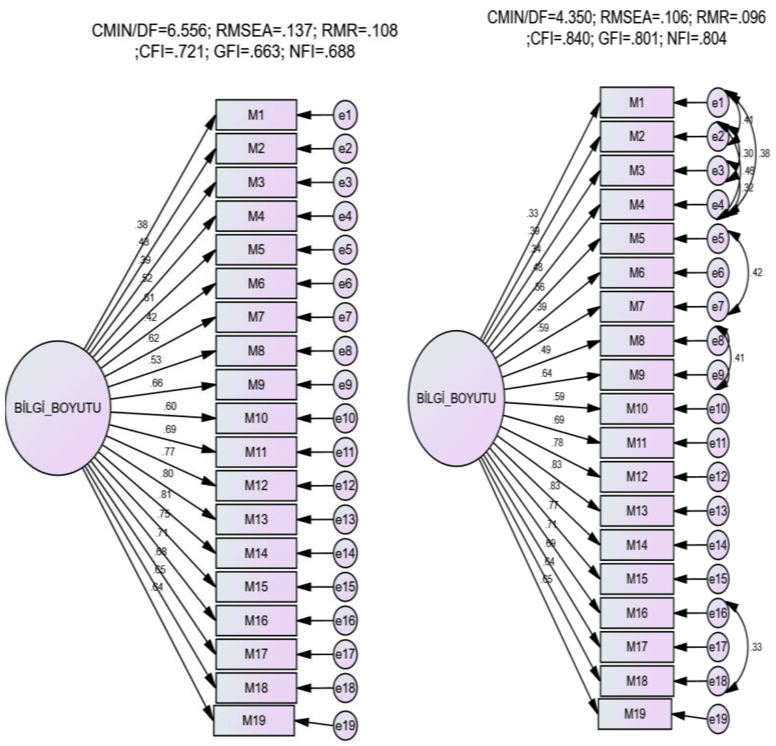


Figure 6. Path diagram before and after covariance assignments (Modification)

The path diagram of the items in each dimension created in Amos is given in figure 6. Since the coefficients of goodness of fit and path coefficients were not at the desired level when the model was first run, a modification was made between the error terms with a high correlation between them. The Amos program gives the correlations between each variable and the improvement that will occur in the goodness of fit and path coefficients after the covariance assignment to be defined between them. The highly correlated error terms were determined from the "modification indices" tab in the output interface, and covariance assignment (modification) was made one by one, starting with the error terms with the highest correlation. The program algorithm rearranges the goodness-of-fit coefficients for each modification. For this reason, the modification process should be done one by one, starting with the error terms with the highest correlation, checking after each start-up. The improvement in the goodness-of-fit coefficients after the modification moves are given in Table 7.

Table 7. Improvement in covariance assignments and cohesiveness coefficients

	CMIN/DF	RMSEA	SRMR	CFI	GFI	NFI
Covariance Assignments						
When the model first runs	6.556	0.137	0.108	0.721	0.663	0.668
e1-e2 between	6.077	0.131	0.106	0.746	0.702	0.713
e2-e4 between	5.886	0.128	0.105	0.757	0.712	0.723
e1-e4 between	5.406	0.122	0.103	0.783	0.744	0.748
e3-e4 between	5.328	0.121	0.102	0.788	0.747	0.753
e2-e3 between	5.183	0.119	0.099	0.796	0.754	0.761
e5-e7 between	4.856	0.114	0.098	0.814	0.776	0.778
e8-e9 between	4.527	0.109	0.098	0.831	0.795	0.794
e16-e18 between	4.350	0.106	0.096	0.840	0.801	0.804

Before and after the modification assignments, the goodness-of-fit coefficients showed correct improvement to acceptable values. A total of 8 modifications were made. As long as this modification process is continued, the model fit coefficients will improve each time. Although the number of modifications is related to the number of items, there is no consensus on how many modifications should be made.

Table 8. Regression coefficients of the items

Structural paths			Regression coefficients	Std. Regression coefficients	Critical Ratio	P value
M1	<---	Unobservable latent variable (information dimension)	1	0.326		
M2	<---		1.118	0.39	5.71	***
M3	<---		1.301	0.342	4.15	***
M4	<---		1.374	0.483	5.988	***
M5	<---		2.018	0.557	5.082	***
M6	<---		1.506	0.387	4.417	***
M7	<---		2.079	0.59	5.169	***
M8	<---		1.733	0.487	4.861	***
M9	<---		2.083	0.644	5.289	***
M10	<---		2.071	0.587	5.163	***
M11	<---		2.338	0.688	5.371	***
M12	<---		2.971	0.78	5.507	***
M13	<---		2.975	0.828	5.564	***
M14	<---		3.202	0.828	5.565	***
M15	<---		3.074	0.767	5.491	***
M16	<---		3.012	0.712	5.41	***
M17	<---		2.673	0.685	5.367	***
M18	<---		2.63	0.643	5.285	***
M19	<---		2.701	0.648	5.298	***

Table 8 shows the path coefficients (regression coefficients) showing the relationship between each item and the latent variable. After the modification processes, it was seen that all of the path coefficients of the items were significant. It can be said that the items with the highest effect among the items related to the latent variable are the 13th and 14th items. There is a positive linear relationship between these two items and the latent variable ($\beta 1= 0.828$, $p<0.01$).

DISCUSSION AND CONCLUSION

For construct validity in the study; outlier (outlier) values analysis, explanatory factor analysis and confirmatory factor analysis, item analysis were performed in order of priority.

Examining the contradictions between the answers, the values outside the answer range and the data with extreme values affect both the factorization structure and the distribution of the data. In order to clear these, outlier value analysis was performed, since in the answers of the participants numbered 288, 276, 265, 213, 178, 169, 145, 131, 113, all questions were marked the same option, leaving all the items blank or answering lightly, the analysis was done. has been excluded. After this extraction, when the program was run again, the normality of the data was determined by the skewness and kutosis values (skewness=-0.559; std.err=0.141 and kurtosis=-0.313; std.err=0.282).

In order to reveal the construct validity, factor analysis was performed by adhering to the original structure of the scale. As a result of the exploratory factor analysis, 3 factors with eigenvalues greater than 1 and explaining 61.072% of the total variance in the Knowledge subscale were obtained. The factor loads of the items in the first factor ranged from 0.555 to 0.830, the items in the second factor ranged from 0.555 to 0.785, and the items in the third factor ranged from 0.596 to 0.818.

Confirmatory factor analysis of the three-factor structure obtained as a result of EFA was used to control the fit indices. According to the results obtained, $\chi^2/df=4.330$; RMSEA=0.106; SRMR=0.096; CFI=0.840; GFI=0.801; NFI = 0.804. The Turkish version of the Sustainable Consciousness Scale used in the study has been used in different studies before. In the study of Yüksel and Yıldız (2019), Confirmatory Factor Analysis was used to reveal whether the scale was validated in the sample of Turkish students, and the goodness of fit coefficients of the knowledge dimension ($\chi^2/df=4.330$; CFI = 0.924; IFI = 0.935; NFI = 0.936; RMSEA). = 0.064) were determined.

Finally, the path (regression coefficients) of each item is given in order to reveal the effect of the items on the knowledge dimension. It has been determined that the item with the highest effect among the items related to the latent variable is the 13th and 14th items ($\beta_{1=}$ 0,828, $p<0,01$).

References

Anderson JC, Gerbing, DW, 1984. The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis, Psychometrika, 49(2), 155–173.

Bartlett MS, 1950 Test of significent in factor analysis, British Journal of Psychology, Statistical Section, 3,77-85.

Byrne BM, 2009. Structural equation modeling with AMOS: basic concepts, applications and programming. 2nd ed., Taylor & Francis Group, London.

Cronbach LJ, 1951. Coeffıcient Alpha and the Internal Structure of Tests, Psychometrika, Vol. 16, 297-334.

George D, Mallery M, 2010. SPSS for Windows Step by Step: A Simple Guide and Reference. 10 ed., Pearson, Boston.

Gerbing DW, Anderson JC, 1993. Monte Carlo evaluations of goodnessof-fit indices for structural equation models. Testing structural equation models. (Editors: Bollen KA, Long JS), Sage publication, pp.40–65, USA.

Harrington D, 2009. Assessing Confirmatory Factor Analysis Model Fit and Model Revision. Confirmatory Factor Analysis. 1st ed., Oxford University Press, New York.

Hooper D, Coughlan J, Mullen, MR, 2008. Structural Equation Modelling: Guidelines for Determining Model Fit, The Electronic Journal of Business Research Methods, 6(1), 53-60.

Hu LT, Bentler PM, 1995. Structural equation modeling: Concepts, issues and application. Evaluating model fit (Editor: Hoyle RH.), Sage publication, pp. 77-99, USA.

Hu LT, Bentler, PM, 1999. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1-55.

Jöreskog KG, Sörbom D, 1993. LISREL 8: Structural equation modeling with the SIMPLIS command language. Scientific Software International, Chicago.

Kelloway EK, 1998. Assessing Model Fit. Using Lisrel for Structural Equation Modeling. 3rd ed., Sage Publications, USA.

Kline RB, 2011. Hypothesis Testing:Principles and Practice of Structural Equation Modeling. 3rd ed., The Guilford Press, New York.

Marsh HW, Balla JR., McDonald RP, 1988. Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size, Psychological Bulletin, 103(3), 391–410.

Özdamar K, 2004. Tabloların Oluşturulması, Güvenirlik ve Soru Analizi. Paket Programlarla İstatistiksel Veri Analizi-1. 5th ed., Kaan Kitabevi, Eskişehir.

Pallant J, 2002. SPSS Survival Manual:A Step by Step Guide to Data Analysis Using IBM SPSS. 7th. Ed., London.

Tabachnick BG, Fidell LS, 2001. Using Multivariate Statistics. 4th Edition, Allyn and Bacon, Boston.

Tanaka JS, 1993. Multifaceted conceptions of fit in structural equation models Testing structural equation models. (Editors: Bollen JA, Long JS), Sage Publictaion, pp. 10–39, Newbury Park.

Yüksel Y, Yıldız B, 2019. Adaptation of sustainability consciousness questionnaire. Erciyes Journal of Education, 3(1), 16-36.

Acknowledgment

The data of this study were used with her consent from a 1943 questionnaire administered by the Middle East Technical University, Department of Primary Science and Mathematics Master's student Duygu Taşcı Ergün through google forms between 17.3.2021-1.6.22021.

Conflict of Interest

The authors have declared that there is no conflict of interest.

Author Contributions

S.K.: initiated the research idea, developed, organized, analyzed and interpreted the data and wrote the manuscript E.Y.: developed, structured the paper and edited the manuscript.

Determining the Effects and Removal Capacity of Malachite Green on Eggshells Using MANOVA

Pınar Karagülle[1*], Berna Yazıcı[1], Semra Malkoç[2], Ece Özgören[1]

[1] Eskişehir Technical University, Faculty of Science, Department of Statistics, 26470, Eskişehir, Turkey
[2] Eskişehir Technical University, Faculty of Engineering, Department of Environmental Engineering, 26555, Eskişehir, Turkey

*Corresponding author e-mail: pinarkaragulle@eskisehir.edu.tr

Abstract

In this project study, adsorption capacity and removal efficiency from eggshells with dyestuff such as malachite green the effects of factors have been researched. The data were obtained as a result of experiments conducted in Eskişehir Technical University Environmental Research Center. These obtained data will be analysed in Minitab package program, MANOVA (Multivariate Analysis of Variance) will be applied and the results will be interpreted. The aim of this project study is to determine the effects of different factor levels on the dyestuff's adsorption capacity and removal efficiency of eggshells with dye materials such as malachite green. There are three factors (mixing time (min), temperature (°C), initial concentration(mg/L)) which is significant for Malachite Green. One of them is initial concentration and there are five levels related to it. These levels are 25, 50, 100, 150 and 200 (mg/L). When the initial concentration levels are investigated, it is seen that the removal is higher at the level of 150 mg/L. In other words, the factor that is significant is the initial concentration and the corresponding removal at the 150 mg/L level is high and should be adopted.

Key words: *Multivariate Data Analysis, Dyestuffs, Eggshell, Adsorption Capacity, Removal Efficiency*

INTRODUCTION

Due to its porous structure, eggshell is a method frequently used in dyestuff studies. Malachite green dyestuff to be used in the experiment can be given as example of basic dye. During the experiment, it is planned to eliminate the factors that may adversely affect the adsorption capacity and removal efficiency, which are dependent variables, and to minimize these factors. Multivariate analysis of variance (MANOVA) will be applied to the data obtained for dependent variables with the Minitab package program. In this way, it is aimed to reduce the effects of damage to living things and the environment and to make necessary analyzes to remove these dyestuffs. The strength of MANOVA is affected by the correlations of the dependent variables and the effect sizes associated with these variables. In addition to this, the size of data reaches quite large proportions beyond imagination and essential analyses can be performed with MANOVA (Stevens JP, 2002).

MATERIAL AND METHODS

Material

Biosorbent Preparation

The eggshells used in the studies were obtained from a local business. Eggshells were washed with tap water and then rinsed with distilled water. It was kept at room temperature for 24 hours. After 24 hours, the shells were kept in an oven at 80 °C for 2 hours to remove moisture. After dehumidification, it was ground in a grinding machine and turned into powder. The resulting eggshell powders were passed through a 250 µm sieve.

Dyestuffs Preparation

While preparing malachite green solutions, 1000 mg/L stock solution was obtained by weighing 1 g of dyestuff and adding 1000 mL of distilled water. Solutions of different concentrations to be used during the experiments were used by dilution from the stock solution.

Methods

The Collection of the Data

In this experiment, it was aimed to investigate the factors and levels affecting the adsorption capacity and removal efficiency of eggshells and dyestuff such as malachite green. By providing the necessary conditions, by using eggshells, the necessary experiments for the adsorption capacity and removal efficiency of the dye malachite green will be carried out with the help of the team.

In the experiment, there are 5 factors (pH, biosorbent amount (g), mixing time (min), dyestuff concentration (mg/L), temperature (°C)) for dyestuff, and it is planned to use different levels for each factor. 5 Factors (pH, amount of biosorbent (g), mixing time (min), concentration of dyestuff (mg/L), temperature (°C)) are included in the experiment for each dyestuff and it is planned to use different levels for each factor. The levels of the pH factor; 2,3,4,5,6,7,8. The amount of biosorbent (g) is 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4. The dyestuff concentration (mg/L) level is 25, 50, 100, 150, 200. Mixing time (min) levels; 1, 5, 10, 20, 30, 40, 50, 60, 90, 120, 150, 180 for malachite green and finally, temperature (°C) levels are planned as 25-35-45 °C.

Statistical Analysis

Two-way multivariate analysis of variance (two-way MANOVA) is generally considered as an extension of two-way ANOVA for situations where there are two or more dependent variables. The main purpose of two-way MANOVA is to understand whether there is an interaction between two independent variables on two or more dependent variables. a two-way MANOVA has generally one primary aim: to understand whether the effect of one independent variable on the dependent variables (collectively) is dependent on the value of the other independent variable. This is called an "interaction effect". However, if no interaction effect is present, you would normally be interested in the "main effects" of each independent variable instead. This is somewhat akin to assessing the effect that an independent variable has on the dependent variables collectively when "ignoring" the value of the other independent variable (Barker HR, Barker BM, 1984).

The mathematical model for the MANOVA is as follows:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk}$$

$$i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, p$$

$\epsilon_{ijk} \sim N(0, \Sigma)$ Σ is variance – covariance matrix

ϵ_{ijk} ‘s are mutually independent.

There are four different statistics in multivariate analysis of variance. The p-values obtained as a result of these tests decided whether MANOVA is significant or not. These tests are Wilks’ Lambda, Lawley – Hotelling’ s Trace, Pillai’s Trace and Roy’s Largest Root values.

Wilks' lambda is a test statistic used in multivariate analysis of variance (MANOVA) to test whether there are differences between the means of identified groups of subjects on a combination of dependent variables. Wilks' test - the most commonly used test because it was the first derived and has a well-known F approximation. Wilks' lambda is a direct measure of the proportion of variance in the combination of dependent variables that is unaccounted for by the independent variable (the grouping variable or factor). If a large proportion of the variance is accounted for by the independent variable, then it suggests that there is an effect from the grouping variable and that the groups have different mean values (Everitt BS, Dunn G, 1991; Polit DF, 1996).

Define Wilks' Lambda as follows:

$$\Lambda_{p,h,e} = \frac{|E|}{|E+H|} = \prod_{j=1}^{p}(1-\theta_j)$$

with $e \geq p$.

The following approximation based on the F-distribution is used to determine significance levels:

$$F_{ph,ft-g} = \frac{(ft-g)\left(1-\Lambda^{1/t}\right)}{ph\Lambda^{1/t}}$$

where

$$f = e - \frac{1}{2}(p-h+1)$$

$$g = \frac{ph-2}{2}$$

$$t = \begin{cases} \sqrt{\dfrac{p^2h^2-4}{p^2+h^2-5}} & if \quad p^2+h^2>5 \\ 1 & otherwise \end{cases}$$

This approximation is exact if p or $h \geq 2$.

RESULTS

Table 1. Data for optimum pH value

pH	Absorption Capacity (qe)	Removal Efficiency (%)
2	7.978	45.588
3	10.570	60.399
4	12.224	69.853
5	14.467	82.668
6	13.456	76.891
7	14.761	84.349
8	14.026	80.147

It was observed that the removal efficiency and absorption capacity of 7 pH was high in the single factor experiment performed in once.

Table 2. Data for optimum amount of biosorbent

Biosorbent Amount (g)	Absorption Capacity (qe)	Removal Efficiency (%)
0.05	23.015	67.024
0.1	13.548	78.908
0.15	8.015	70.021
0.2	6.057	70.557
0.25	5.147	74.946
0.3	3.162	55.246
0.35	3.708	75.589
0.4	3.116	72.591

It was observed that the absorption capacity of 0.05 g amount of biosorbent and the removal efficiency of 0.1 g amount of biosorbent were high in the single factor experiment performed in once.

Table 3. Mixing time and temperature dependent removal efficiency and adsorption capacity

Mixing time (min)	**Absorption Capacity (qe)**			**Removal Efficiency (%)**		
Temperature (°C)	**25°C**	**35°C**	**45°C**	**25°C**	**35°C**	**45°C**
1	6.397	6.085	2.427	48.671	41.118	19.384
5	8.548	8.125	4.485	65.035	54.907	35.83
10	8.732	8.511	7.371	66.434	57.516	58.884
20	9.651	10.129	9.217	73.427	68.447	73.627
30	10.184	10.974	10.028	77.483	74.161	80.103
40	10.515	11.765	10.807	80	79.503	86.329
50	10.662	12.004	11.408	81.119	81.118	91.131
60	10.46	12.279	11.447	79.58	82.981	91.439
90	10.257	12.353	11.412	78.042	83.478	91.16
120	10.864	12.886	11.728	82.657	87.081	93.686
150	10.874	11.451	12.433	83.653	91.334	95.644
180	11.528	11.261	12.521	88.687	89.824	96.323

Table 4. MANOVA Tests for Mixing Time (min)

Criterion	**Test Statistic**	**F**	**Num**	**Denom**	**P**
Wilks'	0.09738	4.209	22	42	0.000
Lawley-Hotelling	7.35984	6.691	22	40	0.000
Pillai's	1.08851	2.388	22	44	0.007
Roy's	7.09063				

*s = 2   m = 4.0   n = 9.5

Since the results of four different multivariate statistics for the mixing time (min) variable are $p<0.05$, it is significant at the 0.05 level. There is significant difference in terms of dependent variables. In other words, the change in the mixing time (min) levels creates a significant difference in the removal efficiency and absorption capacity of malachite green.

Table 5. MANOVA Tests for Temperature (°C)

Criterion	**Test Statistic**	**F**	**Num**	**Denom**	**P**
Wilks'	0.30233	8.596	4	42	0.000
Lawley-Hotelling	2.30714	11.536	4	40	0.000
Pillai's	0.69780	5.895	4	44	0.001
Roy's	2.30694				

*s = 2   m = -0.5   n = 9.5

Since the results of four different multivariate statistics for the temperature (°C) variable are $p<0.05$, it is significant at the 0.05 level. There is significant difference in terms of dependent variables. In other words, the change in the temperature (°C) levels creates a significant difference in the removal efficiency and absorption capacity of malachite green.

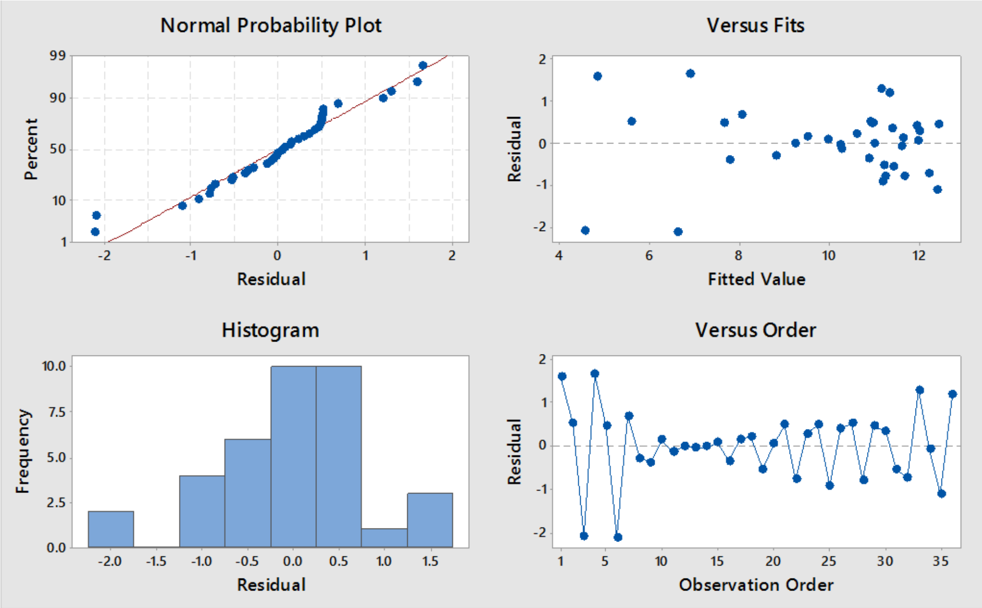


Figure 1. Residual Plots for Absorption Capacity (qe) for Malachite Green

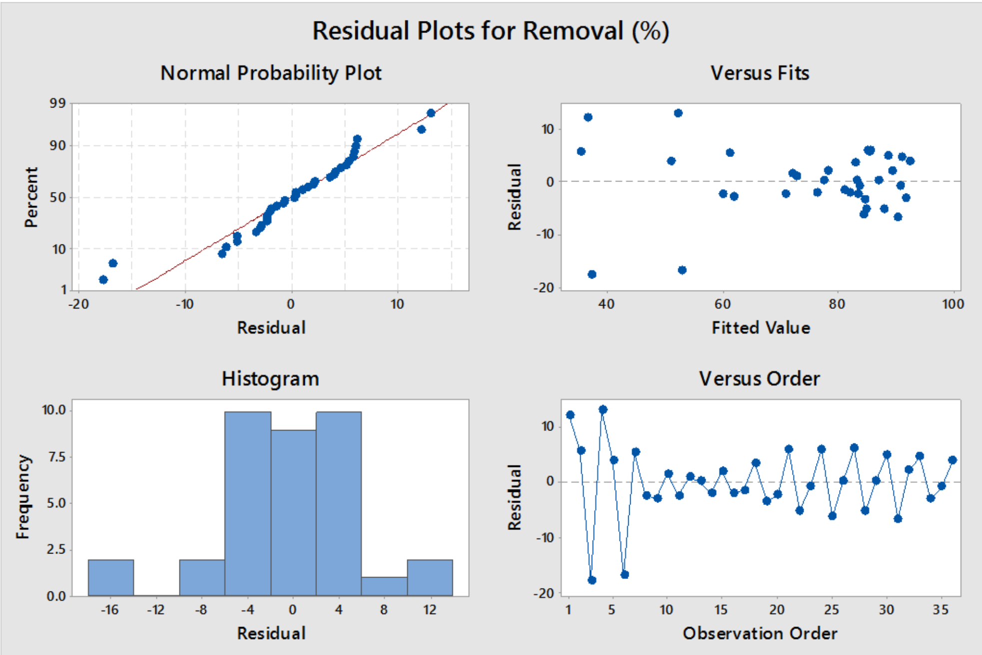


Figure 2. Residual Plots for Removal Efficiency (%) for Malachite Green

Table 6. Means for Mixing Time (min)

Mixing Time (min)	N	Mean	StDev	95% CI
1	3	36.39	15.20	(18.98, 53.81)
5	3	51.92	14.83	(34.93, 68.92)
10	3	60.94	4.80	(55.44, 66.45)
20	3	71.83	2.93	(68.47, 75.20)
30	3	77.25	2.98	(73.83, 80.67)
40	3	81.94	3.81	(77.57, 86.32)
50	3	84.46	5.78	(77.86, 91.05)
60	3	84.67	6.11	(77.67, 91.66)
90	3	84.23	6.59	(76.66, 91.79)
120	3	87.81	5.55	(81.41, 94.20)
150	3	90.21	6.07	(83.27, 97.15)
180	3	91.61	4.12	(86.88, 96.34)

When we analyze the factors as can be seen above, the factor which is significant is the mixing time (min) and there are 12 levels related to it. These levels are 1, 5,10,20,30,40,50,60,90,120,150 and 180. When the temperature levels are investigated, it is seen that the removal efficiency and absorption capacity is higher at the level of 180. In other words, the factor that is significant is the temperature and the corresponding removal efficiency and absorption capacity at the 180 level is high and should be adopted.

Table 7. Means for Temperature (°C)

Temperature (°C)	N	Mean	StDev	95% CI
25	12	75.40	10.81	(70.78, 80.01)
35	12	74.29	15.73	(67.57, 81.00)
45	12	76.13	25.32	(65.32, 86.93)

When we analyze the factors as can be seen above, the factor which is significant is the temperature and there are 3 levels related to it. These levels are 25 °C, 35 °C and 45 °C. When the temperature levels are investigated, it is seen that the removal efficiency and absorption capacity is higher at the level of 45 °C. In other words, the factor that is significant is the temperature and the corresponding removal efficiency and absorption capacity at the 45 °C level is high and should be adopted.

Table 8. Adsorption capacity and removal efficiency of dyestuff concentration and temperature

Initial Concentration (mg/L)	Absorption Capacity (qe)			Removal Efficiency (%)		
Temperature (°C)	**25°C**	**35°C**	**45°C**	**25°C**	**35°C**	**45°C**
25	10.864	12.886	11.728	82.657	87.081	93.686
50	26.011	25.404	24.607	92.726	90.564	97.07
100	50.221	46.967	52.669	89.898	93.865	97.093
150	65.184	74.522	79.564	80.664	95.433	97.396
200	91.581	102.059	97.971	83.563	94.744	97.755

Table 9. MANOVA Tests for Initial Concentration (mg / L)

Criterion	Test Statistic	F	Num	Denom	P
Wilks’	0.00285	31.026	8	14	0.000
Lawley-Hotelling	215.84283	161.882	8	12	0.000
Pillai’s	1.37899	4.441	8	16	0.005
Roy’s	215.22047				

s = 2 m = 0.5 n = 2.5

Since the results of four different multivariate statistics for the initial concentration (mg / L) variable are $p<0.05$, it is significant at 0.05 level. There is significant difference in terms of dependent variables. In other words, the change in the initial concentration (mg / L) levels creates a significant difference in the removal efficiency and absorption capacity of malachite green.

Table 10. MANOVA Tests for Temperature (°C)

Criterion	Test Statistic	F	Num	Denom	P
Wilks’	0.21687	4.016	4	14	0.023
Lawley-Hotelling	3.49924	5.249	4	12	0.011
Pillai’s	0.80739	2.708	4	16	0.068
Roy’s	3.46697				

*s = 2 m = -0.5 n = 2.5

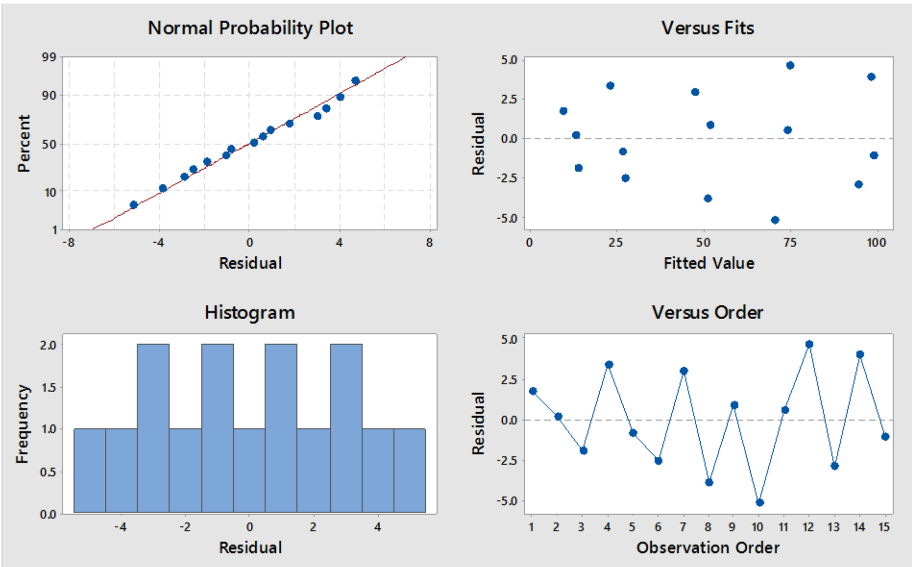


Figure 3. Residual Plots for Absorption Capacity (qe) for Malachite Green

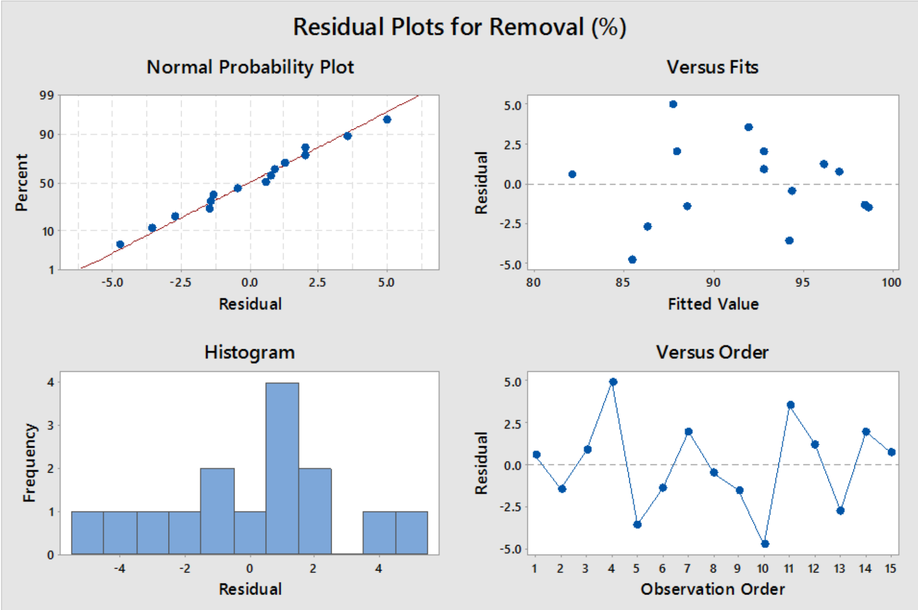


Figure 4. Residual Plots for Removal Efficiency (%) for Malachite Green

Table 11. Means for Initial Concentration (mg / L)

Initial Concentration (mg / L)	N	Mean	StDev	95% CI
25	3	87.81	5.55	(81.41, 94.20)
50	3	93.45	3.31	(89.61, 97.29)
100	3	93.62	3.60	(88.40, 98.83)
150	3	91.16	9.15	(77.90, 104.43)
200	3	92.02	7.48	(81.64, 102.40)

When we analyze the factors as can be seen above, the factor which is significant is the initial concentration and there are 5 levels related to it. These levels are 25, 50, 100, 150 and 200 mg/L. When the initial concentration levels are investigated, it is seen that the removal is higher at the level of 150 mg/L. In other words, the factor that is significant is the initial concentration and the corresponding removal at the 150 mg/L level is high and should be adopted.

Table 12. Means for Temperature (°C)

Temperature (°C)	N	Mean	StDev	95% CI
25	5	85.90	5.15	(62.37, 109.43)
35	5	92.34	3.48	(65.73, 118.94)
45	5	96.600	1.652	(70.186, 123.014)

When we analyze the factors as can be seen above, the factor which is significant is the temperature and there are 3 levels related to it. These levels are 25 °C, 35 °C and 45 °C. When the temperature levels are investigated, it is seen that the removal efficiency and absorption capacity is higher at the level of 45 °C. In other words, the factor that is significant is the temperature and the corresponding removal efficiency and absorption capacity at the 45 °C level is high and should be adopted.

DISCUSSION AND CONCLUSION

In this project study, adsorption capacity and removal efficiency from eggshells with dyestuff as malachite green the effects of factors have been researched. We have determined the dependent variables, factors, and their levels in the experiments. The data obtained from the different factor levels, malachite green dye with eggshells, in order to determine the effects on adsorption capacity and removal were analysed by multivariate analysis of variance in the Minitab package program.

There are three factors which is significant for Malachite Green. The first factor is mixing time (min) and there are nine levels related to it. These levels are 1, 5, 10, 20, 30, 40, 50, 60, 90, 120, 150 and 180. When the temperature levels are investigated, it is seen that the removal efficiency and absorption capacity is higher at the level of 150. In other words, the factor that is significant is the temperature and the corresponding removal efficiency and absorption capacity at the 180 level is high and should be adopted. The second factor is temperature and there are three levels related to it. These levels are 25 (°C), 35 (°C) and 45 (°C). When the temperature levels are investigated, it is seen that the removal efficiency and absorption capacity is higher at the level of 45 (°C). That is, the factor that is significant is the temperature and the corresponding removal at the 45 (°C). level is high and should be adopted. The last factor is initial concentration and there are five levels related to it. These levels are 25, 50, 100, 150 and 200. When the initial concentration levels are investigated, it is seen that the removal is higher at the level of 150 mg/L. In other words, the factor that is significant is the initial concentration and the corresponding removal at the 150 mg/L level is high and should be adopted.

References

Barker HR, Barker BM, 1984. Multivariate Analysis of Variance (Manova): A Practical Guide to Its Use in Scientific Decision Making, University of Alabama Press, USA.

Everitt BS, Dunn G, 1991. Applied Multivariate Data Analysis, Edward Arnold, London. 219 – 220.

Hair JF, Anderson RE, Tatham RL, 1995. Multivariate Data Analysis, 4 th ed., Pearson Collage Div.

Hamutoğlu R, Dinçsoy AB, Cansaran-Duman D, Aras S, 2012. Biyosorpsiyon, Adsorpsiyon ve Fitoremediasyon Yöntemleri ve Uygulamaları. Turk Hij Den Biyol Derg, p:238, p:240.

Howell D, 2002. Statistical Methods for Psychology, Duxbury.

Malkoç, S, Yazıcı, B, Şimşek, M, 2021. Optimization Of Removal Efficiency and Absorption Capacity of Dyestuffs by Aspergillus Sp. Using Manova. International Journal of Engineering and Innovative Research, 3(3):187-200.

Özdamar K, 1997. Paket Programlar ile İstatistiksel Veri Analizi, 649, Kaan Kitabevi, Ankara.

Polit DF, 1996. Data Analysis and Statistics for Nursing Research. Appleton and Lange, Stamford, Connecticut, 320 – 321.

Stevens JP, 2002. Applied Multivariate Statistics for the Social Sciences. Mahwah, NJ: Lawrence Erblaum.

Tatlı İ.A, 2003. Çeşitli Tekstil Boyarmaddelerin Adsorpsiyon /Biyosorpsiyonunun Karşılaştırılmalı Olarak Kesikli Sistemde Incelenmesi, Yüksek Lisans Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Kimya Mühendisliği Anabilim Dalı, Ankara.

Tinsley, H. A. E., Brown, S. D, 2000. Handbook of Applied Multivariate Statistics and Mathematical Modeling. 185 – 188.

MANOVA, Access address: Lesson 8: Multivariate Analysis of Variance (MANOVA) (psu.edu) ; Date of access: 12.01.2022

Two-Way ANOVA, Access address: Two-way analysis of variance - Wikipedia ;Date of access: 17.01.2022

One-Way MANOVA, Access address: One-way MANOVA in Stata | Step-by-step procedure including discussion of assumptions. (laerd.com) ; Date of access: 26.01.2022

Two-Way MANOVA, Access address: How to perform a two-way MANOVA in SPSS Statistics | Laerd Statistics; Date of access: 08.02.2022

Sample size, Access address: Sample Size Determination in Qualitative and Quantitative Research - Project Topics; Date of access: 25.03.2022

Eggshells, Access address: What is Egg Shell Quality and How to Preserve It (archive.org) ; Date of access: 12.02.2022

Dye, Access address: Boya - Vikipedi (wikipedia.org) Date of access: 07.01.2022

Dye, Access address: 12.hafta_Bakterilerin Boyanması ve Metotlar.pdf (ankara.edu.tr) ; Date of access: 17.01.2022

Dye, Access address: mikrobik.net - Bakteriyel Boyalar ve boyama yöntemleri ; Date of access: 16.01.2022

Dyestuffs, Access address: Boyar madde - Vikipedi (wikipedia.org) ; Date of access: 22.02.2022

Malachite green, Access address: https://tr.wikipedia.org/wiki/Malahit_ye%C5%9Fili ; Date of access: 05.01.2022

Minitab, Access address: https://www.veribilimiokulu.com/minitab-destekli-istatistiksel-analiz-egitimleri-ders-1-9-minitab-hakkinda-genel-bilgi-giris-menuler-ve-arac-cubuklari/ Date of access: 11.02.2022

Acknowledgment

This study was supported financially by TUBITAK 2209-A University Students Research Projects Support Program under grant number 1919B012107227 and by the Commission of the Scientific Research Projects of Eskişehir Technical University under grant number 20ADP230 and 22LÖP176.

Conflict of Interest

The authors have declared that there is no conflict of interest.

Author Contributions

%25 Pınar KARAGÜLLE

%25 Berna YAZICI

%25 Semra MALKOÇ

%25 Ece ÖZGÖREN

Analysis of by Some Non-Linear Mathematical Models the Number of Covid-19 Cases and Deaths in Macedonia

Şenol Çelik[1*]

[1] Bingöl University, Faculty of Agriculture, Department of Animal Science, Department of Biometrics and Genetics, Bingöl-Turkey

*Corresponding author e-mail: senolcelik@bingol.edu.tr

Abstract

In this study, it is aimed to determine the best mathematical model to explain the change in the number of covid-19 cases and deaths with some nonlinear models. For this purpose, 790-day cases and death toll records were used between 1 April 2020 and 30 May 2022 in Macedonia. In the study, Gompertz, Brody and Von Bertalanffy models were used. The coefficient of determination (R2) and mean squares of error (MSR) were used as criteria in determining the model that best describes the growth in cases and deaths. The model with a high coefficient of determination (R2) and a low mean square error (MSR) was chosen as the best model describing growth. In the Gompertz, Brody and Von Bertalanffy models, the R2 values for the number of cases were found to be 0.979, 0.976 and 0.982, respectively, while the PCR was found to be 215 028 231.409, 252 903 701.906 and 183 663 526,054, respectively. The R2 values for the death tolls were determined as 0.994, 0.983 and 0.995, and the PLA were 64049.467, 177041.928 and 56184,377, respectively. As a result, the Von Bertalanffy model is the model that best describes the growth in the number of covid-19 cases and deaths in Macedonia. According to the Von-Bertalanffy model, the maximum number of cases and deaths was estimated as 587482 and 12630, respectively. The average rate of increase in the number of cases and deaths was estimated as 0.002 and 0.003, respectively.

Key words: *Gompertz, Brody, Von Bertalanffy, Case, Death.*

INTRODUCTION

The corona virus (Covid-19) was detected in China in December 2019 in patients with severe acute respiratory syndrome as a member of the corona virus family. In a short time, it was understood that the Covid-19 virus was a dangerous epidemic, as it passed from person to person quickly and easily. Within a few months, the disease began to appear in all countries of the world, prompting governments and international organizations to take extraordinary measures as a public health crisis (WHO, 2020).

The Covid-19 pandemic is a unique phenomenon for today's societies, nation states and contemporary international systems. The societies and countries of our age have dealt with a wide variety of economic, political, ecological and military crises in the recent history, but for the first time they have faced an epidemic crisis that is not directly subject to any physical, geographical or socio-cultural borders. In this sense, the epidemic is a brand new situation for today's globalized national societies (Yellow, 2021).

The age profile, density and actions of the population against the disease constitute the basic components of the epidemic process. The corona virus epidemic, in which population mobility was also effective, spread to all continents except Antarctica in the first 6 months (Uzun and Oğlakcı, 2020).

While the loss of life due to Covid-19, which is distributed globally, continues to increase, the impact dimensions of the epidemic at the local and national level have become the focus of many studies. It has been a matter of curiosity which countries will be affected by this epidemic, which has shaken the world deeply. Since the early stages of the epidemic, researchers have resorted to risk analyzes and mathematical models to determine the level of impact of the epidemic despite limited data stores (Gao and Dong, 2020; Medford and Llimós, 2020; Natale et al., 2020; Onder et al., 2020; Çelik. , 2021).

The aim of this study is to establish and predict nonlinear growth models of cases and deaths during the pandemic in Macedonia.

MATERIAL AND METHOD

Material

It was acquired daily updates of the cumulative number of declared approved cases and deaths for the 2019-nCoV pandemic of Macedonia between April 1, 2020 and May 30, 2030, from Worldometer and WHO websites (Worldometer, 2021).

Method

From 1 April 2020 to 30 May 2022 in Macedonia, the number of covid-19 cases and deaths were recorded every day. Time-case and time-death data were analyzed by taking log information. Three different nonlinear growth models (Brody, Gompertz, and Von Bertalanffy) were used in the analyzes (Verhulst, 1838; Gurcan, et. al, 2017; Panik, 2014; Von Bertalanffy, 1957). The equations and inflection points of these models are presented in Table 1.

Table 1. Nonlinear models used to predict growth curves

Model	Equation	IPT	IPW
Brody	$y_t = A * (1 - b * \exp(-k * t))$	---	---
Gompertz	$y_t = A * \exp\,(-b * exp(-k * t))$	$\frac{\ln b}{k}$	A/e
Von Bertalanffy	$y_t = A * (1 - b * \exp(-k * t))^3$	$\frac{\ln 3b}{k}$	$A\frac{8}{27}$

In these models; A, b, k are growth curve parameters. A: Asymptotic magnitude, b: Initial size, k: Growth rate, t: Time, yt: Case/death, IPT: Inflection point case/death, IPW: Case or death at inflection point.

To determine the best model, goodness-of-fit criteria such as Mean Error Squares (MEO), and coefficient of determination (R2), Akaike's Information Criteria (AIC) were used. Mean Error Squares is obtained by dividing the sum of squares of error by the degrees of freedom, and the smallest PCR value is preferred for the most suitable model (Bergerud and Sit, 1994; Nasri et al., 2006). Goodness of fit criteria used in the study are given in Table 2 (Narinç et al, 2010, Eyduran and Akbas, 2015).

Table 2. Goodness-of-fit criteria for growth models

Statistics	Formule
Mean of Squares Error	$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$
Coefficient of Determination (R^2)	$R^2 = \frac{\sum_{t=1}^{n}(\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$

n: Number of observations, k: Number of parameters, y_t: Observed values, $\hat{y}_t$: Expected values

RESULTS

Covid-19 case and death toll values were compared with Brody, Gompertz and Von Bertalanffy models and the results are given in Table 3.

Table 3. Growth model parameters.

	Model	A	b	k
	Brody	1166345375	1	0.0000038
Case	Gompertz	465282	4.47	-0.003
	Von Bertalanffy	587482	0.87	0.002
	Brody	2618516	1	0.0000054
	Gompertz	11225	5.336	-0.004
Death	Von Bertalanffy	12630	1.014	0.003

When Table 3 was examined, in the analysis of both the number of cases and deaths, the maximum number of cases/deaths (A) was estimated by the highest Brody model, and the lowest by the Gompertz model. Von Bertalanffy estimated the k parameter, which expresses the maturity rate, highest in the number of cases, and the Gompertz model was the lowest predictor. In the number of deaths, the k parameter was obtained from the Von Bertalanffy model with the highest and the lowest from the Gompertz model. The b parameter, which shows the ratio of the increase in cases and deaths from the baseline to the number of adult cases, was the Gompertz model that predicted the highest and the Von Bertalanffy model that predicted the lowest. The growth functions of the curves are given in Table 4.

Table 4. Growth model functions.

	Model	Functions
	Brody	$y_t = 1166345375 * (1 - 1 * \exp(-0.0000038 * t))$
Case	Gompertz	$y_t = 465282 * \exp\,(-4.47 * exp(0.003 * t))$
	Von Bertalanffy	$y_t = 587482 * (1 - 0.87 * \exp(-0.002 * t))^3$
	Brody	$y_t = 2618516 * (1 - 1 * \exp(-0.0000054 * t))$
Death	Gompertz	$y_t = 11225 * \exp\,(-5.336 * exp(0.004 * t))$
	Von Bertalanffy	$y_t = 12630 * (1 - 1.014 * \exp(-0.003 * t))^3$

The inflection points of the curves are given in Table 5.

Table 5. Inflection points of growth models

Cinsiyet	Model	IPT	IPW
	Brody	---	---
Case	Gompertz	499	171168
	Von Bertalanffy	480	174069
	Brody	---	---
Death	Gompertz	419	4129
	Von Bertalanffy	371	3742

Gompertz predicted the inflection point day the most in case and death numbers, and the Von Bertalanffy model was the lowest predictor. While the number of inflection point cases was higher in the Von Bertalanffy model, the death toll was estimated higher in the Gompertz model. Goodness of fit statistics are shown in Table 6.

Table 6. Goodness of fit criteria for different models

Event	Model	R^2	MSE
Case	Brody	0.976	252903701.906
	Gompertz	0.979	215028231.409
	Von Bertalanffy	0.982	183663526.054
Death	Brody	0.983	177041.928
	Gompertz	0.994	64049.467
	Von Bertalanffy	0.995	56184.377

As seen in Table 6, R^2 and MSE goodness-of-fit statistics were used to compare the models. R^2 values in Brody, Gompertz and Von Bertalanffy models for the number of cases, respectively; While 0.976, 0.979 and 0.982 were found, 0.983, 0.994 and 0.995 were obtained for the death toll. The highest R^2 value was obtained in the Von Bertalanffy model for both cases and deaths. R^2 values of all models were found to be very high and close to each other. The lowest PCR values were obtained in the Von Bertalanffy model in both the number of cases and the number of deaths. Von Bertalanffy model showed the best fit for both case and death. The results obtained in the number of cases and deaths were found to be very close to each other. This indicates that cases and deaths have similar growth pattern characteristics.

The graphs of the growth curves of the number of cases and deaths in Macedonia for the period from April 1, 2020 to May 30, 2022 are presented in Figure 1 and Figure 2, respectively. When Figure 1 and Figure 2 are examined, it is seen that the observed and predicted values in all models are compatible with each other.

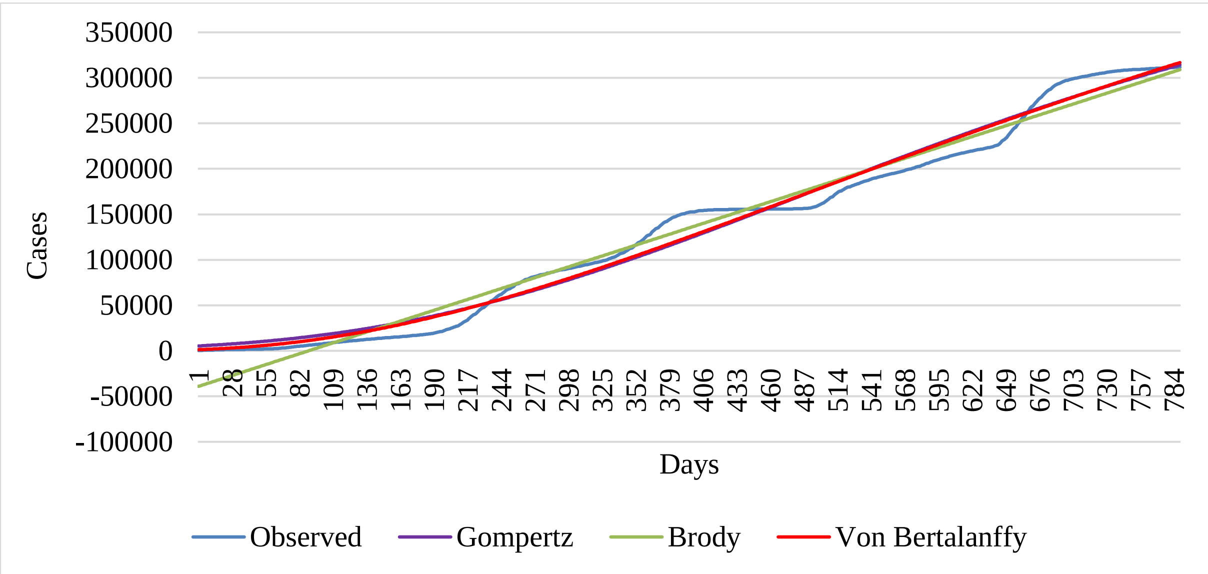


Figure 1. Estimated growth curves of cases

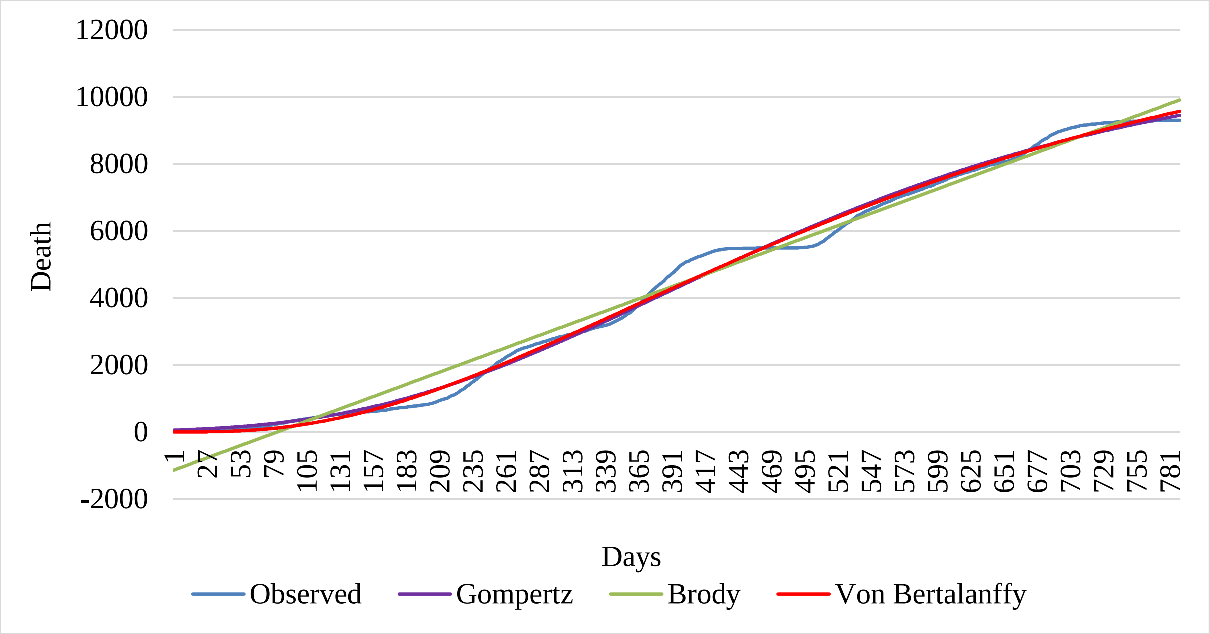


Figure 2. Estimated growth curves of dead numbers

DISCUSSION AND CONCLUSION

In this study, 3 different growth curves of covid-19 cases and deaths in Macedonia were compared. The model that best describes the number of cases and deaths was the Von Bertalanffy model. In the von Bertalanffy model, the coefficient of determination (R^2) in case and dead rabbits was 0.982 and 0.995, respectively; The MSE values were found to be 183663526.054 and 56184.377. According to the von Bertalanffy model, the largest number of cases (A) at the latest time was estimated at 587482 and the death toll at 12630. The daily adult growth rate (k) was 0.002 and 0.003 in cases and deaths, respectively. While the highest increase in the number of cases was on the 480th day, the highest increase in the number of deaths was on the 371st day. During this period, the number of cases was 174069 and the death toll was 3742.

As a result, when the growth curves in Macedonia during the period covering 790 days between April 1, 2020 and May 30, 2022 were examined, it was found that the models were largely similar to each other. It can be said that the Von Bertalanffy model fits better than the Gompertz and Brody models when the R^2 and MSE values are taken into account in the development of case and death numbers.

References

Bergerud, W., Sit, V. 1994. Catalog of curves for curve fitting. Crown Publication Inc. Victoria, B.C., p110.

Çelik, 2021. Modeling of Covid-19 cases and deaths in Uruguay with different, nonlinear growth. Quest Journals. Journal of Research in Applied Mathematics, 7(11): 15-18

Eyduran, E., Akbaş, Y. 2015. Performance of several covariance structures with missing data in repeated measures design. The Journal Animal Plant Science, 25(6):1543-1548.

Gao, X., Dong, Q. 2020. A logistic model for age-specific Covid-19 case-fatality rates. Journal of the American Medical Informatics Association, 3(2): 151–153. https://doi.org/10.1093/jamiaopen/ooaa025

Gurcan, E.K., Çobanoğlu, Ö., Kaplan, S. 2017. Flexible alternatives to models widely used for describing growth in Japanese quail. The Journal of Animal and Plant Sciences, 27(1):48-56.

Medford, A., Llimós, S. T. 2020. Population age structure only partially explains the large number of Covid-19 deaths at the oldest ages. Demographic Research, 43(19): 533-544. https:// dx.doi.org/10.4054/DemRes.2020.43.19

Natale, F., Ghio, D., Tarchi, D., Goujon, A., Conte, A. 2020. Covid-19 cases and case fatality rate by age. European Commission, 52(2): 154–164. https://knowledge4policy.ec.europa.eu/publication/covid-19-cases-case-fatality-rate-age_en

Onder, G., Rezza, G., Brusaferro, S. 2020. Case-fatality rate and characteristics of patients dying in relation to Covid-19 in Italy. Journal of the American Medical Association, 323(18): 1775-1776. https://doi.org/10.1001/jama.2020.4683

Narinc, D., Aksoy, T., Karaman, E. 2010. Genetic parameters of growth curve parameters and weekly body weights in Japanese quail. Journal Animal and Veterinary Advances. 9(3): 501-507.

Nasri, M.H.F., Mesgaran, M.D., France, J., Cant, J.P., Kebreab, E. 2006. Evaluation of models to describe ruminal degradation kinetics from in situ ruminal incubation of whole soybeans. American Dairy Science Association, 89:3087-3095.

Panik, M.J., 2014. Growth curve modelling theory and applications. John Wiley & Sons, Inc., Hoboken, New Jersey.

Sarı, E. 2021. Popüler Basında Covid-19 Haberleri: Yeni Tip Koronavirüs Eski Tip Banal Milliyetçilik. ilef dergisi, 8(1): 63-108. DOI: 10.24955/ilef.798007

Verhulst, P.F. 1838. Notice sur la loi que la population suit dans son accroissement. Corr. Math. Phys. 10: 113–121.

Von Bertalanffy, L. 1957. Quantitative laws for metabolism and growth. Q. Rev. Biol. 32:217-231.

Uzun, A., Oğlakcı, B. 2020. Modernitenin pandemik halleri. Nobel Akademik Yayıncılık.

WHO, 2020. World Health Organization. Coronavirus disease (COVID-19) pandemic https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (Accessed to: 22.11.2020).

Worldometer, 2021. COVID-19 CORONAVIRUS/CASES.

Prediction of Tobacco Production in Macedonia with Trend Analysis and Artificial Neural Networks

Şenol Çelik[1*], Ehlinaz Torun Kayabaşı[2]

[1] Bingöl University, Faculty of Agriculture, Department of Animal Science, Department of Biometrics and Genetics, Bingöl-Turkey
[2] Kocaeli University Faculty of Agriculture Department of Agricultural Economics Kocaeli-Turkey

*Corresponding author e-mail: senolcelik@bingol.edu.tr

Abstract

Tobacco, constituting 3.2% of GDP in Macedonia, is an important agricultural product. Tobacco sector constitutes 3.9% of total agricultural production and 4.1% of employment in the sector. In this study, it is aimed to analyze and predict the amount of tobacco production in Macedonia using artificial neural networks (ANN) and trend analysis. In the development of ANN and trend analysis for the 1992-2020 period, the years parameter was used as the input parameter and the production amount was used as the output parameter. Linear, quadratic, cubic and exponential regression models were used in trend analysis. ANN analysis was carried out with the Hyperbolic Tangent activation function. Model criteria were developed by using statistics such as Mean Squares of Error (MSE) and coefficient of determination (R2). When trend analysis and ANN were compared, the least error squares mean (LCR) and the ANN method, which gave the largest R2 value, gave better results. A prediction for the period of 2021-2025 has been made with the ANN, and as a result of the prediction, it is expected that the amount of tobacco production will be 29405 tons in 2021, 29175 tons in 2022 and 28706 tons in 2025. It is expected that tobacco production will be in an increasing and decreasing course in the coming years. It has been seen that ANN models give better results than trend analysis in production modelling.

Key words: *Artificial Neural Network, Trend Analysis, Production, Tobacco*

INTRODUCTION

Tobacco is a member of the species whose scientific name is Nicotiana, which contains sixty-four genera from the nightshade family (Solanacease), and it is the name given to the leaves of the plants called Nicotiana tabacum and Nicotiana rustica, which are the most well-known in the world (Heiser, 1969).

The first journey of tobacco, known as the origin, was made from America to Europe with the ships belonging to Christopher Columbus and his friends (TEYO, 1978).

It is recorded that Columbus, who saw tobacco and tobacco leaves being smoked with sticks and chewed in the mouth for the first time in San Salvador on October 15, 1492, gave the name "tobacco" to the plant, where the locals smoked tobacco (Barış İzzettin, 1994).

Tobacco production is quite common in the Kocana district of Skopje, the capital of Macedonia. It is understood that relatively more products are obtained from tobacco and rice in the same unit area in Koçana. It can be stated that rice and tobacco, especially in the free trade environment that expanded in the 19th century, were subject to domestic trade and were exported, and in this context, they were capitalist products produced in the region (Karaduman and Tabakoğlu, 2021).

According to the data of FAO (Food and Agriculture Organization of the United Nations) for 2020, the country with the highest tobacco production in the world is China with 2 135 263 tons. India is in the second place with 761 335 tons and Brazil is in the third place with 702 208 tons. These countries were followed by Zimbabwe, Indonesia, United States of America, Mozambique, Pakistan, Argentina and Malawi, respectively, and ranked in the top 10 in the world. Turkey ranks 14th with a production amount of 76 540 tons, while Macedonia ranks 26th with a production of 26 112 tons (FAO, 2020).

The factors affecting the amount of tobacco production in the Black Sea Region during the 1990-2003 period were examined with the Latent Regression model. Cultivation area, yield, temperature, humidity and precipitation were found to be important independent variables affecting production. The price did not make a significant contribution to the model (Terzi et al., 2009).

In the study conducted by Dikmen (2005), the relationship between tobacco production and price in Turkey in the 1982-2003 period was examined with the Koyck model.

In this study, it is aimed to model and predict the amount of tobacco production in Macedonia with artificial neural networks and trend analysis.

MATERIAL AND METHOD

Material

The data used in the study consisted of tobacco production data under the title of “Crops and livestock products” compiled from the www.fao.org website of the Food and Agriculture Organization of the United Nations (FAO). Data from 1992 to 2020 were used in the study and analyzed with artificial neural networks (ANN) and trend analysis. After the appropriate models were determined, the number of ducks for the period 2021-2025 was predicted.

Method

Trend analysis

Least Squares (Least Squares) Method

The functional relationship between time and observation values can be linear or non-linear. The obtained function must be in line with the line or time series graph. Moreover,

Y_t: Observation values, $\widehat{Y_t}$: Trend values (theoretical values),

$$\sum_{t=1}^{n}\left(Y_t - \widehat{Y_t}\right)^2 = min$$

condition must be met. In order to apply the EKK method, the time series graph must first be drawn. The most suitable function type for the development direction of this graph is selected (Serper, 2004). Some trend models used in the EKK method are linear, quadratic and cubic models.

Linear regression model,

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

(Gujarati, 2003; Kadilar, 2009). Normal equations of the linear regression model

$$\sum_{t=1}^{n} Y_t = nb_0 + b_1 \sum_{t=1}^{n} X_t$$

$$\sum_{t=1}^{n} X_t Y_t = b_0 \sum_{t=1}^{n} X_t + b_1 \sum_{t=1}^{n} X_t^2$$

and b_0 and b_1 parameters are estimated by solving this system of equations (Gamgam and Altunkaynak, 2015).

Quadratic regression model,

$$Y_i = b_0 + b_1 X_i + b_2 X_i^2 + \varepsilon_i$$

(Akkaya, 1990).

Cubic regression model,

$$Y_i = b_0 + b_1 X_i + b_2 X_i^2 + b_3 X_i^3 + \varepsilon_i$$

(Montgomery et al., 2012).

Artificial Neural Networks (ANN)

Artificial neural networks (ANNs) learn and generalize by experimentation. Predicting the future is one of the important areas that can be used with ANN. Revealing complex relationships between data can be done with ANN. ANN performs well in learning, optimization, analysis, classification, generalization and association (Öztemel, 2012).

ANN is a very fast definition and understanding of information data, which is usually created in very different ways. It performs non-linear modeling without the need for any prior knowledge and assumptions between input and output variables (Kaastra & Boyd, 1996).

The smallest units that form the basis of ANN's work are called artificial nerve cells. The artificial neuron has five main components. These components are inputs, weights, summation function, activation function and output components.

Inputs consist of information coming from the outside world to an artificial neuron. Information received from the outside world is transferred to artificial nerve cells as input (Özveren, 2006).

Weights express the effect of information coming into an artificial cell on a neuron (Öztemel, 2012). The weights are the coefficients that show the effect of the inputs received through the artificial nerve on the nerve and are expressed with (w1, w2, w3,…, wi) (Elmas, 2003). The addition function calculates the net input to a cell. This function is formulated as follows.

$$z_i = \sum_{İ=1}^{n} (w_{ij}\, x_i + b_j)$$

Where w is the inputs, x is the weights, and n is the number of inputs (process elements).

The value obtained using the sum function and the output of the processing element passed through a linear or non-linear differentiable activation function are expressed as follows (Yavuz and Deveci, 2012).

$$y = f(z_i) = f\left(\sum_{İ=1}^{n} (w_{ij}\, x_i + b_j)\right)$$

Activation function

Activation function is one of the important factors determining neuron behavior (Efe and Kaynak, 2000). In ANN, the output amplitude of the neuron is limited between the desired values. It generally consists of values between [-1, 1] (Sagiroglu et al., 2003).

The Hyperbolic tangent activation function was used as the activation function.

Hyperbolic tangent activation function,

$$F(NET) = \frac{e^{NET} - e^{-NET}}{e^{NET} + e^{-NET}}$$

(Öztemel, 2012; Alp and Öz, 2019). The output values of the hyperbolic tangent activation function are in the range of -1 to 1 (Çayıroğlu, 2015).

Back propagation algorithm selection

Important problems in ANN method are solved by back propagation algorithm. In the back propagation network, the errors of the units in the intermediate layers are determined by backpropagation of the errors of the units in the output layer. For this reason, the method is named as Back Propagation Learning Algorithm. In the back propagation algorithm, more than one non-linearity-free output unit can be used with the sigmoidal transfer function. Output units can generate real numbers in the range of [-1,1] (Kröse and Smagt, 1996).

Model suitability criteria

ANN model performance is usually determined by Mean Error Squares (MSE) and Mean Absolute Error (MAE). MSE is calculated as follows (Singh et al., 2009).

$$MSE = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n}$$

MAE is calculated as follows.

$$MAE = \frac{\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|}{n} = \frac{\sum_{i=1}^{n}|\varepsilon_i|}{n}$$

Burada Y_i: Observed values of the dependent variable, $\hat{Y}$: Estimated values of the dependent variable, n is the number of observations.

RESULTS

The trend analysis of the amount of sainfoin production between 1992 and 2020 in Macedonia was examined. Linear, quadratic and cubic regression analysis were applied. In order to determine the most suitable model, the model with the largest coefficient of determination (R^2) value was evaluated (Table 1).

Table 1. Equation, R^2 and MSE values of the models

Model	Equation	R^2	MSE
Linear	Yt = 21788 +158.99*t	0.113	14799387.664
Quadratic	Yt = 21519 +210.96*t -1.7322*t**2	0.114	15355522.183
Cubic	Yt = 22858 -283.04*t +38.744*t**2 -0.89948*t**3	0.125	15773673.826

In the trend analysis methods in Table 1, the R^2 values of all methods are very low. ANN method was also used to better model the production forecast. The numbers of ANN input, hidden and exit layers were determined as 1-12-1, respectively, and were applied with back propagation learning with 1000 iterations. For ANN, R^2=0.9746, MSE=673530.829 and MAE=697.137. Since MSE value is smaller and R^2 value is larger for ANN method than trend analysis methods, ANN is the most appropriate method to be used.

The graph of the values of tobacco production in Macedonia between 1992 and 2020 is presented in Figure 1.

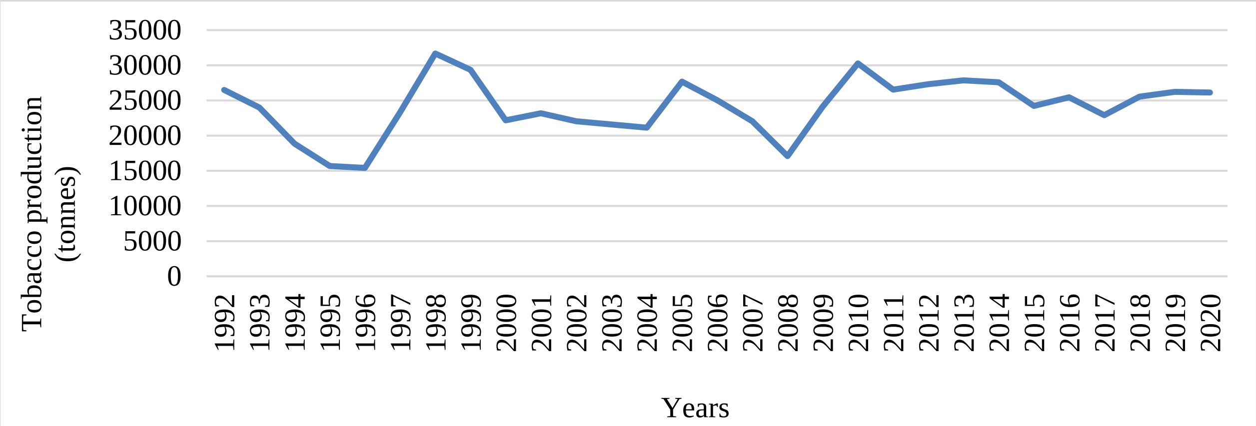


Figure 1. Tobacco production in Macedonia between 1992 and 2020

When Figure 1 is examined, a fluctuating course was observed in tobacco production in the 1992-2020 period. However, since 2011, a lower level of ups and downs has been observed. In the period after 2011, tobacco production was between 22885-27859 tons. The residual values along with the predicted and observed values when ANN was applied are given in Table 2.

Table 2. Observed, estimated and residual values (residual)

Years	Actual	Predicted	Residual
1992	26502	NaN	NaN
1993	24002	NaN	NaN
1994	18862	NaN	NaN
1995	15683	NaN	NaN
1996	15412	NaN	NaN
1997	23308	NaN	NaN
1998	31676	NaN	NaN
1999	29368	NaN	NaN
2000	22175	NaN	NaN
2001	23200	NaN	NaN
2002	22044	NaN	NaN
2003	21592	NaN	NaN
2004	21140	21823.4844	-683.4844
2005	27691	26299.5857	1391.4143
2006	25036	24730.7041	305.2959
2007	22056	22069.5270	-13.5270
2008	17087	18286.4362	-1199.4362
2009	24122	24533.9034	-411.9034
2010	30280	28841.1210	1438.8790
2011	26537	27538.6069	-1001.6069
2012	27333	27180.3069	152.6931
2013	27859	27658.0101	200.9899
2014	27578	28162.0766	-584.0766
2015	24237	23362.5207	874.4793
2016	25443	24655.7809	787.2191
2017	22885	24088.2633	-1203.2633
2018	25547	26340.8852	-793.8852
2019	26234	26442.6467	-208.6467
2020	26112	25511.4712	600.5288

The graph of the actual and estimated values as a result of the ANN application is given in Figure 2, and the graph of the residual values is given in Figure 3.

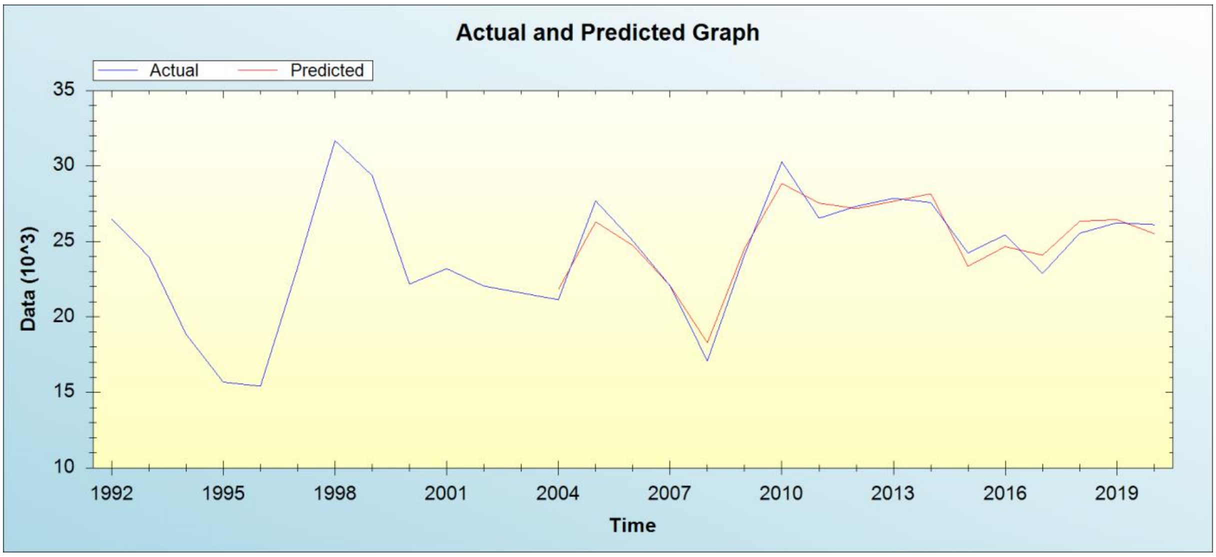


Figure 2. Graph of observed and predicted values together

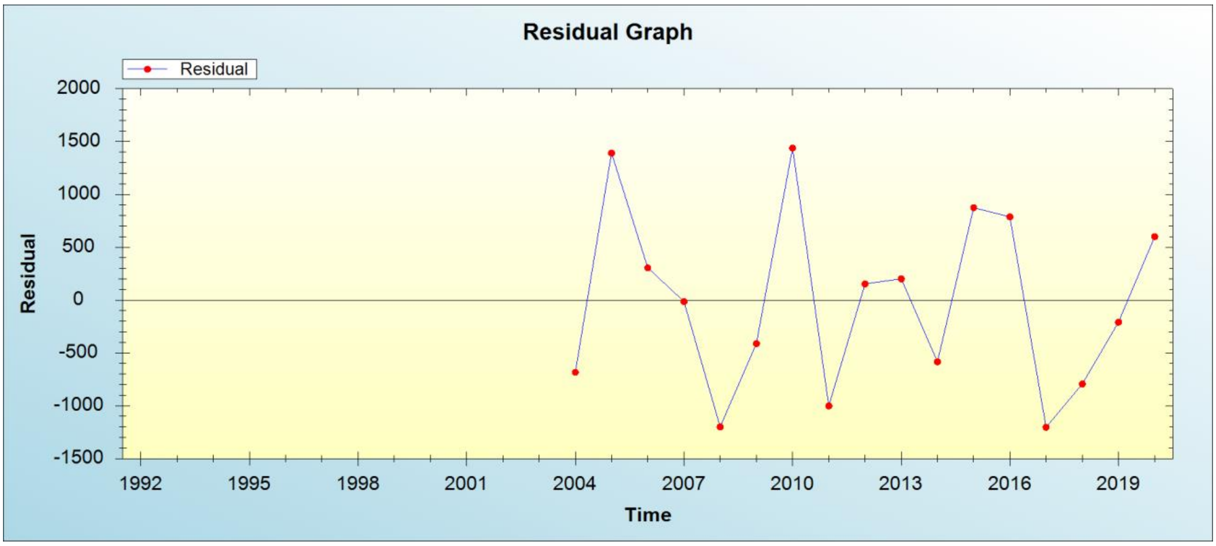


Figure 3. Graph of residuals

When the graph of the residual values in Figure 3 is examined, it is seen that the residual values are independently and randomly distributed. This shows that the important hypotheses of the model are met. The graph of the observed and residual values is presented in Figure 4.

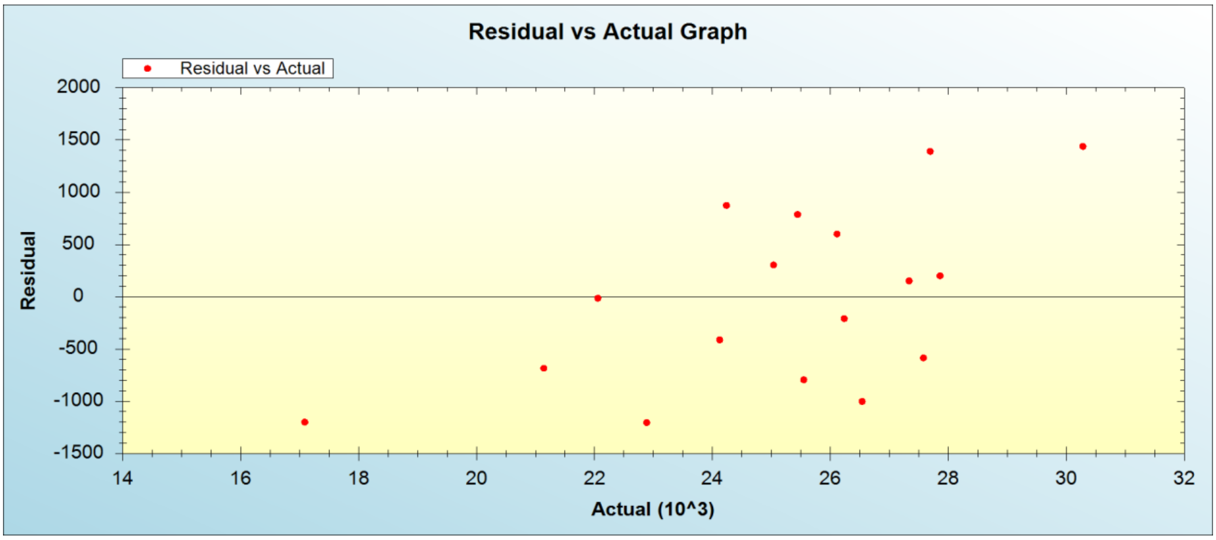


Figure 4. Graph of observed and residual value

Tobacco production projections for the 2021-2025 period are given in Table 3 and Figure 5.

Table 3. Production forecast for future years

Years	Forecast
2021	29404.56
2022	29174.73
2023	29821.13
2024	28293.05
2025	28705.97

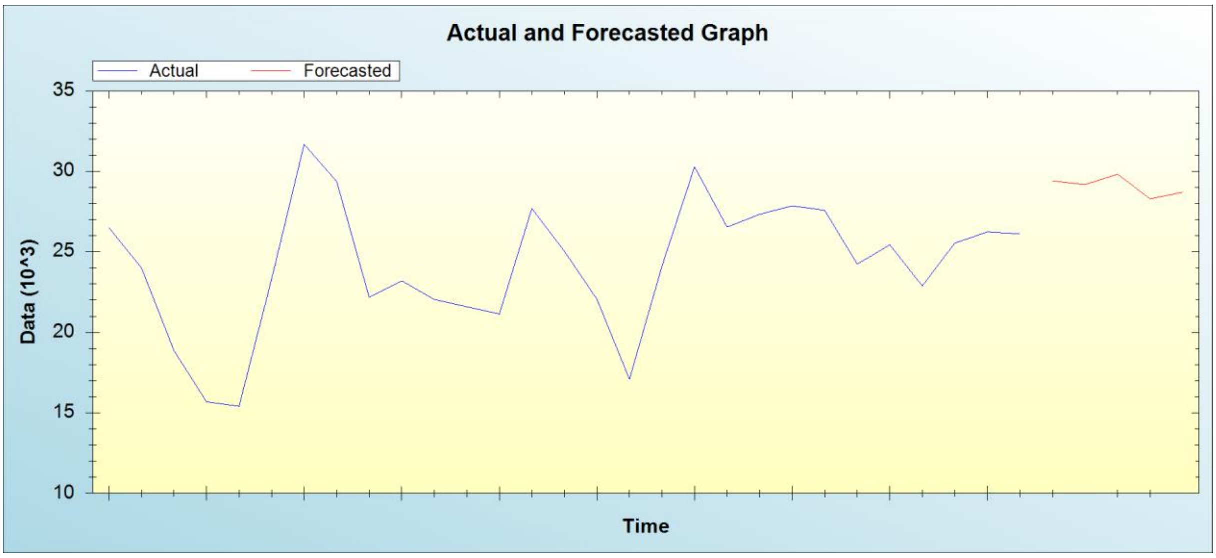


Figure 5. Observed values by years and forecast for the next period

According to the prediction results given in Table 3, it is expected that the amount of tobacco production will increase and decrease, and the production amount values will be close to each other.

DISCUSSION AND CONCLUSION

In one of the studies with agricultural data, Grzesiak et al. (2010) identified the problematic cows in artificial insemination using artificial neural networks and MARS algorithms and showed the elimination of risky ones. Cows are divided into two groups, good and bad. Specific variables were used to identify cows with problems in artificial insemination. In both methods, the most important determining variable is calving interval.

Grzesiak et al. (2003) used artificial neural networks and multiplexes to estimate the 305-day milk yield average, the number of days at milking, the average milk yield in the 1st, 2nd, 3rd, 4th periods and the effect of calving month on milk yield in Holstein-Friesian cows in a barn (2003). used a regression model. The authors explained that artificial neural networks are more suitable for milk prediction.

In the study of Çelik (2021), artificial neural networks and time series analysis were applied to estimate the number of camels in Turkey. For the 1961-2020 period, a network of 1000 iterations with 12 hidden layers, 12 processing elements (12-12-1) and Levenberg–Marquardt back propagation algorithm (trainlm) was established and analyzed. As a result of the analysis, the prediction of the number of camels between 2021 and 2025 was obtained and it is expected that the number of camels will show a fluctuating course in this period.

In this study, the amount of tobacco production in Macedonia was modeled with artificial neural networks. Years (1992-2020) are used as input variable and production amount is used as output variable. Hyperbolic tangent functions from activation functions were used.

The results revealed that the established ANN method gives good predictions. The low MSE and MAE values in the training, testing and validation stages also show this.

As a result of the ANN analyzed using the hyperbolic tangent function, it is expected that the tobacco production in Macedonia will be between 28293 tons and 29405 tons in the 2021-2025 period. In 2024-2025, the amount of tobacco production in Macedonia is expected to decrease, albeit slightly. It is hoped that it will give good results in agricultural data by using artificial neural networks and alternative techniques in future forecasting studies.

References

Akkaya, Ş. 1990. Ekonometri I. Anadolu Matbaacılık, İzmir.

Alp S., Öz E. 2019. Makine öğrenmesinde sınıflandırma yöntemleri ve R uygulamaları. Nobel Akademik Yayıncılık, Ankara.

Barış İzzettin, ed. 1994. Sigara ve sağlık. 1inci baskı. Ankara: MEB yayınları, 11-14

Çayıroğlu, İ. 2015. İleri Algoritma Analizi-5 Yapay Sinir Ağları http://www.ibrahimcayiroglu.com/Dokumanlar/IleriAlgoritmaAnalizi/IleriAlgoritmaAnalizi-5.Hafta-YapaySinirAglari.pdf

Çelik Ş. 2021. Modeling and estimation of camel population in Turkey with time series analysis and Artificial Neural Networks. International Journal of Research in Engineering and Science (IJRES), 9(5):38-44

Dikmen, N. 2005. Koyck-Almon Yaklaşımı İle Tütün Üretimi ve Fiyat İlişkisi. VII. Ulusal Ekonometri ve İstatistik Sempozyumu, 26-27 Mayıs 2005, İstanbul Üniversitesi. http://www.ekonometridernegi.org/bildiriler/o16s1.pdf, Accessed to: Ağustos 2006.

Efe Ö., Kaynak O. 2000. Yapay sinir ağları ve uygulamaları, Boğaziçi Üniversitesi Yayını, İstanbul

Elmas Ç. 2003. Yapay sinir ağları, Birinci Baskı, Seçkin Yayıncılık, Ankara

FAO, 2020. Food and Agriculture Organization of the United Nations. Crops and livestock products. https://www.fao.org/faostat/en/#data/QCL (Accessed to: 21.05.2022).

Gamgam, H., Altunkaynak, B. 2015. SPSS Uygulamalı Regresyon Analizi. Seçkin Yayıncılık, Ankara.

Grzesiak W., Lacroix R., Wójcik J., Blaszczyk P. 2003. A comparison of neural network and multiple regression prediction for 305-day lactation yield using partial lactation records. Canadian Journal of Animal Science, 83:307-310.

Grzesiak W., Zaborski D., Sablik P., Zukiewicz A., Dybus A., Szatkowska I. 2010. Detection of cows with insemination problems using selected classification models. Computers and Electronics in Agriculture, 74(2): 265-273

Gujarati, D. N. 2003. Basic Econometrics. McGraw-Hill, North America.

Heiser, C.B. 1969. Night shades: The Paradoxical Plants, San Francisco, s.12-13.

Kaastra I., Boyd M. 1996. Designing a neural network for forecasting financial and economic time series. Neurocomputing, 10: 215-236

Kadılar, C. 2009. SPSS Uygulamalı Zaman Serileri Analizine Giriş. Bizim Büro Yayınevi, Ankara.

Karaduman, G., Tabakoğlu, A. 2021. 19. Yüzyıl Temettuat Defterleri'ne Göre Üsküp/Koçana Kazası'nın Lorenz Eğrisi ve Gini Katsayısı ile Gelir Dağılımının Analizi. Eskişehir Osmangazi Üniversitesi İİBF Dergisi, 16(1): 146-165. Doi: 10.17153/oguiibf.860980

Kröse, B., Smagt, P. V. D. 1996. An Introduction to Neural Networks. Eight Ed., The University of Amsterdam.

Montgomery, D.C., Peck, E.A., Vining, G.G. 2012. Introduction to Linear Regression Analysis. John Wiley and Sons, Inc., USA.

Öztemel E. 2012. Yapay sinir ağları. Papatya Yayıncılık, İstanbul.

Özveren, U. 2006. Pem Yakıt Hücrelerinin Yapay Sinir Ağları ile Modellenmesi. Yayınlanmamış Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul

Sağıroğlu, Ş., Beşdok E., Erler M. 2003. Mühendislikte yapay zeka uygulamaları-I Yapay Sinir Ağları, Ufuk Kitap Yayıncılık.

Serper, Ö. 2004. Uygulamalı İstatistik 2. Ezgi Kitabevi, Bursa.

Singh KP., Basant, A, Malik A, Jain G. 2009. Artificial neural network modeling of the river water quality-A case study. Ecological Modelling, 220(6): 888-895.

Terzi, E., Şenel, T., Terzi, Y., Cengiz, M.A. 2009. Latent Regresyon Modellerin Gerçek Veri Kümesine Uygulaması. e-Journal of New World Sciences Academy, 4(4): 104-109. Article Number: 3A0013

Tütün Eksperleri Yüksek Okulu. 1978. Tütüncülüğe giriş. İstanbul: TEYO yayını, 918

Yavuz, S., Deveci, M. 2012. İstatistiksel Normalizasyon Tekniklerinin Yapay Sinir Ağı Performansına Etkisi. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 40: 167-187

Unit Burr-Hatke Distribution with A New Quantile Regression Model

Şule Sağlam[1*], Kadir Karakaya[1]

[1] Selçuk University, Department of Statistics, 42130, Konya, Turkey

*Corresponding author e-mail: sulesaglam75@gmail.com

Abstract

In this study, a new distribution in the range between 0 and 1 is derived. The Burr-Hatke distribution is considered the baseline distribution. Since the baseline distribution has one parameter, the new unit distribution also has one parameter. Some mathematical properties of the new distribution are studied. Five estimators are examined to estimate the unknown parameter of the new model. The performances of the estimators are analyzed according to the bias and mean square error criteria calculated by Monte Carlo simulation. Two numerical data analyses are performed. A new quantile regression model is also introduced based on the new distribution as alternative to beta and Kumaraswamy regression. A Monte Carlo simulation is also conducted for the new quantile regression model.

Key words: *Unit distribution, Burr-Hatke distribution, maximum likelihood estimation, Monte Carlo simulation, Quantile Regression.*

An Examination of the Performance of Logistic Regression and Artificial Neural Networks Algorithms from Machine Learning Methods on Disabled Employment Quota Method

Özge Nisa Yegül[1*], Berna Yazıcı[1], Yener Şişman[2]

[1] Eskişehir Technical University, Faculty of Science, Department of Statistics, 26470, Eskişehir, Turkey
[2] 2Anadolu University, Faculty of Economics and Administrative Science, Department of Labor Economics and Industrial Relations, 26210, Eskişehir, Turkey

*Corresponding author e-mail: ozgenisayegul@eskisehir.edu.tr

Abstract

In recent years, the accelerating of technological developments increased data, faster and more efficient in getting results which are growing popularity machine learning, one of the sub-branches of Artificial Intelligence (AI), is the most common method. The main reason why machine learning shines so much is to train machine-learning algorithms with the existing database related to the problem/research to provide identify it, following, by establishing a model, whether the relevant matches are provided or not according to the desired result is obtained. Nowadays, the most commonly used machine learning algorithm is divided into supervised learning, regression, and classification. This study, using "Logistic Regression" and "Artificial Neural Networks", which are classification methods under the supervised learning, Disabled Employment Quota Method in Turkey survey data and the quota system stipulated in Article 30 of the Labor Law No. 4857, in order to determine whether private workplaces have the quota requirement or not, the completed ready-made survey data was collected between September 2019 and March 2020 for businesses whose socio-demographic structure is within the scope of the quota method across Turkey. Studied to determine the parameters that affected the requirements of compliance with this law in private sector workplaces. Besides, the performance of the two methods was compared on this data set. According to the results obtained at the end of the comparison, it was found that the performance of Artificial Neural Networks Algorithms outperformed Logistic Regression Algorithms.

Key words: *Disabled Employment, Machine Learning, Binary Logistic Regression, Logistic Regression, Artificial Neural Networks*

Prediction of Tobacco Production in Macedonia with Trend Analysis and Artificial Neural Networks

Şenol Çelik[1*], Ehlinaz Torun Kayabaşı[2]

[1] Bingöl University, Faculty of Agriculture, Department of Animal Science, Department of Biometrics and Genetics, Bingöl-Turkey
[2]Kocaeli University Faculty of Agriculture Department of Agricultural Economics Kocaeli-Turkey

* Corresponding author: senolcelik@bingol.edu.tr

Abstract

Tobacco, constituting 3.2% of GDP in Macedonia, is an important agricultural product. Tobacco sector constitutes 3.9% of total agricultural production and 4.1% of employment in the sector.
In this study, it is aimed to analyze and predict the amount of tobacco production in Macedonia using artificial neural networks (ANN) and trend analysis. In the development of ANN and trend analysis for the 1992-2020 period, the years parameter was used as the input parameter and the production amount was used as the output parameter. Linear, quadratic, cubic and exponential regression models were used in trend analysis. ANN analysis was carried out with the Hyperbolic Tangent activation function. Model criteria were developed by using statistics such as Mean Squares of Error (MSE) and coefficient of determination (R2). When trend analysis and ANN were compared, the least error squares mean (LCR) and the ANN method, which gave the largest R2 value, gave better results. A prediction for the period of 2021-2025 has been made with the ANN, and as a result of the prediction, it is expected that the amount of tobacco production will be 29405 tons in 2021, 29175 tons in 2022 and 28706 tons in 2025. It is expected that tobacco production will be in an increasing and decreasing course in the coming years. It has been seen that ANN models give better results than trend analysis in production modelling.
Keywords: *Artificial Neural Network, Trend Analysis, Production, Tobacco.*

INTRODUCTION

Tobacco is a member of the species whose scientific name is Nicotiana, which contains sixty-four genera from the nightshade family (Solanacease), and it is the name given to the leaves of the plants called Nicotiana tabacum and Nicotiana rustica, which are the most well-known in the world (Heiser, 1969).

The first journey of tobacco, known as the origin, was made from America to Europe with the ships belonging to Christopher Columbus and his friends (TEYO, 1978).

It is recorded that Columbus, who saw tobacco and tobacco leaves being smoked with sticks and chewed in the mouth for the first time in San Salvador on October 15, 1492, gave the name "tobacco" to the plant, where the locals smoked tobacco (Barış İzzettin, 1994).

Tobacco production is quite common in the Kocana district of Skopje, the capital of Macedonia. It is understood that relatively more products are obtained from tobacco and rice in the same unit area in Koçana. It can be stated that rice and tobacco, especially in the free trade environment that expanded in the 19th century, were subject to domestic trade and were exported, and in this context, they were capitalist products produced in the region (Karaduman and Tabakoğlu, 2021).

According to the data of FAO (Food and Agriculture Organization of the United Nations) for 2020, the country with the highest tobacco production in the world is China with 2 135 263 tons. India is in the second place with 761 335 tons and Brazil is in the third place with 702 208 tons. These countries were followed by Zimbabwe, Indonesia, United States of America, Mozambique, Pakistan, Argentina and Malawi, respectively, and ranked in the top 10 in the world. Turkey ranks 14th with a production amount of 76 540 tons, while Macedonia ranks 26th with a production of 26 112 tons (FAO, 2020).

The factors affecting the amount of tobacco production in the Black Sea Region during the 1990-2003 period were examined with the Latent Regression model. Cultivation area, yield, temperature, humidity and precipitation were found to be important independent variables affecting production. The price did not make a significant contribution to the model (Terzi et al., 2009).

In the study conducted by Dikmen (2005), the relationship between tobacco production and price in Turkey in the 1982-2003 period was examined with the Koyck model.

In this study, it is aimed to model and predict the amount of tobacco production in Macedonia with artificial neural networks and trend analysis.

MATERIAL AND METHOD

Material

The data used in the study consisted of tobacco production data under the title of "Crops and livestock products" compiled from the www.fao.org website of the Food and Agriculture Organization of the United Nations (FAO). Data from 1992 to 2020 were used in the study and analyzed with artificial neural networks (ANN) and trend analysis. After the appropriate models were determined, the number of ducks for the period 2021-2025 was predicted.

Method

Trend analysis

Least Squares (Least Squares) Method

The functional relationship between time and observation values can be linear or non-linear. The obtained function must be in line with the line or time series graph. Moreover,

Y_t: Observation values, $\widehat{Y_t}$: Trend values (theoretical values),

$$\sum_{t=1}^{n}\left(Y_t - \widehat{Y_t}\right)^2 = min$$

condition must be met. In order to apply the EKK method, the time series graph must first be drawn. The most suitable function type for the development direction of this graph is selected (Serper, 2004). Some trend models used in the EKK method are linear, quadratic and cubic models.

Linear regression model,

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

(Gujarati, 2003; Kadilar, 2009). Normal equations of the linear regression model

$$\sum_{t=1}^{n} Y_t = nb_0 + b_1 \sum_{t=1}^{n} X_t$$

$$\sum_{t=1}^{n} X_t Y_t = b_0 \sum_{t=1}^{n} X_t + b_1 \sum_{t=1}^{n} X_t^2$$

and b_0 and b_1 parameters are estimated by solving this system of equations (Gamgam and Altunkaynak, 2015).

Quadratic regression model,

$$Y_i = b_0 + b_1 X_i + b_2 X_i^2 + \varepsilon_i$$

(Akkaya, 1990).

Cubic regression model,

$$Y_i = b_0 + b_1 X_i + b_2 X_i^2 + b_3 X_i^3 + \varepsilon_i$$

(Montgomery et al., 2012).

Artificial Neural Networks (ANN)

Artificial neural networks (ANNs) learn and generalize by experimentation. Predicting the future is one of the important areas that can be used with ANN. Revealing complex relationships between data can be done with ANN. ANN performs well in learning, optimization, analysis, classification, generalization and association (Öztemel, 2012).

ANN is a very fast definition and understanding of information data, which is usually created in very different ways. It performs non-linear modeling without the need for any prior knowledge and assumptions between input and output variables (Kaastra & Boyd, 1996).

The smallest units that form the basis of ANN's work are called artificial nerve cells. The artificial neuron has five main components. These components are inputs, weights, summation function, activation function and output components.

Inputs consist of information coming from the outside world to an artificial neuron. Information received from the outside world is transferred to artificial nerve cells as input (Özveren, 2006).

Weights express the effect of information coming into an artificial cell on a neuron (Öztemel, 2012). The weights are the coefficients that show the effect of the inputs received through the artificial nerve on the nerve and are expressed with (w1, w2, w3,…, wi) (Elmas, 2003). The addition function calculates the net input to a cell. This function is formulated as follows.

$$z_i = \sum_{\mathrm{i}=1}^{n} (w_{ij}\, x_i + b_j)$$

Where w is the inputs, x is the weights, and n is the number of inputs (process elements).

The value obtained using the sum function and the output of the processing element passed through a linear or non-linear differentiable activation function are expressed as follows (Yavuz and Deveci, 2012).

$$y = f(z_i) = f\left(\sum_{\mathrm{i}=1}^{n} (w_{ij}\, x_i + b_j)\right)$$

Activation function

Activation function is one of the important factors determining neuron behavior (Efe and Kaynak, 2000). In ANN, the output amplitude of the neuron is limited between the desired values. It generally consists of values between [-1, 1] (Sagiroglu et al., 2003).

The Hyperbolic tangent activation function was used as the activation function.

Hyperbolic tangent activation function,

$$F(NET) = \frac{e^{NET} - e^{-NET}}{e^{NET} + e^{-NET}}$$

(Öztemel, 2012; Alp and Öz, 2019). The output values of the hyperbolic tangent activation function are in the range of -1 to 1 (Çayıroğlu, 2015).

Back propagation algorithm selection

Important problems in ANN method are solved by back propagation algorithm. In the back propagation network, the errors of the units in the intermediate layers are determined by backpropagation of the errors of the units in the output layer. For this reason, the method is named as Back Propagation Learning Algorithm. In the back propagation algorithm, more than one non-linearity-free output unit can be used with the sigmoidal transfer function. Output units can generate real numbers in the range of [-1,1] (Kröse and Smagt, 1996).

Model suitability criteria

ANN model performance is usually determined by Mean Error Squares (MSE) and Mean Absolute Error (MAE). MSE is calculated as follows (Singh et al., 2009).

$$MSE = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n}$$

MAE is calculated as follows.

$$MAE = \frac{\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|}{n} = \frac{\sum_{i=1}^{n}|\varepsilon_i|}{n}$$

Burada Y_i: Observed values of the dependent variable, $\hat{Y}$: Estimated values of the dependent variable, n is the number of observations.

RESULTS

The trend analysis of the amount of sainfoin production between 1992 and 2020 in Macedonia was examined. Linear, quadratic and cubic regression analysis were applied. In order to determine the most suitable model, the model with the largest coefficient of determination (R^2) value was evaluated (Table 1).

Table 1. Equation, R^2 and MSE values of the models

Model	Equation	R^2	MSE
Linear	Yt = 21788 +158.99*t	0.113	14799387.664
Quadratic	Yt = 21519 +210.96*t -1.7322*t**2	0.114	15355522.183
Cubic	Yt = 22858 -283.04*t +38.744*t**2 -0.89948*t**3	0.125	15773673.826

In the trend analysis methods in Table 1, the R^2 values of all methods are very low.

ANN method was also used to better model the production forecast. The numbers of ANN input, hidden and exit layers were determined as 1-12-1, respectively, and were applied with back propagation learning with 1000 iterations. For ANN, R^2=0.9746, MSE=673530.829 and MAE=697.137. Since MSE value is smaller and R^2 value is larger for ANN method than trend analysis methods, ANN is the most appropriate method to be used.

The graph of the values of tobacco production in Macedonia between 1992 and 2020 is presented in Figure 1.

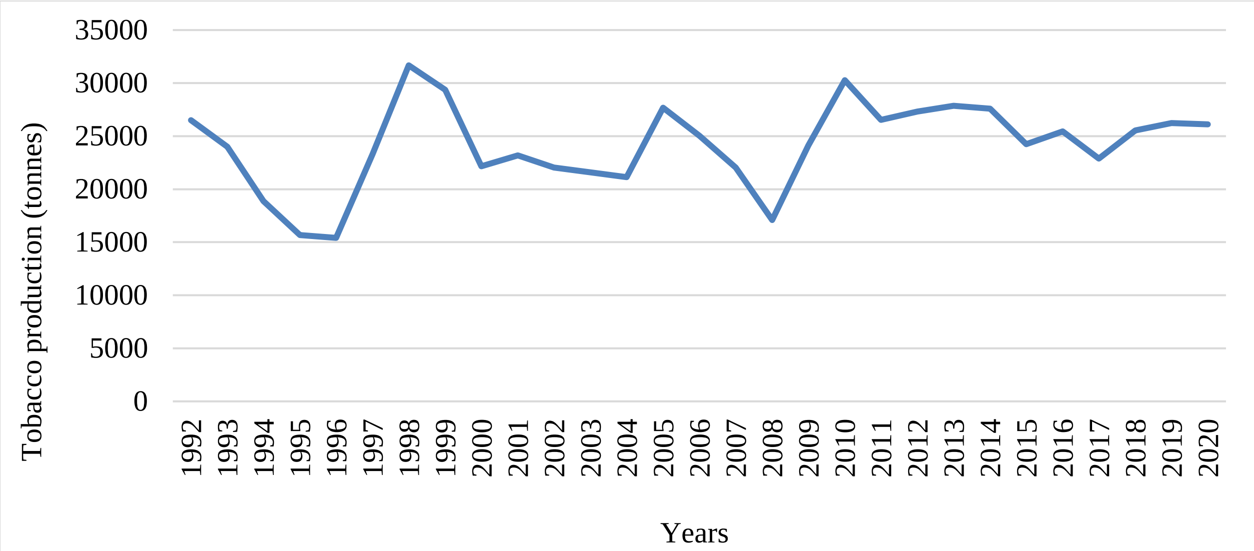


Figure 1. Tobacco production in Macedonia between 1992 and 2020

When Figure 1 is examined, a fluctuating course was observed in tobacco production in the 1992-2020 period. However, since 2011, a lower level of ups and downs has been observed. In the period after 2011, tobacco production was between 22885-27859 tons. The residual values along with the predicted and observed values when ANN was applied are given in Table 2.

Table 2. Observed, estimated and residual values (residual)

Years	Actual	Predicted	Residual
1992	26502	NaN	NaN
1993	24002	NaN	NaN
1994	18862	NaN	NaN
1995	15683	NaN	NaN
1996	15412	NaN	NaN
1997	23308	NaN	NaN
1998	31676	NaN	NaN
1999	29368	NaN	NaN
2000	22175	NaN	NaN
2001	23200	NaN	NaN
2002	22044	NaN	NaN
2003	21592	NaN	NaN
2004	21140	21823.4844	-683.4844
2005	27691	26299.5857	1391.4143
2006	25036	24730.7041	305.2959
2007	22056	22069.5270	-13.5270
2008	17087	18286.4362	-1199.4362
2009	24122	24533.9034	-411.9034
2010	30280	28841.1210	1438.8790
2011	26537	27538.6069	-1001.6069
2012	27333	27180.3069	152.6931
2013	27859	27658.0101	200.9899
2014	27578	28162.0766	-584.0766
2015	24237	23362.5207	874.4793
2016	25443	24655.7809	787.2191
2017	22885	24088.2633	-1203.2633
2018	25547	26340.8852	-793.8852
2019	26234	26442.6467	-208.6467
2020	26112	25511.4712	600.5288

The graph of the actual and estimated values as a result of the ANN application is given in Figure 2, and the graph of the residual values is given in Figure 3.

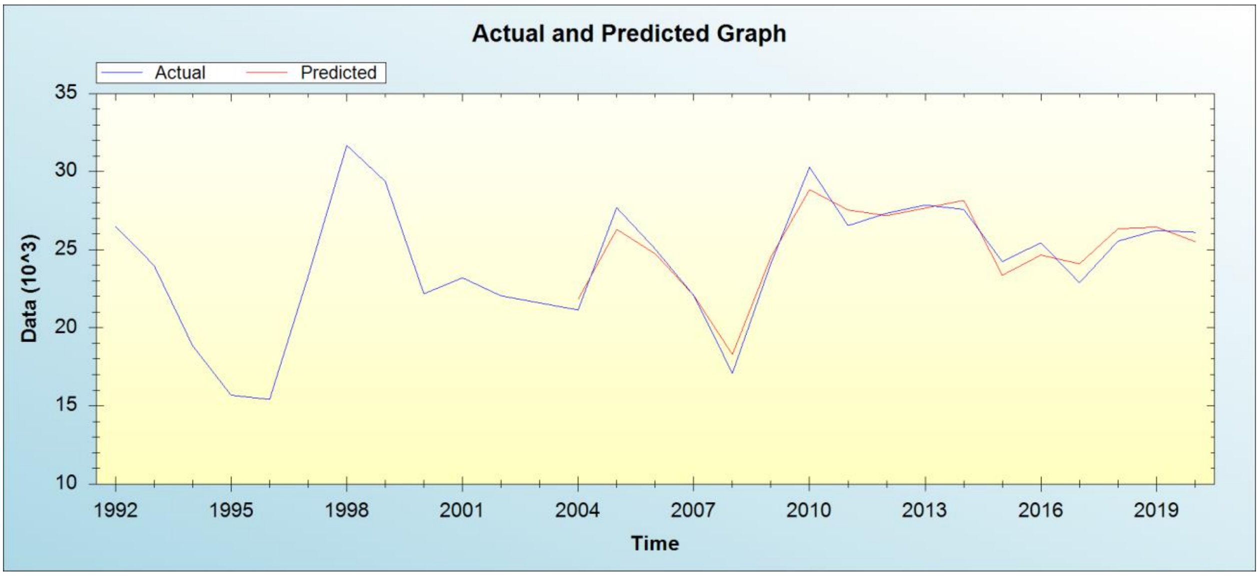


Figure 2. Graph of observed and predicted values together

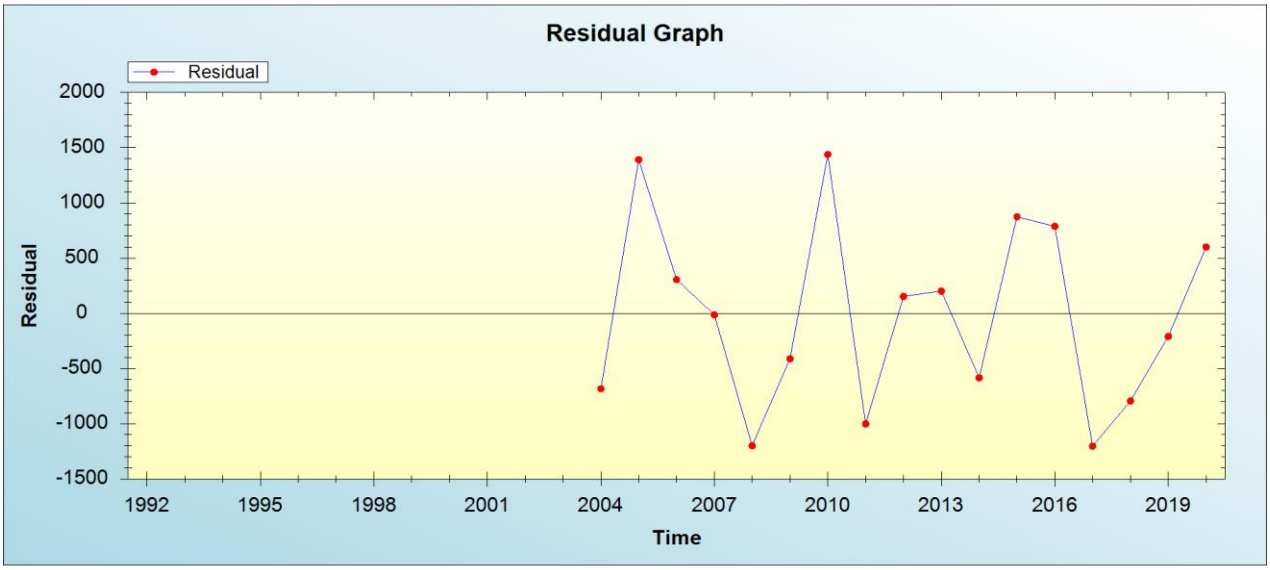


Figure 3. Graph of residuals

When the graph of the residual values in Figure 3 is examined, it is seen that the residual values are independently and randomly distributed. This shows that the important hypotheses of the model are met. The graph of the observed and residual values is presented in Figure 4.

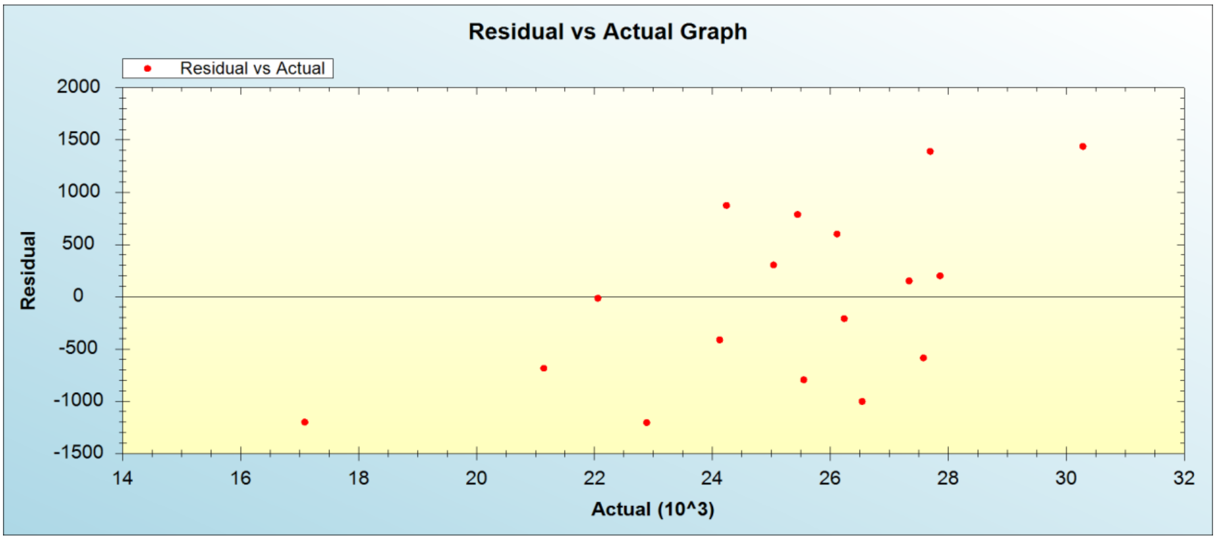


Figure 4. Graph of observed and residual values

Tobacco production projections for the 2021-2025 period are given in Table 3 and Figure 5.

Table 3. Production forecast for future years

Years	Forecast
2021	29404.56
2022	29174.73
2023	29821.13
2024	28293.05
2025	28705.97

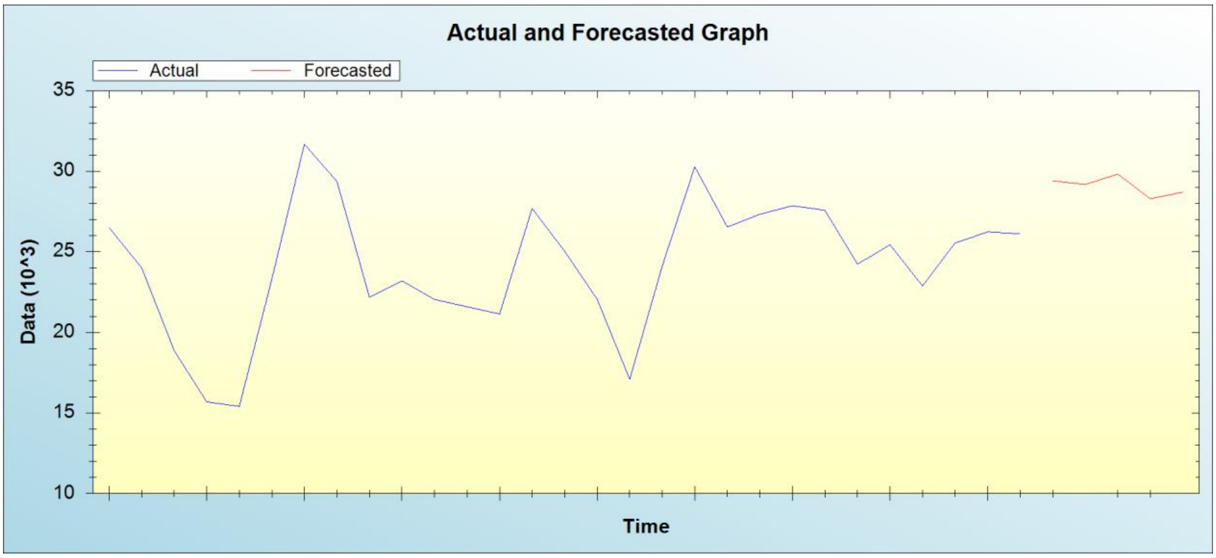


Figure 5. Observed values by years and forecast for the next period

According to the prediction results given in Table 3, it is expected that the amount of tobacco production will increase and decrease, and the production amount values will be close to each other.

DISCUSSION AND CONCLUSION

In one of the studies with agricultural data, Grzesiak et al. (2010) identified the problematic cows in artificial insemination using artificial neural networks and MARS algorithms and showed the elimination of risky ones. Cows are divided into two groups, good and bad. Specific variables were used to identify cows with problems in artificial insemination. In both methods, the most important determining variable is calving interval.

Grzesiak et al. (2003) used artificial neural networks and multiplexes to estimate the 305-day milk yield average, the number of days at milking, the average milk yield in the 1st, 2nd, 3rd, 4th periods and the effect of calving month on milk yield in Holstein-Friesian cows in a barn (2003). used a regression model. The authors explained that artificial neural networks are more suitable for milk prediction.

In the study of Çelik (2021), artificial neural networks and time series analysis were applied to estimate the number of camels in Turkey. For the 1961-2020 period, a network of 1000 iterations with 12 hidden layers, 12 processing elements (12-12-1) and Levenberg–Marquardt back propagation algorithm (trainlm) was established and analyzed. As a result of the analysis, the prediction of the number of camels between 2021 and 2025 was obtained and it is expected that the number of camels will show a fluctuating course in this period.

In this study, the amount of tobacco production in Macedonia was modeled with artificial neural networks. Years (1992-2020) are used as input variable and production amount is used as output variable. Hyperbolic tangent functions from activation functions were used.

The results revealed that the established ANN method gives good predictions. The low MSE and MAE values in the training, testing and validation stages also show this.

As a result of the ANN analyzed using the hyperbolic tangent function, it is expected that the tobacco production in Macedonia will be between 28293 tons and 29405 tons in the 2021-2025 period. In 2024-2025, the amount of tobacco production in Macedonia is expected to decrease, albeit slightly. It is hoped that it will give good results in agricultural data by using artificial neural networks and alternative techniques in future forecasting studies.

References

Akkaya, Ş. 1990. Ekonometri I. Anadolu Matbaacılık, İzmir.
Alp S., Öz E. 2019. Makine öğrenmesinde sınıflandırma yöntemleri ve R uygulamaları. Nobel Akademik Yayıncılık, Ankara.
Barış İzzettin, ed. 1994. Sigara ve sağlık. 1inci baskı. Ankara: MEB yayınları, 11-14
Çayıroğlu, İ. 2015. İleri Algoritma Analizi-5 Yapay Sinir Ağları http://www.ibrahimcayiroglu.com/Dokumanlar/IleriAlgoritmaAnalizi/IleriAlgoritmaAnalizi-5.Hafta-YapaySinirAglari.pdf
Çelik Ş. 2021. Modeling and estimation of camel population in Turkey with time series analysis and Artificial Neural Networks. International Journal of Research in Engineering and Science (IJRES), 9(5):38-44
Dikmen, N. 2005. Koyck-Almon Yaklaşımı İle Tütün Üretimi ve Fiyat İlişkisi. VII. Ulusal Ekonometri ve İstatistik Sempozyumu, 26-27 Mayıs 2005, İstanbul Üniversitesi. http://www.ekonometridernegi.org/bildiriler/o16s1.pdf, Accessed to: Ağustos 2006.
Efe Ö., Kaynak O. 2000. Yapay sinir ağları ve uygulamaları, Boğaziçi Üniversitesi Yayını, İstanbul
Elmas Ç. 2003. Yapay sinir ağları, Birinci Baskı, Seçkin Yayıncılık, Ankara
FAO, 2020. Food and Agriculture Organization of the United Nations. Crops and livestock products. https://www.fao.org/faostat/en/#data/QCL (Accessed to: 21.05.2022).
Gamgam, H., Altunkaynak, B. 2015. SPSS Uygulamalı Regresyon Analizi. Seçkin Yayıncılık, Ankara.

Grzesiak W., Lacroix R., Wójcik J., Blaszczyk P. 2003. A comparison of neural network and multiple regression prediction for 305-day lactation yield using partial lactation records. Canadian Journal of Animal Science, 83:307-310.

Grzesiak W., Zaborski D., Sablik P., Zukiewicz A., Dybus A., Szatkowska I. 2010. Detection of cows with insemination problems using selected classification models. Computers and Electronics in Agriculture, 74(2): 265-273

Gujarati, D. N. 2003. Basic Econometrics. McGraw-Hill, North America.

Heiser, C.B. 1969. Night shades: The Paradoxical Plants, San Francisco, s.12-13.

Kaastra I., Boyd M. 1996. Designing a neural network for forecasting financial and economic time series. Neurocomputing, 10: 215-236

Kadılar, C. 2009. SPSS Uygulamalı Zaman Serileri Analizine Giriş. Bizim Büro Yayınevi, Ankara.

Karaduman, G., Tabakoğlu, A. 2021. 19. Yüzyıl Temettuat Defterleri’ne Göre Üsküp/Koçana Kazası’nın Lorenz Eğrisi ve Gini Katsayısı ile Gelir Dağılımının Analizi. Eskişehir Osmangazi Üniversitesi İİBF Dergisi, 16(1): 146-165. Doi: 10.17153/oguiibf.860980

Kröse, B., Smagt, P. V. D. 1996. An Introduction to Neural Networks. Eight Ed., The University of Amsterdam.

Montgomery, D.C., Peck, E.A., Vining, G.G. 2012. Introduction to Linear Regression Analysis. John Wiley and Sons, Inc., USA.

Öztemel E. 2012. Yapay sinir ağları. Papatya Yayıncılık, İstanbul.

Özveren, U. 2006. Pem Yakıt Hücrelerinin Yapay Sinir Ağları ile Modellenmesi. Yayınlanmamış Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul

Sağıroğlu, Ş., Beşdok E., Erler M. 2003. Mühendislikte yapay zeka uygulamaları-I Yapay Sinir Ağları, Ufuk Kitap Yayıncılık.

Serper, Ö. 2004. Uygulamalı İstatistik 2. Ezgi Kitabevi, Bursa.

Singh KP., Basant, A, Malik A, Jain G. 2009. Artificial neural network modeling of the river water quality-A case study. Ecological Modelling, 220(6): 888-895.

Terzi, E., Şenel, T., Terzi, Y., Cengiz, M.A. 2009. Latent Regresyon Modellerin Gerçek Veri Kümesine Uygulaması. e-Journal of New World Sciences Academy, 4(4): 104-109. Article Number: 3A0013

Tütün Eksperleri Yüksek Okulu. 1978. Tütüncülüğe giriş. İstanbul: TEYO yayını, 918

Yavuz, S., Deveci, M. 2012. İstatistiksel Normalizasyon Tekniklerinin Yapay Sinir Ağı Performansına Etkisi. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 40: 167-187

Kümeleme Analizinde Kullanılan Yeni Bir Veri Görselleştirme Yaklaşımı

Emrah Akdamar[1*]

[1] Bandırma Onyedi Eylül Üniversitesi, Denizcilik Fakültesi, Denizcilik İşletmeleri Yönetimi Bölümü, 10200, Bandırma/Balıkesir, Türkiye

*Corresponding author e-mail: eakdamar@bandirma.edu.tr

Özet

Bu çalışmada, kümeleme analizinde verilerin görselleştirilmesi için kullanılan dendogram yönteminin dezavantajlarını gidermek üzere, Öklid uzaklığına dayalı kümeleme analizleri ve çok boyutlu ölçekleme analizi kullanılarak yeni bir kümeleme görselleştirme yöntemi önerilmiştir. Öklid uzaklığına dayalı yöntemlerden k-ortalamalar yöntemi ile küme merkezi sayısı ve kümelere atanan gözlemler belirlenmiş, çok boyutlu ölçekleme ile boyut indirgenerek iki boyutlu düzlemde her bir gözlemin koordinatları elde edilmiştir. Belirlenen küme merkezleri ile gözlemlerin koordinatlarına göre gözlemler kümelere atanmış ve elde edilen yeni kümeleme görseli sunulmuştur. Önerilen yöntem ile elde edilen yeni görselin küme içi ve kümeler arası ilişkilerin anlaşılması ve yorumlanmasında dendogram grafiğine kıyasla daha etkili olduğu ve dendogram grafiğinin kısıtlarını giderdiği görülmüştür.

Key words: *Veri Görselleştirme, Kümeleme Analizi, K- Ortalamalar, Çok Boyutlu Ölçekleme, Dendogram*

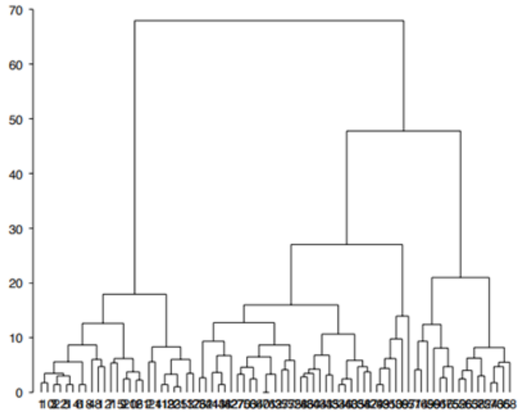
GİRİŞ

Çok değişkenli verilerin kümelenmesi ve kümelere ait bilgilerin belirlenmesi önemli ve zor problemlerdendir. Verilerin anlamlı kümelere (alt gruplara) bölünmesi için birçok farklı kümeleme yöntemi geliştirilmiştir. Kümeleme işlemi, heterojen yapıya sahip bir kitleyi daha homojen birkaç alt gruba ya da kümeye bölme işlemidir. Sınıflandırma ile kümelemeyi birbirinden ayıran en önemli fark, kümeleme işleminin sınıflandırma işleminde olduğu gibi önceden belirlenmiş sınıflara göre gruplandırma yapmamasıdır. Sınıflandırmada her bir veri, önceden gruplandırılmış belli sınıflar üzerinde yapılan bir eğitim neticesinde ortaya çıkan bir modele göre, önceden belirlenmiş olan bir sınıfa atanmaktadır. Kümeleme işleminde ise önceden tanımlanmış sınıflar ya da örnek sınıflar bulunmamaktadır. Sınıflandırmaları hakkında ön bilgi bulunmayan durumlarda, topluluğa ilişkin tahminlerin yapılmasında yararlanılan bir yöntemler grubu olarak tanımlanan kümeleme analizi, üzerinde çalışılan herhangi bir veri setindeki benzer (homojen) gözlem gruplarını bulma, kendi içinde homojen fakat kümeler arası heterojen olacak şekilde kümelere ayırmada kullanılan çok değişkenli istatistiksel analiz tekniğidir.

Kümeleme analizinde, birbirine yakın nesne ya da gözlemlerin oluşturdukları gruplar küme olarak tanımlanır. Kümeleme analizinde gözlemler arasındaki uzaklıkları hesaplamak için en yaygın kullanılan uzaklık ölçüsü olan Öklid uzaklığı, iki gözlem arasına çizilecek en kısa düz doğru uzunluğunu temel alır. Bu uzaklık ölçüsü dışında Minkowski uzaklığı, Ölçekli Öklid uzaklığı, İkili Öklid uzaklığı, Pearson uzaklığı, Manhattan (City Blok) uzaklığı, Mahalonobis uzaklığı, Hotelling T2 uzaklığı ve Canberra uzaklığı gibi diğer uzaklık ölçüleri de kullanılmaktadır (Lorr, 1983).

Kümeleme analizi için birçok algoritma öne sürülmüştür. Ancak literatürde bu algoritmalar Hiyerarşik kümeleme teknikleri, Hiyerarşik olmayan kümeleme tekniği olmak üzere iki başlık altında toplanmıştır (Ketchen & Shook, 1996). Her iki yöntemde de ortak amaç kümeler arasındaki farklılıkları ve kümeler içi benzerlikleri optimize etmektir.

Kümelere ait bilgilerin sunulmasında ve bilgi kaybının önlenmesinde veri görselleştirme oldukça önemlidir. Kümelemede en sık kullanılan görselleştirme araçlarından biri dendogram grafiğidir. Ancak dendogram grafiğinin iki özelliğinden dolayı yeterince kullanışlı bir araç olmadığı söylenebilir. İlk olarak, dendogram grafiğinde her bir gözlem birimi bir yaprak olarak göründüğünden bu grafik yalnızca az sayıda gözlem olması durumunda etkili bir görsel sağlamaktadır. Çok sayıda gözlem bulunan veri setlerinde dendogram grafiği küme yapılarını tam olarak yansıtamamaktadır. Şekil 1'de 75 gözlemden oluşan dendogram grafiği görülmektedir.



Şekil 1. Hiyerarşik kümeleme yönteminde kullanılan dendogram grafiği (Schonlau, 2002)

İkinci olarak, dendogram grafiği yalnızca hiyerarşik kümelemede kullanılan bir görsel araçtır (Schonlau, 2002). Bu nedenle literatürde dendograma alternatif olarak çeşitli görselleştirme araçları önerilmiştir (Schonlau, 2002; Schonlau, 2004).

Çok boyutlu ölçekleme (ÇBÖ) ülkedeki yerleşim yerlerinin haritada kümelenmesi için kullanılan yöntemlerin başında gelmektedir. Yerleşim yerleri arasındaki mesafe veya yolculuk sürelerine dayalı ölçekleme ile haritalandırma yapılabilmektedir. Cox ve Cox İngiltere'nin 12 şehir arasında yolculuk sürelerini baz alarak ÇBÖ ile şehirlerin yerlerinin İngiltere haritası ile örtüştüğünü göstermiştir (Cox & Cox, 2001). Jin, vd. (2004), veri görselleştirme ve kümeleme analizi için kullanılan kendi kendini düzenleyen haritalara kıyasla genişletilmiş kendi kendini düzenleyen haritaların daha iyi sonuçlar verdiğini göstermiştir. Garg & Rani (2017), Twitter verilerini k-ortalamalar yöntemiyle kümelemiş ve görselleştirmiştir. Coğrafi bir harita üzerinde gerçekleştirilen kümelemede bir R paketi olan ggplot kullanılmıştır. Twitter verileri ile ilgili bir başka çalışmada Sechelea, vd. (2016) verilerin kümelenmesi ve görselleştirmesi ile ilgili bir sistem tasarlamıştır. Oğuzlar (2005), Suçlu profilinin belirlenmesi amacıyla kendi kendini düzenleyen haritalar ile kümeleme analizi gerçekleştirmiştir. Wu vd. (2010) çok boyutlu verilerin kümelenmesi için Java tabanlı bir matris görselleştirme yazılımı olan GAP'ı tanıtmıştır. Bildirici & Afacan (2018) kümeleme analizi bulgularının tematik haritalar yöntemiyle görselleştirilebileceğini göstermiştir. Lasue, vd. (2011) veri görselleştirmek ve 2 boyutlu uzayda kümeleri belirlemek için doğrusal ve doğrusal olmayan yöntemleri karşılaştırmıştır. Vatansever & Büyüklü (2009) Görsel veri madenciliği tekniklerine yönelik yeni bir yöntem önermiştir.

Bu çalışmada, dendogram grafiğinin belirtilen kısıtlarından dolayı yeterince kullanışlı olmaması nedeniyle yeni bir kümeleme görselleştirme yaklaşımı önerilmektedir. Kümelemede küme içi ve kümeler arası ilişkilerin (uzaklıkların) daha açık görünmesinin öneminden hareketle geliştirilen bu yaklaşım, k-ortalamalar yöntemi ile ÇBÖ'nün ortak olarak kullandığı öklid uzaklığına dayalı olarak geliştirilmiştir. Geliştirilen görsel, tüm kümeleme teknikleriyle uyumludur ve çok sayıda gözlem birimi olması durumunda da kullanılabilmektedir. Bu özellikleri dikkate alındığında literatüre önemli bir katkı sağlayacağı düşünülmektedir.

Bu çalışma dört bölümden oluşmaktadır. İkinci bölümde, kullanılan yöntem detaylı olarak anlatılmıştır. Üçüncü bölümde, önerilen yöntem Iris gerçek veri seti üzerinde uygulanmış ve yeni geliştirilen görsel ile ilgili bulgular paylaşılmıştır. Dördüncü ve son bölümde ise sonuç ve tartışmaya yer verilmiştir.

MATERYAL VE METOD

Materyal

Fisher tarafından 1936'da sunulmuş olan Iris veri seti (UCI, 2022) kümeleme analizinde yaygın olarak kullanılmaktadır. Çanak yaprağı uzunluğu, çanak yaprağı genişliği, taç yaprağı uzunluğu ve taç yaprağı genişliği olmak üzere her değişkeninde 150 gözlem bulunan dört değişkenli bir veri setidir. Setosa, Versicolor ve Virginica olarak adlandırılan üç kümeye ayrılmaktadır.

Metod

K- Ortalamalar Yöntemi

Hiyerarşik olmayan ve denetimli bir kümeleme yöntemi olan k-ortalamalar algoritması en çok kullanılan kümele yöntemleri arasındadır. k-ortalamalar yönteminde belli sayıdaki k küme için toplam ortalama hata küme içi homojenliği sağlamak için minimize edilmeye çalışılır. K boyutlu uzayda N örnekli kümelerin verildiği varsayıldığında uzay $\{C_1, \dots, C_k\}$ biçiminde k kümeye ayrılsın, o zaman $\sum n_r = N,\ k = 1, \dots, r$ olmak üzere C_k kümesinin ortalama vektörü μ_k gözlemlere dayalı olarak,

$$\mu_k = \frac{1}{n_r} \sum_{i=1}^{n_r} X_{ik} \tag{1}$$

şeklinde elde edilir. Burada X_{ik} değeri C_k kümesine ait i. nesneyi temsil etmektedir. C_k kümesi hata kare, küme içi değişim değerini belirlemede her bir C_k değeri ile küme merkezi arasındaki Öklid uzaklıkların toplamı olarak,

$$e_i{}^2 = \sum_{i=1}^{n_r} (x_{ir} - \mu_r)^2 \tag{2}$$

şeklinde elde edilir. k. kümeyi içeren tüm kümeler uzayı için hata kare değeri,

$$E_k{}^2 = \sum_{k=1}^{K} e_k{}^2 \tag{3}$$

şeklinde hesaplanır. Hata kare kümeleme yönteminin amacı, verilen k değeri için $E_k{}^2$ değerini minimize eden k kümelerini iterasyonlarla değişiklik kalmayana kadar elde etmektir.

Algoritma 1: k-ortalamalar ile küme sayısı ve gözlemlerin belirli kümelere atanması

Girdi: $\boldsymbol{c_k}$: Veri setindeki küme sayısı (ön bilgi), $(\boldsymbol{k = 1, \dots, r})$

Çıktı: Küme içi minimum uzaklık ile kümeler arası maksimum uzaklığın sağlandığı kümeleme dağılımı.

1. $\boldsymbol{k = 1, \dots, r}$ olmak üzere önceden belirlenen sayıda küme merkezi seçilir.

2. $\boldsymbol{e_i, i = 1, \dots, k}$ küme içi gözlem değerleri arasındaki değişimler hesaplanır.

3. $\boldsymbol{e_i{}^2}$ küme içi değişim uzaklıkları ile küme içi değişim toplamı olan $\boldsymbol{E_k{}^2}$ değeri elde edilir.

4. Küme içi değişimler kalmayana kadar merkez seçimi için 1. ve 2. adımlar tekrarlanır. Gözlemler belirlenen merkezlerden en yakın olan kümeye atanır.

Çok Boyutlu Ölçekleme Analizi

Çok boyutlu ölçekleme analizi çok değişkenli verilerin istatistiksel yöntemlere dayalı boyut indirgeme tekniği olarak kullanılmaktadır. ÇBÖ aynı zamanda nesneler arasındaki benzerlikleri ve bağımlılık yapısını belirleyen bir istatistiksel yöntemdir. Çok boyutlu ölçeklemede Öklid uzaklığı, Minkowski uzaklığı, Manhattan (City Blok) Uzaklığı, Chebychev Uzaklığı gibi ölçüler yaygın kullanılmaktadır (Bülbül & Köse, 2010).

Çok değişkenli verinin her bir noktasının konfigürasyonu k boyutlu bir uzayın elemanı olarak kabul edilir. ÇBÖ'de a ve b gibi iki nesne arasındaki benzemezlik δ_{ab} ile gösterilsin, bu k boyutlu uzaydaki nokta bir nesneye denk gelir ve nesneler arası uzaklıklar d_{ab} ile gösterilir. ÇBÖ'deki esas amaç $\delta_{ab} = d_{ab}$ benzemezlik uzaklıkları ile nesnelerin uzaklıklarını mümkün olan en az hata ile belirlemektir.

Bir metriğe dayalı ÇBÖ analizinin temel yaklaşımı, nxp boyutlu veri matrisinden hesaplanan birimler arasındaki δ_{ab} uzaklıklarını 2 boyutlu uzayda temsil etmektir. f sürekli ve monoton bir fonksiyon olmak üzere, d_{ab} uzaklığı için,

$$d_{ab} = f(\delta_{ab}) \tag{4}$$

eşitliği ile çok değişkenli veride boyut indirgenir.

Veri matrisindeki a ve b noktaları arasındaki uzaklıklar ile konfigürasyon uzaklıkları arasındaki farkı minimum yapmak için optimizasyon teknikleri kullanılır. Nesneler arasındaki uzaklıkların toplamına dayalı ölçek faktörü,

$$\ddot{o}l\c{c}ek\ fakt\ddot{o}r\ddot{u} = \sum_{a=1}^{n-1}\sum_{b=a+1}^{n} {d_{ab}}^2 \tag{5}$$

ile elde edilir. Nesneler arasındaki uzaklıkların uyumu ölçek faktörüne dayalı olarak stres göstergesi,

$$stress = (\ddot{o}l\c{c}ek\ fakt\ddot{o}r\ddot{u})^{-1} \sum_{a=1}^{n-1}\sum_{b=a+1}^{n}\big(d_{ab} - f(\delta_{ab})\big)^2 \tag{6}$$

denklemi ile elde edilir (Martinez & Martinez, 2005).

Amaç Fonksiyonu

Çok değişkenli verilerin boyut indirgeme yöntemiyle kümeleme analizi kullanılarak elde edilen kümelerin görselleştirilmesi için k- ortalamalar ve ÇBÖ analizinin karma yapısı kullanılmaktadır. $f(.): \mathbb{R}^n \to \mathbb{R}^2$ sürekli ve monoton fonksiyon $(x_1, \dots, x_n) \to (x_1, x_2)$ boyut indirgeme ile kümelerin düzlemdeki küme adreslerinin belirlenmesini sağlamaktadır.

KÜMELEME ANALİZİ VE ÇBÖ'YE DAYALI VERİ GÖRSELLEŞTİRME YÖNTEMİ

Çok değişkenli veri setinin her bir değişkendeki gözlemlerin ön bilgi olarak elde edilen k_1, k_2, k_3 kümelerine atanması için k-ortalamalar algoritması ile kümeleme analizi yapılmıştır. Belirlenen kümelere düşen gözlem sayıları Tablo 1.'de gösterilmiştir.

Tablo 1. Yeni kümeleme görseli için elde edilen veri setinin yapısı

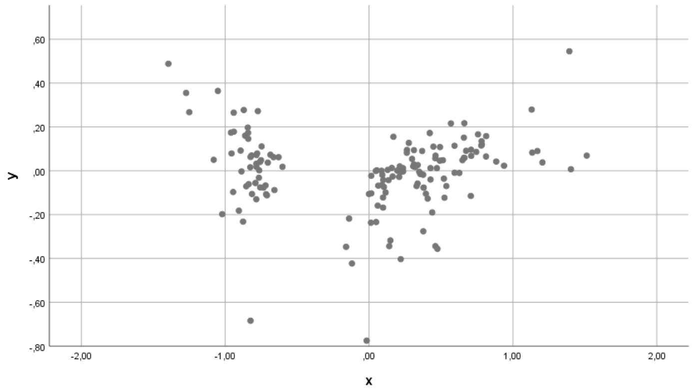
Küme merkezleri	k_1	k_2	k_3
Gözlem sayıları	$n_1 = 38$	$n_2 = 50$	$n_3 = 62$

Dört değişkenli Iris veri setinin gözlemleri ÇBÖ ile iki boyuta indirgenmiş ve her bir gözleme ait düzlemdeki koordinatlar belirlenmiştir. ÇBÖ ve k-ortalamalar ile elde edilen ve yeni görseli oluşturmak için kullanılan verinin teorik yapısı Tablo 2'de gösterilmiştir.

Tablo 2: Yeni kümeleme görseli için elde edilen veri setinin yapısı

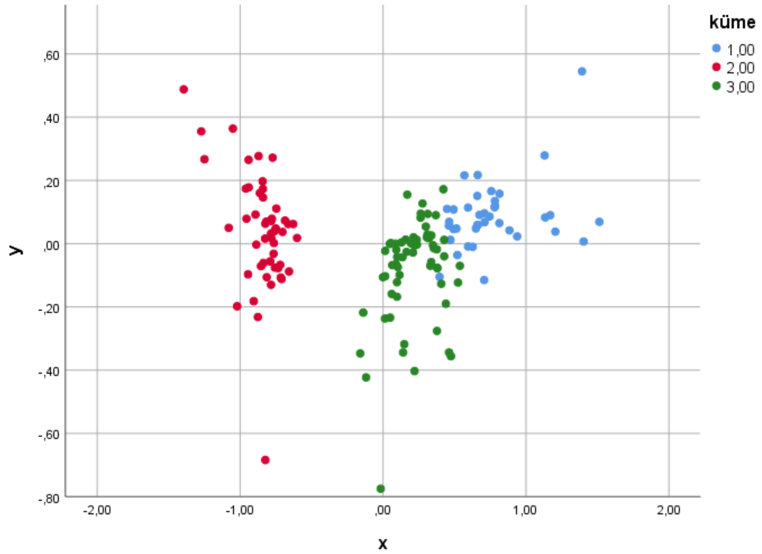
Kümelere atanacak gözlemler	ÇBÖ ile elde edilen ikili koordinatlar		k-ortalamalar ile elde edilen küme aidiyetleri
GÖZLEM	**x**	**y**	**KÜME**
1	x_1	y_1	$x_i \in c_k$, k=1,2,..,K
2	x_2	y_2	$x_i \in c_k$, k=1,2,..,K
.	.	.	.
.	.	.	.
.	.	.	.
n	x_n	y_n	$x_i \in c_k$, k=1,2,..,K

Dört değişkenli Iris veri setindeki gözlemlerin ÇBÖ ile iki boyutlu düzleme indirgenerek elde edilen koordinatları ve kümelere düşen gözlemler Ek 1'de gösterilmiştir. Çok değişkenli Iris veri setinden ÇBÖ kullanılarak elde edilen Tablo 2 deki (x,y) ikililerinin saçılım diyagramı ile iki boyuttaki izdüşümü elde edilmiş ve Şekil 2 de gösterilmiştir.



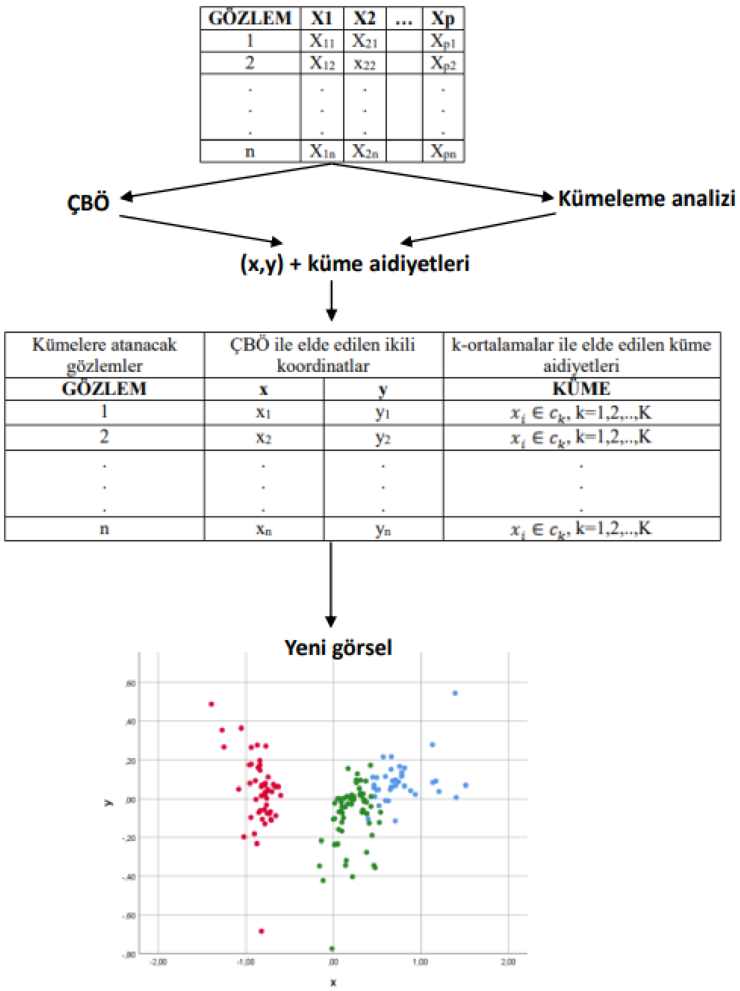
Şekil 2. Gözlemlerin saçılım diyagramı üzerinde gösterilmesi

K-ortalamalar algoritması ile elde edilen küme elemanları ÇBÖ ile indirgenen ve koordinatları elde edilen gözlemlerle birleştirilmiştir. k-ortalamalar ve ÇBÖ çıktılarının birleştirilmesi ile elde edilen yeni küme görseli Şekil 3'de verilmiştir.



Şekil 3. Saçılım diyagramı üzerinde yeni veri görseli

Çok değişkenli veri setleri için Irıs veri seti üzerinde uygulaması anlatılan, ÇBÖ ve kümeleme analizine dayalı olarak geliştirilen yeni görselin algoritması Şekil 4.'de gösterilmiştir.



Şekil 4. ÇBÖ ve kümeleme analizi ile elde edilen yeni görsel algoritması

SONUÇ VE TARTIŞMA

Çok değişkenli verilerin kümelenmesinde küme içi ve kümeler arası ilişkilerin daha net elde edilebilmesi için veri görselleştirme önemli bir araçtır. Literatür incelendiğinde, kümelerin görselleştirilmesine yönelik clustergram, kohonen ağları, tematik haritalar vb. araçlar önerilmiştir. Bu çalışmada, hiyerarşik kümelemede kullanılan dendogram grafiğine ve hiyerarşik olmayan kümelemede çok boyutlu verilerin indirgenmesine alternatif bir veri görselleştirme aracı önerilmiştir. ÇBÖ'de boyut indirgenirken Öklid uzaklığı kullanıldığından önerilen görsel araç aynı uzaklığı kullanan kümeleme analizleri ile uyumludur.

Hiyerarşik kümeleme yöntemlerinde kullanılan dendogram grafiğinin çok sayıda gözlem içeren verilerde kullanıcının yorumlamasını zorlaştıran karmaşık bir yapısı mevcuttur. Ayrıca, dendogram grafiği sadece hiyerarşik kümeleme analizinde kullanılmaktadır. Diğer taraftan, dendogram grafiği küme içi ve kümeler arası ilişkilerin yansıtılmasında yetersiz kalmaktadır. Önerilen yöntem, çok sayıda gözlem içeren veri setlerini görselleştirmesi, hiyerarşik olmayan kümeleme analizlerinde de kullanılabilmesi ve kümelerin yapısını göstermede kolaylık sağlaması bakımından dendogram grafiğine göre daha kullanışlıdır.

Yeni yöntemin bahsedilen bu avantajlarının yanında farklı iki yöntemin birleştirilerek elde edilmesi birtakım zorlukları beraberinde getirmektedir. Önerilen yeni görsel ancak ÇBÖ ve kümeleme analizlerinin ayrı ayrı uygulanması sonucu elde edilen çıktılardan hareketle bir kaç adımdan sonra çizdirilebilmektedir. Önerilen yöntemin R, SPSS, Matlab, Phyton vb. programlama dilleri ve/veya paket programlar için kodları yazılması önerilmektedir.

Kaynakça

Bildirici, İ. Ö., & Afacan, N. (2018). Kümeleme Analizi Sonuçlarının Tematik Haritalar ile Görselleştirilmesi. Geomatik Dergis, 3(1), 92-99.

Bülbül, S., & Köse, A. (2010). Türkiye'de Bölgelerarası İç Göç Hareketlerinin Çok Boyutlu Ölçekleme Yöntemi ile İncelenmesi. İstanbul Üniversitesi İşletme Fakültesi Dergisi, 39(1), 75-94.

Cox, T., & Cox, M. (2001). Multidimensional Scaling. USA: Chapman&Hall/CRC.

Garg, N., & Rani, R. (2017). Analysis and Visualization of Twitter Data using k-means Clustering. (s. 670-675). Madurai: IEEE.

Jin, H., Shum, W.-H., Leung, K.-S., & Wong, M.-L. (2004). Expanding Self-Organizing Map for data visualization and cluster analysis. Information Sciences, 163, 157-173.

Ketchen, D., & Shook, C. (1996). The Application of Cluster Analysis in Strategic Management Research: An Analysis snd Critique. Stratagic Management Journal, 17(6), 441-458.

Lasue, J., Wiens, R. C., Stepinski, T. F., Forni, O., Clegg, S. M., & Maurice, S. (2011). Nonlinear mapping technique for data visualization and clustering assessment of LIBS data: application to ChemCam data. Anal Bioanal Chem, 400, 3247-3260.

Lorr, M. (1983). Cluster Analysis for Social Scientists. London: Josey-Bass Publishers.

Martinez, W., & Martinez, A. (2005). Exploratory Data Analysis with Matlab. USA: Chapman&Hall/CRC.

Oğuzlar, A. (2005). Kümeleme Analizinde Yeni Bir Yaklaşım: Kendini Düzenleyen Haritalar (Kohonen Ağları). İktisadi ve İdari Bilimler Dergisi, 19(2), 94-107.

Schonlau, M. (2002). The clustergram: A graph for visualizing hierarchical and nonhierarchical cluster analyses. The Stata Journal, 2(4), 391-402.

Schonlau, M. (2004). Visualizing non-hierarchical and hierarchical cluster analyses with clustergrams. Computational Statistics, 95-111.

Sechelea, A., Huu, T. D., Zimos, E., & Deligiannis, N. (2016). Twitter Data Clustering and Visualization. 23rd International Conference on Telecommunications (ICT) (s. 1-5). IEEE.

UCI. (2022, Ocak 10). UCI Machine Learning Repository. http://archive.ics.uci.edu/ml adresinden alındı

Vatansever, M., & Büyüklü, A. H. (2009). USING VISUAL DATA MINING TECHNIQUES IN CLUSTERING ANALYSIS AND AN APPLICATION. Journal of Engineering and Natural Sciences, 27, 83-104.

Wu, H.-M., Tien, Y.-J., & Chen, C.-h. (2010). GAP: A graphical environment for matrix visualization and cluster analysis. Computational Statistics and Data Analysis, 54, 767-778.

Essay on the Relationship Between Income Inequality and Inclusive Growth: Evidence From Morocco

Nahi Brahim[1*], Ritahi Oussama[2], Echaoui Abdellah[3]

[1] Faculté des Sciences Juridiques, Economiques et Sociales-Souissi/ Université Mohamed V de Rabat / Maroc

*Corresponding author e-mail: brahinahi@yahoo.fr

Abstract

For more than a century, the relationship between inequality and economic growth has been the main concern of economists. Yet the nature of this relationship remains elusive. Ideally, if economic growth is assumed to benefit the poor on the same terms as the rich. However, the process of growth and wealth creation is always accompanied by dynamic changes in the distribution of its dividends. Some studies have recently shown that the effects of growth are not automatically translated into a fair improvement in living conditions for all parts of society. Its effects have not systematically manifested themselves in terms of paid employment and increased social welfare. The reality and trajectory of some countries, particularly in Asia, over the past decades, show that despite the observation of a remarkable increase in income levels due to the growth rates achieved, socio-economic inequalities have been accentuated at the same time. Several empirical studies have contributed to this policy shift. While Stiglitz (2012) and Piketty (2014) have done most to shed new light on the growth-equity link, the trigger probably came from an unexpected source, an IMF study, which found that "lower net inequality is strongly correlated with faster and sustainable growth" (Ostry et al., 2014). This, however, led to the realization that what matters is not growth per se but the type and pattern of growth. Thus, the relationship between growth and inequality remains complex and multi-dimensional. This study aims to determine the nature of links between income inquality and inclusive growth in the Moroccan context. To do so, we started by literature review with the most recent studies, then we conducted an empirical analysis on Morocco over the period 1990-2020. Before running regression analysis, stationarity and cointegration tests were used, Jonhansen cointegration test shows that there is evidence on the long run relationship between the variables. Using Vector Error Correction Model (VECM), the analysis results show that there is a negative relationship between income inequality and economic growth. Granger Causality test shows that there is no causality between the two variables.

Key words: *Inclusive Growth, Income Inequality, Income Distribution, VECM*

Investigation of Gender in Turkey by Clustering Analysis on Provinces

Ceren Yaman Yilmaz[1*]

[1] Ankara Haci Bayram Veli University, Faculty of Economics and Administrative Sciences, Forensic Management, 06570, Ankara, Turkey

*Corresponding author e-mail: ceren.yaman@hbv.edu.tr

Abstract

Gender defines roles, responsibilities and behaviors that are appropriate to women and men by society. These roles, responsibilities and behaviors may differ from society to society, and may vary over time or in times of crisis. This differentiates the participation of people in social life by gender. The purpose of this study is to examine gender in terms of the provinces, using indicators that indicate inequality in social, economic and cultural issues between the male and female citizens living in Turkey. 18 indicators were selected in IBBS Level 3 of 2021 in light of the "Gender Statistics Data Set" published by the Turkish Statistical Institute and the data obtained from the institution's website. These are fertility rates, women's representation rate, birth rate based on the mother's marriage period (less than 1 years in marriage and more than 20 years in marriage), unmarried mother rate, foreign-population female ratio, first marriage age, one-family households rate, extended-family households rate, the marriage rate at which a man is 11 years older (age difference of spouses), the rate of consanguineous marriage, marriage rate, crude divorce rate, house sales rate by genders, the ratio of women who have never been married, the rate of birth under 17, and the average age of the mother. The study uses clustering analysis from multivariate statistical analysis techniques. Analysis was performed with SPSS (Statistical Package for the Social Sciences) v.25. In the determination of the inequalities of opportunities caused by gender patterns and the drawing of a roadmap related to what can be done, the determination of provinces that are similar and/or unique in the context of gender indicators is primarily important for policy makers and non -governmental organizations. This study is expected to raise awareness in this sense.

Key words: *Gender Statistics, Multivariate Statistics, Cluster Analysis*

Comparison of Liu, Ridge and Least Square Estimators on Relative Feed Value Estimation

Burcu Kurnaz[1*], Hasan Önder[1]

[1] Ondokuz Mayıs University, Agricultural Faculty, Department of Animal Science, 55139, Samsun, Türkiye

*Corresponding author e-mail: burcu2039@hotmail.com

Abstract

The least square estimation (LSE) method is frequently and reliably used in constructing the regression model under its assumptions. In cases where the assumptions cannot be met, the reliability of the model and its power to predict the model coefficients are weakened. When encounter with the multicollinearity problem, which is one of these assumptions, the LSE method will lose its importance. Alternative Ridge Regression and Liu estimators to EKK give more reliable results. In this study was aimed least squares method, ridge regression and Liu regression to estimate the parameters of multiple regression model in situation that the underlying assumptions of least squares estimation are untenable because of multicolinearity. In this study, it is aimed to determine the method that gives the best result with the LSE, Ridge and Liu estimators of the regression on estimating the relative feed value on continuous data. As a result of the analysis, the highest adj-R2 (0.38) value and the lowest RMSE (3.256), RRMSE (2.299), MAPE (2.249) and AIC (15.804) values were taken into account, and the most reliable result was obtained with the Liu estimator.

Key words: *Regression, Multicollinearity, Liu, Ridge, Least Square Estimator*

Adaptation of S Test for Selection of Dependent Features with Ordered Patterns in High-Dimensional Data

Deniz Cebeli[1*], **Bülent Altunkaynak**[1]

[1] Gazi University, Science Faculty, Department of Statistics, 06500, Ankara, Turkey

*Corresponding author e-mail: denizcebeli@gmail.com

Abstract

Feature selection in high-dimensional data is one of the critical steps in machine learning. High-dimensional data are data structures that contain many attributes but few observations (samples). Especially in studies on gene data, such data are frequently encountered. With the widespread use of machine learning techniques in recent years, the number of genome-wide association studies (GWAS) has increased. In such studies, the relationship between the increase in the SNP (Single-nucleotide polymorphism) level and the patterns of increase or decrease in the Marker values are tried to be determined. In classical statistics, such patterns are examined with ordered alternative tests such as Jonckheere-Terpstra (JT), Terpstra-Magel (TM), Ferdhiana Terpstra-Magel (FTM), KTP, Modified JT and S test. However, the use of these tests for high-dimensional data is not economical in terms of computation time. Therefore, these tests need to be adapted for high-dimensional data. Lin et al. (2019) proposed the fastJT algorithm for high-dimensional data. On the other hand, power test statistics than JT statistics are available in the literature, especially in extremely skewed distributions and/or convex/concave alternative hypothesis situations (Shan et al., 2014, Altunkaynak & Gamgam, 2020). In this study, we propose an adapted algorithm of S statistics for high-dimensional data, which gives better results than JT test in extreme skewed distributions and/or convex/concave alternative hypothesis situations. The results show that the S statistic is more useful for high-dimensional data.

Key words: *High-Dimensional Data, Ordered Alternatives, Jonckheere-Terpstra Test, SNP, GWAS*

INTRODUCTION

Ordinal alternative tests are used to evaluate whether a quantitative feature is associated with an ordinal feature. Such studies are especially common in the field of health. The relationship between ammonia levels and the severity of hepatic encephalopathy (Ong et al., 2003), the relationship between abnormal MRI findings and bone marrow disease (Bredella et al., 2006), the relationship between single-nucleotide polymorphisms in human genes (SNP, Single-nucleotide polymorphism) and quantitative phenotypes. Relationship (Hoffmeyer et al., 2000; Cheng et al., 2005; Kawaguchi et al., 2012; Uchiyama et al., 2012; Tan et al., 2014; Yorifuji et al., 2018) can be given as examples of such studies. Especially in SNP studies, it is tried to determine the increase or decrease patterns in the SNP level and the increase or decrease in the Marker values and to choose among thousands of SNPs.

The JT test statistic is the first test statistic used in ordered alternatives (Jonckheere, 1954). However, under different distribution structures and in the case of different alternative hypotheses, tests that are stronger than the JT statistics are also included in the literature (Altunkaynak & Gamgam, 2020). Especially Shan test statistic gives better results than other tests in extremely skewed distributions and convex/concave alternative hypothesis structures (Shan et al., 2014).

However, in studies with high-dimensional data such as genome-wide association studies (GWAS), sequential alternative tests are costly in terms of computational time. Therefore, Lin et al. (2019) proposed the fastJT algorithm. Since studies on SNP data may have different data structures, we propose an accelerated version of the Shan test for high-dimensional data in this study.

MATERIAL AND METHODS

Ordered Alternative Tests

Let $X_{i1}, X_{i2}, ..., X_{in},\ i = 1, ..., k$ be random independent samples with size n_i from k populations with continuous cumulative distribution function $F_i(x) = F((x-\theta_i)/\sigma_i),$ where $-\infty < \theta_i < +\infty$ and $\sigma_i > 0$ are location and scale parameters, respectively. H_0 hypothesis to identify whether the populations have common continuous cumuslative distribution function can be expessed as

$$H_0 : F_1(x) = F_2(x) = ... = F_k(x) \qquad \forall x. \tag{1}$$

A number of test statistics have been proposed to test the H_0 hypothesis in (1) under certain assumptions and for different forms of H_1. The ordered alternative states that the distributions are stochastically ordered, i.e.,

$$H_1 : F_1(x) \geq F_2(x) \geq ... \geq F_k(x) \quad \exists x : F_1(x) > F_k(x). \tag{2}$$

Under H_1, X_i tends to be smaller than $X_{i+1},$, $i = 1, 2, ..., k-1$, since $F_i(x) \leq F_{i+1}(x)$ implies that $P(X_i \geq X_{i+1}) \geq 1/2$. For the special case of the location model, (2) is equivalent to (Terpstra et al.,2011)

$$H_1 : \theta_1 \leq \theta_2 \leq ... \leq \theta_k \qquad (\theta_1 < \theta_k). \tag{3}$$

Similarly, the ordered alternative hypothesis

$$H_1 : F_1(x) \leq F_2(x) \leq ... \leq F_k(x) \qquad \exists x : F_1(x) < F_k(x) \tag{4}$$

states that X_i tends to be larger than X_{i+1}, $i = 1, 2, ..., k-1$, since $F_i(x) \leq F_{i+1}(x)$ implies that $P(X_i \geq X_{i+1}) \geq 1/2$ under H_1 given in (4). For the location model, (4) is equivalent to

$$H_1 : \theta_1 \leq \theta_2 \leq ... \leq \theta_k \quad (\theta_1 > \theta_k). \tag{5}$$

JT Test

The Jonckheere-Terpstra test (JT) was proposed by Terpstra (1952) and Jonckheere (1954) to test ordered alternatives in the k-group situation. This non-parametric test statistic is defined as the sum of $k(k-1)/2$ Mann-Whitney statistics.

$$JT = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} U_{ij}$$

where n_i and n_j are the sample sizes for the *i*th and *j*th groups, respectively. In this case *i*th and *j*th the Mann-Whitney statistic (U_{ij}) for groups is defined as follows.

$$U_{ij} = \sum_{l=1}^{n_i} \sum_{m=1}^{n_j} I(X_{il} < X_{jm})$$

Here $I(X_{il} < X_{jm}) = 1$ if $X_{il} < X_{jm}$ is true and 0 otherwise.

The statistic JT is approximately normally distributed under H_0. The mean and variance of this statistic are calculated as follows, respectively.

$$E(JT) = \frac{N^2 - \sum_{i=1}^{k} n_i^2}{4}$$

and

$$V(JT) = \frac{N^2(2N+3) - \sum_{i=1}^{k} n_i^2(2n_i+3)}{72}.$$

As mentioned before, the JT statistic is the sum of the U statistics obtained over all possible groups i and j (i < j). As can be seen, it is necessary to make $n_i \times n_j$ queries in the calculation of each U_{ij} statistic, and $\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} n_i \times n_j$ queries in the calculation of the JT statistic.

Especially when it comes to high-dimensional data such as gene expression data, computation time becomes very costly. To overcome this situation, Lin et al. (2019) proposed the fast JT algorithm for calculating the JT statistic. Before applying the algorithm, the observation values within each group should be ordered from smallest to largest. According to the algorithm, the steps given in Figure 1 are applied in the calculation of each U_{ij} value.

Algorithm: Calculation of U_{ij} statistics

l ← starting index of *i*th group
m ← starting index of *j*th group
while $l<n_i$ and $m<n_j$
    if ($X_{il} < X_{jm}$) **then**
        U ← $U + n_j - m + 1$
        l ← $l + 1$
    else if ($X_{il} = X_{jm}$) **then**
        starting at 1, increase *a* until the condition $X_{i(l+a)} != X_{il}$ is satisfied.
        starting at 1, increase *b* until the condition $X_{j(m+b)} != X_{jm}$ is satisfied.
        U ← $U + 0.5(a + 1)(b + 1)$
        l ← l + a
        m ← m + b
    else if ($X_{il} > X_{jm}$) **then**
        m ← $m + 1$
    end if
end while
return U

Figure 1. Accelerated algorithm for calculating U_{ij} values

The logic of the algorithm is based on the case that when two ordered vectors are taken, if the observation *i*th of the first vector is smaller than the observation *j*th of the second vector, it is also smaller than the observations *j*+1., *j*+2.,…. In this case, instead of comparing the *i*th observation of the first vector with each of the observations *j*., *j*+1., *j*+2.,… performing a single addition operation by accepting that all of them fit the ordered pattern reduces the number of operations significantly.

S Test

Shan et al. (2014) proposed the S-ordered alternative test, taking into account the size of the observations. S testinde $I(X_{il} < X_{jm})$ ifadesi R_{jm} ve R_{il} sıra sayıları arasındaki fark ile ağırlıklandırılmaktadır. By this means, the proposed test statistic is also sensitive to the size of the observation values. In addition, since it considers the ordinal numbers instead of the observation values, it preserves its feature of being robust against outliers and skewness in the distribution. In the calculation of the S statistic, first ordinal numbers are assigned to the combined sample (without group difference), and then the following formula is applied.

$$S = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} D_{ij} \ ,$$

$$D_{ij} = \sum_{l=1}^{n_i} \sum_{m=1}^{n_j} Z_{ijlm}$$

$$Z_{ijlm} = (R_{jm} - R_{il}) I(X_{jm} > X_{il})$$

Here $R_{il}(R_{jm})$ is the rank of observation $X_{il}(X_{jm})$ in the combined data.

Under H_0 , the statistics S has a normal distribution with the mean and variance.

$$E(S) = \frac{N+1}{6} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i n_j$$

$$V(S) = \frac{N^2+N}{12} - \frac{(N+1)^2}{36} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i n_j$$

$$+2\left[\sum_{i=1}^{k-1} n_i \binom{\sum_{j=i+1}^{k} n_j}{2} + \sum_{i=2}^{k} n_i \binom{\sum_{j=2}^{i-1} n_j}{2}\right] CovA \ ,$$

$$+2\left(\sum_{i=1}^{k-2} \sum_{j=i+1}^{k-1} \sum_{l=j+1}^{k} n_i n_j n_l\right) CovB$$

Where $CovA = \dfrac{2N^2+N-1}{90}$ and $CovB = \dfrac{-7N^2-11N-4}{360}$.

fastS Test

Motivation

The opposite hypothesis for position parameters in ordered alternative tests is given as follows.

$$H_1 : \theta_1 \le \theta_2 \le \ldots \le \theta_k \qquad (\theta_1 < \theta_k).$$

In case the null hypothesis is rejected, the opposite hypothesis may have different pattern structures such as linear, convex or concave, as given in the figure below.

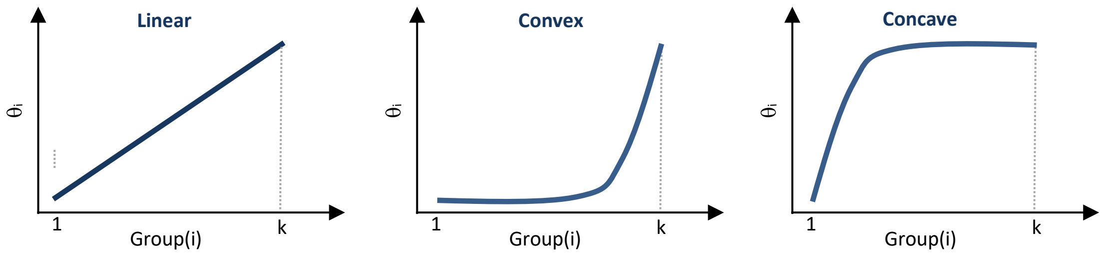


Figure 2. Linear, convex and concave alternative hypothesis structures

Considering the literature studies, it has been seen that the S statistic is stronger than the JT statistic according to the different pattern structures of the alternative hypothesis. For example, Shan et al. (2014) showed that the S test is generally more powerful than the JT test in their simulation study, taking into account different alternative hypothesis patterns, different distributions, and different sample sizes. Similar results are also found in the study by Altunkaynak and Gamgam (2020) for the comparison of multiple sequential alternative tests.

Since too many variables will be considered in high-dimensional data, different alternative hypothesis structures and different distributions will be more likely to emerge. For this reason, it can be said that the use of S statistics would be more appropriate to detect ordered patterns in high-dimensional data. However, as in the JT test, the computation time remains a problem in the S test. Because in the S test, it is necessary to question as much as in the JT test. For this reason, fastS test is proposed in this thesis study.

Algorithm

Before obtaining the test statistics, the values in each group should be ordered from smallest to largest. The steps of the algorithm for obtaining the D_{ij} value of the *i*th and *j*th groups for the test statistic are given below.

Algorithm: Calculation of U_{ij} statistics

```
l ← starting index of ith group
m ← starting index of jth group
while l<n_i and m<n_j
    if ( X_il < X_jm ) then
        k ← m
        R ← 0
        while k<n_j
            R ← R + R_jk
            k ← k + 1
        end while
        D ← D + R – (n_j – m + 1)R_il
        l ← l + 1
    else if ( X_il > X_jm ) then
        m ← m + 1
    end if
end while
return D
```

Figure 3. Accelerated algorithm for calculating D_{ij} values

Generally speaking, when $X_{il} < X_{jm}$ is provided, $\sum_{k=m}^{n_j} R_{jk} - (n_j + m + 1)R_{il}$ value is added to D_{ij}, otherwise D_{ij} value does not change. In order to implement the algorithm, R functions named fastS, FastSG and generateData have been prepared.

Simulation Study

Problem Definition

For genome-wide studies, a data structure containing p SNPs and q Markers can be given as follows.

Table 1. Data structure for genome-wide studies

Geno	SNP1	SNP2	…	SNPp	MARK1	MARK2	…	MARKq
1	0	2	…	2	-0.784	0.415	…	1.451
2	1	1	…	1	2.157	-2.011	…	-2.137
3	1	0	…	1	1.135	-1.072	…	1.099
…	…	…	…	…	…	…	…	…
100	2	0	…	1	0.212	0.473	…	-0.541

Here, SNPs are factors with different levels (0, 1, 2,…). The measurements obtained from the markers serve as the dependent variable and more than one SNP for each marker can have a pattern structure at the same time (dependent SNPs).

Design

The simulation design for the comparison of fastJT and fastS tests was prepared as follows.
Two symmetric (Normal and t) and two skewed distributions (chi-square and exponential) were taken into account in generating the marker data. Different scenarios were produced for each distribution, taking into account different location parameters. Position parameters have been created in linear, convex and concave structures, including different sequential alternative patterns.
The scenarios used in the simulation study are given in the table below (k=4).

Table 2. The scenarios used in the simulation study

Distribution	Constant (a)	Location Parameters	Sample Size
Normal(0,1)+a t_3+a	(0, 0.2, 0.4, 0.8) (0, 0.5, 0.5, 0.5) (0, 0.5, 1, 1) (0, 0, 0, 1) (0, 0, 1, 1)	θ=(0, 0.2, 0.4, 0.8) θ=(0, 0.5, 0.5, 0.5) θ=(0, 0.5, 1, 1) θ=(0, 0, 0, 1) θ=(0, 0, 1, 1)	(5, 5, 5, 5) (10, 10, 10, 10) (15, 15, 15, 15) (20, 20, 20, 20)
Chi-square(1)+a Üstel(1)+a	(0, 0.2, 0.4, 0.8) (0, 0.5, 0.5, 0.5) (0, 0.5, 1, 1) (0, 0, 0, 1) (0, 0, 1, 1)	θ=(1, 1.2, 1.4, 1.8) θ=(1, 1.5, 1.5, 1.5) θ=(1, 1.5, 2, 2) θ=(1, 1, 1, 2) θ=(1, 1, 2, 2)	(5, 5, 5, 5) (10, 10, 10, 10) (15, 15, 15, 15) (20, 20, 20, 20)

RESULTS

Table 3. Power values with data selected from Normal(0,1)+a distribution

Location parameters(a)	Sample Size	fastJT	fastS
(0, 0.2, 0.4, 0.8)	(5, 5, 5, 5)	0.4027	0.5503
	(10, 10, 10, 10)	0.5988	0.7328
	(15, 15, 15, 15)	0.6384	0.7640
	(20, 20, 20, 20)	0.7087	0.8183
(0, 0.5, 0.5, 0.5)	(5, 5, 5, 5)	0.3197	0.4426
	(10, 10, 10, 10)	0.4222	0.5614
	(15, 15, 15, 15)	0.4188	0.5459
	(20, 20, 20, 20)	0.4537	0.5781
(0, 0.5, 1, 1)	(5, 5, 5, 5)	0.5151	0.6754
	(10, 10, 10, 10)	0.6747	0.7952
	(15, 15, 15, 15)	0.7434	0.8446
	(20, 20, 20, 20)	0.8807	0.9379
(0, 0, 0, 1)	(5, 5, 5, 5)	0.4410	0.6055
	(10, 10, 10, 10)	0.5332	0.6741
	(15, 15, 15, 15)	0.6556	0.7814
	(20, 20, 20, 20)	0.7596	0.8560
(0, 0, 1, 1)	(5, 5, 5, 5)	0.4866	0.6467
	(10, 10, 10, 10)	0.6952	0.8162
	(15, 15, 15, 15)	0.8569	0.9250
	(20, 20, 20, 20)	0.7684	0.8534

For k=4, it was observed that the fastS statistic gave better results when the sample size was 20 and 40, when data were generated from the standard normal distribution. When the power values are examined with the data selected from the normal distribution, it is seen that the power values of the fastS statistics are higher for all scenarios.

Table 4. Power values with data selected from t_3+a distribution

Location parameters	Sample Size	fastJT	fastS
(0, 0.2, 0.4, 0.8)	(5, 5, 5, 5)	0.3434	0.4718
	(10, 10, 10, 10)	0.4702	0.6066
	(15, 15, 15, 15)	0.5671	0.6999
	(20, 20, 20, 20)	0.6212	0.7426
(0, 0.5, 0.5, 0.5)	(5, 5, 5, 5)	0.2950	0.3996
	(10, 10, 10, 10)	0.3607	0.4803
	(15, 15, 15, 15)	0.3711	0.4900
	(20, 20, 20, 20)	0.4415	0.5715
(0, 0.5, 1, 1)	(5, 5, 5, 5)	0.4558	0.6131
	(10, 10, 10, 10)	0.5104	0.6728
	(15, 15, 15, 15)	0.7030	0.8148
	(20, 20, 20, 20)	0.8140	0.8967
(0, 0, 0, 1)	(5, 5, 5, 5)	0.4017	0.5569
	(10, 10, 10, 10)	0.5142	0.6598
	(15, 15, 15, 15)	0.6022	0.7399
	(20, 20, 20, 20)	0.6644	0.7885
(0, 0, 1, 1)	(5, 5, 5, 5)	0.4438	0.6035
	(10, 10, 10, 10)	0.6540	0.7893
	(15, 15, 15, 15)	0.6646	0.7804
	(20, 20, 20, 20)	0.8122	0.8926

For k=4, it was observed that the fastS statistic gave better results when the sample size was 20 and 40, when data were generated from the t distribution. When the power values are examined with the data selected from the t distribution, it is seen that the power values of the fastS statistics are higher for all scenarios.

Table 5. Power values with data selected from Chi-square(1)+a distribution

Location parameters	Sample Size	fastJT	fastS
(1, 1.2, 1.5, 2)	(5, 5, 5, 5)	0.5469	0.6915
	(10, 10, 10, 10)	0.8540	0.9186
	(15, 15, 15, 15)	0.9046	0.9509
	(20, 20, 20, 20)	0.9131	0.9525
(1, 1.5, 1.5, 1.5)	(5, 5, 5, 5)	0.4509	0.6029
	(10, 10, 10, 10)	0.5263	0.6680
	(15, 15, 15, 15)	0.5228	0.6556
	(20, 20, 20, 20)	0.6143	0.7404
(1, 1.5, 2, 2)	(5, 5, 5, 5)	0.7062	0.8258
	(10, 10, 10, 10)	0.8046	0.8879
	(15, 15, 15, 15)	0.8944	0.9452
	(20, 20, 20, 20)	0.9306	0.9665
(1, 1, 1, 2)	(5, 5, 5, 5)	0.5228	0.7070
	(10, 10, 10, 10)	0.6229	0.7554
	(15, 15, 15, 15)	0.7676	0.8765
	(20, 20, 20, 20)	0.8521	0.9248
(1, 1, 2, 2)	(5, 5, 5, 5)	0.5406	0.6946
	(10, 10, 10, 10)	0.7259	0.8348
	(15, 15, 15, 15)	0.6901	0.7861
	(20, 20, 20, 20)	0.8778	0.9390

For k=4, it was observed that the fastS statistic gave better results when the sample size was 20 and 40, when data were generated from the chi-square distribution. When the power values are examined with the data selected from the chi-square distribution, it is seen that the power values of the fastS statistics are higher for all scenarios.

Table 6. Power values with data selected from the exponential(1)+a distribution

Location parameters	Sample Size	fastJT	fastS
(1, 1.2, 1.4, 1.8)	(5, 5, 5, 5)	0.3874	0.4898
	(10, 10, 10, 10)	0.6414	0.7563
	(15, 15, 15, 15)	0.7853	0.8734
	(20, 20, 20, 20)	0.8721	0.9284
(1, 1.5, 1.5, 1.5)	(5, 5, 5, 5)	0.3777	0.5138
	(10, 10, 10, 10)	0.4788	0.6153
	(15, 15, 15, 15)	0.5384	0.6715
	(20, 20, 20, 20)	0.6007	0.7290
(1, 1.5, 2, 2)	(5, 5, 5, 5)	0.5595	0.6958
	(10, 10, 10, 10)	0.6742	0.7827
	(15, 15, 15, 15)	0.9008	0.9525
	(20, 20, 20, 20)	0.9255	0.9638
(1, 1, 1, 2)	(5, 5, 5, 5)	0.6381	0.8338
	(10, 10, 10, 10)	0.7575	0.8809
	(15, 15, 15, 15)	0.8203	0.9081
	(20, 20, 20, 20)	0.9176	0.9672
(1, 1, 2, 2)	(5, 5, 5, 5)	0.6178	0.7583
	(10, 10, 10, 10)	0.7241	0.8241
	(15, 15, 15, 15)	0.9570	0.9829
	(20, 20, 20, 20)	0.9574	0.9818

For k=4, it was observed that the fastS statistic gave better results when the sample size was 20 and 40, when data were generated from the exponential distribution. When the power values are examined with the data selected from the exponential distribution, it is seen that the power values of the fastS statistics are higher for all scenarios.

CONCLUSION

In this study, the S test statistic is adapted to detect ordered alternatives in GWAS data with high-dimensional data structure. A simulation study was conducted to compare the proposed fastS statistic with the current fastJT statistic. In the simulation study, the increase in the SNP level and the increase in the Marker values were produced for different scenarios in accordance with the real data structure. It has been observed that the fastS statistic gives better results than the fastJT statistic in all scenarios. In addition, R functions have been created for the fastS statistic. Therefore, fastS statistics can be used for the determination of ordered alternatives (feature selection) in GWAS data, which has become an important field of study in recent years.

References

Terpstra, JT, Chang CH, Magerl RC, 2011. On the use of Spearman's Correlation Coefficient for testing ordered alternatives. Journal of Statistical Computation and Simulation, 81(11):1381–1392.

Terpstra JT, 1952. The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. Indagationes Mathematicae, 14(3):327–333. [p147, 148]

Jonckheere AR, 1954. A distribution-free k-sample test against ordered alternatives. Biometrika, 41(1): 133–145.

Buning H, Kossler W, 1996. Robustness and efficiency of some tests for ordered alternatives in the c-sample location problem. Journal of Statistical Computation and Simulation, 55(4):337–352.

Neuhäuser M, Liu PY, Hothorn LA, 1998. Nonparametric tests for trend: Jonckheere's Test, a modification and a maximum test. Biometrical Journal, 40(8):899–909.

Shan GG, Young D, Kang L, 2014. A new powerful nonparametric rank test for ordered alternative problem. Plos One, 9(11):1–10.

Ong JP, Aggarwal A, Krieger D, Easley KA, Karafa MT, Van Lente F, Arroliga AC, Mullen D, 2003. Correlation between ammonia levels and the severity of hepatic encephalopathy. The American Journal of Medicine, 114(3):188-193.

Bredalla MA, Steinbach LS, Morgan S, Ward M, Davis JC, 2006. Mri of the sacroiliac joints in patients with moderate to severe ankylosing spondylitis. American Journal of Roentgenology, 187(6):1420-1426.

Hoffmeyer S, Burk O, Von Richter O, Arnold HP, Brockmöller J, Johne A, Cascorbi I, Gerloff T, Roots I, Eichelbaum M, Brinkmann U, 2000. Functional polymorphisms of the human multidrug resistance gene: Multiple sequence variations and correlation of one allele with p glycoprotein expression and activity in vivo. Proceedings of the National Academy of Sciences of the United States of America, 97(7):3473–3478.

Cheng Q, Yang W, Raimondi SC, Pui CH, Relling MV, Evans WE, 2005. Karyotypic abnormalities create discordance of germline genotype and cancer cell phenotypes. Nature Genetics, 37(8):878–882.

Kawaguchi T, Sumida Y, Umemura A, Matsuo K, Takahashi M, Takamura T, Yasui K, Saibara T, Hashimoto E, Kawanaka M, Watanabe S, Kawata S, Imai Y, Kokubo M, Shima T, Park H, Tanaka H, Tajima K, Yamada R, Matsuda F, Okanoue T, Japan D, 2012. Study Group of Nonalcoholic Fatty Liver. Genetic polymorphisms of the human PNPLA3 gene are strongly associated with severity of non-alcoholic fatty liver disease in Japanese. PloS one, 7(6):e38322–e38322.

Uchiyama T, Kanno H, Ishitani K, Fujii H, Ohta H, Matsui H, Kamatani N, Saito K, 2012 An snp in CYP39A1 is associated with severe neutropenia induced by docetaxel. Cancer Chemotherapy and Pharmacology, 69(6):1617–1624.

Tan HL, Zain SM, Mohamed R, Rampal S, Chin KF, Basu RC, Cheah PL, Mahadeva S, Mohamed Z, 2014. Association of glucokinase regulatory gene polymorphisms with risk and severity of non-alcoholic fatty liver disease: An interaction study with adiponutrin gene. Journal of Gastroenterology, 49(6):1056–1064.

Yorifuji K, Uemura Y, Horibata S, Tsuji G, Suzuki Y, Miyagawa K, Nakayama K, Hirata K.-i, Kumagai S, Emoto N, 2018. CHST3 and CHST13 polymorphisms as predictors of bosentan-induced liver toxicity in Japanese patients with pulmonary arterial hypertension. Pharmacological Research, 135:259–264.

Altunkaynak B, Gamgam H, 2020. npordtests: An R Package of Nonparametric Tests for Equality of Location Against Ordered Alternatives. Gazi University, Ankara, Turkey.

Potter RW, Sturm GW, 1981. The Power of Jonckheere's Test. The American Statistician, 35, 249- 250.

Conflict of Interest

The authors have declared that there is no conflict of interest.

An Evaluation of Supermarkets From the Lens of Multiple Criteria: The Intuitionistic Fuzzy TOPSIS Method

Feride Tuğrul[1*]

[1] Kahramanmaraş Sütçü İmam University, Faculty of Science and Arts, Department of Mathematics, 46100, Kahramanmaraş, Turkey

*Corresponding author e-mail: feridetugrul@gmail.com

Abstract

Following the definition of the fuzzy logic, the intuitionistic fuzzy logic, namely which is the extension of fuzzy logic, was defined. Thanks to the intuitionistic fuzzy sets, the degree of hesitation gains meaning and a new perspective is presented to the researchers' study. For a long time, supermarket chains have an important role in people's lives for shopping. This study is an attempt to evaluate supermarkets by both the consumer and the market researcher through the intuitionistic fuzzy TOPSIS. In this regard, the decision makers' views were expressed in linguistic terms and ranked from the most preferred supermarket to the least preferred supermarket. The satisfaction of using linguistic terms is that it gives decision makers the chance to express views that they cannot express in numerical values. Furthermore, undecided situations were interpreted by means of the intuitionistic fuzzy sets. Thanks to the TOPSIS method, the positive ideal and the negative ideal solutions were calculated and the best alternative was determined. This method, which may be easily used in many application areas, offers researchers more consistent results than the others due to its advantages. Besides, the main goal of this study is to show an innovative implementation of the intuitionistic fuzzy TOPSIS method on its algorithm and mathematical basis.

Key words: *Intuitionistic Fuzzy Set, TOPSIS Method, Multi Criteria Decision Making.*

The Relation of the Intuitionistic Fuzzy Sets with Multi-Criteria Decision Making Methods

Feride Tuğrul[1*]

[1] Kahramanmaraş Sütçü İmam University, Faculty of Science and Arts, Department of Mathematics, 46100, Kahramanmaraş, Turkey

*Corresponding author e-mail: feridetugrul@gmail.com

Abstract

Multi-criteria decision-making aims to choose the best of all the multiple alternatives when there are many criteria. In the relevant literature, numerous multi-criteria decision-making methods such as PROMETHEE, TOPSIS, ELECTRE, AHP, etc. were described. Moreover, multi-criteria decision-making methods were redefined through using the intuitionistic fuzzy sets, studies emerged that will shed light on many application areas. Thanks to the intuitionistic fuzzy sets, membership, nonmembership and hesitation degrees are evaluated simultaneously. The intuitionistic fuzzy sets and multi-criteria decision-making methods provide satisfactory results in many application areas such as medicine, education, economics, banking, engineering, renewable energies and selection and ranking decision-making processes, etc. In this study, the results through the intuitionistic fuzzy based-methods and using other methods were compared and interpreted. This study will make a great contribution to presenting a new perspective to decision makers who benefit from multi-criteria decision-making together with the intuitionistic fuzzy sets approach.

Key words: *Fuzzy set, Intuitionistic Fuzzy Set, Multi Criteria Decision Making.*

Parameter Estimation for Geometric Process with Exponential Distribution Under Censored Data

Ömer Altındağ[1*]

[1] Bilecik Şeyh Edebali University, Department of Statistics and Computer Sciences,
11230, Bilecik, Turkey

*Corresponding author e-mail: omer.altindag@bilecik.edu.tr

Abstract

In statistical analysis of a series of consecutive events, a general approach is to use a stochastic counting process model. The homogenuous Poisson process (HPP) is utilized when the process holds Markov property. If the inter-arrival times of the process are assumed to be independent and identically distributed, a general approach is to use renewal process (RP) as model. The HPP and its generalization, i.e., the RP, are only used when the inter-arrival times are identically distributed. If the inter-arrival times follow a trend, they can't be assumed as identically distributed. In this case some altervative counting processes should be considered. One of the counting processes having capacity to model trend of the inter-arrival times is geometric process (GP). This process can be preferred when the inter-arrival times follow a monotonic trend. The GP is a generalization of RP allowing to model monotonic trend of inter-arrival times. This process is defined as follows. Let {N(t), t≥0} be a counting process, giving the number of events occurred up to time t, and Xk, k=1,2,… be inter-arrival times of the process. The counting process {N(t), t≥0} is called as GP if there exist a real number a>0 such that ak-1Xk, k=1,2,… are independent and identically distributed with a distribution function F. Here, the parameter a is called as trend parameter of the process. Let Fk denote the distribution function of Xk, then Fk(x)=F(ak-1x), k=1,2,… . The process is stochastically decreasing when a>1 and stochastically increasing when a<1. It reduces to a RP when a=1. Parametric and non-parametric estimation of the trend parameter a and Fk are well studied in the literature. However, the inter-arrival times of consecutive events may be observed as censored. In this study, we consider parametric estimation of the trend parameter a and Fk under censored data when F is assumed to be exponential with mean θ, that is F(x)=1-exp{-x/θ}, x>0; θ>0. We obtain maximum likelihood estimators of model parameters and derive their asymptotic distributions. A simulation study is carried out to observe small sample performance of the estimators.

Key words: *Geometric Process, Trend Parameter, Censored Data, Parametric Estimation*

Partition of Pearson's Chi-Square Statistic to Its Components

Rabia Albayrak Delialioğlu[1*], **Zahide Kocabaş**[1]

[1]Ankara University, Faculty of Agriculture, Department of Animal Science, Biometrics and Genetics Unit, 06110, Ankara, Turkey

* Corresponding author e-mail: ralbayrak@ankara.edu.tr

Abstract

The data obtained in studies in medicine, biology and social sciences are usually discrete and do not show normal distribution. Cross tables are used in the evaluation of these data, and "categorical data analysis methods" are applied in the analysis. Statistical methods that are frequently performed for this purpose are Pearson's chi-square test, G-test and Fisher's exact test. In most studies, the researcher may wish to explore the relationship between three or more categorical variables, each with at least two levels. However, when the number of categorical variables is three or more in Pearson chi-square statistics, the variables are considered and analyzed in pairs. The results obtained in this way will lead to loss of information. For this reason, in this study, the division of Pearson's chi-square statistics into components in multi-way tables is emphasized. Thus, it is possible to examine the data obtained from the study in more detail and to test the relationships between the categorical variables discussed in a more sensitive way.

Key words: *Pearson's chi-square statistic, Multi-way tables, Components of chi-square statistic*

Çok Kriterli Karar Verme Modellerinde AHP Yönteminin İncelenmesi

Aysu Yaşar[1*], Mehmet Kenan Terzioğlu[2]

[1] Nişantaşı Üniversitesi, İktisadi İdari ve Sosyal Bilimler Fakültesi, İşletme Bölümü, Edirne, Turkey
[2] Trakya Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, Edirne, Turkey

*Corresponding author e-mail: yasar.aysu@nisantasi.edu.tr

Özet

Çok kriterli karar verme metodolojileri, ikili karşılaştırmaları bulanık kümeler veya tek bir olasılık dağılımı (örneğin, tek tip, üçgen) olarak modelleyerek kesin olmayan ikili karşılaştırmaları ele almaktadır. Analitik Hiyerarşi Süreci, karar vericiler ve araştırmacılar tarafından yaygın kullanıma sahip olan ve ikili karşılaştırmaları ele alan bir yöntem olmaktadır. Yöntem kapsamında alternatifleri değerlendirmek için kriterlerin tanımı, ağırlıkların hesaplanması ve tutarlılık şartının sağlanması gerekmektedir. Özellikle karşılaştırılacak çok fazla alternatif olduğunda, geleneksel AHP tarafından sağlanan ikili karşılaştırma yöntemi karmaşıklaştığı için yöntemin işlem adımlarının anlaşılması önemli olmaktadır. Çalışma kapsamında, bilgi güvenliği risk değerlendirmeleri çerçevesinde AHP yönteminin etkin kullanılabilmesi için işlem adımlarının ortaya konması amaçlanmaktadır.

Key words: *Bilgi Güvenliği, Analitik Hiyerarşi Süreci (AHP), Çok Kriterli Karar Verme*

GİRİŞ

Bilgi güvenliği risk değerlendirmesi, olası tehditler ve güvenlik açıkları ile risklerin gerçekleşmesini ve etkilerini ilgili teknoloji standartlarına göre incelemeyi amaçlayan bir değerlendirmedir. Potansiyel tehditler, sistem açıkları için olası risk oluşturduğunda tehdidin etki boyutu kuruluşlar için bilgi güvenliği risklerini belirlemektedir. Kuruluşlar için bilgi sistemi güvenliğinin sağlanmasında, sistemdeki belirsizlik durumunda bile varlıkların risk değerlendirmeleri önemli olmaktadır.

Risk değerlendirmesinin ilk adımı tehditlerin tanımlanmasıdır. Olası tehditlerin analizi, karmaşık organizasyon yapılarını, sistemler ve insan davranışları arasındaki etkileşimi, alt sistemler arasındaki birbiriyle ilişkili faktörleri ve belirli toplumların güvenliğe olan tutumlarını kapsamalıdır. Risklerin değerlendirmesinde kullanılan karar verme yöntemleri; öncelikleri birbirlerine eklenen faktörler olarak ele alan, geometrik ya da aritmetik ortalamayla çoğunluk tercihini değerlendiren, riskin kalitatif değerlendirmesiyle birlikte kritik güvenlik risk alanlarını ve azaltma politikalarını belirlemek amacıyla kullanılmaktadır. Risk derecesini, risk faktörlerin olasılığını ve etkisini ele alan OCTAVE gibi yöntemlerin risk değerlendirmesinden sonra ortaya koyması gereken belirli koşulları bulunmaktadır. Doğrusal olmayan risk analizi yöntemlerinde, yinelenebilir olması, tehdit profilleri gerektirmesi ve güvenlik açıklarının bir dökümanının oluşturulması gibi gereksinimlerin sağlanması önemli olmaktadır. Bununla birlikte, bu yöntemlerde matris analizinin fikri, önce risk olayının ortaya çıkma olasılığının ve etkisinin değerlendirilmesi daha sonra da her bir risk olayının risk derecesinin sezgisel ve basit matris değerlendirilmesine dayanmaktadır. Hiyerarşi sürecinin analizi ve karar verme yöntemi olan AHP, risk değerlendirme uygulamalarında sistemin belirsizliğini ele almada nitel değerleri nicel değerlere dönüştürelebildiğinden diğer yöntemlerle karşılaştırıldığında daha yaygın bir kullanıma sahiptir.

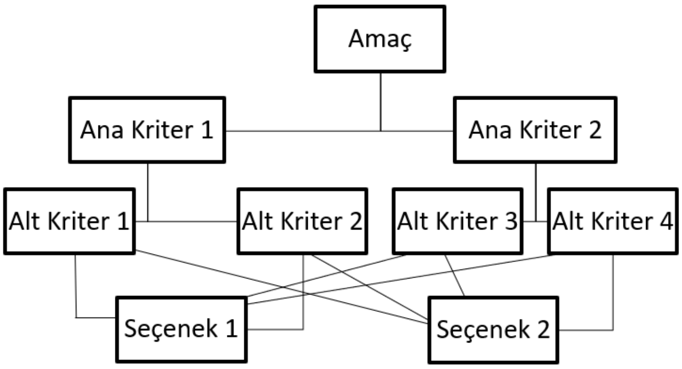
Günümüzde bilgisayar ve ağ teknolojisi, modern bilgi edinme, depolama, işleme, durum analizi, karar destek, yönetim ve kontrolün temel aracı haline gelmektedir. Ancak bilgisayarların ve ağ yapılarının zayıflıkları ve eksiklikleri nedeniyle bilgi ağının kendisi büyük bir güvenlik riski taşımaktadır. Bilgi ağı, varlıklardan (güvenlik politikalarıyla korunan bilgi veya değerli kaynaklar) oluşmakta ve her bir varlığın riski, tüm bilgi ağı riskinin temelini oluşturduğundan varlıkların her birinin riskinin değerlendirilmesi gerekmektedir. Bu kapsamda da her bir varlığın riski için varlık değeri, tehdit düzeyi ve güvenlik açığı düzeyi dikkate alınması önemli olmaktadır. Bilgi ağı karmaşık bir insan-bilgisayar sistemi olduğundan yalnızca sistem yapısı ve bileşimi değil aynı zamanda sistemdeki çok sayıda faktör de doğru bir şekilde ölçülmelidir. Bilgi varlığı risk analizlerinde faktörlerin ölçümünde kullanılan AHP yöntemi, nitel ve nicel analizin bir kombinasyonu olan sistematik ve hiyerarşik bir karar analizi yöntemi olup zor ve nicel değerlendirme öğelerinin değerlendirilmesi için uygun olmaktadır. AHP yaklaşımı diğer yöntemlerle bütünleştirilerek uygulanmasında artış görülmekte ve birçok farklı alanda kullanıma sahip olmaktadır. Sum (2015), bir sigorta şirketinin faaliyetleri kapsamında risk değerlendirmesi için AHP yöntemini kullandığı çalışmasında; AHP'nin öznel yargılar üzerinde tutarlılık kontrolü sağlayabildiğini, çok sayıda riski yapılandırılmış bir çerçeve içinde düzenleyebildiğini, risk yöneticilerine açık risk değişimleri yapmalarında yardımcı olduğunu ve anlaşılması kolay ve sistematik bir risk değerlendirme süreci sağlayabildiğini göstermektedir. Dhir vd. (2015), İngiltere’de altyapı projesi için en uygun raylı taşıt üreticisini belirlemek için maliyet-fayda analizi ve çok kriterli analiz gibi metodolojilerin kullanımını göz önünde bulundurarak SWOT analizi ile AHP yöntemini kullanmaktadır. Yavuz (2017), limanlardaki iş güvenliği sorunlarını çözmek ve iş güvenliğine farklı bir bakış açısı kazandırmak amacıyla risk değerlendirme yöntemlerinden oransal risk değerlendirme tekniğini (PRAT) kullanılarak limanlara özgü risk değerlendirme yaptığı çalışmasında öznel yargılardan ortaya çıkabilecek sorunları giderebilmek için AHP (Analitik Hiyerşi Prosesi) yaklaşımından faydalanarak risklerin birbirleri arasında önceliklendirilmesinde kullanmaktadır. Chunmei Su vd. (2018), bilgi ağı güvenliği risk değerlendirme yöntemlerini ve modellerini analiz ettikleri çalışmasında, risk yargı matrisi tutarlılığı problemini etkin bir şekilde çözmek amacıyla varlıkların değerini atamak için geliştirilmiş olan AHP yöntemini önermektedir.

ANALİTİK HİYERARŞİ SÜRECİ

Saaty (1997) tarafından karmaşık problemlerin çözümlenmesi için geliştirilen Analtik Hiyerarşi Süreci, ekonomik, sosyal ve teknik birçok alanda kullanılan çok kriterli karar verme yöntemlerinden biridir. AHP yaklaşımı, birden fazla seçeneğin arasından karar verici(lerin) belirlemiş olduğu kriterler kapsamında karar seçeneklerini yüksek önemden az öneme sıralayan bir tekniğe sahiptir. AHP, karar almada hem nicel hem de niter kriterleri değerlendiren, grup veya bireylerin tercihlerini, bilgilerini veya sezgilerini de karar verme sürecine dahil etmektedir (Saaty, 1980).

AHP’de öncelikle amaç belirlenerek amacı etkileyen faktör ve alt faktörlerin ortaya konması gerekmektedir. Faktörlerin belirlenmesinde farklı kurum, uzman görüşleri, anket çalışmalarından faydalanılmaktadır. Problemin her biri için en az bir elemandan oluşan hiyerarşik yapı oluşturularak altta yer alan elamanın üstte yer alan elemanı etkilediği varsayımı ile ikili karşılaştırmalar yapılmaktadır. Yapılan karşılaştırma kapsamında karşılıklı etkilerin oranı belirlenmektedir. AHP’nin temelinde systemin çeşitleri ögelerinin tüm system üzerindeki etkilerini değerlendirip göreceli önemlerini ortaya koyma yaklaşımı bulunmaktadır.

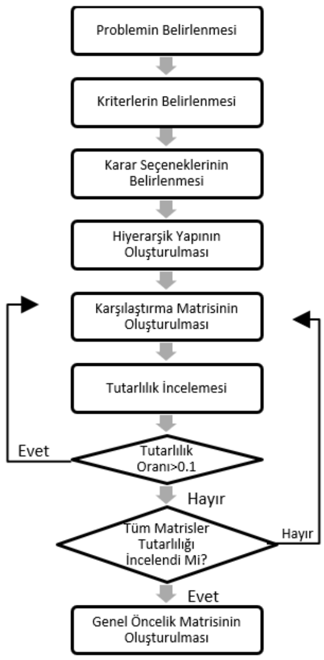
AHP yaklaşımı problemin hiyerarşik bir yapıya dönüştürüldüğü ve daha sonra karar vericinin yargılarına dayalı olarak alternatiflerin sıralandığı en popüler çok kriterli karar verme tekniklerinden biridir. AHP modelinin hiyerarşik yapısı Şekil 1.’de gösterilmektedir.



Şekil 1. AHP Yapısı

AHP yaklaşımında, genel karar hedefi (en iyi alternatif) en üst seviyede, kriterler (gerekirse, alt kriterler) orta seviye(ler)de ve seçenekler en alt seviyede yer almaktadır. AHP'nin özellikle nitel değerlendirmeler için kullanılmasının nedeni, nitel kriterlerin nicel veri olarak ifade edilememesidir. AHP yaklaşımının sağladığı analizlerde karar vericilerin kabul edilebilirliği ve güveni diğer çok kriterli karar yaklaşımlarına göre daha yüksektir. Bununla birlikte AHP'nin diğer avantajları ise: öznel kararlar için sistematik bir metodoloji sağlamak, duyarlılık analizi yapılabilmesi, değerlendirme kriterlerinin ağırlıkları hakkında bilgi sunması, farklı koşullarda uygulanabilir olması şeklinde sıralanabilir.

AHP yaklaşımında bir problem çözmek için Şekil 2'de gösterin işlem adımları uygulanmaktadır (Özbek vd., 2013)



Şekil 2. AHP İşlem Adımları

Adım 1: Problemin tanımlanması: Karar verilmesi gereken sorunun AHP yaklaşımıyla çözülüp çözülemeyeceği belirlenmelidir. Geçmiş deneyim ve uzman görüşleri doğrultusunda problemin çözümü belirlenerek çözüm algoritması oluşturulur.
Adım 2: Hiyerarşinin oluşturulması: Hiyerarşinin en tepe noktasında amaç yer almak üzere bir alt seviyede amacı etkileyen kriterler ve varsa alt kriterler belirlenerek sonucu belirleyen seçenekler yapısı oluşturulur.
Adım 3: İkili karşılaştırma matrislerinin oluşturulması: İkili karşılaştırma matrisi, hiyerarşik yapıda bir düzeyde yer alan kriterlerin bir üst faktör kapsamında ikili olarak birbirleriyle karşılaştırılmasıyla elde edilmektedir. Matriste yer alan ham veriler Tablo 1'deki 1-9 ölçeği kullanılarak seçilen nitel kriterlere göre önem kıyası yapılarak oluşturulmaktadır.

Tablo 1. Karşılaştırma Ölçeği

Önem Yoğunluğu	Tanım
1	Eşit Önem
3	Bir Faktörün Diğerine Kıyasla Orta Önemi
5	Bir Faktörün Diğerine Kıyasla Güçlü Önemi
7	Bir Faktörün Diğerine Kıyasla Çok Güçlü Önemi
9	Bir Faktörün Diğerine Kıyasla Aşırı Derecede Önemli Olması
2,4,6,8	Ara Değerler
$1/a_{ij}$	Ters Karşılaştırma Için Karşılıklar

Kaynak: Saaty, T. L. (1980)

İkili karşılaştırma yargısı olan a_{ij}, bir üst düzeydeki faktöre göre i. ve j. kriterlerinin göreceli önemini göstermek üzere A matrisi,

$$A = A(a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \tag{1}$$

belirlenmektedir. Bir düzeyde n sayıda eleman bulunduğundan n(n-1)/2 adet karşılaştırma yapılması gerekmektedir. Karşılaştırmalar, ikili karşılaştırma matrisinin tüm değerleri 1 olan köşegenin üstünde kalan elemanlar için yapılarak a_{ij}, i. kriter ile j. kriterin karşılaştırma değeri; a_{ji} değerinin karşılıklı karşılaştırma değeri ise $1/a_{ij}$ eşitliğinden elde edilmektedir (Saaty, 1980). İkili karşılaştırma matrisi Tablo 2'de gösterilmektedir.

Tablo 2. Karşılaştırma Matrisi

A	Kriter 1	Kriter 2	Kriter 3	…	Kriter $_n$
Kriter 1	1	a_{12}	a_{13}	…	a_{1n}
Kriter 2	$a_{12} = 1/a_{12}$	1	a_{23}	…	a_{2n}
Kriter 3	$a_{13} = 1/a_{13}$	$1/a_{23}$	1	…	a_{3n}
…	…	…	…	1	…
Kriter $_n$	$a_{n1} = 1/a_{1n}$	$a_{n2} = 1/a_{2n}$	$a_{n3} = 1/a_{3n}$	…	1

Adım 4: İkili karşılaştırma matrislerinin normalize edilmesi: İkili karşılaştırma marisi oluşturulduktan sonra,

$$a'_{ij} = \frac{a_{ij}}{\sum_{i=1}^{n} a_{ij}} \tag{2}$$

denklemi ile normalize edilmektedir.

Adım 5: Öncelik vektörünün hesaplanması: Normalize edilen matriste herbir satır toplamı matrisin boyutuna bölünerek ortalaması hesaplanmakta ve elde edilen değerler her bir kritere ait önem ağırlıkları olmaktadır. Öncelik vektörü olarak adlandırılan önem ağırlıkları w_i,

$$w_i = \left(\frac{1}{n}\right)\sum_{j=1}^{n} a'_{ij} \qquad i,j = 1,2,\dots,n \tag{3}$$

şeklinde elde edilmektedir.

Adım 6: Tutarlılık oranının hesaplanması: Kriterler arasında karşılaştırma sonucunu belirleyen değerlerin tutarlı olup olmadığının kontrol edilmesi gerekmektedir. λ_{max} özdeğer olmak üzere tutarlılık endeksi CI;

$$\lambda_{max} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{j=1}^{n} a_{ij}w_j}{w_i}\right] \tag{4}$$

$$CI = \frac{\lambda_{max}-n}{n-1} \tag{5}$$

olarak hesaplanmaktadır. Hesaplanan tutarlılık endeksinin değerlendirilmesi için “Random Indeks” (RI) değerlerinin bilinmesi gerekmektedir. Her matrisin boyutuna karşılık gelen bir random indeks değeri bulunmakta ve bu değerler Tablo 3’te gösterilmektedir (Kwiesielewicz vd., 2004).

Table 3. Random Indeksi

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.53	1.56	1.57	1.59

CI ve RI değerleri belirlendikten sonra tutarlılık oranı (CR),

$$CR = CI/RI \tag{6}$$

hesaplanmakta ve elde edilen değerin 0.10’dan küçük olduğunda oluşturulan karar matrisinin tutarlılığına karar verilmektedir.

Adım 7: Seçeneklerin ikili karşılaştırması: Seçeneklerin tümü için bir kritere göre önemini belirlemek amacıyla Eşitlik 1’de formüle edilen ikili karşılaştırma matrisleri tekrardan oluşturulmaktadır. Kriterlerin ağırlığının belirlendiği süreçte Eşitlik (2) ve (3) numaraları kullanılarak tutarlılık oranları kontrol edilmelidir.

Adım 8: Seçeneklerin Sıralanması: Seçeneklerin amaca göre sıralamasının belirlenmesinde ana kriterler, varsa alt kriterlerin kendi aralarında ve seçeneklerin ise her bir kriterle karşılaştırmaları oluşturularak öncelik vektörlerinin belirlenmesi gerekmektedir. Her bir kritere ait genel ağırlık değerleri ile seçeneklerin o alt kritere göre olan tercih değerleri çarpılarak seçeneklerin toplam ağırlıkları elde edilmekte ve her seçeneceğin ağırlık değerleri toplanarak o seçeneğin sıralaması bulunmaktadır (Özbek, 2021).

TARTIŞMA VE SONUÇ

Bilgi güvenliği risk değerlendirmeleri sadece risk seviyesini belirtmek amacıyla değil, aynı zamanda risk hesaplamalarına ve değerlendirme sonuçlarına, daha yüksek olasılıklı tehditler tarafından etkilenebilecek zayıflıkların ortaya çıkartılmasına dayanarak tüm sistemin güvenliğini etkileyen temel onayları tanımlamak amacıyla da gerçekleştirilmektedir. Bilgi güvenliğini her yönüyle kontrol altına almak amacıyla öncelikle uygulanacak sistemde uygun risk kontrol önlemlerini belirlenmeli ve eksiksiz bilgi güvenliği yönetim sistemi kurulması ile olası tehditleri kontrol altına almak mümkün olabilmektedir.

Risk değerlendirme analizlerinde olası tehditleri sınıflandırmak için çeşitli çok kriterli teknikler kullanılmaktadır. AHP yöntemiyle tutarlılığı etkin bir şekilde iyileştirebilen matrisler belirlenerek sinir ağları sistemlerine dayalı bilgisayar tabanlı risk matrisi ölçüm yöntemleri birleştirilerek risk değerlendirme analizleri başarılı bir biçimde otomatikleştirilebilir.

Kaynakça

C Su, Y Li, W Mao, S Hu, 2018. "Information Network Risk Assessment Based on AHP and Neural Network," *2018 10th International Conference on Communication Software and Networks (ICCSN)*, pp. 227-231, doi: 10.1109/ICCSN.2018.8488314.

Kwiesielewicz M, Van Uden E, 2004. Inconsistent and Contradictory Judgements in Pairwise Comparison Method in The AHP. Computer & Operations Research, 31(5), 713-719.

Özbek A, 2015. Çok Kriterli Karar Verme Yöntemleri ve Excel ile Program Çözümü. Ankara: Seçkin Akademik ve Mesleki Yayınlar.

S, Dhir, MV, Marinov, Worsley D, 2015. "Application of the analytic hierarchy process to identify the most suitable manufacturer of rail vehicles for High Speed 2. Case Studies," Transport Policy,vol. 3(4), pp. 431-448.

Saaty TL, 1980. The analytic hierarchy process: planning, priority setting. New York: McGraw Hill International Book Co.

Sum R, 2015. “Risk Prioritisation Using The Analytic Hierarchy Process”, Proceedings of the and Innovation and Analytics Conference & Exhibition.

Yavuz H, 2017. Liman sektöründe iş güvenliği analizi ve uygulamaları: risk analizinde PRAT tekniği, bulanık mantık ve AHP yaklaşımı (Master's thesis, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü).

The Role of Clinical Significance, Statistical Significance, and Effect Size in Estimating the Minimum Required Sample Size

E. Arzu Kanık[1], **Elif Ertaş**[1]

[1] MedicReS Turkey Mersin University Techoscope, 33190, Mersin, Turkey

*Corresponding author e-mail: arzukanik@gmail.com

Abstract

*Türkiye In medical and health research, researchers have difficulty estimating the required sample size for their studies and finding the minimum values that can be clinically meaningful. Determining which value is clinically significant for which statistic is one of the most difficult issues for young researchers. While experienced people can estimate these values more easily, researchers who have just started in this field do not have intellectual knowledge about which relative risk value, which correlation value, or how much difference between means can be considered both statistically and clinically significant. For this reason, we think that this study is important to support researchers who want to plan their research themselves. In this paper, Cohen effect sizes were used for summary statistics with confidence intervals in the z, t, chi-square, and F test criteria, which are frequently used by researchers. The recommended small effect for Pearson and Spearman correlation is 0.3, the medium effect is 0.5, and the large effect is 0.7. For Differences between two independent proportions the Smallest Cohen's h is 0.2, Medium Cohens'h is 0.5, and large Cohen's h is 0.8. Limits of Cohen's F for ANOVA, ANCOVA, and Univariate regression analysis are small 0.10, medium, 0.25, large 0.40 For the differences between two independent means, the student's test, Suggested Cohen's d limits small is 0.2, medium 0.5, and large effect is 0.8. Cohen's ω is used for all chi-*squared tests, *according to the contingency coefficient, the small effect limit is 0.1, the medium is 0.3 and the large effect limit is 0.5 The last suggestions for the researchers are about risk and odds ratios. For the odds ratio, the suggested effects are small is 1.2, the medium is 2.5, and the large is 4. For the risk ratio, the suggested effects are small is 1.2, the medium is 1.4, and the large is 1.6. The effect sizes suggested in this study are used in all modules by E-PICOS, a software that calculates clinical significance and Cohen's effect sizes, at www.e-picos.com. Beginners and researchers who do not want to make mistakes can find this information in the E-PICOS software.*

Key words: *Artificial intelligence, biostatistics, medical research, clinical research, animal research, Clinical Significancy, Cohen's Effect Size, E-PICOS, MedicReS*

Millî Eğitim Bakanlığında Çalışan Rehber Öğretmenlerin Çocukları Tanımada Kullandıkları Tekniklerin Değerlendirilmesi

Rıdvan Karabulut[1*], Esra Karabulut[2]

[1] Kayseri University, Health Faculuty Department of Child Development, 38028, Kayseri, Turkiye
[2], Minister of National Education, 38028, Kayseri, Turkiye

*Correspondingauthor e-mail: ridvan_karabulut0@yahoo.com

Özet

Çocukların davranışları içinde bulunduğu çevreye göre değişmektedir bu değişim oldukça hızlı ilerleyen bir süreçtir. Bundan ötürü çocuklara uygun bir eğitim ortamı sunabilmek için onları tanıma teknikleri önem arz etmektedir. Ayrıca çocuklarda var olan problemleri ortaya çıkarmak için tanıma ve değerlendirme tekniklerinin oluşturulması önemli bir basamaktır. Bir çocuğu tanımak için tek bir yol yoktur. Çeşitli tekniklerden yararlanılarak çocukla ilgili daha fazla bilgi elde edilir. Bu tarz araştırmalarda çocuk çevresinin bir parçası olarak düşünülmeli ve birlikte ele alınmalıdır. Kullanılan değerlendirme teknikleri bilimsel standartlara ters düşmemelidir. Değerlendirme öznellikten uzak olmalıdır. Objektif olmalıdır. Ele geçirilen verilerde gizlilik ilkesi esas alınmalıdır. Değerlendirmelerin ve gözlemlerin güncel olmasına dikkat edilmelidir. Çocukla ilgili elde edilen düşüncelerin bir eğitimcinin gözlemlerinden daha fazla kişinin düşüncelerine de dayandırılması gerekmektedir. Anektod, gözlem: çocukları kendi doğal ortamlarında bir hedefe yönelik olarak planlı ve sistemli inceleme yöntemidir. Çocukların günlük etkinlikleri sırasında gözlem yapılır. Gözlemin sonucunu kaydetmek amacıyla bir form hazılanabilirveya hazır bir form kullanılabilir. Bu formlar hazırlanırken çocukların içinde bulundukları gelişim dönemleri göz önünde bulundurulmalıdır. Vaka incelemesi: bu teknik çoğunlukla uyum sorunu yaşayan bireylerde kullanılır. Hedef bireyi tanımak ve uyumsuzluk sebeplerini belirlemektir. Bu teknik uyum sorunlarını ortadan kaldırmak için oldukça etkili bir tekniktir. *Görüşme: bu yöntemin hedefi çocukla ilgili ayrıntılı bilgi toplamaktır görüşme ebeveynler ile yapılır. Görüşme isteğinde görüşmeyi yapacak kişi veya ailede yapabilir. Görüşmenin gerçekleştirileceği ortamsakin ve sessiz olmalıdır. Bu yöntem yardımı ile çocuğun okul dışındaki yaşamı ile ilgili de ayrıntılı bilgiler alınır. Projektif teknikler: bu yöntemde bireye karmaşık, belirsiz uyaranlar verilir. Bu uyaranlarla ilgili düşüncelerini paylaşması istenir. Kendisinde ne gibi hisler uyandırdığı sorulur. Duygu ve düşüncelerini dile getirmesiistenilir. Uygulanan birey düşüncelerinin nasıl yorumlanacağını bilmez. Bu teknikler uygulayan eğitimcinin yorumlarına bağlıdır. Bu sebeple uygulayan bireyin bu konuda yeterli ve yetkin olması önemlidir. Bu testlerden bazıları şunlardır Cümle tamamlama testi: bireye tamamlanmamış cümleler verilerek bu cümleleri aklına geldiği şekilde tamamlaması istenir. Böylelikle o bireyin bilinçli veya bilinçsiz kimi duygularına erişilmiş olunur. Tematik algı testi: bireye beyaz siyahbirkaç resim* gösterilir ve bu *resimleri yorumlaması istenir. Yorumlamasına göre bireyin belli başlı duygularına, değer yargılarına, düşüncelerine, tutumlarına ulaşılır. Mürekkep lekesi (rorschach) testi: bu test erişkinler için kullanılmaktadır. Çocuklara uygun değildir. Değerlendirme ölçeği: bu ölçek birey ile ilgili birden fazla kişinin fikir beyan etmesi esas alınır. Bilgi almak istenilen kişinin etrafında olan kişilere "bu kişiyi nasıl bilirsiniz?" sorusu sorularak kişi hakkında bilgi toplanır. Bu yöntemde Milli Eğitim Bakanlığında görev alan Rehber Öğretmenlerin çocukları tanımada kullandıkları yöntemler arasında değerlendirilmektedir.*

Key words: *Çocuk, Bireyi Tanıma, Psikolojik Danışmanlık, Rehberlik*

E-Picos Helps Handle the Challenge of Teaching Biostatistics

E. Arzu Kanık[1*], **Ö. Burak Akıcıer**[2*]

[1] MedicReS Turkey Mersin University Techoscope, 33190, Mersin, Turkey
[2] South America One World Trade Center 85th Floor Suite 8500 New York United States

*Corresponding author e-mail: arzu.kanik@medicres.org

Abstract

E-PICOS is an ai and web-based, smart, interactive software for making statistical calculations based on MedicReS Good Biostatistical Practice for creating Good Medical Research. This program has 7 modules, Ethics, Calculator, Mydata, Power, Designer Assistant, and Reviewer E-PICOS is an intelligent program that minimizes mistakes in ethical scientific research and biostatistics for medical and health science students, medical residency students, specialist physicians who take responsibility for clinical research, and professional researchers. Thanks to E-PICOS, the number of waste literature that has been mentioned in the literature for many years can be reduced because E-PICOS selects the study design suitable for the purpose of the researcher with E-PICOS Designer, determines the sample size suitable for the outcome variable with E-PICOS Power, the statistical method suitable for its sample and purpose E-PICOS It selects with the assistant, ensures that the data is saved and stored securely on the web with E-PICOS My Data, it offers the most appropriate data analysis to work without errors with a single button. It is the only software that clearly writes the research hypotheses to the user and interprets the results in terms of both clinical and statistical significance. In addition to all basic classical parametric, non-parametric z, t, Chi-Square F tests, and advanced, Survival, Cox Regression, and ROC Curve tests, they can also be used in the E-PICOS Calculator menu with minimal errors and correct interpretations with the assistant. E-PICOS is also helpful in clinical trials. Which Ethics Committee should I contact and Randomization table creation menus are also available. For professionals, E-PICOS also offers a module called E-PICOS Reviewer, an article evaluation wizard, heterogeneity control, and mixed random effect model suggestions in systematic review meta-analysis software. www.e-picos.com

Key words: *Power, Artificial intelligence, biostatistics, medical research, clinical research, animal research, Clinical Significancy, Cohen's Effect Size, E-PICOS, Biostatistics, SoftwareMedicReS*

Comparison of Parameter Estimation Methods for the Kavya-Manoharan Exponential Distribution

Hasan Hüseyin Gül[1*]

[1] Giresun University, Faculty of Arts and Sciences, Statistics, 28100, Giresun, Turkey

*Corresponding author e-mail: hasan.huseyin@giresun.edu.tr

Abstract

Many researchers have studied different methods to get new distributions. The most important point in proposing a new method is to increase the flexibility of the existing distribution. For the past several years, there are several methods to propose new distribution by using baseline distribution. For example; exponentiated transformation (Gupta et al. 1998), quadratic rank transformation map (QRTM) (Shaw and Buckley, 2007) and many more. All of these methods involve adding new parameter(s) to the existing distribution. Additional parameter(s) will provide flexibility but will also cause computational difficulties. In recent years, some transformation techniques without additional parameters have been studied by authors. For example; Dinesh-Umesh-Sanjay (DUS) transformation (Kumar et al. 2015), sine-function transformation (Kumar et al. 2015) and logarithmic transformation (LT) (Maurya et al. 2016). The main advantage of these methods is that they don't contain any additional parameter(s) while generating the new distribution and they are computationally easy in the parameter estimation process. Recently, Kavya and Manoharan (2021) introduced a Kavya-Manoharan (KM) transformation for producing new distribution without any additional parameter(s). Also, they proposed KM exponential (KME) distribution using the exponential distribution as the baseline distribution. In this study, various statistical properties and shape of the KME distribution have been examined. Five estimators of the unknown parameter of the KME distribution are compared in terms of their biases and mean square errors. The estimators considered for this comparison are maximum likelihood (ML) estimators, least square (LS) estimators, weighted least square (WLS), Anderson-Darling (AD) estimators and Cramer-von-Mises (CVM) estimators. A Monte-Carlo simulation study is implemented by using MATLAB with 10.000 repetitions. From the simulation study, it is clear that WLS and AD estimators are the most efficient estimators in terms of bias. According to the mean square error, ML estimator has the best performance among the five estimators.

Key words: *Exponential Distribution, Maximum Likelihood, Least Square, Parameter Estimation, Monte-Carlo Simulation*

Maximum Likelihood Estimation Based on Ranked Set Sampling of the Unit-Weibull Distribution

Hasan Hüseyin Gül[1*]

[1] Giresun University, Faculty of Arts and Sciences, Statistics, 28100, Giresun, Turkey

*Corresponding author e-mail: hasan.huseyin@giresun.edu.tr

Abstract

In many applied scenarios, a continuous distribution with a bounded domain is needed to describe the uncertainty of a bounded phenomenon such as a proportion of a certain characteristic. Moreover, this bounded interval happens to be (0,1). Some of the well-known distributions having supports in (0,1) discussed extensively in the literature are uniform, beta, Kumaraswamy and Topp-Leone. Unit distributions bounded by the range (0,1) are obtained by using a transformation of the type X=e^(-Y), where Y is the baseline distribution. In recent years, numerous unit distributions have been suggested in the literature. For example; unit-Logistic distribution (Tadikamalla and Johnson, 1982), unit-Gamma distribution (Mazucheli et al., 2017), unit-Birnbaum-Saunders distribution (Mazucheli et al., 2018), unit-Gompertz distribution (Mazucheli et al., 2019), unit-Lindley distribution (Mazucheli et al., 2019) and unit-Weibull distribution (Mazucheli et al., 2018). Ranked set sampling (RSS) was first introduced by McIntyre (1952) as an alternative to simple random sampling (SRS). An important advantage of RSS over SRS is that it improves the efficiency of estimators of the population parameters. Dell and Clutter (1972) showed that, regardless of ranking errors, RSS is more efficient than SRS. In this paper, maximum likelihood (ML) estimators of the unknown parameters of the unit-Weibull distribution are provided using SRS and RSS. The behavior of the ML estimates for the model parameters is studied using Monte-Carlo simulation. The biases and mean squared errors of estimators in SRS and RSS with different set and cycle sizes. It is found that the estimators based on RSS are more efficient than SRS in terms of comparison criteria.

Key words: *Ranked Set Sampling, Maximum Likelihood Estimation, Unit-Weibull Distribution, Simple Random Sampling, Monte-Carlo Simulation*

Robust and Efficient Estimation of Parameters in Regression Model: Application on Energy Data

Mehmet Niyazi Çankaya[1*]

[1] Usak University, Faculty of Applied Sciences, International Trading and Finance, 64200, Uşak, Turkey

*Corresponding author e-mail: mehmet.cankaya@usak.edu.tr

Abstract

Production and consumption of energies still show theirself as an important issue. There are different types of energy production and consumption. In this study, energy production and consumption data are modeled by years using a robust and effective modeling technique. It can be seen that energy consumption and production can be in balance for the years 2010-2021. In addition, exponential power distribution was chosen among the models used. The estimation process of the parameters of the regression model was done in the R environment. By giving the prediction performance, information criteria, it was tried to obtain the best estimates not only on the best regression model, but also on the distribution of the dependent variable in the regression model. Forward-looking projections are also provided.

Key words: *Energy Consumption, Exponential Power Distribution, Regression, Robust Estimation.*

Yazılım Maliyet Tahmininde Sık Kullanılan Veri Setlerinin İstatistiksel Olarak İncelenmesi ve Yapay Zekâ Yöntemleri ile Yapılan Yazılım Maliyet Tahmininde Kullanılan Değerlendirme Ölçütleri

Şükran Ebren Kara

Şırnak Üniversitesi, Cizre Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, 73200, Şırnak, Türkiye

Sorumlu Yazar e-mail: sukranebren@hotmail.com

Özet

*Yazılım projelerinin maliyet tahmini, yazılım projelerinin gerçekleştirilmesi esnasında gerekli olan kaynak değerlerinin parasal olarak tahmin edilmesidir. Yazılım maliyet tahmini, yazılımı talepeden yazılım projesini yöneten ve yazılımı gerçekleştiren her kesim için çok önemlidir. Yanlış yapılan yazılım maliyet tahminleri ek bütçelere, zaman aşımına hatta projelerin başarısızlıkla sonuçlanmasına neden olmaktadır. Yazılım maliyetinin doğru tahmin etmek için geçmiş proje verilerine ihtiyaç duyulmaktadır. Çok miktarda gerçek proje verileri üzerinde yapılan tahmin yöntemleri ile elde edilen değerlendirmeler, proje yöneticisine doğru bir kestirimde bulunmasına yardımcı olmaktadır. Bu çalışmada literatürde yazılım maliyet tahmini için en sık kullanılan veri setlerinin karşılaştırmalı bir analizi verilmiştir. Maliyet tahmininin değerlendirilmesi için kullanılan değerlendirme ölçütleri; korelasyon katsayısı, MAE (*Mean Absolute Error – Ortalama Mutlak Hata*), RAE (*Relative Absolute Error – Bağıl Mutlak Hata*), RMSE (Root Mean Squared Error – Kök Ortalama Kare Hata), RRSE (Root Relative Squared Error – Kök Bağıl Kare Hata), MAPE (Mean Absolute Percentage Error – Ortalama Mutlak Hata Yüzdesi) detaylı olarak ele alınmıştır.*

Anahtar Kelimeler: *Yazılım Maliyet Tahmini, COCOMO, Maliyet Faktörleri, Hazır Veri Seti*

GİRİŞ

Yazılım maliyeti, yazılım projesinin gerçekleştirilmesi süresince gerekli olan kaynakların parasal değeridir. Maliyet faktörleri olarak ta bilinen bu kaynaklar; kullanılan donanım araçları, çalışan personel harcamaları, bina giderleri gibi birçok faktör olabilmektedir.
Yazılım maliyet tahmini, yazılım projesinin gerçekleşmesi sırasında gerekli olan bütçenin ön görülmesidir. Yazılım projesini gerçekleştirenlerin ve yazılım projesini yönetenlerin ortak ve en temel amacı belirlenen zamanda ve ön görülen bütçe ile projeyi tamamlamaktır. Tahmin doğruluğunu belirlemek için proje tamamlandığında, tahmin edilen yazılım maliyet tahmini ile gerçek yazılım maliyet tahmini karşılaştırılmaktadır.
Gerçek maliyete en yakın maliyet tahminini gerçekleştirmek için değişik yazılım maliyet tahmin yöntemleri geliştirilmiştir. Bu tahmin yöntemlerinin bazıları Yapay Zekâ temelli tahmin yöntemleridir. Bu tahmin yöntemlerinde çok miktarda veri seti ve çeşitli algoritmalar kullanılmaktadır. Kullanılacak veri setinin güvenirliği, kalitesi, çıkış kaynağı gibi özellikleri Yapay Zekânın yapacağı tahmin doğruluğunu doğrudan etkilemektedir. Bu çalışmada yazılım maliyet tahmininde kullanılan geçmiş proje verileri detaylı bir şekilde incelenmiştir.
Yazılım maliyetini gerçeğee en yakın tahmin edebilmek için geçmiş proje verisine ihtiyaç duyulmaktadır. Geçmiş proje verilerinin tutulduğu ücretli ya da ücretsiz veri depları mevcuttur. Bunlardan bazıları; PROMISE (Predictor Models in Software Engineering – Yazılım Mühendisliğinde Tahmin Modelleri), ISBSG (International Software Benchmarking Standards Group – Uluslararası Yazılım Kıyaslama

Standartları Grubu), SDR (SoftLab Data Repository), UCI (University of California, Irvine – Kaliforniya Univsersitesi, Irvine) veri depolarıdır. Bu veri depolarında, geçmiş proje verilerinin tutulduğu veri setleri mevcuttur.
Bosu ve MacDonell (2019) çalışmasında, yazılım maliyet tahmini üzerinde yapılan araştırmalarda kullanılan 13 veri setinin kalitesi değerlendirilmiştir. Bu sayede yaygın olarak kullanılan veri setlerinin amaca uygun olup olmadıkları incelenmiştir.
Bu çalışmada literatürde yazılım maliyet tahmini için en sık kullanılan veri setleri incelenmiştir. Bu veri setlerinde bulunan maliyet faktörlerinin istatisksel değerleri verilmiştir. Yazılım maliyet tahmini için kullanılan değerlendirme ölçütleri ele alınmış förmüleri ile birlikte detaylı bir şekilde anlatılmıştır.

MATERYAL VE METOT

Materyal

Bu çalışmada, yazılım maliyet tahmin yöntemleri araştırmalarında literatürde çok sık kullanılan Albrecht, China, COCOMO81, COCOMONASA, COCOMONASA2, Deshernais, Finnish, Kemerer, Kitchenham, Maxwell, Miyazaki94 ve NASA18 veri setleri incelenmiştir.

Tablo 1. İncelenen veri setlerinin bilgileri.

Veri Seti	Kayıt Sayısı	Öznitelik Sayısı	Büyüklük	Maliyet
Albrecht	24	8	Fonksiyon Noktası	Adam - Saat
China	499	19	Fonksiyon Noktası	Adam - Saat
COCOMO81	63	17	LOC	Adam - Ay
COCOMONASA	60	17	LOC	Adam - Ay
COCOMONASA2	93	24	LOC	Adam - Ay
Deshernais	81	12	Fonksiyon Noktası	Adam - Saat
Finnish	38	9	Fonksiyon Noktası	Adam - Saat
Kemerer	15	8	KSLOC	Adam - Ay
Kitchenham	145	10	Fonksiyon Noktası	Adam - Saat
Maxwell	62	27	Fonksiyon Noktası	Adam - Saat
Miyazaki94	48	9	KSLOC	Adam - Ay
NASA18	18	3	KLOC	Adam - Ay

Farklı veri depolarından alınan her biri farklı sayıda proje verisi barındıran bu veri setlerinde bağlı ve bağımsız öznitelikler bulunmaktadır. Eğer bir öznitelik gerçek maliyet değerini veriyorsa bağlı öznitelik; act_effort (Actual effort – yazılım geliştirme çabası), maliyetle alakalı değerleri veriyorsa; bağımsız öznitelik olarak adlandırılmaktadır.

Albrecht: Albrecht veri seti, IBM veri işleme hizmetlerinde gerçekleştirilen projelerden toplanan 24 kayıttan oluşur. Projeler COBOL, PL/I ve DMS programlama dilleri kullanılarak geliştirilmiştir. Projelerin boyutu ve karmaşıklığı, Albrecht tarafından önerilen fonksiyon noktası yaklaşımı kullanılarak ölçülmüştür (Albrecht ve Gaffney, 1983). Albrecht veri kümesinin lnquiry özniteliğinde 5 kayıtta eksik değerler vardır. Bu sayı Inquiry özniteliğin %21'ine denk gelmektedir. Bunların yerine sıfır yazılmıştır. Tablo 2'de Albrecht maliyet faktörleri açıklamaları ile birlikte verilmiştir.

Tablo 2. Albrecht veri seti maliyet faktörleri.

Sno	Öznitelik	Tanımlama	
1	Input	No of inputs	Giriş sayısı
2	Output	No of outputs	Çıkış sayısı
3	Inquiry	No of inqueries	Sorgu sayısı
4	File	No of master files	Ana dosya sayısı
5	FPAdj	Function points adjustment	Fonksiyon noktaları ayarı
6	RawFPcounts	Count of raw function points	Ham fonksiyonlarının sayısı
7	AdjFP	Adjusted function points	Ayarlanmış fonksiyon noktaları
8	Effort	Person hours	Adam saat

China: China veri seti, 2010 yılında PROMISE deposuna eklenmiş diğer veri setlerine göre daha yeni bir veri setidir. China veri seti, Çin'deki çeşitli yazılım şirketleri tarafından farklı iş alanları için geliştirilen 499 projeye ait verileri içermektedir. China veri seti 18'i bağımsız değişken ve 1 tanesi bağımlı (Effort) değişken olmak üzere 19 öznitelikten oluşmaktadır. Veri setindeki eksik değerlerin yerine bazen "0" bazen de "-" yazılmıştır (Bosu ve MacDonell, 2019). Bu çalışmada eksik değerlerin yerine "0" yazılmıştır.

COCOMO81: COCOMO81 veri seti 1981 yılında önerilmiştir. Veri setinde 63 yazılım projesinin kaydı ile 17 öznitelik bulunmaktadır. Bu öznitelikler; projenin kod satır sayısı, projenin gerçek geliştirme maliyeti ve 15 adet maliyet çarpanıdır.

COCOMONASA: COCOMONASA veri setinde 1980'ler ve 1990'larda farklı merkezlerden toplanmış 60 NASA projesine ait kayıt ve 17 öznitelik bulunmaktadır.

COCOMONASA2: COCOMONASA2 veri seti bazı kayıtlarda NASA93 veri seti olarak geçmektedir (Bosu ve MacDonell, 2019), NASA tarafından 1971 ile 1987 yılları arasında üretilen 93 proje verisinin beş farklı geliştirme merkezinden toplanmasıyla oluşturulmuştur. Bu veri seti 24 öznitelikten oluşan 93 NASA projesine ait veri içermektedir. Tablo 2'de COCOMO81, COCOMONASA ve COCOMONASA2 veri setlerinde 17'si (bağlı öznitelik; act_effort, bağımsız öznitelikler; rely, data, cplx, time, stor, virt, turn, acap, aexp, pcap, vexp, lexp, modp, tool, sced, loc) aynı olan 24 maliyet faktörü verilmiştir (Boehm, 2000).

Tablo 3. COCOMONASA2 maliyet faktörleri

Sno	Öznitelik	Tanımlama	
1	recordnumber	Unique id	Benzersiz kimlik numarası
2	projectname	Project name	Proje ismi
3	cat2	Cagetory of application	Uygulama kategorisi
4	forg	Flight or ground system	Uçuş mu yer sistemi mi?
5	center	center	Hangi NASA merkezi
6	year	Year of development	Geliştime yılı
7	mode	Development mode	Geliştirme modu
8	rely	Required software reliability	Gerekli yazılım güvenliği
9	data	Database size	Veritabanı büyüklüğü
10	cplx	Software product complexity	Ürün karmaşıklığı
11	time	Execution time constraint	Çalışma süresi kısıtı
12	stor	Main storage constraint	Temel depolama kısıtı
13	virt	Virtual machine volatility	Sanal makine geçiciliği
14	turn	Computer turn around time	Bilgisayar yanıt süresi
15	acap	Analist capability	Çalışan analistin kapasitesi
16	aexp	Application experience	Proje takımının uygulama tecrübesi
17	pcap	Programmer capability	Programcı kapasitesi
18	vexp	Virtual machine experience	Takımın Sanal makine tecrübesi
19	lexp	Language experience	Takımın programlama dili tecrübesi
20	modp	Use of modern programming practices	Modern programlama uygulamaları
21	tool	Use of software tools	Kullanılan yazılım araçları
22	sced	Development schedule constraint	İş takvimi kısıtı
23	equivphyskloc	Equivalent physical kilo line of code	Eşdeğer fiziksel kilo kod satırı
24	act_effort	Actual effort	Gerçek maliyet

Desharnais: Bu veri seti Kanadalı bir yazılım evinin oluşturduğu yazılım projelerini içermektedir. Sarro (2011) çalışmasında Desharnais veri setindeki Language özniteliğini baz alarak veri setini Desharnais1, Desharnais2 ve Desharnais3 olarak kullanmıştır.
Bu veri setinde 4 kayıtta eksik veri bulunmaktadır. Bu yüzden çoğu araştırmacı 81 kayıttan oluşan veri setinin 77 kaydını kullanmıştır (Bosu ve MacDonell, 2019). Eksik değerlerden 2'si TeamExp (%2 oranında kayıp) özniteliğinde, 3 tanesi ManagerExp (%4 oranında kayıp) özniteliğinde mevcuttur. Veri setine ait istatiksel değerler bulunurken kayıp değerlerin bulunduğu veri satırları silinmemiştir bunun yerine hangi öznitelik ile ilgili kayıp değerler varsa sadece o özniteliklerde kayıp değerler hesaplamalara katılmamıştır. Desharnais veri setine ait maliyet faktörleri Tablo 4'te belirtilmiştir.

Tablo 4. Desharnais veri seti maliyet faktörleri.

Sno	Öznitelik	Tanımlama	
1	Project	Project no	Proje numarası
2	TeamExp	Team experience	Takım deneyimi
3	ManagerExp	Manager experience	Yönetici deneyimi
4	YearEnd	Year end	Proje bitiş yılı
5	Length	Length	Proje süresi(ay)
6	Effort	Effort (person – hours)	Çaba (adam – saat)
7	Transactions	Transactions	İşlemler
8	Entities	Entities	Varlıklar
9	PointsAdjust	Adjusted function points	Ayarlanmış fonksiyon noktaları
10	Envergure	Envergure	Envergure
11	PointsNonAjust	Unadjusted function points	Ayarlanmamış fonksiyon noktaları
12	Language	Languaage	Programlama dili

Finnish: Finnish veri seti, TIEKE organizasyonu tarafından Finlandiya'daki 9 firmadan toplanan 40 proje verisi ve 9 öznitelikten oluşmaktadır. Projelerin boyutu ve karmaşıklığı fonksiyon noktası yaklaşımı kullanılarak ölçülmüştür. Veri setinde bulunan iki proje verisinde eksik değerler mevcuttur (Kitchenham ve Kansala, 1993). Bu çalışmada eksik değerlere sahip veri satırları çıkarılmış 38 veri satırı üzerinde çalışılmıştır. Tablo 5'te Finnish veri seti maliyet faktörleri verilmiştir.

Tablo 5. Finnish veri seti maliyet faktörleri.

Sno	Öznitelik	Tanımlama	
1	ID	Project no	Proje numarası
2	dev,eff,hrs	Development effort hours	Geliştirme çabası saati
3	hw	Hardware type	Donanım tipi
4	at	Application type	Uygulama tipi
5	FP	Function point data	Fonksiyon noktası verileri
6	co	Application area	Uygulama alanı
7	prod	Project duration (calender months)	Proje süresi (takvim ayları)
8	lnsize	System requirements size in raw Albrecht function points	Ham Albrecht fonksiyon noktalarında sistem gereksinimleri boyutu
9	lneff	Effort provided by application user	Uygulama kullanıcısı tarafından sağlanan çaba

Kemerer: Kemerer veri seti (Kemerer, 1987), veri işleme yazılımı geliştiren bir Amerikan firmasından toplanmıştır. Bu veri seti 8 özniteliğe sahip 12'si COBAL dilinde yazılımış 15 yazılım projesinin verisinden oluşmaktadır (Sarro, 2011). Veri setindeki en eski proje 1981'de başlamış olup projelerin çoğu 1983'te başlamıştır. Veri setindeki proje verileri 1985'te toplanmıştır. Tablo 6'da Kemerer veri seti maliyet faktörleri verilmiştir.

Tablo 6. Kemerer veri seti maliyet faktörleri.

Sno	Öznitelik	Tanımlama	
1	ID	Project ID	Proje kimliği
2	Language	Software used	Kullanılan yazılım
3	Hardware	Hardware used	Kullanılan donanım
4	Duration	Duration	Süre
5	KSLOC	Number of source lines code in thousands	Bin olarak kaynak kod satır sayısı
6	AdjFP	Adjusted function points	Ayarlanmış fonksiyon noktaları
7	RAWFP	Raw function points	Ham fonksiyon noktaları
8	EffortMM	Effort Man Months	Adam ay çaba

Kitchenham: Kitchenham veri seti merkezi Amerika'da olan CSC (Computer Sciences Corporation – Bilgisayar Bilimleri Şirketi)'den toplanmıştır. Bu veri seti 10 tane öznitelikten oluşan 145 yazılım projesinin verisini içermektedir. Yazılım projelerinin büyüklüğü fonksiyon noktaları ile ölçülmüştür. Bu projeler 1994 ve 1999 yılları arasında hazırlanmış projelerdir (Kitchenham ve diğ., 2002).
Bu veri setinde Project.type özniteliğinde 10 tane (%7 oranında) kayıp ve Estimated.completion.date özniteliğinde 3 tane (%2 oranında) kayıp değer vardır. Kayıp değerler yerine "?" kullanılmıştır. Veri setine ait istatiksel değerler bulunurken kayıp değerlerin bulunduğu veri satırları silinmemiştir bunun yerine hangi öznitelik ile ilgili kayıp değerler varsa sadece o özniteliklerde kayıp değerler hesaplamalara katılmamıştır. Kitchenham veri setine ait maliyet faktörleri Tablo 7'de belirtilmiştir.

Tablo 7. Kitchenham veri seti maliyet faktörleri.

Sno	Öznitelik	Tanımlama	
1	Project	Project no	Proje numarası
2	Client.code	Client code	İstemci kodu
3	Project.type	Project type	Proje türü
4	Actual.start.date	Actual start date	Gerçek başlangıç tarihi
5	Actual.duration	Actual duration	Gerçek süre (gün)
6	Actual.effort	Actual effort (pearson – hours)	Gerçek çaba (adam – saat)
7	Adjusted.function.points	Adjusted function points	Ayarlanmış fonksiyon noktaları
8	Estimated.completion.date	Estimated completion date	Tahmini tamamlanma tarihi
9	First.estimate	First estimate	İlk tahmin
10	First.estimate.method	First estimate method	İlk tahmin metodu

Maxwell: Maxwell veri seti bir Fin ticari bankasından toplanmıştır. 27 öznitelik ile temsil edilen 62 proje verisinden oluşmaktadır (Maxwell, 2002; Singh ve Misra, 2012). Projelerin başlangıç yılları 1985 ile 1993 yılları arasındadır. Tablo 8'de Maxwell veri seti maliyet faktörleri verilmiştir.

Tablo 8. Maxwell veri seti maliyet faktörleri.

Sno	Öznitelik	Tanımlama	
1	Syear	Year	Geliştirme yılı
2	App	Application type	Uygulama çeşidi
3	Har	Hardware platform	Donanım platformu
4	Dba	Database	Veritabanı
5	Ifc	User interface	Kullanıcı arayüzü
6	Source	Where developed	Nerede geliştirildi
7	Telonuse	Telon use	Telon kullanımı
8	Nlan	# of development languages	Geliştirme dili
9	T01	Customer participation	Müşteri katılımı
10	T02	Development Env, adequacy	Geliştirme ortamı, yeterlilik
11	T03	Staff availability	Personel durumu
12	T04	Standards use	Standartlar kullanımı
13	T05	Methods use	Yöntem kullanımı
14	T06	Tools use	Araçlar
15	T07	Software logical complexity	Yazılım mantıksal karmaşıklığı
16	T08	Requirements volatility	Gereksinim oynaklığı
17	T09	Quality requirements	Kalite gereksinimleri
18	T10	Efficiency requirements	Verimlilik gereksinimleri
19	T11	Installation requirements	Kurulum gereksinimleri
20	T12	Staff analysis skills	Personel becerisi analizi
21	T13	Staff application knowledge	Personel uygulama becerisi
22	T14	Staff tool skills	Personel araç becerisi
23	T15	Staff team skills	Personel takım becerileri
24	Duration	Duration	Süre
25	Size	Function points	Fonksiyon noktaları
26	Time	Time	Zaman
27	Effort	Work hours Effort	Çalışma saati çabası

Miyazaki94: Miyazaki94 veri seti, Fujitsu Büyük Sistem Kullanıcıları Grubu tarafından toplanmıştır. Veriler, 20 farklı Japon şirketindeki birden fazla bölümünde geliştirilen 48 COBOL projesinden elde edilmiştir. Her proje, 1 bağımlı değişken (MM) ve 8 bağımsız değişken olmak üzere 9 öznitelikten oluşturulmuştur (Miyazaki ve diğ., 1994; Sarro, 2011). Tablo 9'da Miyazaki94 veri setinin öznitelikleri verilmiştir.

Tablo 9. Miyazaki94 veri seti maliyet faktörleri.

Sno	Öznitelik	Tanımlama	
1	ID	Project ID	Proje kimliği
2	KLOC	Number of COBOL source lines in thousands	Binlerce COBOL kaynak satırı sayısı
3	SCRN	Number of different input or output screens	Farklı giriş veya çıkış ekranlarının sayısı
4	FORM	Number of different (report) forms	Farklı (rapor) form sayısı
5	FILE	Number of different record formats	Farklı kayıt biçimlerinin sayısı
6	ESCRN	Total number of data elements in all the screens	Tüm ekranlardaki toplam veri öğesi sayısı
7	EFORM	Total number of data elements in all the forms	Tüm formlardaki toplam veri öğesi sayısı
8	EFILE	Total number of data elements in all the files	Tüm dosyalardaki toplam veri öğesi sayısı
9	MM	Effort measured in man-months	Adam-ay olarak ölçülen çaba

Nasa18: Bu veri seti Bailey ve Basili tarafından efor tahmini için NASA/Goddard Uzay Uçuş Merkezi'ndeki Yazılım Mühendisliği Laboratuvarından toplanmıştır. Yazılım Mühendisliği Laboratuvarı 1976'da kurulmuştur. Veri setinde bulunan yazılımlar çoğunlukla çeşitli uzay aracı projeleri için yer destek rutinleridir. Yazılımlar altı ay ile iki yıl arasındaki sürelerde, iki ile on programcı tarafından geliştirilmişlerdir. Yazılımlarda kullanılan programlama dili Assembly ve FORTRAN'dır (Bailey ve Basili, 1981). Bu veri seti, üç tane öznitelikten oluşan 18 yazılım proje verisi içermektedir. Bu öznitelikler yazılımlar için geliştirilen kod satırları, yazılımın geliştirildiği metodoloji ve yazılım geliştirme çabasıdır (Singh ve Misra, 2012). Tablo 10'da Nasa18 veri seti maliyet faktörleri verilmiştir.

Tablo 10. Nasa18 veri seti maliyet faktörleri.

Sno	Öznitelik	Tanımlama	
1	KLOC	Kilo line of code	Bin kod satır sayısı
2	Methodology	Methodology	Yazılım geliştirme metodolojisi
3	Act_effort	Actual Effort	Yazılım efor değeri

Kayıp Değerler

Bu çalışmada Kitchenham ve Desharnais veri setlerinde eksik değerler yerine "?"; Albrecht ve Çin veri setlerinde eksik değerler yerine "0" kullanılmıştır. Mevcut çalışmada mümkün oldukça veri setindeki değerler üzerinde oynama yapılmamıştır. Kayıp değerler için "?" kullanılmış ise sadece o öznitelik için o değer silinmiştir. Kayıp değerler için veri setinde hali hazırda zaten "0" kullanılmışsa o şekilde bırakılmıştır.

Metot

Yazılım maliyet tahmini için kullanılan veri setlerinin İstatistiksel değerlerinin elde edilmesi için Excell ve WEKA programları kullanılmıştır.

Yazılım Maliyet Tahmininde Kullanılan Değerlendirme Ölçütleri

Korelasyon Katsayısı: İki farklı değişken arasındaki ilişkinin gücünü ve yönünü belirtir. Korelasyon katsayısı ilişkinin durumuna göre -1 ile 1 arasında bir değer almaktadır. Katsayı değerinin negatif bir değer olması değişkenler arasında ters bir ilişkinin olduğunu göstermektedir. Katsayı 1'e yaklaştıkça ilişkinin değeri artmaktadır. Katsayı 0'a yaklaştıkça ilişkinin değeri azalmaktadır.

RMSE (Root Mean Squared Error – Kök Ortalama Kare Hatası):

$$RMSE = \sqrt{\frac{1}{\mathrm{n}}\sum_{i=1}^{n}(T_i - G_i)^2} \tag{1}$$

Burada T_i = tahmini değer, G_i= gerçek değer, n = örnek sayısıdır.

MAE (Mean Absulate Error – Ortalama Mutlak Hata):

$$MAE = \frac{1}{\mathrm{n}}\sum_{i=1}^{n}|T_i - G_i| \tag{2}$$

Burada T_i = tahmini değer, G_i = gerçek değer, n = örnek sayısı'dır

RAE (Relative Absulate Error – Bağıl Mutlak Hata)

$$RAE = \frac{\sum_{i=1}^{n}|T_i - \mathrm{G}_i|}{\sum_{i=1}^{n}|\mathrm{G}_m - \mathrm{G}_i|} \tag{3}$$

Burada T_i = tahmini değer, G_i = gerçek değer, G_m= gerçek değerlerin ortalaması, n = örnek sayısıdır.

RRSE (Root Relative Squared Error – Kök Ortalama Kare Hata)

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(T_i - \mathrm{G}_i)^2}{\sum_{i=1}^{n}(\mathrm{G}_m - \mathrm{G}_i)^2}} \tag{4}$$

Burada T_i = tahmini değer, G_i = gerçek değer, G_m= gerçek değerlerin ortalaması, n = örnek sayısıdır.
MMRE (Mean Magnitude Of Relative Error – Göreceli Hatanın Ortalama Büyüklüğü)

$$MMRE = \frac{1}{\mathrm{n}}\sum_{i=1}^{n}\left|\frac{T_i - G_i}{G_i}\right| \tag{5}$$

Burada T_i = tahmini değer, G_i= gerçek değer, G_m= gerçek değerlerin ortalaması, n = örnek sayısıdır.
MAPE (Mean Absolute Percentage Error – Ortalama Mutlak Hata Yüzdesi)
Bazı çalışmalarda MAPE, MMRE olarak kabul edilmektedir (Tofallis, 2015). MAPE çok sık kullanılan tahmin değerlendirme ölçütüdür. MAPE çoğu çalışmada %MMRE olarak da kullanılmıştır. Formülü Denklem 6'da verilmiştir.

$$MAPE = \frac{1}{\mathrm{n}}\sum_{i=1}^{n}\left|\frac{T_i - G_i}{G_i}\right| \mathrm{x}100 \tag{6}$$

Burada T_i = tahmini değer, G_i= gerçek değer, G_m= gerçek değerlerin ortalaması, n = örnek sayısıdır.

BULGULAR

Veri Setlerinde Bulunan Kayıp Değerler

Veri depolarından temin edilen hazır veri setlerinde bilinmeyen bazı nedenlerden dolayı kayıp veriler bulunmaktadır.

Tablo 11'de veri setlerinde kayıp değerlerin olup olmadığı varsa hangi özniteliklerde olduğu, veri setini hangi oranda etkilediği ve bu kayıp değerler yerine veri setinde hangi değerlerin yazıldığı bilgisi verilmiştir.

Tablo 11. Veri setlerinde kayıp değerler

Veri Seti	Kayıp	Öznitelik	Kayıp sayısı	Oran	Yerine yazılan
Albrecht	Var	Inquiry	5	%2	0
China	Var	-	-	-	0
COCOMO81	Yok				
COCOMONASA	Yok				
COCOMONASA2	Yok				
Deshernais	Var	TeamExp	2	%2	?
		ManagerExp	3	%4	?
Finnish	Yok				
Kemerer	Yok				
Kitchenham	Var	Project.type	10	%7	?
		Estimated.completion.date	3	%2	?
Maxwell	Yok				
Miyazaki94	Yok				
NASA18	Yok				

Veri Setlerinin İstatistiksel Değerleri
Çalışmanın bu kısmında ilgili veri setlerinin istatistiksel değerleri verilmiştir. Bu değerler özniteliğin (maliyet faktörünün) en küçük ve en büyük değerleri ile özniteliğin ortalaması ve standart sapması hesaplanarak Tablo 12, Tablo 13, Tablo 14, Tablo 15, Tablo 16, Tablo 17, Tablo 18, Tablo 19, Tablo 20 ve Tablo 21'de verilmiştir.

Tablo 12. Albrecht veri seti istatistikleri.

Öznitelik	EnKüçük	EnBüyük	Ortalama	Standart Sapma
Input	7	193	40,25	36,914
Output	12	150	47,25	35,169
Inquiry	0	75	16,875	19,338
File	3	60	17,375	15,522
FPAdj	0,75	1,2	0,99	0,135
RawFPcounts	189,52	1902	638,54	452,654
AdjFP	199	1902	647,625	487,995
Effort	0,5	105,2	21,875	28,418

Tablo 13. China veri seti istatistikleri.

Sno	Öznitelik	EnKüçük	EnBüyük	Ortalama	Standart Sapma
1	ID	1	499	250	144,193
2	AFP	9	17518	487	1059,171
3	Input	0	9404	167	486,339
4	Output	0	2455	113,601	221,274
5	Enquiry	0	952	61,601	105,423
6	File	0	2955	91,234	210,271
7	Interface	0	1572	24,234	85,041
8	Added	0	13580	360,355	829,842
9	Changed	0	5193	85,062	290,357
10	Deleted	0	2657	12,353	124,224
11	PDR_AFP	0.3	83.8	11,771	12,106
12	PDR_UFP	0.3	96.6	12,08	12,819
13	NPDR_AFP	0.4	101	13,27	14,01
14	NPDU_UFP	0.4	108,3	13,626	14,843
15	Resource	1	4	1,459	0,824
16	Dev.Type	0	0	0	0
17	Duration	1	84	8,719	7,347
18	N_effort	31	54620	4277,641	7071,248
19	Effort	26	54620	3921,048	6480,856

Tablo 14. COCOMO maliyet faktörlerinin standart istatistiksel değerleri.

Öznitelik	Çok Düşük	Düşük	Normal	Yüksek	Çok Yüksek	Ekstra Yüksek	Ortalama	Standart Sapma
acap	1,46	1,19	1,00	0,86	0,71		0,905	0,152
pcap	1,42	1,17	1,00	0,86	0,70		0,937	0,167
aexp	1,29	1,13	1,00	0,91	0,82		0,949	0,119
modp	1,24	1,10	1,00	0,91	0,82		1,004	0,131
tool	1,24	1,10	1,00	0,91	0,83		1,017	0,086
vexp	1,21	1,10	1,00	0,90			1,005	0,093
lexp	1,14	1,07	1,00	0,95			1,001	0,052
sced	1,23	1,08	1,00	1,04	1,10		1,049	0,076
stor			1,00	1,06	1,21	1,56	1,144	0,179
data		0,94	1,00	1,08	1,16		1,004	0,073
time			1,00	1,11	1,30	1,66	1,114	0,162
turn		0,87	1,00	1,07	1,15		0,972	0,081
virt		0,87	1,00	1,15	1,30		1,008	0,121
cplx	0,70	0,85	1,00	1,15	1,30	1,65	1,091	0,203
rely	0,75	0,88	1,00	1,15	1,40		1,036	0,193

Tablo 15. Desharnais veri seti istatistikleri.

Öznitelik	EnKüçük	EnBüyük	Ortalama	Standart Sapma
Project	1	81	41	23,527
TeamExp	0	4	2,266	1,337
ManagerExp	0	7	2,667	1,518
YearEnd	83	88	85,79	1,148
Length	1	39	11,716	7,4
Effort	546	23940	5046,309	4418,767
Transactions	9	886	179,901	143,315
Entities	7	387	122,333	84,882
PointsAdjust	73	1127	302,235	179,677
Envergure	5	52	27,63	10,592
PointsNonAjust	62	1116	287,049	185,108
Language	-	-	-	-

Tablo 16. Finnish veri seti istatistikleri.

Öznitelik	EnKüçük	EnBüyük	Ortalama	Standart Sapma
ID	1	38	19,5	11,113
dev,eff,hrs	460	26670	7678,289	7135,28
hw	1	3	1,263	0,644
at	1	5	2,237	1,497
FP	65	1814	763,579	510,834
co	2	10	6,263	2,728
prod	1,473	29,473	10,074	7,086
lnsize	4,17	7,5	6,357	0,835
lneff	6,131	10,191	8,397	1,179

Tablo 17. Kemerer veri seti istatistikleri.

Öznitelik	EnKüçük	EnBüyük	Ortalama	Standart Sapma
ID	1	15	8	4,472
Language	1	3	1,2	0,561
Hardware	1	6	2,333	1,676
Duration	5	31	14,267	7,545
KSLOC	39	450	186,573	136,817
AdjFP	99,9	2306,8	999,14	589,592
RAWFP	97	2284	993,867	597,426
EffortMM	23,2	1107,31	219,25	263,055

Tablo 18. Kitchenham veri seti istatistikleri.

Öznitelik	EnKüçük	EnBüyük	Ortalama	Standart Sapma
Project	1	145	73	42,002
Actual.duration	37	946	206,448	134,092
Actual.effort	219	113930	3113,117	9598,009
Adjusted.function.points	15,36	18137,48	527,669	1521,995
First.estimate	121	79870	2855,972	6789,287
First.estimate.method				

Tablo 19. Maxwell veri seti istatistikleri.

Öznitelik	EnKüçük	EnBüyük	Ortalama	Standart Sapma
Syear	85	93	89,58	2,131
App	1	5	2,355	0,993
Har	1	5	2,613	0,998
Dba	0	4	1,032	0,442
Ifc	1	2	1,935	0,248
Source	1	2	1,871	0,338
Telonuse	0	1	0,242	0,432
Nlan	1	4	2,548	1,019
T01	1	5	3,048	0,999
T02	1	5	3,048	0,711
T03	2	5	3,032	0,886
T04	2	5	3,194	0,698
T05	1	5	3,048	0,711
T06	1	4	2,903	0,694
T07	1	5	3,242	0,9
T08	2	5	3,806	0,955
T09	2	5	4,065	0,744
T10	2	5	3,613	0,894
T11	2	5	3,419	0,984
T12	2	5	3,823	0,69
T13	1	5	3,065	0,956
T14	1	5	3,258	1,007
T15	1	5	3,339	0,745
Duration	4	54	17,21	10,651
Size	48	3643	673,306	784,085
Time	1	9	5,58	2,131
Effort	583	63694	8223,21	10499,903

Tablo 20. Miyazaki94 veri seti istatistikleri.

Öznitelik	EnKüçük	EnBüyük	Ortalama	Standart Sapma
ID	-	-	-	-
KLOC	6,9	417,6	70,792	87,568
SCRN	0	281	33,688	47,236
FORM	0	91	22,375	20,548
FILE	2	370	34,813	53,365
ESCRN	0	3000	525,604	626,058
EFORM	0	1566	460,667	396,816
EFILE	57	45000	1854,583	6398,605
MM	5,6	1586	87,475	228,76

Tablo 21. Nasa18 veri seti istatistikleri.

Öznitelik	EnKüçük	EnBüyük	Ortalama	Standart Sapma
KLOC	2,1	100,8	33,589	32,595
Methodology	19	35	27,778	5,386
Act_effort	5	138,3	49,472	45,726

TARTIŞMA VE SONUÇ

Yazılım projelerinin maliyetinin doğru tahmin edilmesi konusunda çok fazla araştırma yürütülmektedir. Geçmişten günümüze yapılan araştırmalar incelendiğinde çoğu çalışmada ortak kullanılan geçmiş veri setlerine rastlanmaktadır. Yeni geliştirilen bir tahmin yönteminin test edilmesi ve tahmin doğruluğunun incelenmesi için önceki yazılım projelerinin verilerine ihtiyaç duyulmaktadır. Bu kapsamda geçmiş projelerin tutulduğu veri depoları mevcuttur. Araştırmacılar geliştirdikleri tahmin yöntemini test etmek için bu veri setlerini kullanmaktadır.

Bu çalışmada yazılım maliyet tahmini için kullanılan geçmiş proje veri setleri incelenmiştir. Veri setlerinin toplandığı şirketler, içerdikleri proje sayıları, proje tarihleri gibi birçok bilgiye ulaşılmıştır. Bunun yanında yazılım projelerinin maliyetini etkileyen faktörlerin neler olabileceği bunların en küçük, en büyük değerleri, ortalamaları ve standart sapmaları hesaplanmıştır.

Bu çalışma, yazılım maliyeti tahmin yöntemi üzerinde çalışacak araştırmacılara kullanacakları hazır veri setleri hakkında geniş bir bilgi sunmuştur. Kullanılacak veri seti hakkında bilgi sahibi olmak çalışmanın kalitesini artıracaktır.

Bu çalışmada sunulan veri setleri dışında başka hazır veri setleri de mevcuttur. SDR veri seti bunlardan bir tanesidir. SDR, Boğaziçi Üniversitesi Yazılım Mühendisliği Araştırma Laboratuvarı Softlab tarafından toplanmıştır. SDR veri seti diğer veri setlerine göre daha yenidir ve Türkiye'deki yazılım firmalarının verilerini içermektedir. Gelecek başka bir çalışmada SDR veri seti incelenebilir.

Kaynaklar

Albrecht AJ, Gaffney JE, 1983. Software function, source lines of code, and development effort prediction: a software science validation. IEEE Transaction on Software Engeneering 9(6): 639-648.

Bailey JW, Basili VR, 1981. A meta-model for software development resource expenditures. In Proceedings of the 5th international conference on Software engineering, 9 March, 1981, Page: 107-116.

Baskeles B, Turhan B, Bener A, 2007. Software effort estimation using machine learning methods. In 2007 22nd international symposium on computer and information sciences, November, 2007, Page: 1-6.

Boehm B, Abts CA, Chulani S, Clark BK, Horowitz E, Madachy R, Reifer DJ, Steece B, 2000. Software cost estimation with COCOMO II, Prentice Hall.

Bosu MF, Macdonell SG, 2019. Experience: Quality benchmarking of datasets used in software effort estimation. Journal of Data and Information Quality (JDIQ), 11(4), 1-38.

Kemerer C F, 1987. An empirical validation of software cost estimation models. Communications of the ACM, 30(5), 416-429.

Kitchenham B, Kansala K, 1993. Inter-item correlations among function points. In [1993] Proceedings First International Software Metrics Symposium, May, 1993, Page: 11-14.

Kitchenham B, Pfleeger S L, McColl B & Eagan S, 2002. An empirical study of maintenance and development estimation accuracy. Journal of systems and software, 64(1): 57-77.

Maxwell KD, 2002. Applied statistics for software managers. Applied Statistics for Software Managers.

Miyazaki Y, Terakado M, Ozaki K, Nozaki H, 1994. Robust regression for developing software estimation models. Journal of Systems and Software 27(1): 13-16.

Sarro F, 2011. Search-based approaches for software development effort estimation. In Proceedings of the 12th International Conference on Product Focused Software Development and Process Improvement, June, 2011, Page: 38-43.

Singh BK, Misra AK, 2012. Software effort estimation by genetic algorithm tuned parameters of modified constructive cost model for nasa software projects. International Journal of Computer Applications, 59(9): 22-26.

Tofallis C, 2015. A better measure of relative prediction accuracy for model selection and model estimation, Journal of the Operational Research Society, 66: 1352–1362.

(k,m)-type Slant Helices for Symplectic Curve in 4-Dimensional Symplectic Space

<u>Esra Çiçek Çetin</u>[1*], Mehmet Bektaş[2*]

[1] Firat University, Department of Mathematics, 23119 Elazığ/Türkiye
[2] University, Faculty or Vocational school name, Department, Postal Code, City, Country
[3] University, Faculty or Vocational school name, Department, Postal Code, City, Country

*Corresponding author e-mail: esracicek23@gmail.com

Abstract

In the present work, we introduce new slant helices called (k,m)-type. Also, we define the notion of (k,m)-type slant helix in 4- symplectic space and we express brief related to symplectic space. Also, we define some equations for (k,m)- type slant helix of symplectic regular curves. We give some definition about symplectic space such as symplectic arc lenght, non-degenerate condition. On the other hand, we obtain some characterizations for (k,m)- type slant helix of symplectic regular curves by. using Frenet equations for different type slant helices. Also, we find that there do not exist (1,3) and (2,4) -type slant helices by using definition and Frenet equations in $\mathbb{R}^4$. Later, we obtain some theorem and their proof about these conclusions.

Key words: *Symplectic Curve, Helix , (K,M)- Type Slant Helix .*

Determination of Impact Size by Canonic Correlation Analysis of the Factors Affecting the Buying or Selling Agricultural Lands

Mustafa Şahin[1], İsmail Gök[2*], Tolga Tolun[2]

[1] Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Department of Agricultural Biotechnology, 46100, Kahramanmaraş, Turkey.
[2] Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Department of Bioengineering, 46100, Kahramanmaraş, Turkey.

*Corresponding author e-mail: gkisoo1995@gmail.com

Abstract

The aim of this study is to determine the relationship between the influencing factors in the data set created with the help of canonical correlation in the factors affecting the purchase or sale of agricultural lands in certain neighborhoods in the Dulkadiroğlu district of Kahramanmaraş province and the most effective factor or factors compared to the others. For this, a survey was conducted with the parcel owners in the determined neighborhoods and a data set of 1000 people was created. As a result of the data obtained, the first data set of the study, in which the canonical correlation method was used, was the slope and irrigation status of the land, the average productivity rate of the land, and the presence of tractors and equipment used in the land, among the factors affecting the purchase or sale of agricultural lands; land parcel size, land purchase or sale price and for what purpose the land was bought or sold were also the second data set. The rate of canonical correlation coefficient in the data sets created according to the estimation results was found to be 40.32%. In the data set we compared, the average productivity of the land was determined in the first data set with the highest effect, while in the second data set it was determined as the purpose for which the land was bought or sold.

Key words: *Agricultural Land, Factor, Survey and Land Trading.*

INTRODUCTION

Land within the factors of production; It is the part of the earth that affects the climate, soil, topography and living things in various sizes. The land, which is only a means of production without the possibility of increasing, as well as the plant production area; It is a source of livelihood for people in various fields such as beekeeping, greenhouse cultivation, sheep and cattle breeding. It is also known as the security guarantee and investment area of most people (Başer, 2015). Due to the infrastructure and superstructure investments implemented by private and public areas in Turkey, the expropriation work is in the first place, and various areas such as consolidation works, the irrigation planning project of the land and making the land open to the zoning area are closely related to the determination of the current price of the land. However, the fact that studies such as taxation and guarantees in banks are related to the current market prices of the land increases the importance of the research applied to determine the land precedent value. Valuation; It is the real and complete presentation of the prices of business area parcels, agricultural enterprises and other rural goods and rights.(Guide, 2012).

In this study, considering the geographical shape of the lands in the determined neighborhoods, it can be possible to scientifically analyze which degree of influence is dominant when buying or selling land in the survey conducted for our farmers. For this reason, the factors that have had an impact on the sale or purchase of land and the effect sizes of these factors have been classified with each other, and it is necessary to determine the degree of impact of these factors in the sale or purchase of land. It is thought that what are the factors affecting the sale or purchase of land for various targets and it will be beneficial to reduce these factors.

The study is expected to be a guide in determining the current value in the sale or purchase of the land of the landowners in the selected areas and determining the extent of the effect of the factors affecting this activity. As a result of the results obtained in this study, the real and scientific appraisal of the valuation taken by the landowners in the land acquisition or sale of the land in the region, taking into account the factors such as the structure of the land, its roughness, wet-dry state, its location, productivity rate, and the distance to the nearest settlement area. will determine whether it is based on fundamentals. It is foreseen that it will also contribute to the investment planning in the region where the research is carried out.

The aim of the research is to try to determine the effect sizes of the factors affecting the sale or purchase of agricultural lands in the determined neighborhoods in the Dulkadiroğlu district of Kahramanmaraş province. For this, the effect sizes of the factors were classified from most to least with the help of the canonical correlation method using the data obtained through the survey study.

MATERIAL AND METHODS

Material

This research refers to the 27-year period when the land was sold or bought with the land owners in 9 determined neighborhoods in the Dulkadiroğlu district of Kahramanmaraş between the years 1995-2021. The most important factor in determining the places where the research is applied is to consider that the general data on the purchase or sale of agricultural land in the region will be met.

In this research, a questionnaire was applied with a total of 1000 participants who sold or bought land in 9 neighborhoods, whose sole purpose of participation was determined. In the survey conducted with the land owners in question, it was determined that there is a total of 15913,000 m2 of land that has been subject to purchase and sale. The total land parcel sizes of the neighborhoods are given in Table 1.

Table 1. The total parcel size of the lands in the determined neighborhoods in Dulkadiroğlu district.

Neighbourhood	1995-2021 In Trading Parcel Size
Çınarlı	1395.000 m^2
Çiğli	2838.000 m^2
Sivricehüyük	1100.000 m^2
Kapıçam	1964.000 m^2
Abbaslar	1241.000 m^2
Alibeyuşağı	2680.000 m^2
Kocalar	936.000 m^2
Yeniyurt	2266.000 m^2
Tevekkeli	1493.000 m^2
Total	15913.000 m^2

A survey was conducted randomly and by reaching as many people as possible and it was determined that 91.07% of the people who participated in the survey were male and 8.93% were female. In addition, the average number of children in the family was determined as 4, and their education level was determined to be 97.48% primary school or literate.

Table 2. Wet and dry parcel sizes of the lands in the determined neighborhoods in Dulkadiroğlu district.

Neighbourhood	AQUEOUS Land Size	ANHYDROUS Land Size
Çınarlı	1286.000 m^2	109.000 m^2
Çiğli	2486.000 m^2	352.000 m^2
Sivricehüyük	1100.000 m^2	0 m^2
Kapıçam	1964.000 m^2	0 m^2
Abbaslar	1241.000 m^2	0 m^2
Alibeyuşağı	2680.000 m^2	0 m^2
Kocalar	936.000 m^2	0 m^2
Yeniyurt	1866.000 m^2	400.000 m^2
Tevekkeli	1493.000 m^2	0 m^2
Total	15052.000 m^2	861.000 m^2

As a result of the answers given by the land owners whose lands were the subject of the survey between 1995-2021, it was determined that there were 15052.000 m2 of irrigated land and 861.000 m2 of non-watery land. In addition, while Yeniyurt has the highest amount of waterless land in the determined neighborhoods, Alibeyuşağı District has the highest amount of wet land. The products planted in this region; barley, cotton, beet, wheat, cucumber, corn, watermelon, pepper and chickpea. In addition, per 1000 m2 in neighborhoods: The average productivity of irrigated lands in Çınarlı Mahallesi; beet=7000-12000 kg, corn=1000-1500 kg, cotton=500-700 kg, wheat=500-600 kg, barley=400-500 kg, cucumber=350-400 kg, pepper=400-500 kg and chickpeas =300-400 kg of land productivity without water; wheat=300-400 kg, barley=250-300 kg and chickpeas=200-300 kg; Average productivity of irrigated lands in Çiğli District; beet=6000-12000 kg, corn=1000-1500 kg, cotton=500-700 kg, wheat=500-600 kg, barley=400-500 kg and chickpeas=300-400 kg, while the land productivity without water; wheat=300-400 kg, barley=250-300 kg and chickpeas=200-300 kg; Average productivity of irrigated lands in Sivricehüyük District; beet=4000-7000 kg, corn=800-1000 kg, cotton=300-500 kg, wheat=500-600 kg, barley=400-500 kg, cucumber=300-400 kg and pepper=250-400 kg; Kapicam District average irrigated land productivity; beet=4000-8000 kg, corn=700-1000 kg, cotton=300-500 kg, wheat=400-500 kg barley=300-400 kg and cucumber=300-500 kg; Average productivity of irrigated lands in Abbaslar District; beet=10000-13000 kg, corn=1000-1700 kg, cotton=400-600 kg, wheat=700-800, barley=400-500 kg and chickpeas=300-400 kg; Average productivity of irrigated lands in Alibeyuşağı Mahallesi; beet=6000-12000 kg, corn=1000-1500 kg, cotton=500-700 kg, wheat=500-600 kg, barley=400-500 kg, cucumber=300-400 kg and pepper= 450-500 kg; Average productivity of irrigated lands in Kocalar Mahallesi; beet=6000 12000 kg, corn=1000-1500 kg, cotton=500-900 kg, wheat=600-700 kg, barley=500-600 kg, cucumber=400-700 kg and pepper=300-500 kg; Average productivity of irrigated lands in Yeniyurt Mahallesi; beet=4000-8000 kg, corn=800-1000 kg, cotton=400-600 kg, wheat=400-500 kg, barley=300-400 kg, cucumber=300-500 kg, pepper=250-300 kg and chickpeas = 250-300 kg, while the average productivity of the lands without water; wheat=300-400 kg, barley=200-300 kg and chickpeas=200-300 kg and average productivity of irrigated lands in Tevekkeli District; beet=4000-8000 kg, corn=700-1000 kg, cotton=300-500 kg, wheat=400-500 kg, barley=300-400 kg and cucumber=300-500 kg.

While it is possible to produce all of them in irrigated agricultural lands, wheat, barley and chickpea production is common in non-irrigated agricultural lands. In line with the answers given by the landowners who participated in the survey, there is a production of once a year in the irrigated lands, while this production is carried out twice in the irrigated agricultural lands. Since the structural shape is not uneven in the region where the land is located, it was determined that the soil is stony in lands without water and normal in wet lands. In addition, the distance range to the nearest residential area; Average 19-21 km in Tevekkeli district, 19-22 km in Çınarlı district on average, 16-18 km in Kapıçam district, 20-22 km in Kocalar district on average, 16-18 km in Çiğli district on average, 25-27 km in Abbaslar district on average, Yeniyurt district on average It has been determined as a result of the answers given by the landowners in the survey that it is 19-20 km on average, 20-23 km on average in the Sivricehüyük district and 24-25 km on the Alibeyuşağı district.

Table 3. The presence of tractors and equipment in the designated neighborhoods in Dulkadiroğlu district.

Region	Number of People Who Participated in the Survey	Number of Own Tractors and Equipment	Number of Those Without Tractors and Equipment
Çınarlı	179	113	66
Çiğli	138	90	48
Sivricehüyük	40	40	0
Kapıçam	178	104	74
Abbaslar	71	71	0
Alibeyuşağı	110	100	10
Kocalar	44	34	10
Yeniyurt	120	80	40
Tevekkeli	120	60	60
Total	1000	692	308

As a result of the survey conducted with the landowners in the sale or purchase of land between 1995-2021, it was determined that 69.20% of the land owners included in the survey had tractors and equipment, while 30.80% did not have tractors and equipment.

While the factors affecting the land owners who were included in the survey by buying or selling land within the specified years were investigated with the help of the survey, 19.23% of them became landowners in Çınarlı Mahallesi with the aim of animal husbandry activities and 80.77% of them with the aim of having an investment relationship; 83.02% of them sold their lands due to expropriation of the land, 1.30% of them quitting agriculture, 2.61% of them due to migration and 13.07% of them due to financial insufficiency. In Çiğli Neighborhood, 12% becomes land owner with the aim of relations between heirs, 4% of the transfer of the land from the treasury to private property, 12% of livestock activities and 72% of investment relations; 52.30% of them sold their lands due to expropriation of the land, 6.81% due to relations between heirs, 6.81% due to migration, 4.54% to quit agriculture and 29.54% due to financial insufficiency. In Sivricehüyük Neighborhood, 30% becomes the owner of land with the aim of relations between heirs, 10% of livestock activities and 60% of investment relations; 50% of them sold their lands due to relations between heirs, 10% due to immigration and 40% due to financial insufficiency. In Kapıçam Neighborhood, 48.07% becomes land owner with the aim of transferring the land from the treasury to private property, 13.47% with the aim of relations between heirs, 1.93% with livestock activities and 36.53% with the aim of investment relationship; 8.10% of them sold their lands due to relations between heirs, 21.62% due to expropriation of the land, 2.70% due to migration and 67.56% due to financial insufficiency. In Abbaslar Neighborhood, 11.42% becomes land owner with the aim of relations between heirs, 5.71% of livestock activities and 82.87% of investment relations; 8.33% of them left the agriculture, 11.11% of them sold their lands due to relations between heirs, 16.16% due to changes in land productivity and 63.90% due to financial insufficiency. While 26.67% of them become landowners in Alibeyuşağı Mahallesi with livestock activities and 73.33% with the target of investment relationship; 13.21% of them sold their lands due to quitting agriculture, 13.21% due to immigration and 73.58% due to financial insufficiency. While 36.37% of them become landowners in

Kocalar Mahallesi with the aim of animal husbandry activities and 63.63% of them with the aim of investment relationship; 9.10% of them left the agriculture, 27.27% of them migrated and 63.63% of them sold their lands due to financial insufficiency. In Yeniyurt Mahallesi, 17.85% becomes land owner with the aim of relations between heirs, 17.85% of livestock activities and 64.30% of investment relations; 14.25% of them sold their lands due to relations between heirs, 12.50% due to immigration, 12.53% due to quitting agriculture and 60.72% due to financial insufficiency. In Tevekkeli Neighborhood, 50% will become landowners with the aim of relations between heirs, 5.88% with livestock activities and 44.12% with the aim of investment relations; 23.12% of them sold their lands due to relations between heirs, 15.35% to quit agriculture, 15.38% to immigration and 46.15% to financial insufficiency reasons.

The survey was applied in 9 neighborhoods determined in Dulkadiroğlu District of Kahramanmaraş Province in 2021. Dulkadiroglu District; A survey was conducted with 1000 landowners whose only condition of participation was the purchase or sale of agricultural lands in the neighborhoods of Çınarlı, Çiğli, Sivricehüyük, Kapıçam, Abbaslar, Alibeyuşağı, Kocalar, Yeniyurt and Tevekkeli between the years 1995-2021 and for all parcels where the survey was conducted in the periods covering this research. It is assumed that agricultural technology does not change.

Method

The aim of this research is; The aim of this study is to determine the factors that are effective in the region and the degree of impact of these factors, in line with the information obtained from the land owners who were included in the survey by purchasing or selling land between the years 1995-2021 in 9 selected neighborhoods in the Dulkadiroğlu district of Kahramanmaraş province. In this, the canonical correlation method, which is a statistical analysis method, was used.

Canonical correlation analysis

Canonical correlation analysis, which is the most basic and most complex interaction analysis, deals with the connections between two-way datasets taken from a multi-dimensional population. A method related to the general problems of the connection between the two-way dataset taken from it is envisaged. In this method, which is called canonical correlation analysis, the highest level correlation of the variables in each class and combination pairs with unit variances are found at the beginning, and the process continues by finding the second linear combination pair. Canonical correlation analysis is a special part of multiple regression analysis (Çemrek, 2012). While multiple regression analysis examines the relationship between one dependent and more than one independent variable, canonical correlation analysis examines the relationship between p dependent and q independent variables (Özçomak, Demirci, 2010). In canonical correlation analysis, the goal is not to develop or derive a model for the variables, but to analyze the connection between the variable sets (Sümbüloğlu, Akdağ, 2009).

Canonical correlation analysis can be shown in general as follows (Albayrak, 2006):

$$Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + \cdots + Y_p = X_1 + X_2 + X_3 + X_4 + X_5 + \cdots + X_k \quad (1)$$

In the above equation, the variables can take measured or unmeasured values. in the first group (Y_p) p(1-p)/2 among the variables, whereas in the second group (X_k) there are q(1-q)/2 correlations between the variables and p.q correlations between the two variable groups. When the number is large, it is very difficult to analyze the correlation coefficients one by one. For this, canonical correlation analysis aims to minimize these correlation coefficients (Özçomak, Gündüz, Demirci, Yakut, 2012). If there is only one dependent variable in canonical correlation analysis, multiple regression analysis is used instead of canonical correlation analysis. Since ANOVA and two-class dicirminant analysis are special cases of multiple regression analysis, these two methods are also special cases of canonical correlation analysis. If the dependent variable is a versatile nominal variable, canonical correlation analysis refers to multiple discriminant analysis (Albayrak, 2006).

In canonical correlation analysis, one of the variable groups can be called the independent variable group and the other the dependent variable group. If it is a group of independent and dependent variables, the goal in canonical correlation is to analyze whether and to what extent the independent variable group affects the dependent variable group (Dillon and Goldstein, 1984). However, there is no obligation to define this in variable groups. In this analysis, it is aimed to maximize the correlation between variable groups, and towards this goal, canonical variable dichotomies are derived from linear combinations of variables in both variable groups (Keskin and Özsoy, 2004).
The objectives of canonical correlation analysis are as follows (Çemrek, 2012):

a) Testing that two classes of variables obtained from identical individuals are statistically free from each other,

b) Identifying the variables in both variable classes that provide the greatest benefit among the classes,

c) Determining the linear combination that maximizes the correlation in the classes of dependent and independent variables,

d) Determining the extent to which variable sets can explain each other among themselves,

e) Determining how much a canonical variable benefits the explanatory power of its class in the variable class,

f) Determining the relative power of various canonical factors to predict or explain relationship size.

1. Canonical correlation analysis assumptions

Results of canonical correlation analysis; covariance, multiple normal distribution, multiple linear relationship assumptions and linearity should be analyzed. The linear assumption affects canonical correlation analysis in two ways. First, the link between the two variables is assumed to be linear, and if the link between the two variables is nonlinear, analysis is performed until it becomes linear. Second, the canonical correlation coefficients show the linear relationship of the two canonical variables. For this reason, nonlinear connections between variables cannot be explained (Kalaycı, 2006). The multiple normal distribution condition must be minimized or provided, the variables must meet a normal distribution condition and there must be no multicollinearity problem. The minimum number of data variables is 20

it should be solid and the data should not have extreme values (Karagöz, 2014). There should be no unnecessary variables in variable groups (Sagın, Koğar, Çakan, 2012). Since identical variance reduces the covariance between variables, it should also be analyzed in canonical correlation analysis (Haır, Anderson, Tatham, & Black, 1998).

2. Significance of canonical correlation coefficients

As a result of canonical correlation analysis, it should be decided how important the canonical variable dichotomies are (Tatlıdil, 1996). In this method, the goal is to analyze how many of the obtained canonical correlation binaries are significant. While it is predicted that all canonical correlations are equal to zero in Wilk's Lambda method, the H_0 hypothesis is analyzed against the H_1 hypothesis, which predicts that the lowest canonical correlation coefficient is different from zero (Oktay and Kaynak, 2007). The hypotheses to be written to determine the meaning levels of the coefficients are as follows (Özçomak, Gündüz, Demirci, Yakut, 2012):

$$H_0 : P_1 = P_2 = P_3 = \cdots P_n = 0 \quad (2)$$

$$H_1 : P_1 \neq P_2 \neq P_3 \neq \cdots P_n \neq 0 \quad (3)$$

Wilk's Lambda analyzes canonical correlation coefficients together and in identical time. Failure to accept the null hypothesis suggests that the lowest first canonical correlation coefficient is significant, while the remaining n-1 canonical correlation coefficients may be meaningless or significant. The significance of the second canonical correlation coefficient can be analyzed independently of the first. In the calculation of the first canonical correlation coefficient, the significance level of Wilk's Lambda statistic conforms to the p×q degrees of freedom χ^2 distribution. Secondly, $(p \times q) \times (q-1)$ degrees of freedom χ^2 distribution and i. Inside $(p \times i) \times (q-i)$ degrees of freedom χ^2 fits the distribution. If the probability value is less than 0.05 at the 5% confidence level, the null hypothesis is rejected. That is, the correlation coefficients are found to be significant and significant. The second canonical correlation coefficient is analyzed independently of the first canonical correlation coefficient. This process is continued until we find the insignificant one in the canonical correlation coefficient (Karagöz, 2014).

RESULTS

The purpose of this research is to determine the land that is thought to affect the landowners while buying or selling land in line with the answers given by the landowners as a result of the data obtained from the survey, where the only participation condition was the sale or purchase of land within the specified period in 9 neighborhoods determined in the Dulkadiroğlu district of Kahramanmaraş province. The aim of this study is to analyze the relationship and effect sizes between the factors that affect the slope and roughness of the land, the average productivity of the land, the presence of tractors and equipment of the land owners, the size of the land parcel and the reason why the land owners sell or sell the land, with the help of the canonical correlation method.

The main goal is to determine the degree of influence and size of the factors that affect the land owners while buying or selling land, by scientific methods, within the specified period. The sample of the research was obtained with the help of a questionnaire, which was the purchase or sale of land by landowners between the years 1995-2021 in 9 neighborhoods in the Dulkadiroğlu district of Kahramanmaraş province, where the only participation condition was determined. As a result of the data obtained from the questionnaire, the canonical correlation method was used to indicate the relationships between the factors and their degree of influence. According to the estimation results obtained in the research, the average productivity of the land in the first data set, which is thought to be under the influence of the landowners while purchasing or selling the land within the specified years and the degree of impact was determined as the purpose for which the land was bought or sold in the second data set. The rate of canonical correlation coefficient in the data sets created according to the estimation results was found to be 40.32%. It has been determined by the scientific analysis that the impact factors obtained with the help of the survey covering the years 1995-2021 and the only participation condition is the purchase or sale of land by the landowners during these years.

Açıl (1976), Chicoine (1981), First (1997), Öztürk et al. (2013), Baser et al. (2016), Karaca et al. (2016), Engindeniz et al. (2017), Cinar et al. (2018) and Başer et al. (2019)' is in harmony with their work.

In 9 neighborhoods selected from the Dulkadiroğlu district of Kahramanmaraş, the only condition of participation was specified. The fact that landowners do not buy or sell land each year has led to disruptions in the analysis. In future articles or thesis research, in order to avoid these and similar problems, villages or neighborhoods that have been subject to land sales or purchases every year or where there are land owners who bought or sold land in the same years should be preferred, land expropriation and transition assets from treasury to private property should be considered. should not be overlooked.

DISCUSSION AND CONCLUSION

In this research, first of all, data groups were created as a result of the data obtained from the land owners who bought or sold land between the years 1995-2021 with the sole participation condition in 9 determined neighborhoods in the Dulkadiroğlu district of Kahramanmaraş province. Ranking in the obtained data set;

X1: The slope and unevenness of the lands of the surveyed landowners

X2: Average productivity of the lands of the surveyed landowners involved in the land purchase and sale between 1995-2021.

X3: Tractor and equipment availability of surveyed landowners between 1995-2021. X4: Parcel sizes of the lands of the surveyed land owners that were included in the land purchase and sale between 1995-2021.

X5: The current value of the land purchase and sale of the land owners included in the survey between the years 1995-2021.

X6: The reason why the land owners included in the survey sold or bought land.

is in the form. However, for the canonical correlation analysis, the first data set was X1, X2 and X3, while the second data set was analyzed as X4, X5, X6.

Table 4. Spss pearson correlation probability value.

	X1	X2	X3	X4	X5	X6
X1	1	-,401	-,002	,027	,098	,004
		,000	,946	,394	,002	,893
X2	-,401	1	,163	,002	,232	-,161
	,000		,000	,953	,000	,000
X3	-,002	,163	1	,150	-,011	,246
	,946	,000		,000	,729	,000
X4	,027	,002	,150	1	-,010	-,080
	,394	,953	,000		,764	,011
X5	,098	,232	-,011	-,010	1	-,153
	,002	,000	,729	,764		,000
X6	,004	-,161	,246	-,080	-,153	1
	,893	,000	,000	,011	,000	

In Table 4., the pearson probability values are given between the determined values.

Table 5. Spss canonical correlation result.

	Correlation	Eigenvalue	Wilk's statistic	F	Number D.F.	Name D.F.	Probability
1	,403	,194	,785	28,113	9,000	2419,285	,000
2	,250	,067	,937	16,363	4,000	1990,000	,000
3	,012	,000	1,000	,139	1,000	996,000	,710

Table 5. In addition, 3 different correlational relationships were determined. Among them, it was found that the correlational relationship of the first and second in probability values was significant and the third was insignificant. It was determined that the first correlation in the eigenvalues was probably higher than the others. In Wilk's statistics, the relationship of the first correlation is closer to 0.05, so it is probably more

significant than the others. It can be said that two of the three correlation scores that occur together are statistically significant. Canonical correlation value is 40.3%.

Table 6. Standardized canonical correlation coefficient in the spss first dataset.

Values	1	2	3
X1	,466	-,601	-,787
X2	,906	-,549	,330
X3	-,715	-,693	,204

Since the first correlation scores are found to be more significant, a model can be written over the first scores that are significant in Table 6 and more statistically significant models can be established.

$$0.466 \times X1 + 0.906 \times X2 - 0.715 \times X3 \quad (4)$$

form a meaningful model.

Table 7. Standardized canonical correlation coefficient in the spss second dataset.

Values	1	2	3
X4	-,283	-,531	,803
X5	,540	-,798	-,310
X6	-,733	-,505	-,489

Since the first correlation scores are found to be more significant, a model can be written over the first scores that are significant in Table 7 and more statistically significant models can be established.

$$-0.283 \times X4 + 0.540 \times X5 - 0.733 \times X6 \quad (5)$$

Form a meaningful model.

Table 8. Result of canonical loads in spss first dataset.

Values	1	2	3
X1	,105	-,380	-,919
X2	,602	-,421	,679
X3	-,568	-,781	,260

In the first data set in Table 8., it was determined that the perceived safety was 0.602, which has the highest effect among the first canonical scores values. Therefore, in the first data set, X2 (the average productivity of the land owners included in the survey included in the land purchase and sale between the years 1995-2021) is the perceived reliability, which has the strongest explanatory power.

Table 9. Result of canonical loads in spss second dataset.

Values	1	2	3
X4	-,230	-,483	,845
X5	,655	-,715	-,242
X6	-,793	-,341	-,505

In the second data set in Table 9, it was determined that the perceived security was the highest with -0.793 among the first canonical scores values. Therefore, in the second data set, X6 (the reason why the land owners in the survey sell or buy land) is the perceived reliability, which is negative but has the strongest explanatory power.

Table 10. Cross-correlation result in spss first dataset.

Values	1	2	3
X1	,042	-,095	-,011
X2	,243	-,105	,008
X3	-,229	-,195	,003

The variable showing the strongest canonical effect in the first data set in Table 10. was X2 with -0.243. In other words, the variable X2 (the average productivity of the land owners involved in the land purchase and sale between the years 1995-2021) that established the strongest relationship with the second data set in the first data set.

Table 11. Cross-correlation result in spss second dataset.

Values	1	2	3
X4	-,093	-,121	,010
X5	,264	-,179	-,003
X6	-,320	-,085	-,006

The variable showing the strongest canonical effect in the second data set in Table 11 was -0.320 to X6. In other words, the variable X6 (the reason why the land owners included in the survey sold or bought land) had the strongest relationship with the first data set in the second data set.

Table 12. Explained rate of variance of canonical correlation in spss.

Canonical Variable	Set 1 to set 1	Set 1 to set 2	Set 2 sets to 2	Set 2 sets to 1
1	,232	,038	,370	,060
2	,310	,019	,287	,018
3	,458	,000	,343	,000

In Table 12., the strongest explanation value for the first data set was 0.458, the second data set for the first data set was found to be 0.038, the second data set for the first data set was found to be 0.370, and the first data set for the second data set was found to be 0.060.

References

Açıl AF, 1976. The Values of Various Agricultural Land Types in Turkey Between 1950-1974 and Their Changes, Ankara University Faculty of Agriculture Publications, No: 619.

Albayrak S, 2006. Applied Multivariate Statistical Techniques, Asil Yayın Dağıtım, p500, Ankara, Turkey.

Başer U, 2015. Determination of the factors affecting the land price in Lâdik district of Samsun province, Master's thesis, Ondokuz Mayıs University, Institute of Science and Technology, Samsun.

Başer U, Kılıç O, 2016. Determination of Factors Affecting Land Price, Samsun Province Lâdik District Example, Turkey XII. National Agricultural Economics Congress, 25-27 May, 273-280, Isparta, Turkey.

Başer U, Kılıç O, Abacı H, 2019. Determination of Direct and Indirect Effects of Factors Affecting Land Price by Path Analysis, Samsun Province Lâdik District Example, Turkish Journal of Agricultural Research, 6(1):50-56.

First A, 1997. A Research on the Factors Affecting the Values of Field Lands in Erzurum and Erzincan Provinces, PhD Thesis, Atatürk University, Institute of Science, Erzurum.

Chicoine DL, 1981. Farmland Values at the Urban Fringe, an Analysis of Sale Prices, Land Econ, 57(3):353 -362.

Çemrek F, 2012. Investigation of the Relationship Between Income and Welfare Level Variables of Provinces in Turkey with Canonical Correlation Analysis, Eskişehir Osmangazi University Journal of Economics and Administrative Sciences, 7(2):197-215.

Çınar G, Altınok AC, Özcan H, Aslan F, 2018. Estimation of Factors Affecting Agricultural Land Value in Aydın Province with Hedonic Pricing Model, Ahtamara I. International Multidisciplinary Studies Congress, 25-26 August, Van, Turkey.

Dillon WR, Goldstein M, 1984. Multivariate Analysis Methods and Applications, John Wiley and Sons Inc., New York.

Engindeniz S, Bayraktar Ö, Örük G, 2017. Analysis of Factors Affecting the Value of Irrigable Agricultural Lands in İzmir: İzmir Case, Selçuk Journal of Agriculture and Food Sciences, 31(3): 75-87.

Har JF, Anderson RE, Tatham RL, Black WC, 1998. Multivariate Data Analysis, Prentice Hall, New Jersey.

Kalaycı Ş, 2006. SPSS Applied Multivariate Statistical Techniques, Asil Yayın Dağıtım, Ankara.

Karagöz Y, 2014. SPSS Applied Biostatistics-For Medicine, Pharmacy, Dentistry and Health Sciences, Nobel Yayın Dağıtım, (2): p760, Ankara, Turkey.

Karaca D, İbraimi S, Engindeniz S, Akyüz Y, Çınar G, 2016. Analysis of Factors Affecting the Land Purchase-Sale Decisions of Producers in Rural Areas: The Example of Kemalpaşa District of İzmir, Journal of Ege University Faculty of Agriculture, 53(4): 481-489 .

Keskin S, Özsoy AN, 2004. Canonical Correlation Analysis and an Application, Journal of Agricultural Sciences, 10(1):67-71.

Oktay E, Kaynak S, 2007. Investigation of Canonical Relationship Between Information Economy Input and Output Variables of Turkey and European Union Countries, Atatürk University Journal of Social Sciences Institute, 2(10):419-440.

Özçomak MS, Demirci A, 2010. Investigation of the Relationship Between Social and Economic Indicators of African Union Countries by Canonical Correlation Analysis, Atatürk University Journal of Economics and Administrative Sciences, 14 (1):261-274.

Özçomak MS, Gündüz M, Demirci A, Yakut E, 2012. Investigation of the Relationship Between Various Climate and Crop Data with Canonical Correlation Analysis and Data Envelopment Analysis Methods, Atatürk University Journal of the Faculty of Economics and Administrative Sciences, 26 (1):111-131.

Öztürk KC, Engindeniz S, 2013. Hedonic Analysis of Agricultural Land Values: The Example of Menemen District of İzmir, Ege University Faculty of Agriculture Journal, 50(3):241-250.

Guide E, 2012. Agricultural Valuation and Expertise, Third Edition, Ekin Publishing House, Bursa.

Mr. A, Koğar H, Çakan M, 2012. Examination of Relationships Between Graded Courses with Canonical Correlation Technique: The Example of Classroom Teaching, Journal of Measurement and Evaluation in Education and Psychology, 3(1):210- 220.

Sümbüloğlu K, Akdağ B, 2009. Advanced Biostatistical Methods, Hatipoğlu Publications, p330, Ankara, Turkey.

Tatlıdil H, 1996. Applied Multivariate Statistical Methods, Cem Web Ofset, Ankara, Turkey.

Contribution Rate Statement Summary of Researchers

The authors declare that they have contributed equally to the article.

Conflict of Interest Statement

The authors of the article declare that there is no conflict of interest between them.

Some Properties of the Helices on De-Sitter Space S_1^3

Esra Çiçek Çetin[1*], Mehmet Bektaş[1]

[1] Firat University , Department of Mathematics, 23119 Elazığ/Türkiye

*Corresponding author e-mail: esracicek23@gmail.com

Abstract

It is well known that there exist characterizations for curve in Euclidean space. Also, a lot of authors extended these characterizations for de Sitter space S_1^3 and obtained very different results. In the present work, first of all, we give some definition and theorem about 4-dimensional Minkowski space and then we express basic conceptsand main result on geometry of space-like curves in de sitter S_1^3- space. On the other hand, we obtain some characterizations of a general helix in de Sitter space S_1^3 and we research characterizations of helices in terms of circular helices and general helices with respect to Frenet frame in de Sitter space S_1^3. So, we obtain some theorem and their proof by using some definition and Frenet equations. Accordingly, we have obtained some results in de Sitter space S_1^3

Key words: *Helices, De Sitter Space, Space-Like Curve.*

Determination of Agricultural Mechanization Level in Kahramanmaraş with Poission Regress By District

Mustafa Şahin[1], İsmail Gök[2*], Tolga Tolun[2]

[1] Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Department of Agricultural Biotechnology, 46100, Kahramanmaraş, Turkey.
[2] Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Department of Bioengineering, 46100, Kahramanmaraş, Turkey.

*Corresponding author e-mail: gkisoo1995@gmail.com

Abstract

In this study, agricultural mechanization level indicators of the districts of Kahramanmaraş between the years 2012-2021 were calculated. A model has been established for this. While the dependent variable in the model is the agricultural areas planted in the districts according to the years, the independent variables are the number of tractors in the districts according to the years and other determined tools. In this area where Poisson regression test is used, there is a correlation of 25.022 chi-square probability between the agricultural areas planted in Kahramanmaraş districts between 2012 and 2021, with the number of tractors in the districts and other determined equipment by years. At the level of agricultural mechanization, it was determined that there was a 3.32 decrease in the average number of tractors per 1000 hectares of land in Kahramanmaraş districts between 2012 and 2021, and a decrease of 12.45 in the number of specific equipment per tractor over the years. Among this decrease, it was thought that reasons such as climate change, input costs and job change were among the main reasons for leaving agriculture. In addition, as a result of the trend analysis value applied, it was concluded that if the average absolute percentage errors are the average number of tractors per 1000 hectares of districts over the years, the average number of tractors is 4.10443, and the number of specific equipment per tractor in the districts is 1.96718, according to the years.

Key words: *Agricultural land, Tractor, District and Years.*

INTRODUCTION

Agricultural sector in Turkey; It is an economic and social business line with its impact on livelihood and labor force, its contribution to general national income, and the raw materials it provides to the industry.

In order to meet the needs of the rapidly increasing human population, a better quality and more measured production area in the agricultural sector working areas is one of the main goals of agricultural production in today's agriculture. This goal makes it inevitable to use technology to accelerate and improve agricultural production. Within the production work areas, agricultural mechanization covers a special area. Agricultural mechanization studies should be used in production in order to obtain higher quality and higher efficiency in unit area and unit time. Agricultural mechanization methods are divided into two basic clusters as power and machines used in the job site. Tractors, which are called force machines in the field of agricultural mechanization, can give power to agricultural-construction machines that do not have the ability to move themselves in various situations in order to fulfill their mobility (Sümer et al., 2003; Sümer et al., 2004).

Agricultural mechanization, use of advanced technological networks in the agricultural sector; It is an important production method that allows widespread use of inputs such as irrigation, fertilization, soil cultivation and maintenance (Korucu et al., 2015). Agricultural mechanization level can show awareness from region to region in terms of both economic and technical levels. Determination of its level with the help of signs of agricultural mechanization; It is important in terms of comparing and interpreting the agricultural mechanization levels of countries, regions and provinces. The most basic power source used in the operation of the fields is the tractor. For this reason, the tractor used by the landowners in the land operation is one of the most important factors in determining the agricultural mechanization levels of the regions. The agricultural mechanization level, which is based on the widely used tractor, is the tractor power per the planted land for production, the number of tractors per 1000 hectares, the production area per one tractor and the number of equipment per one tractor (Sabancı et al., 2003).

The aim of this research is to determine the signs of agricultural mechanization and to determine the effect of mechanization based on the size of the agricultural land planted in the districts of Kahramanmaraş between the years 2012-2021, the number of tractors and the determined tractor equipment. is to detect the change that may occur in the level.

MATERIAL AND METHODS

Material

The data of this research was obtained from the Turkish Statistical Institute (TUIK). According to the data obtained from TUIK, tractor horsepower is 35-50 horsepower, 51-70 horsepower, 70 and more horsepower, lug tractor plow, arc plow and cultivator, which are estimated to be widely used in the equipment of the tractor, and fruits in the planted agricultural lands, Size of agricultural land planted between 2012-2021 in Afşin, Andırın, Dulkadiroğlu, Ekinözü, Elbistan, Göksun, Nurhak, Onikisubat, Pazarcık, Türkoğlu and Çağlayancerit districts of Kahramanmaraş province, based on beverage and spicy plants, vegetable fields and cereals and other herbal products areas. A database was created about the specified tractor equipment and the number of tractors. A model has been established to determine the level of mechanization. While the dependent variable in the model is the agricultural areas planted in the districts according to the years, the independent variables are the number of tractors in the districts and other determined tools according to the years. It was tested with the poisson distribution method in the SPSS program and a comparison was made between the districts as a result of the years and data determined. Agricultural land cultivated by years in all districts between 2012-2021 Table 1.A. and it is given in Table 1.B.

Table 1.A. Total processed parcel size in districts by year.

	Total Parcel Size Processed by Years				
Districts	**2012**	**2013**	**2014**	**2015**	**2016**
Afşin	465279	571975	591262	585041	520283
Andırın	123581	139869	133982	137441	137897
Dulkadiroğlu	-	260721	278840	270480	271755
Ekinözü	54880	57287	57913	59476	53050
Elbistan	651569	789037	675844	812548	768263
Göksun	432367	395509	408389	409696	401786
Nurhak	22151	21111	19785	20110	17812
Onikişubat	-	285647	287291	282687	283001
Pazarcık	339190	492012	495116	483087	365206
Türkoğlu	131039	185061	177950	170392	157967
Çağlayancerit	55461	38694	41877	41722	38043

Table 1.B. Total processed parcel size in districts by year.

	Total Parcel Size Processed by Years				
Districts	**2017**	**2018**	**2019**	**2020**	**2021**
Afşin	515353	495761	524733	542292	534596
Andırın	141226	184605	193800	195221	193722
Dulkadiroğlu	280649	296847	305961	320840	258641
Ekinözü	42304	46613	42722	48429	50093
Elbistan	742876	714372	813482	813605	808605
Göksun	402286	401382	383421	394269	393365
Nurhak	21622	21875	22521	23812	28297
Onikişubat	281186	281788	272667	302150	296269
Pazarcık	356573	387593	360641	365816	357829
Türkoğlu	184275	184486	218148	221055	237495
Çağlayancerit	43668	44431	45118	48317	59623

The total processed parcel size in the districts in 2012 is 2275517 m2, in 2013 3236923 m2, in 2014 3168249 m2, in 2015 3272680 m2, in 2016 3015063 m2, in 2017 3012018 m2, in 2018 3059753 m2, in 2019 3183214 m2, in 2020 3275806 m2 and In 2021, it was determined to be 3218535. In addition, when the total processed parcel size in the districts in 2012 and the total processed parcel size in the districts in 2021 are compared, an increase in the total parcel size processed in the districts was observed by 41.44%.

In this research, the total number of tractors used in agricultural lands cultivated in all of the districts between 2012 and 2021 is given in Table 2.

Table 2. The total number of tractor equipment in the districts by years.

	Total Number of Tractor Equipment by Years.									
Districts	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
Afşin	2810	2820	2830	2830	2900	2910	2960	3030	3030	3030
Andırın	1618	1625	1625	1635	1635	1685	1685	1691	1691	1724
Dulkadiroğlu	-	1828	2180	1920	1924	1975	2230	2260	2375	2430
Ekinözü	183	185	187	187	187	187	191	194	202	202
Elbistan	6550	6500	6500	6510	7500	7500	7450	7600	8210	9400
Göksun	1445	1455	1480	1490	1504	1521	1521	1530	1537	1544
Nurhak	195	201	201	196	205	212	219	222	227	228
Onikişubat	-	1352	1390	1390	1400	1401	1413	1424	1443	1514
Pazarcık	2428	2428	2428	2428	2430	2438	2438	2440	2440	2440
Türkoğlu	920	940	997	1018	1029	1115	1137	1152	1171	1205
Çağlayancerit	294	294	294	294	308	313	312	312	308	308

The total number of tractor equipment in the districts in 2012 was determined to be 16443, 19628 in 2013, 20112 in 2014, 19868 in 2015, 21022 in 2016, 21257 in 2017, 21556 in 2018, 21855 in 2019, 22634 in 2020 and 24025 in 2021. In addition, when the total number of tractor equipment in all districts in 2012 and the total number of tractor equipment in all districts in 2021 are compared, an increase in the total number of tractor equipment was observed by 46.11%.

Method

The aim of this research is; The aim of this study is to determine the degree of impact between the number of tractors and other determined tools, which are thought to be connected to the agricultural areas processed in the districts of Kahramanmaraş between the years 2012-2021. In this, the poisson distribution, which is a statistical analysis method, was used.

1. Poisson regression

Poisson regression is a mathematical analysis that determines the random degree of a data set, is useful in making the data sets created for analysis fit the institutional curve, and is used in the stage of executing certain predictions from the main data sets (Gerlough; Schuhl, 1955). It is applied to the problems that have occurred in the explanatory variables detected in the cases that have emerged in the period sections that are desired to be analyzed. Model; It sets out by estimating that the logarithm of possible numbers is a linear function of the explanatory variables, since they contain discontinuous and non-negative countable data. Special explanatory variables can be mentioned as the logarithm of the degrees of certain numbers before and after a one-unit increase that may occur in the explanatory variable as soon as all other explanatory variables are assumed to be constant (Köleoğlu, 2006). There are times when the answer variable takes a non-continuous value such as 0,1,2,…,n, but is not categorical. The model with continuous and non-categorical response variables associated with rare events is called the Poisson regression model based on certain assumptions. This model, called the Poisson regression model, is generally used to analyze counting data (Akın, 2002). Since the model is an exponential model, even if it is a negative situation that it causes difficulty and complexity in the interpretation of the coefficients of the model, linear regression analysis is a model that can be alternated in events consisting of the counting data of the answer variable (Deniz, 2005). Poisson regression model; In multiplicative models containing approximate data, when there is a conditional dependence between the explanatory variables, it provides an advantage to those who work because it is an elastic model (Lloyd, 1999). The Poisson regression model is the most used model after the logistic regression model. The most striking feature of this model is that there is an equality between the variance and the mean. However, in many applications it is not possible to achieve this equality. In the Poisson distribution; If the obtained variance is less than the obtained mean, it is called under-dispersion, and if the obtained variance is greater than the obtained mean, it is called over-dispersion. The dependent variable Y, which is defined as the number of related cases in the Poisson regression;

$$X_1, X_2, \cdots, X_n \tag{1}$$

If independent variables are given, it is assumed to obtain the Poisson distribution. Thus, it is assumed that there is a linear factor hypothesis derived from the independent variables of the logarithm of μ, which is the Poisson mean (SAS, 2005; Yeşilov et al, 2006). Poisson regression model according to the function in the subject;

$$Log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \tag{2}$$

given in the form. in the model μ becomes an exponential factor of the independent variables. $\mu'y\ddot{u}$,

$$\mu = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m) \tag{3}$$

can be written as

The Poisson distribution is a type of distribution that can determine the number of events or situations that are likely to occur in a specified place or volume in a specified period. This distribution was introduced by a French mathematician in the 18th century, and he named this distribution the poisson. In the following times, it has taken its current form in the research and analysis of different scientists (Gürsakal, 1997).

The Poisson distribution formula is:

$$P(y;\ \mu) = \frac{\mu^{y} e^{-\mu}}{y_i} \quad y=0,1,2\ldots \tag{4}$$

İn equality y_i, the number of occurrences of expected event events, and μ_i' i is the distribution parameter. Well $E(y_i) = \mu_i$ is in the form.

μ_i'in Poisson regression model is obtained when it is dependent on the explanatory variables. $\mu - i$ generally $\mu_i = e^{x'\beta}$ can be obtained as a supplement. Here x is in the position of the explanatory variable vector. β' a indicates the parameter vector to be evaluated. The approximate value of the Poisson distribution is:

$$\mu_i = E(\frac{y_i}{x_i}) = \exp(x'_i\beta) \tag{5}$$

given in the form. In the statistics branch, the above-mentioned model is referred to as the linear logarithmic model. In the Poisson distribution, the approximate variance is equal.

$$\mu_i = E(\frac{y_i}{x_i}) = V(\frac{y_i}{x_i}) \tag{6}$$

The state of equality between the mean and the variance is called equal spread. However, in the analysis, the counting variables indicate overdispersion as their variances are larger than the mean in most periods. Since the number of analyzed zero values is overdispersed, the poisson regression model causes situations or events such as unobserved heterogeneity if it exceeds the obtained zero values (Kibar, 2008).

The overdispersion problem occurring in the model is not likely to affect the predictions made in the coefficient or coefficients. However, it has the possibility to have an effect on the estimated standard error area in the model, and within this, it increases the reliability level of the model (AL-Ghirbal; AL-Ghamdi, 2006).

2. Trend analysis

It is the representation of the long-term fundamental trend in a time series with a curve or a straight line (Yavuz, 2016). In the trend analysis method, the trend equation that best describes the time series is determined and predictions are made for the next times. It can be applied with trend analysis method estimation, moving average analysis method, semester analysis method and least squares analysis method. However, in order to obtain the equation of the trend line, it is necessary to observe various types of equations (logarithmic, exponential, etc.) apart from the least squares analysis method (Witt and Witt, 1992).

RESULTS

The aim of this research is to determine the signs of agricultural mechanization and to determine the effect of mechanization based on the total land size, number of tractors and determined tractor equipment in the districts of Kahramanmaraş between the years 2012-2021. is to determine the change that may occur in the level of mechanization to the region.

The sample of this study was obtained from the Turkish Statistical Institute (TUIK). According to the data obtained from TUIK, tractor horsepower is 35-50 horsepower, 51-70 horsepower, 70 and more horsepower, lug tractor plow, arc plow and cultivator, which are estimated to be widely used in the equipment of the tractor, and fruits in the planted agricultural lands, Size of agricultural land planted between 2012-2021 in Afşin, Andırın, Dulkadiroğlu, Ekinözü, Elbistan, Göksun, Nurhak, Onikisubat, Pazarcık, Türkoğlu and Çağlayancerit districts of Kahramanmaraş province, based on beverage and spicy plants, vegetable fields and cereals and other herbal products areas. A database was created about the specified tractor equipment and the number of tractors. For the estimation results obtained in the research, poisson regression analysis and trend analysis methods were used. According to the estimation results, there is a 25.022 chi-square probability relationship between the agricultural areas planted in Kahramanmaraş districts between 2012 and 2021, with the number of tractors in the districts by years and other specified equipment. At the level of agricultural mechanization, it was determined that there was a 3.32 decrease in the average number of tractors per 1000 hectares of land in Kahramanmaraş districts between 2012 and 2021, and a decrease of 12.45 in the number of specific equipment per tractor over the years. Among this decrease, it was thought that reasons such as climate change, input costs and job change were among the main reasons for leaving agriculture. In addition, as a result of the trend analysis value applied, it has been concluded that the average absolute percentage errors in the districts are 4.10443, the average number of tractors per 1000 hectares by years, and the number of specific equipment per tractor in the districts is 1.96718, according to the years.

Root (1993), Lule et al. (2012), Bilim et al. (2014), Korucu et al. (2015), Bozkurt (2016), Akar et al. (2017), Aslantürk et al. (2018) and Abdikoğlu (2019) are in harmony with their work.

The reasons for not obtaining full good results in the districts of Kahramanmaraş as a result of the data obtained from TUIK are that the specified time interval was kept short and the criteria under the preference sub-title were chosen more in the data obtained. . In order to avoid these and similar problems in future article or thesis research, the specified period should be kept wider, and regions or regions dealing with agriculture should be preferred.

DISCUSSION AND CONCLUSION

In this research, first of all, a model was established in the districts of Kahramanmaraş as a result of the data set obtained from TUIK between the years 2012-2021, the dependent variable is the agricultural areas planted in the districts according to the years, the number of tractors in the districts according to the years in the independent variables and the other determined equipments, and the hypotheses of the model are:

H0: There is no relationship between the agricultural lands planted in Kahramanmaraş districts between 2012-2021, the number of tractors in the districts by years and other specified equipment.

H1: : There is a relationship between the agricultural lands planted in Kahramanmaraş districts between 2012-2021, the number of tractors in the districts by years and other specified equipment.

model:

$$Y_{it=}\beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + u_{\ it} \tag{7}$$

Yit: The size of the agricultural areas planted in Kahramanmaraş districts between 2012-2021 by years.

X1it: Number of tractors found in Kahramanmaraş districts between 2012-2021 by years.

X2it: The number of specific tractor equipment found in Kahramanmaraş districts by years between 2012-2021.

Table 3. Categorical variable outcome of X1 poisson regression.

Categorical Variable Information		
X1	**N**	**Percent**
0-500	30	%27.8
501-1000	22	%20.4
1001-1500	26	%24.1
1501-2000	20	%18.5
2501-3000	4	%3.7
3001-3500	3	%2.8
3501-4000	3	%2.8
Total	108	%100

Between the years 2012-2021, the number of tractors found in Kahramanmaraş districts according to the years is 30, as a result of the years specified in the range of 0 to 500 in the districts, and this determined range explains 27.8%. Tractor assets as a result of the years specified in the range of 501 to 1000 are 22, and this specified range explains 20.4%. As a result of the years specified in the range of 1001-1500, the tractor asset is 26 and this specified range explains 24.1%. Tractor assets as a result of the years specified in the range of 1501-2000 are 20, and this specified range accounts for 18.5%. Tractor assets as a result of the years specified in the range of 2501-3000 are 4 and this explains 3.7% of the specified range. The tractor asset as a result of the years specified in the range 3001-3500 is 3, and this specified range explains 2.8%. The tractor asset as a result of the years specified in the range 3501-4000 is 3, and this specified range explains 2.8%.

Table 4. Categorical variable outcome of x2 poisson regression.

Categorical Variable Information		
X2	**N**	**Percent**
0-1000	33	%30.6
1001-2000	40	%37.0
2001-3000	22	%20.4
3001-4000	3	%2.8
6001-7000	4	%3.7
7001-8000	4	%3.7
8001-9000	1	%0.9
9001-10000	1	%0.9
Total	108	%100

Between the years 2012-2021, the number of tractors found in Kahramanmaraş districts by years is 33, as a result of the years specified in the range of 0 to 1000 in the districts, and this determined range explains 30.6%. As a result of the years specified in the range of 1001 to 2000, the tractor asset is 40, and this specified range explains 37%. The number of tractors as a result of the years specified in the 2001-3000 range is 22, which explains 20.4% of the specified range. Tractor assets as a result of the years specified in the range of 3001-4000 are 3, and this specified range explains 2.8%. Tractor assets as a result of the years specified in the range 6001-7000 are 4 and this explains 3.7% of the specified range. Tractor assets as a result of the years specified in the range 7001-8000 are 4 and this explains 3.7% of the specified range. Tractor assets as a result of the years specified in the range 8001-9000 are 1 and this explains 0.9% of the specified range. The tractor asset as a result of the years specified in the range 9001-10000 is 1 and this explains 0.9% of the specified range.

Table 5. Poisson regression chi-square likelihood result.

Chi-Square Likelihood Ratio	Degrees of Freedom	Probability
25.022	7	0.001

As stated in Table 5., since the probability value is less than 0.05, the null hypothesis is rejected. In other words, in the established model, there is a 25,022 chi-square probability relationship between the agricultural areas planted in Kahramanmaraş districts between 2012 and 2021, with the number of tractors in the districts by years and other specified equipment.

In addition, at the agricultural mechanization level, the average number of tractors per 1000 hectares of districts by years Table 6. and the number of specific equipment per tractor in the districts by years is given in Table 7.

Table 6. Average number of tractors per 1000 hectares of districts by years.

	Average Number of Tractors Per 1000 Hectares by Years.									
District	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
Afşin	3.32	2.72	2.63	2.66	3.01	3.04	3.19	2.97	2.85	2.86
Andırın	10.89	9.63	10.05	9.95	9.92	9.75	7.46	7.17	7.12	7.19
Dulkadiroğlu	-	2.96	4.42	4.62	4.60	4.49	4.37	4.34	4.22	5.45
Ekinözü	2.00	1.92	1.98	1.93	2.16	2.71	2.55	2.90	2.99	3.05
Elbistan	4.45	3.80	4.43	3.69	4.16	4.33	4.65	4.36	4.37	4.77
Göksun	1.94	2.14	2.10	2.13	2.19	2.26	2.27	2.38	2.31	2.30
Nurhak	7.90	8.19	8.99	9.69	10.61	8.92	9.18	9.14	7.18	6.11
Onikişubat	-	4.56	7.22	7.34	7.33	7.37	7.36	7.61	6.90	7.39
Pazarcık	4.84	3.33	3.31	3.40	4.51	4.64	4.28	4.61	4.54	4.68
Türkoğlu	4.76	3.41	3.70	3.96	4.46	4.01	4.01	3.41	3.48	3.72
Çağlayancerit	0.50	0.72	0.66	0.67	0.76	0.66	0.65	0.70	0.68	0.53

The average number of tractors per 1000 hectares between the years 2012-2021 in all districts by years; It was determined that it was 4.51 in 2012, 3.94 in 2013, 4.49 in 2014, 4.54 in 2015, 4.88 in 2016, 4.74 in 2017, 4.54 in 2018, 4.50 in 2019, 4.24 in 2020 and 4.36 in 2021. In addition, when the average number of tractors per 1000 hectares of land in 2012 is compared with the average number of all districts by years and the average number of tractors per 1000 hectares of land in 2021, it has been determined that the average number of all districts decreases by 3.32%. Among the reasons for the decrease compared to these years, climate change, input costs, job change were among the main reasons for leaving agriculture. In addition, the trend analysis result of the average number of tractors per 1000 hectares of districts by years is given in Figure 1.

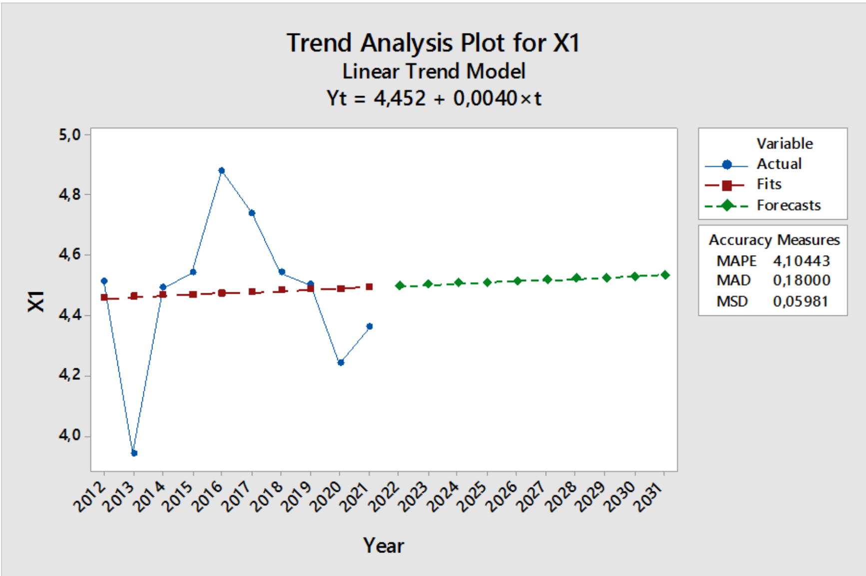


Figure 1. Trend analysis of the average number of tractors (X1) per 1000 hectares of districts by years.

Mean absolute percentile errors (MAPE) measure how much the values estimated by this measure deviate from the true values. From this point of view, a low MAPE value means a better estimation has been made. In Figure 1., since the MAPE value is 4.10443, it is concluded that a good estimation has been made.

Table 7. The number of specific equipment per tractor in the districts by years.

	Average Number of Certain Equipment Per Tractor by Years.									
District	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
Afşin	1.81	1.81	1.81	1.81	1.84	1.85	1.86	1.94	1.95	1.98
Andırın	1.20	1.20	1.20	1.19	1.19	1.22	1.22	1.21	1.21	1.23
Dulkadiroğlu	-	2.36	1.76	1.53	1.53	1.56	1.71	1.69	1.75	1.72
Ekinözü	1.66	1.68	1.62	1.62	1.62	1.62	1.60	1.56	1.39	1.32
Elbistan	2.25	2.16	2.16	2.17	2.34	2.32	2.24	2.14	2.30	2.43
Göksun	1.72	1.71	1.72	1.70	1.70	1.66	1.66	1.67	1.68	1.70
Nurhak	1.11	1.16	1.12	1.10	1.08	1.09	1.08	1.07	1.32	1.31
Onikişubat	-	1.03	0.66	0.66	0.67	0.67	0.68	0.68	0.69	0.69
Pazarcık	1.47	1.47	1.47	1.47	1.47	1.47	1.46	1.46	1.46	1.45
Türkoğlu	1.47	1.48	1.51	1.50	1.45	1.50	1.53	1.54	1.51	1.36
Çağlayancerit	10.50	10.50	10.50	10.50	10.62	10.79	10.75	9.75	9.33	9.62

All district averages of the number of specific equipment per tractor in the districts between 2012-2021; It was determined that it was 2.57 in 2012, 2.41 in 2013, 2.32 in 2014, 2.29 in 2015, 2.31 in 2016, 2.34 in 2017, 2.34 in 2018, 2.24 in 2019, 2.23 in 2020 and 2.25 in 2021. In addition, it has been determined that the number of specific equipment per tractor in the districts in 2012 compared to the average number of all districts and the number of specific equipment per tractor per year in the districts in 2021, compared to the average number of all districts, decreased by 12.45%. Among the reasons for the decrease compared to these years, climate change, input costs, job change were among the main reasons for leaving agriculture. In addition, the trend analysis result of the number of specific equipment per tractor in the Districts by years is given in Figure 2.

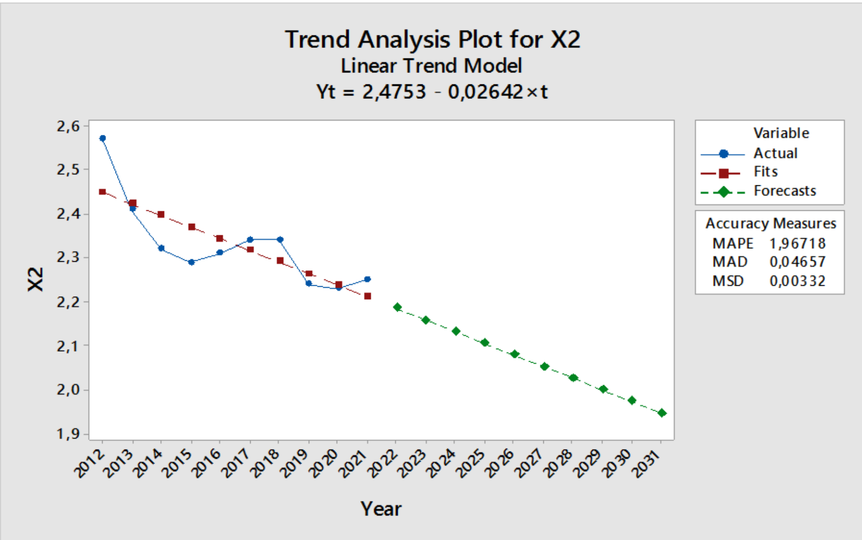


Figure 2. Trend analysis of the number of specific equipment per tractor (X2) of the districts by years.

Mean absolute percentile errors (MAPE) measure how much the values estimated by this measure deviate from the true values. From this point of view, a low MAPE value means a better estimation has been made. As the MAPE value is 1.96718 in Figure 2., it is concluded that a good estimation has been made.

References

Abdikoğlu Dİ, 2019. Determination of Agricultural Mechanization Level by Provinces in Thrace Region, Journal of Agriculture and Nature, 22(6):865-871.

Akar M, Çelik A, 2017. Agricultural Mechanization Characteristics of Muş Plain Agricultural Enterprises, Turkish Journal of Agricultural and Natural Sciences, 4(4):491-498.

Akın F, 2002. Qualitative analysis of preference models, Ekin Bookstore, p135, Bursa, Turkey.

Al-Ghirbal AS, Al-Ghamdi AS, 2006. Predecting severe accidents rates at roundabouts using Poisson distribution. TRB Annual Meeting, TRB 06-1684.

Aslantürk B, Altuntaş E, 2018. Agricultural Mechanization Level of Malatya Province, Gaziosmanpaşa Scientific Research Journal, 7(2):15-26.

Bilim C, Korucu T, Semerci T, 2014. Agricultural Mechanization Characteristics of Gaziantep Province, KSU Journal of Natural Sciences, 17(2):14-23.

Bozkurt M, 2016. Agricultural Structure Mechanization Properties in Şanlıurfa Province Harran Plain Determination of Production Techniques and Technologies, Master Thesis, Kahramanmaraş Sütçü İmam University, Institute of Science, Biosystem Engineering, Kahramanmaraş, Turkey.

Deniz Ö, 2005. Poisson Regression Analysis, Istanbul Commerce University, Journal of Science, 4(7):59-72.

Gerlough DL, Schuhl A, 1955. Use of Poisson Distribution in Highway Traffic – The Probability Theory Applied to Distribution of Vehicles on Two-Lane Highways, Saugatuck, Connecticut, USA: The ENO Foundation for Highway Traffic Control, 1-58.

Gürsakal N, 1997. Computer Applied 1, Marmara Bookstore, p527, Bursa, Turkey.

Kibar FT, 2008. Traffic accidents and Establishment of Accident Prediction Model in the Example of Trabzon Divided Coastal Road, Master Thesis, Karadeniz Technical University, Institute of Science and Technology, Trabzon, Turkey.

Korucu T, Aybek A, Sivrikaya F, 2015. Creation and Evaluation of Local Change Maps of Mechanization Level on the Basis of Agricultural Regions of Turkey, KSU Journal of Natural Sciences, 18(4):77-90.

Korucu T, Aybek A, Sivrikaya F, Gürlek E, Mert C, Kozak B, 2015. Mapping and Evaluation of Agricultural Mechanization Level of Kahramanmaraş Province, KSU Journal of Natural Sciences, 18(2):10-24.

Kök M, 1993. A Research on Agricultural Mechanization Characteristics of Tunceli Province and Examination of These Features in Terms of Agricultural Potential of the Region, Master Thesis, Çukurova University, Institute of Science, Department of Agricultural Machinery, Adana, Turkey.

Köleoğlu N, 2006. Random Effects in Event Time Analysis Investigation of Unobservable Heterogeneity with Poisson Regression Model, PhD Thesis, Marmara University, Institute of Social Sciences, Istanbul, Turkey.

Lloyd C J, 1999. Statistical Analysis of Categorical Data, New York.

Lüle F, Koyuncu T, Engin KE, 2012. Agricultural Mechanization Level of Adıyaman Province, 27th Agricultural Mechanization National Congress, 5-7 September, Samsun, Turkey.

Sabancı A, Sümer SK, Say MS, Has M, 2003. Economic Tractor Park and Its Development in Turkey, 21st Agricultural Mechanization National Congress, 3-5 September, Konya, Turkey.

SAS., 2005. SAS/STAT Software: Hangen and Enhanced. SAS, Inst. Inc., USA.

Sümer SK, Say SM., Has M., Sabancı A, 2003. Economic Tractor Park and Its Development in Turkey, Çukurova University, Journal of the Faculty of Agriculture, 18(4):45-52.

Sümer SK, Has M, Sabancı A, 2004. Technical Specifications of Agricultural Tractors Produced in Turkey, Çukurova University, Journal of the Faculty of Agriculture, 19(1):17-26.

Yeşilova A, Yılmaz A, Kaki B, 2006. The Use of Nonlinear Regression Models in the Analysis of Some Discrete Reproduction Behavioral Traits of Norduz Male Lambs, Yüzüncü Yıl University, Faculty of Agriculture, Journal of Agricultural Science, 16(2):87-92.

Yavuz H, 2016. Evaluation of Livestock Support Policies in Terms of Agricultural Enterprises, Master Thesis, Gaziosmanpaşa University, Institute of Science, Department of Agricultural Economics, Tokat, Turkey.

Witt SF, Witt CA, 1992. Modeling and Forecasting Demand in Tourism, Academic Press, London.

Contribution Rate Statement Summary of Researchers

The authors declare that they have contributed equally to the article.

Conflict of Interest Statement

The authors of the article declare that there is no conflict of interest between them.

Modeling of Egg Weight, Width and Length Values in Partridges

Tolga Tolun[1*], Esra Yavuz[2], Mustafa Şahin[3]

[1] Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Bioengineering and Sciences, 46000, Kahramanmaraş, Turkey.
[2] Şırnak University, Çizre Melek High School, Accounting and Tax Practices, 73200, Şirnak, Turkey.
[3] Kahramanmaraş Sütçü İmam University, Department of Agriculture, Agricultural Biotechnology, 46000, Kahramanmaraş, Turkey.

*Corresponding author e-mail: tolga_tolun@hotmail.com

Abstract

In this study, Cubic Spline (two-knotted), Gompertz, Logistic, Richard, Schunute and Quadratic Spline models were examined comparatively in modeling the egg weights, width and length distributions of henna partridges. The egg weight, width and length values obtained in the study were obtained from a flock of 60 individuals. The mean squares of error, determination coefficient, Akaike Information Criterion and Durbin-Watson Autocorrelation values were used in the comparison of egg weights, width and length models. It has been determined that the Logistic (HKO: 0.025, R^2: 0.991, AIC: 198.96, DW: 2.01), Richard and Gompertz models are the best models for modeling the curves of egg length values and give more or less similar results. It was determined that the Quadratic Spline (HKO: 0.026, R^2: 0.942, AIC: 312.56, DW: 1.51) model had the worst results. In modeling the curves of the egg width, the best model is the Cubik Spline (HKO: 0.015, R^2: 0.927, AIC: 43.7, DW: 2.03) model and the worst model is the Schnute model (HKO: 0.023, R^2: 0.858, AIC: 184.59, DW: 1.23). Finally, in modeling the curves of egg weight values, the best model belongs to the Cubik Spline (HKO: 0.01, R^2: 0.993, AIC: 42.69, DW: 2.09) model, and the worst model is the Schnute model (HKO: 0.049, R^2: 0.991, AIC: 125.56, DW: 1.45).

Key words: *Partridge, Egg Weight, Egg Width, Egg Length, Modeling.*

INTRODUCTION

Mathematical equations are widely used in modeling point distributions in many fields such as automotive, forestry, textile, ovine and poultry farming, and to reveal and interpret cause-effect relationships.

As in laying hens and Japanese quail (Coturnix coturnix japonica), egg production in partridges is a process influenced by environmental factors such as genotype, care and nutrition. Naturally, the effect of race, age, hatching, moult, feeding style and other environmental factors on egg production amount, egg weight, width and length is high. No matter how different these care and environmental factors are, the distribution of egg production, egg weight, width and length will show a more or less similar distribution from population to population. The model was created and used. Modeling studies of these curves are used to predict the egg production and weight of the current flock in early periods and to establish breeding flocks. For this reason, correct modeling of curves in breeding studies is extremely important in terms of shortening the required time in breeding studies and increasing the degree of accuracy in selection.

In addition to linear, quadratic and cubic models, many non-parametric models such as Gamma, McMillan, Richard, Schunute, Adam and Bell, Logistics, Gloor, etc. are widely used in modeling curves of egg yield and weight in laying hens, quails and partridges (Cason et al. Ware, 1990; Miyoshi et al., 1996; Narushin and

Takma, 2003; Savegnago et al., 2011). In general, although studies that model egg yield curves in poultry breeding are frequently encountered, studies in which modeling of egg weight, width and length are very few in the literature. Features such as egg weight (egg white-yolk), shell thickness, and number of shell pores affect the future performance of poultry. Egg weight; It affects the hatchability of the egg, the hatching time and the weight of the fry. Therefore, it also has an effect on offspring mortality and egg laying in later periods (Balkan and Recep, 2007). An increase in egg weight causes thinning of the eggshell, and if the weight is too low, it causes an increase in the thickness of the eggshell. In both cases, it will cause a decrease in hatchability. For the reasons stated, it is extremely important for the breeder and the breeder to know the change in egg sizes (weight, width and height) during the laying season.

MATERIAL AND METHODS

Material

The values of egg weight, width and length obtained in this study were obtained from henna partridges sold commercially in pet shops in Kahramanmaraş province. For this purpose, 60 henna partridges (55 females-5 males) were obtained and measurements were taken from the eggs obtained from the end of January to the beginning of May. Standard commercial poultry layer feed was used for feeding. Eggs with deformity (extremely sharp or blunt) and eggs with contralateral weights (too light and too heavy) were excluded from the experiment. During the dormitory period, the partridges were housed in the ground compartment covered with wire mesh. In the experiment, the partridges were given free feed and water. The arithmetic mean, standard deviation, maximum and minimum values of 13-week-old egg length, width and weight are given in Table 1.

Table 1. Descriptive statistics of height, width and weight values.

Week	Size				Width				Weight			
	$\bar{X}$	S	Mak	Min	$\bar{X}$	S	Mak	Min	$\bar{X}$	S	Mak	Min
1	39.19	0.19	43.24	36.4	29.48	0.13	31.59	26.8	18.40	0.18	21.93	16.20
2	41.13	0.15	44.44	39.1	30.65	0.07	32.49	28.5	21.57	0.14	24.56	20.30
3	41.94	0.14	44.15	39.2	30.97	0.07	32.15	28.6	22.03	0.15	25.57	21.60
4	41.29	0.15	44.25	39.5	30.74	0.08	32.56	29.3	21.32	0.15	25.79	20.80
5	41.65	0.15	44.17	39.7	30.88	0.07	32.15	29.1	21.98	0.15	26.95	20.50
6	41.71	0.17	45.12	38.5	30.87	0.07	32.17	28.9	21.87	0.15	25.99	20.40
7	41.24	0.15	44.01	38.4	30.68	0.07	32.56	29.1	21.39	0.13	26.01	20.60
8	41.63	0.15	44.23	38.6	30.72	0.07	32.58	29.3	21.64	0.15	25.58	20.30
9	41.07	0.16	44.56	38.2	30.49	0.08	32.14	29.5	20.90	0.16	26.45	20.10
10	41.15	0.16	45.87	38.1	30.58	0.08	32.47	29.6	21.14	0.16	26.86	20.40
11	41.29	0.16	44.61	38.7	30.64	0.09	32.54	29.3	21.38	0.17	25.63	20.70
12	41.48	0.15	44.25	38.4	30.64	0.07	32.56	29.2	21.46	0.14	25.87	20.10
13	41.33	0.16	44.89	38.2	30.49	0.08	32.89	29.7	21.05	0.15	25.19	20.30

Methods

In this study, Segmented Cubic Spline (two-knotted), Gompertz, Logistic, Richard, Schunute and Quadratic Spline models were used to model the egg weights, width and length curves of henna partridges. Egg weights, curves of width and length, and model parameter estimations were made in the SAS package program (SAS, 2011).

Equations of these models (1.1, 1.2, 1.3, 1.4, 1.5 and 1.6) and their expansions are as follows.

Cubic Spline, $w_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (t-\alpha)^3 + \beta_5 (t-b)^3$ (1.1)

Logistic, $w_t = \beta_0/(1+\beta_1 e(-\beta_2 \mathrm{t}))$ (1.2)

Gompertz, $w_t = \beta_0 e(-\beta_2 e(-\beta_3 t)$ (1.3)

Richard, $w_t = 1/(\beta_0 + \beta_1 e(\beta_2 t)^{(-\beta_3)}$ (1.4)

Schunute, $Z_1 = {\beta_4}^{(\beta_2)} - {\beta_3}^{(\beta_2)}, Z_2 = {\beta_3}^{(\beta_2 + Z_1)}$,

$Z_3 = (1 - e(-\beta_1(X - X_1)/(1 - e(-\beta_1(X_2 - X_1))^{(\frac{1}{\beta_2})}$

$w_t = Z_1 Z_3$ (1.5)

Quadratic spline, $w_t = \beta_0 + \beta_1 t + \beta_2 t^2$ (1.6)

is in the form. Here w_t t. egg weight, width and length per week, $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4\, ve\, \beta_5$ constants defined for models, a and b; In the piecewise regression, the nodal points, e: 2.7182, t: represent the week (Mazi, 2021).

In the study, Coefficient of Determination, Mean Squares of Error, Durbin-Watson and AIC (Akaike Information Criteria) were taken into account in the comparison of the model fit of the equations used in modeling the curves of egg weight, width and length in partridges (Kazancı, 2019, Şengül et al., 2016). The values obtained for egg weight, width and length were averaged on a weekly basis, and modeling was performed on these average values.

RESULTS

When the curves of egg length, width and weight values in henna partridges are examined, the Mean Error Squares, Coefficient of Determination, Akaike Information Criteria and Durbin-Watson Autocorrelation values are examined, respectively, the Logistic, Richard and Gompertz models yield similar and good results for egg length, respectively. It has been concluded that the Quadratic Spline model has the best results in egg width, the Cubic Spline model has the best results, the worst model is the Schnute model, and the best model in egg weight values is the Cubik Spline model and the worst model is the Schnute model. Especially in Cubic Spline model, it should be considered that the goodness of fit of the model will increase with increasing the position and number of nodes. It should be noted that the Logistic, Richard, Gompertz and Cubic Spline models used in the modeling study give close values to each other and this situation may vary from population to population.

DISCUSSION AND CONCLUSION

As a result of the study, the Coefficient of Determination, Mean Error Squares, Durbin-Watson and AIC (Akaike Information Criteria) used in the comparison of the model goodness of fit of the equations used in modeling the curves of egg weight, width and length are given in Table 2, Table 3 and Table 4, respectively.

Table 2. Mean error squares of egg length curves, determination coefficient, akaike information criterion and durbin-watson autocorrelation values.

Model	HKO	R^2	AIC	DW
Richard	0,025	0,991	147,25	1,96
Logistic	0,025	0,991	198,96	2,01
Gompertz	0,024	0,989	148,57	2,11
Schnute	0,075	0,990	196,58	1,75
Cubic Spline	0,048	0,927	221,58	2,22
Quadratic Spline	0,026	0,942	312,56	1,51

Table 3. Mean error squares of egg width curves, determination coefficient, akaike information criterion and durbin-watson autocorrelation values.

Model	HKO	R^2	AIC	DW
Richard	0,014	0,902	145,74	1,78
Logistic	0,014	0,902	156,89	2,25
Gompertz	0,014	0,902	161,53	2,19
Schnute	0,023	0,858	184,59	1,23
Cubic Spline	0,015	0,927	43,7	2,03
Quadratic Spline	0,014	0,901	89,56	1,11

Table 4. Mean error squares of egg weight curves, determination coefficient, akaike information criterion and durbin-watson autocorrelation values.

Model	HKO	R^2	AIC	DW
Richard	0,054	0,998	89,56	1,91
Logistic	0,051	0,991	91,57	2,25
Gompertz	0,052	0,998	102,25	2,31
Schnute	0,049	0,991	125,56	1,45
Cubic Spline	0,010	0,993	42,69	2,09
Quadratic Spline	0,044	0,958	115,61	1,67

When the Mean Error Squares, Coefficient of Determination, Akaike Information Criteria and Durbin-Watson Autocorrelation values of the egg length curves in Table 2 are examined, it is seen that the Logistic, Richard and Gompertz models give more or less similar results. It is seen that the Quadratic Spline model has the worst results in terms of coefficient of determination and AIC values. When the values of the egg width in Table 3 are examined, it is seen that the Cubic Spline model is the best model in terms of Mean Error Squares, Coefficient of Determination, Akaike Information Criteria and Durbin-Watson Autocorrelation values, while the worst model is the Schnute model. On the other hand, when the Mean Error Squares, Coefficient of Determination, Akaike Information Criteria and Durbin-Watson Autocorrelation values of egg weight values in Table 4 are examined, it is seen that the best model belongs to the Cubic Spline model, the worst model is the Schnute model, and Richard, Logistic and Gompertz. It is seen that the models give very close values to each other. In Figure 1, Figure 2 and Figure 3, the curves obtained for all models are given for egg length, width, and weight. The visual results obtained in Figure 1, Figure 2 and Figure 3 support the findings obtained in Table 2, Table 3 and Table 4.In the modeling of egg width and egg weights, it is seen that the best model is Cubik Spline and the worst model is Schnute model. On the other hand, Logistic, Richard and Gompertz models give close values to each other and to the best models in terms of all three features. It should not be forgotten that the goodness of fit of the model will increase further with increasing the number of nodes, especially in the Cubic Spline model. It is obvious that the curves obtained from the egg-related measurements will show slight differences depending on the population and environmental-care conditions, although they do not go out of the general pattern. For this reason, the use of more than one model and appropriate comparison criteria is of great importance in modeling studies.

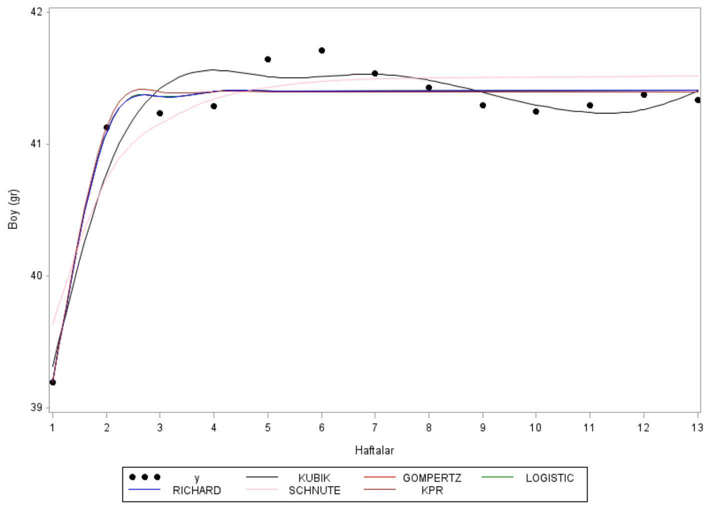


Figure 1. The curves of the 13-week egg length values.

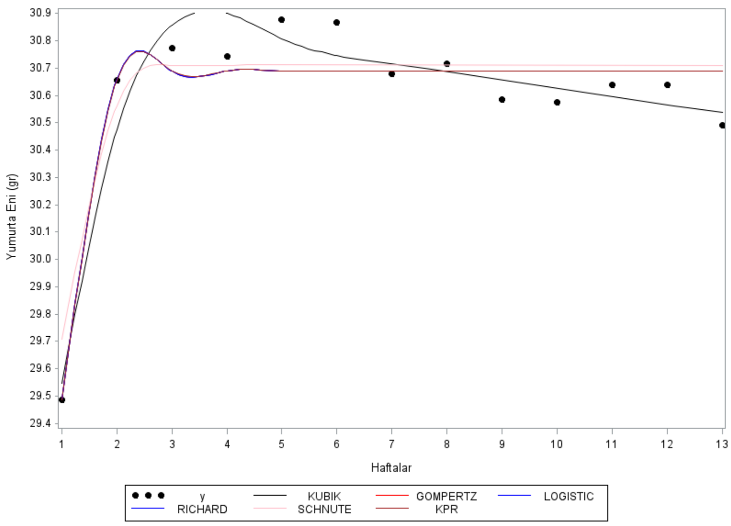


Figure 2. 13-week egg entry curves.

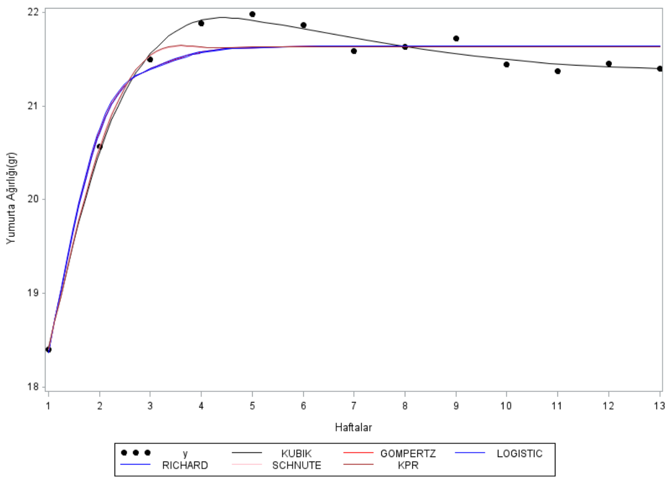


Figure 3. The curves of the 13-week egg weight values.

References

Balkan M, Karataş R, 2007. Some Characteristics of Quail (coturnix coturnix japonica) Eggs. Journal of DU Ziya Gökalp Education Faculty (8):112-119.

Cason JA, 1990. Comparison of linear and curvilinear decreasing terms in logistics flock egg production models, Poultry Science,(69):1467-1470.

Kazancı ON, 2019. Comparative Analysis of Some Models Used in Individual Growth Curves of Japanese Quail, Master Thesis, Kahramanmaraş Sütçü İmam University, Institute of Science, Kahramanmaraş, Turkey.

Mazı A, 2021. Modeling of Egg Yield and Weight in Japanese Quail, Master Thesis, Kahramanmaraş Sütçü İmam University, Institute of Science, Kahramanmaraş, Turkey.

Miyoshi S, Luc MK, Kuchida K, Mitsumoto T, 1996. Application of Nonlinear Models to Egg Production Curves in Chickens, Jpn. Poultry Science, (33):178-184.

Narushi VG, Nickname C, 2003. Sigmoid Model for the Evaluation of Growth and Production Curves in Laying Hens, Biosystems Engineering, (84):343-348.

SAS Institute, 2011. SAS/STAT User Guide, Version 9.3 edition. SAS Institute Inc.

Savegnago RP, Nunes BN, Caetano SL, Ferraduo AS, Schmidt GS, Ledur MC, Munari DP 2011. Comparison of Logistics and Neural Network Models to Fit to the Egg Production Curve of White Leghorn Hens, Poultry Science, (90):705- 711.

Şengül T, Çelik Ş, İnci H, Söğüt B, Şengül A, 2016. Investigation of Egg Production in Henna Partridges with Some Nonlinear Models, YYU Tar. know Journal, 26(1):33-39.

Author Contributions

The authors declare that they have contributed equally to the article.

Conflict of Interest

The authors of the article declare that there is no conflict of interest between them.

Modeling Egg Curves in Partridges

Tolga Tolun[1*], Esra Yavuz[2], Mustafa Şahin[3], İsmail Gök[1]

[1] Kahramanmaraş Sütçü İmam University, Faculty of Agriculture, Bioengineering and Sciences, 46000, Kahramanmaraş, Turkey.
[2] Şırnak University, Çizre Melek High School, Accounting and Tax Practices, 73200, Şirnak, Turkey.
[3] Kahramanmaraş Sütçü İmam University, Department of Agriculture, Agricultural Biotechnology, 46000, Kahramanmaraş, Turkey.

*Corresponding author e-mail: tolga_torun@hotmail.com

Abstract

In this study, a modeling study was carried out on the weekly average values of henna partridge egg production. For this purpose, weekly average of henna partridge egg yields of 148 days from the herd consisting of 320 breeders were taken and cubic spline, Gompertz, Logistic, Richard, Schunute, Quadratic Spline, McNally and Gamma equations were used in the modeling. In the comparison of the goodness of fit of the equations, The Coefficient of Determination, Mean Square Error, Durbin-Watson and AIC (Akaike Information Criteria) values were taken into account. As a result of the study, Logistic (HKO: 12.4, R^2: 0.994, AIC: 43.56, DW: 2.09), Cubic Spline (HKO: 10.56, R^2: 0.996, AIC: 46.55, DW: 1.95) and McNally (HKO: 11.02, R^2: 0.996, AIC: 48.67, DW: 2.11) models were found to have the best results with similar results. It was concluded that the Schnute (HKO: 11.24, R^2: 0.990, AIC: 136.51, DW: 0.49), Gamma (HKO: 24.67, R^2: 0.991, AIC: 69.89, DW: 2.95) and Quadratic Spline (HKO: 10.43, R^2: 0.946, AIC: 149.34, DW: 2.97) models had the worst results.

Key words: *Partridge, Egg Production, Curve Modeling.*

INTRODUCTION

Henna partridges are extremely important both in terms of economy and in terms of controlling harmful species due to the fact that they are materials for hunting tourism and the pests (mainly ticks) they consume in nature. Partridges produced in the partridge production centers of the Ministry of Agriculture and Forestry are released both to nature and to private hunting grounds. For this reason, it is extremely important to obtain quality eggs in terms of maximum and incubation yield in one season. This will be possible by establishing breeding flocks that are superior in terms of egg production.

As in laying hens, turkeys and Japanese quails, egg production in partridges is a complex cycle influenced by genotype and environmental factors. The effects of race, season, age, hatching conditions, moult, feeding style and other environmental factors on egg production in poultry are too high to be ignored. Although these factors are different from population to population, the curve they will form will show a similar distribution. The main purpose of modeling egg yield curves is to estimate the egg yield of the flock in the early stages, and to use it for selection and creation of breeding flocks. The aim is to create a flock with maximum productivity in terms of egg production. Of course, in the selection to be made on the basis of these models, the decision-making and process will be shortened. As a result, the degree of accuracy in selection will also increase. In particular, the approach of modeling individual egg production will increase the probability of selecting individuals with high genetic capacity on a flock basis. However, it is not considered possible in this application, especially in henna partridge. Because in practice the ratio of male to female is usually 1:4 (1 male: 4 females) and it is difficult to determine the individual eggs of females kept in the same compartment. As the number of females per male increases, the fertility rate decreases (Alkan et al., 2008). On the other hand, the ratio of 1:1 is not preferred as it will cause the necessity of feeding more male individuals. For this reason, egg curves are mostly obtained and modeled over cage or flock averages.

MATERIAL AND METHODS

Material

In this study, 148 days (December 17-13 May) henna partridge egg yields were obtained from the Kapiçam Henna Partridge Production Center, which is affiliated to the Republic of Turkey Ministry of Agriculture and Forestry, Kahramanmaraş Nature Conservation and National Parks Branch Office. Daily egg numbers of 320 breeders in 2021 were obtained and a modeling study was carried out on the weekly average values of egg yields corresponding to 20 weeks. The arithmetic mean, standard deviation, maximum and minimum values for 20-week-old egg yields are given in Table 1**.**

Table 1. Descriptive statistics of egg yield values.

Week	Number of eggs			
	$\bar{X}$	S	Mak	Min
1	15,762	6,934	8	26
2	27,042	1,676	29	21
3	26,619	1,471	20	31
4	26,952	2,344	19	33
5	29,905	2,884	26	35
6	44,233	3,244	39	52
7	50,952	1,595	43	57
8	54,857	1,986	49	29
9	56,095	0,474	51	19
10	61,762	3,372	59	71
11	54,143	2,181	51	69
12	51,667	1,113	47	61
13	52,048	1,844	49	65

Methods

In this study, segmented Cubic Spline (two-knotted), Gompertz, Logistic, Richard, Schunute, Quadratic Spline, McNally and Gamma models were used to model the egg production of henna partridges. The curve and model parameter estimations of egg yields were made in the SAS package program (SAS, 2011).

Equations of these models (1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7 and 1.8) and their expansions are as follows.

Cubic Spline, $w_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (t-\alpha)^3 + \beta_5 (t-b)^3$ (1.1)

Logistic, $w_t = \beta_0/(1 + \beta_1 e(-\beta_2 \mathrm{t}))$ (1.2)

Gompertz, $w_t = \beta_0 e(-\beta_2 e(-\beta_3 t)$ (1.3)

Richard, $w_t = 1/(\beta_0 + \beta_1 e(\beta_2 t)^{(-\beta_3)}$ (1.4)

Schunute, $Z_1 = {\beta_4}^{(\beta_2)} - {\beta_3}^{(\beta_2)}, Z_2 = {\beta_3}^{(\beta_2 + Z_1)},$

$$Z_3 = (1 - e(-\beta_1(X - X_1)/(1 - e(-\beta_1(X_2 - X_1))^{(\frac{1}{\beta_2})}$$

$$w_t = Z_1 Z_3 \tag{1.5}$$

Quadratic spline, $$w_t = \beta_0 + \beta_1 t + \beta_2 t^2 \tag{1.6}$$

McNally, $$w_t = \beta_0 t^{(\beta_1)} e(-\beta_2 t + \beta_3 t^{0.5} \tag{1.7}$$

Gamma, $$w_t = \beta_0 t^{(\beta_1)} e(-\beta_2 t) \tag{1.8}$$

is in the form. Here w_t: t. number of eggs per week $\beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4\ ve\ \beta_5$ constants defined for models, a and b; In the piecewise regression, the nodal points, e: 2.7182, t: represent the week (Mazi, 2021 and Yavuz et al., 2019).

Coefficient of Determination, Mean Error Squares, Durbin-Watson and AIC (Akaike Information Criteria) were taken into account in the comparison of the goodness of fit of the equations used in the modeling (Kazancı, 2019, Şengül et al., 2016¸ Narinç et al., 2014). The average of the obtained values of egg production on a weekly basis was taken and modeling was done on the average values of 20 weeks.

RESULTS

In this study, in which the egg production curves of henna partridges were modeled, it was concluded that the Logistic, Cubic Spline and McNally models were similar and gave the best values in terms of comparison criteria. It was determined that Schnute, Gamma and Quadratic Spline models had the worst results in terms of goodness of fit. In the literature, it is seen that different models are applied in the modeling of egg yield curves in layer hens, quails and henna partridges. Considering that many environmental factors affect egg production such as breed, age, temperature, humidity, feeding style, etc., although there are small changes in the standard egg yield curve in each egg production period, it is necessary to give priority to the models obtained from this and similar studies and to have more than one model in the modeling. The use of a model is extremely important in terms of breeding. On the other hand, it must be taken into account that the determined model can be interpreted as biologically meaningful.

DISCUSSION AND CONCLUSION

As a result of the study, the Coefficient of Determination, Mean Error Squares, Durbin-Watson and AIC (Akaike Information Criteria) used in the comparison of the model fit of the equations used in the modeling of the egg production curve are given in Table 2. In Figure 1, the point distribution of egg yields of 320 breeder henna partridges at 148 days, and in Figure 2, the curves created for 20-week egg yields are given.

Table 2. Mean Error Squares, Coefficient of Determination, Akaike Information Criteria and Durbin-Watson Autocorrelation Values of Egg Yield Curves.

Model	HKO	R^2	AIC	DW
Richard	16,1	0,994	89,71	2,99
Logistic	12,4	0,995	43,56	2,09
Gompertz	13,9	0,994	71,52	2,13
Schnute	11,24	0,990	136,51	0,49
Cubic Spline	10,56	0,996	46,55	1,95
Quadratic Spline	10,43	0,946	149,34	2,97
McNally	11,02	0,996	48,67	2,11
Gamma	24,67	0,991	169,89	2,95

When the Mean Error Squares, Coefficient of Determination, Akaike Information Criteria and Durbin-Watson Autocorrelation values of the egg yield curves in Table 2 are examined, it is seen that the Logistic, Cubic Spline and McNally models give similar results. All three models have the best results compared to other models. It is seen that Schnute, Gamma and Quadratic Spline models have the worst results. It is seen that especially the AIC values of these three models are quite high compared to the other models, and the DW values are quite close to the negative or positive autocorrelation limit values. When the curves created for all models in Figure 2 are examined, it supports the findings obtained.

Şengül et al., (2016), non-linear Gamma, McNally, Modified Compartmental and Adams-Bell models were used to model egg production in henna partridges and it was concluded that McNally model is the best model to describe egg production in partridges, while AdamsBell model is the least descriptive model. they have put. Yalçınöz and Şahin (2020), two different Cubic Piecewise Regression (two and three node), Logistics, MMF, Gamma, McNally, Modified Compartmental and Quadratic Piecewise Regression models used in modeling egg production curves in layer hens, and the best results are from the Modified Compartmental model. They reported that Cubic Piecewise Regression (two and three nodes), Logistics, MMF and McNally models gave very close values to the Modified Compartmental model in terms of model comparison criteria. Thuja (2021) used Cubic, Gompertz, Logistics, Gamma, Richard, Schunute, Quadratic Spline and McNeally equations to model the average egg yield curves in Japanese quails and reported that the best results in average egg yield were obtained from the Bifid Cubic Piecewise Regression model. The results obtained from this study Sengul et al. (2016), Yalçınöz and Şahin (2020), Üçkardeş and Narinç (2014) and Mazı (2021).

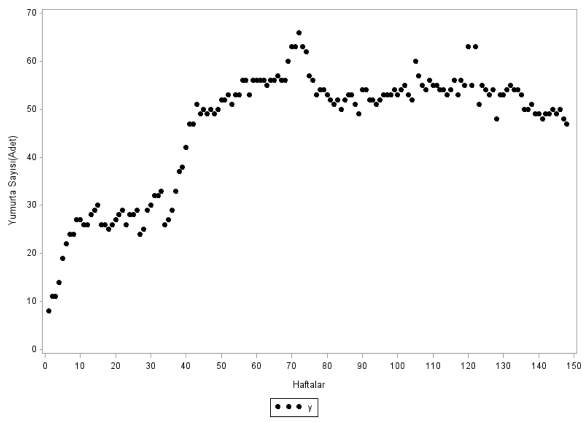


Figure 1. Point distribution of egg yields for 320 breeding henna partridges for 148 days.

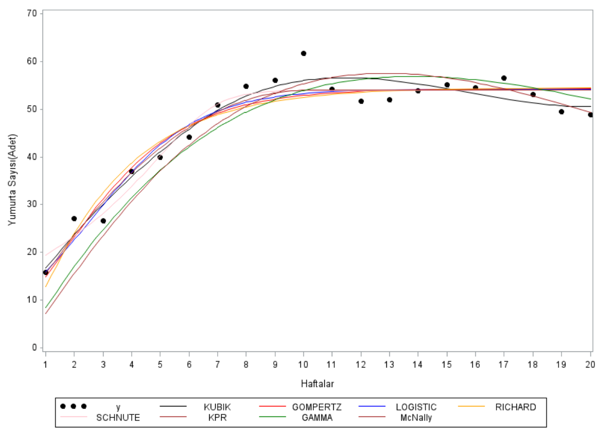


Figure 2. Curves created for 20-week egg yields.

References

Alkan S, Karabağ K, Galiç A, Balcıoğlu MS, 2008. Effects of Male-Female Ratio on Hatching Traits in Henna Partridges (alectoris chukar). Lalahan Hay. Journal of Research Institute 48(1): 45 –50.

Kazancı ON, Comparative Analysis of Some Models Used in Individual Growth Curves of Japanese Quail, Master Thesis, Kahramanmaraş Sütçü İmam University, Institute of Science, Kahramanmaraş, Turkey.

Mazı A, 2021. Modeling of Egg Yield and Weight in Japanese Quails, Master Thesis, Kahramanmaraş Sütçü İmam University, Institute of Science, Kahramanmaraş, Turkey.

Narinç D, Üçkardeş F, Aslan E, 2014. Egg production curve analyzes in poultry science, World's Poultry Science Journal, 70(04): 817-828.

SAS Institute, 2011. SAS/STAT User Guide, Version 9.3 edition. SAS Institute Inc.

Şengül T, Çelik Ş, İnci H, Söğüt B, Şengül A, 2016. Investigation of Egg Production in Henna Partridges with Some Nonlinear Models, YYU Tar. know Journal, 26(1): 33-39.

Uckardes F, Narinç D, 2014. An application of modified Logistic and Gompertz growth models in Japanese quail, Indian Journal of Animal Sciences, 84(8):903-907.

Yalçınöz E, Şahin M, 2020. Modeling of Egg Yield Curves in Layer Chickens, KSU Journal of Agriculture and Nature, 23(5): 1373-1378.

Yavuz E, Önem AB, Fahrettin K, Çanga D, Şahin M, 2018. Modeling of Individual Growth Curves in Japanese Quails, Black Sea Journal of Engineering and Science, 2(1):11-15.

Author Contributions

The authors declare that they have contributed equally to the article.

Conflict of Interest

The authors of the article declare that there is no conflict of interest between them.

Erlang Strength Against to Weibull Stress

Mine Doğan[1*], **Mehmet Gürcan**[1]

[1] Fırat University, Faculty of Science, Statistics, 23000, Elazığ, Turkey

*Corresponding author e-mail: mine.dogan@firat.edu.tr

Abstract

In reliability analysis, the stress-strength model is often used to describe the life of a component which has a random strength (X) and is subjected to a random stress (Y). The stress–strength models have been widely used for reliability design of systems. This model defines the reliability R of a component as the probability that strength of a component (X) is greater than the stress (Y) imposed on it, that is, R=P(X>Y). In this paper, a general reliability model R=P(X>Y) for the stress- strength problem has been discussed. The model has been illustrated by considering the stress to follow Weibull distribution and the strength to follow the Erlang distribution. First, we estimated parameters of Weibull distribution. The maximum likelihood estimator (MLE), stress–strength reliability is obtained when the shape parameter is known. Then, we estimate the stress-strength parameters and R using maximum likelihood. Finally, MATLAB programming was used to obtain both simulation and estimation parameters. A numerical estimate of the reliability under different values of the distribution parameters (for both the strength and stress distributions) has been presented and shown in the table.

Key words: *Stress-Strength Model, Reliability, Maximum Likelihood Estimation, Weilbull Distribution, Erlang Distribution.*

Covid 19 Salgını Öncesinde ve Salgın Sürecinde Üniversite Öğrencilerinin Başarı Durumlarının Değerlendirilmesi

Burçin Öner[1*], **Yüksel Öner**[2]

[1] Bilecik Şeyh Edebali Üniversitesi, Bozüyük Meslek Yüksek Okulu, Finans, Bankacılık ve Sigortacılık Bölümü, 11300, Bilecik, Türkiye
[2] Ondokuz Mayıs Üniversitesi, Fen - Edebiyat Fakültesi, İstatistik Bölümü, 55139, Samsun Türkiye

*Corresponding author e-mail: burcin.oner@bilecik.edu.tr

Özet

Bütün dünyada olduğu gibi ülkemizde de her yönüyle yoğun etkilerini hissettiğimiz Covid 19 salgını, Yükseköğretim Kurulu'nun almış olduğu kararlar kapsamında yükseköğretimde eğitim öğretimin yürütülmesinde farklı uygulamalara geçilmesine yol açmıştır. Yükseköğretim kurumlarında 2019-2020 Eğitim-Öğretim yılı Güz Yarıyılı dâhil olmak üzere bu döneme kadar yüz yüze eğitim uygulanmaktaydı. Covid 19 salgının yaygınlaşması ve salgının etkilerinin ağırlaşması üzerine üniversiteler, 2019-2020 Eğitim-Öğretin yılı Bahar Yarıyılı itibariyle hemen hemen bütün ülkelerde olduğu gibi ülkemizde de uzaktan eğitim sürecine geçilmiş ve bu süreç 2020-2021 Eğitim-Öğretim yılı boyunca da devam etmiştir. 2021-2022 Eğitim-Öğretim yılı itibariyle ise hibrit sistem (derslerin veya herhangi bir dersin %40'ı uzaktan eğitim ve %60'ı yüz yüze eğitim) uygulanmaya başlanmıştır. Bu çalışmada Yükseköğretim Kuruluna bağlı iki devlet üniversitesine 2018-2019 ve 2019-2020 Eğitim Öğretim yıllarında kayıt yaptıran Meslek Yüksek Okulu öğrencilerinin salgın öncesine ve salgın sürecine ait başarı durumları öğrenci transkriptleri dikkate alınarak değerlendirilmesi amaçlanmıştır. Bu amaçla oluşturulan örnekleme planlarına göre söz konusu üniversitelerden elde edilen öğrenci başarı durum belgelerinden oluşturulan veriler, parametrik ve parametrik olmayan istatistiksel analiz teknikleri kullanılarak analiz edilmiştir. Elde edilen bulgulara dayanılarak öğrencilerin başarı durumları hem söz konusu üniversitelerin kendi içerisinde hem de üniversiteler arasında yüz yüze eğitim ve uzaktan eğitim yönüyle karşılaştırmalı olarak değerlendirilmiştir.

Key words: *Yüz Yüze Eğitim, Uzaktan Eğitim, İstatistiksel Analiz, Öğrenci Transkripti*

Exchange Rate Forecasting with Artificial Intelligence

Lorena Saliaj[1*], **Albina Toçilla**[2]

[1] Università degli Studi "G. D'Annunzio", Dipartimento di Economia, 65127, Pescara, Italy
[2] Università degli Studi di Ferrara, Dipartimento di Architettura, 44121, Ferrara, Italy

*Corresponding author e-mail: lorena.saliaj@unich.it

Abstract

This study concerns the problem of forecasting the exchange rate between the official currency of EU member states, Euro and Albanian Lek, aiming to identify the best predictive model for financial time series future trend prediction. We compare the forecasting performance of linear and nonlinear forecasting models using monthly data for the period between January 2002 until January 2022. We discuss various forecasting approaches, including an Autoregressive Integrated Moving Average model, a Nonlinear Autoregressive Neural Network model, a BATS model and Exponential Smoothing on the collected data and compare their accuracy using error term measuring indicators, choosing the model with the lowest Mean Absolute Percentage Error value. Finding the most accurate forecasting model would help improve monetary and fiscal politics, as well as orient future personal investments.

Key words: *NARNN, ARIMA, Artificial Intelligence, Time Series Forecasting*

Different Method for Appraisal on Functional Data Analysis

Tuba Şekerci[1*], **Mehmet Gürcan**[2]

[1]Firat University, Vocational school of social sciences, Accounting, 23000, Elazığ, Turkey
[2] Firat University, Faculty of science, Statistics, 23000, Elazığ, Turkey

*Corresponding author e-mail: tsekerci@firat.edu.tr

Abstract

*The data obtained when repeated measurements depend on time are commonly referred to as "longitudinal data" in the literature. The main purpose in the analysis of longitudinal data is to examine the difference between the trials (group factor), the change over time (time factor), and whether the change over time is similar in the groups (group*time interaction). If the functional structure of the distribution of the population from which the data were taken is unknown, it is appropriate to use non-parametric approaches in the estimation of the probability density function. One of the methods used in this context is the kernel estimation method. The kernel estimation of the probability density function was first proposed by Rosenblatt (1956) and its statistical properties were studied in detail by Parzen (1962). Kernel estimation methods are frequently preferred methods. Functional data analysis involves infinite-dimensional processes and/or data and shows how L-spline theory can support linear modeling generalizations and principal component analysis from random functions to samples. The most common type of data suitable for analysis with functional data analysis is time series that do not have longitudinal observations or equally spaced observations. Finally, using the matlab program, intermediate decimation values were obtained with the help of some functions.*

Key words: *Functional Data Analysis, Regression Analysis, Kernel Estimators, Longitudinal Data.*

Three-factor experimental design as a tool in Applied Statistics

Vesna Antoska Knights[1], Jetmira Millaku[2*]

[1] University "St Kliment Ohridski"- Bitola, Faculty or Technology and Technical Science -Veles, 7000, Bitola, Republic of North Macedonia
[2] "Ss. Cyril and Methodius University in Skopje" Faculty of Philosophy- Department of Psychology. Goce Delchev Blvd. 9A. 1000 Skopje. Republic of North Macedonia

*Corresponding author e-mail: vesna.knights@uklo.edu.mk

Abstract

The purpose of this research is the implementation of a Three-factor experimental design as a tool in Applied Statistics. In this case, it is used as the optimal definition of indicators of the effectiveness of meditation training and yoga exercise training in reducing stress levels and anxiety in students. Using this three-factor model three parameters were monitored: blood pressure, duration of sleeping, and smoking. This system of indicators is called a three-factor experimental design with two levels of variation (2^3) for the process of stress and anxiety. A way to obtain simplified models is using full factorial designs, which are important means to evaluate the influence of the factors on response. All 30 students have been randomly selected, in the three-factor experiment, and analyzed, and in each case, regardless of gender, the effect of the factors. The result obtained from these analyses using the three-factor experiment were very similar for all the respondents. By applying this model, it will show which has the greatest effect from the three factors blood pressure, duration of sleeping, and smoking, highest influence has a coefficient of factor smoking, then is hours of sleeping on response y (stress and anxiety). This model is a systematic method to determine the relationship between factors affecting a process and the output of that process. In other words, it is used to find cause-and-effect relationships. This information is needed to manage process inputs to optimize the output.

Key words: *Three-factor experimental design, Applied Statistics, Stress levels and anxiety*

INTRODUCTION

Statistical design of experiments refers to the process of planning the experiment so that appropriate data will be collected and analyzed by statistical methods, resulting in valid and objective conclusions. The statistical approach to experimental design is necessary if we wish to draw meaningful conclusions from the data. When the problem involves data that aresubject to experimental errors, statistical methods are the only objective approach to analysis. Thus, there are two aspects to any experimental problem: the design of the experiment and the statistical analysis of the data. These two subjects are closely related because the method of analysis depends directly on the design employed. The three basic principles of experimental design are randomization, replication, and blocking. We add the factorial principle to these three. (Montgomery, 2013).

A full factorial designed experiment consists of all possible combinations of levels for all factors. The total number of experiments for studying k factors at 2-levels is 2^k. A way to obtain simplified models is using full factorial designs, which are important means to evaluate the influence of the factors on response. Design of experiments has lately been applied to various research areas.

The purpose of this research is the optimal definition of indicators of effectiveness of meditation training and yoga exercises training in reducing the stress levels and anxiety in students.

İn this paper is considered three-factor model with two levels of variation (2^3).

This model is a systematic method to determine the relationship between factors affecting a process and the output of that process. In other words, it is used to find cause-and-effect relationships. This information is needed to manage process inputs to optimise the output. (Knights al., 2022)

Using this three-factor model three parameters were monitored: blood pressure, duration of sleeping, smoking. This system of indicators is called a three-factor experimental design with two levels of variation (2^3). This model is implemented for determing the indicators of stress levels and anxiety. A way to obtain simplified models is using full factorial designs, which are important means to evaluate the influence of the factors on response.

MATERIAL AND METHODS

Material

This system of indicators will be performed on 30 random chosen students from meditation group&yoga exercise group from Kosovo at the age from 18 to 30 old. Half of them are males and ohrer 15 are females. All 30 students attending inensive course of joga exercises.

Student who decides to be part of this research had duty to have diary of their sleeping hours and measuring their blood pressure at the morning and in the evening. Also, in diary they were taking notice how many cigaretes they were smoking per day.

Methods

Measure for stress and anxiety is taken for condudted Beck Anxiety Inventory (BAI) Questionare to the students every week in period of 8 weeks. BAI is a short list describing 21 anxiety symptomssuch (Beck et al., 1988). Internal consistency for the BAI = (Cronbach's α=0.92) Test-retest reliability (1 week) for the BAI = 0.75. (Creswell, 2008). The total score has a minimum of 0 and a maximum of 63. Score of 0 – 21 = low anxiety Score of 22 – 35 = moderate anxiety Score of 36 and above = potentially concerning levels of anxiety. (Vogt, 2005)

In this three-factor model the three parameters are considered as important to the operation: blood pressure (factor X1), duration of sleeping (factor X2), and smoking (factor X3). A way to obtain simplified models is using full factorial designs, which are important means to evaluate the influence of the factors on response. We want to ascertain the relative importance of each of these factors on response which is stress and anxiety (*Y*). This type of analysis will be performed in Excel in addition to using regression equations plus Cochran's test, Student t-test and Fisher test with SPSS.

RESULTS

At first for this experimental design with two levels of variation (2^3) we should determine the number of series i.e., various experiments with the relations: $N = n^k = 2^3 = 8$

In the table 1 is given experimental planning matrix and results of the factorial experiment. The data for the first experiment is taken at random from one of the 30 monitored students, in period of 8 weeks (number of series.) For example, student No1 is chosen from meditation group&yoga exercise group. He/she has beginning stress level 30,(using the results form BAI Questionnaire) and after an eight-weeks a reduction of 12 was achieved or the total 18 units.

Table 1. Matrix plan of the experiment for student No1.

No	X_1	X_2	X_3	x_1 blood pressure (mmHg)	x_2 duration of sleeping (h/day)	x_3 smoking (cigarettes /day)	y_{i1}	y_{i2}	y_{sr}	S_j^2
1	1	1	1	120	8	5	30	28	29	2.00
2	-1	1	1	70	8	5	28	27	27.5	0.50
3	1	-1	1	120	6	5	27	26	26.5	0.50
4	-1	-1	1	70	6	5	26	24	25	2.00
5	1	1	-1	120	8	3	24	22	23	2.00
6	-1	1	-1	70	8	3	22	21	21.5	0.50
7	1	-1	-1	120	6	3	21	19	20	2.00
8	-1	-1	-1	70	6	3	19	17	18	2.00

For each series for each patient a combination of factors for high and low levels has been made, along with measurements of the questionnaire and we have two responses y_{i1} and y_{i2}. y_{sr} is an average of responses y_{i1} and y_{i2}. S^2j is the variance of each response (Trochim, 2022). The average value of all variances is $S_j^2 = 1{,}43$.

To determine that the order of dispersions is considered homogeneous, it is necessary to calculate Cochren's criterion. Cochrane's test is a statistical method for testing for differences between three or more matched sets of frequencies or proportions. Cochran's test is the traditional test for checking heterogeneity in meta-analyses. Using Cochren's criterion we are testing hypothesis for reproductive experiments.

$$G_p = \frac{max\ S_j^2}{\sum_{j=1}^{N} S_j^2} = 0{,}17$$

G_p is calculated Cochren for the experiments and it is present as a ratio of maximal value of variance which is $max\ S_j^2 = 15{,}86$ and the sum of all variances of dispersion $\sum_{j=1}^{N} S_j^2 = 1.43$

$G_{\alpha,f,N} = 0{,}6798$ is the value of Cochren, which have been read from table for *95%* confidence interval and degree of freedom f *=N(k-1)=8,* where N-is the number of experiments which is 8 and k=2 which is the number of levels of variation. If criteria $G_{p.} \geq G_{\alpha,f,N}$ is satisfied then statistical heterogeneity is determined, but if $G_{p.} \leq G_{\alpha,f,N}$, is presented then the order of variances of dispersions is considered as homogeneous. (Levine, 2014)

Fulfillment of the above condition indicates the reproducibility of the experiments, and the dispersion estimate is homogeneous.

After calculating the dispersion of the experiments, it is necessary to calculate the coefficients of regression: β0, β1, β2, β12. Some coefficients may be negligibly small or insignificant. The determination of the significance of the regression coefficients is done with the help of the Student test criterion. In order to determine whether they are significant or not, first of all, the variance in which they are determined should be assessed:

$$S_\beta = \sqrt{\frac{S_j^2}{N}} = \sqrt{\frac{1{,}43}{8}} = 0{,}42$$

The regression coefficients are significant if the condition is: $|\beta| \geq S_\beta * t$, where t is the calculated value of Student t test for all responses y_i in the table 1, and it is t = 2,36.

$$|\beta| \geq S_\beta t \geq 0{,}42 \cdot 2{,}36 \geq 1{,}0024$$

t_k = 2.31 is the critical value that has been read from the table for the Student criterion, for confidence interval p = 0.95 and degree of freedom $f = N(k-1) = 8(2-1) = 8$

In case the criterion, $|\beta| \geq S_\beta * t$, is not satisfied, then the regression coefficient is taken as insignificant, ie. non-influential coefficient in the regression equation or in other words will not be taken in calculations in formula.

Table 2 presents the Matrix plan of the experiment for determining the coefficient of regression according to the given equations:

$$\beta_i = \frac{1}{N}\sum_{j=1}^{N} x_{ij}y_i \, ; \quad \beta_{im} = \frac{1}{N}\sum_{j=1}^{N} x_{ij}x_{im}y_i \; (i \neq m)$$

Table 2. Matrix plan of the experiment for determining the coefficient of regression.

No	x0	x1	x2	x3	x12	x13	x23	x123	ysr
1	1	1	1	1	1	1	1	1	29
2	1	-1	1	1	-1	-1	1	-1	27.5
3	1	1	-1	1	-1	1	-1	-1	26.5
4	1	-1	-1	1	1	-1	-1	1	25
5	1	1	1	-1	1	-1	-1	-1	23
6	1	-1	1	-1	-1	1	-1	1	21.5
7	1	1	-1	-1	-1	-1	1	1	20
8	1	-1	-1	-1	1	1	1	-1	18

The linear three-factor model is applied is given by the equation below:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + \; b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3$$

where y is factor of stress, for the corresponding measurements yi1 and yi1, the factors x_1, x_2, x_3, their units are given in the table 1, while b_i is represent coefficients of regression and b_{ij} coefficient of interaction between factors. The results of the calculated values of the coefficients are given in Table 4.

Table 3. Calculated values for regression coefficients

bo	b1	b2	b3	b12	b13	b23	b123
23.8125	0.8125	1.4375	3.1875	-0.0625	-0.0625	-0.1875	0.0625

From the above condition in equation (3), the coefficients that do not satisfy the condition are neglected and the model equation (1) and new equation with coded variables will be:

$$y = 23.8125 + 0.8125\, x_1 + 1.4375x_2 + 3.1875x_3$$

The adequacy of the obtained equation has been checked by using the Fisher criterion. The equation is adequate if the condition is met:

$$F_p \leq F_t \qquad F_p = \frac{S_{ad}^2}{S_j^2} \qquad S_{ad}^2 = \frac{\sum_{j=1}^{N}\left(\underline{y_j} - y_j\right)^2}{N-k-1}$$

In this equation yj denotes the response values calculated according to the regression equation.

These are the response values obtained from the model by substituting the corresponding values of the variables, +1 and -1, i.e. with their values when at lower and upper levels, respectively. The values obtained from the ysr experiment with the values obtained from the model y are given in the following table:

Table 4. Calculations to determine compliance of the Fisher criterion

No	y_{sr}	y	y_{sr} - y	$(y_{sr} - y)^2$
1	29	28.4375	0.56	0.316
2	27.5	28.4375	-0.94	0.879
3	26.5	25.5625	0.94	0.879
4	25	25.5625	-0.56	0.316
5	23	22.0625	0.94	0.879
6	21.5	22.0625	-0.56	0.316
7	20	19.1875	0.81	0.660
8	18	19.1875	-1.19	1.410

The sum of all values in the last column is determined: $\sum_{j=1}^{N}\left(\underline{y}_i - y_j\right)^2 = 5{,}56$, for the number of degrees of freedom $f = N - k - 1 = 8 - 2 - 1 = 5$

So, we can determine variance of adequacy in equation: $S_{ad}^2 = \frac{1{,}13}{5} = 1{,}13$, where from table above, already is calculated an average of all variances is $S_j^2 = 1{,}43$, then is calculated Fisher criteria form equation: $F_p = \frac{1{,}13}{1{,}43} = 0{,}78$

F_t = 3.69 is critical value which is read from the table for the Fisher criteria, for confidence interval p = 0.95 and degree of freedom $f = 5$. If criteria $F_p \leq F_t$ is satisfied, it is concluded that **model is adequate**.
The next step is to determine the model in natural units, to get the response in real units it is necessary to convert the coded variables into natural variables. For that purpose, into the regression equation, we involve the following relation:

For each of the three factors, an appropriate replacement of the coded factors (small x) to natural (capital X) was performed according to the following relations:

$$x_1 = \frac{X_1 - 95}{25}; \quad x_2 = \frac{X_2 - 7}{1}; \quad x_3 = \frac{X_3 - 4}{1}$$

where: X - natural variable (factor), X_{i0} - mean level of natural variable, ΔX - the interval of change of x (standard deviation), x - coded variable.

We need these coded values to be able to determine the sign before the coefficients (b_i).

Table 5. Variables need for the conversion relation

Factor	Low level	high level	ΔX Standard deviation	X_{i0} mean level (average)
x_1	70	120	25	95
x_2	6	8	1	7
x_3	3	5	1	4

In the equation with coded factors, a substitution is made and an equation in natural units is obtained:

$$y = 23{,}8125 + 0.8125\frac{X_1 - 95}{25} + 1{,}4375\frac{X_2 - 7}{1} + 3{,}1875\frac{X_3 - 4}{1}$$

The equation in natural units has the following form:

$$y = 23.8 + 0.8X_1 + 1{,}437X_2 + 3{,}19X_3$$

As we can see from equation (11), in this case the highest influence has coefficient of factor X3 (smoking), then is X2(hours of sleeping) on response y, which is stress and anxiety.

The process has been repeated for all other random selected students. The result obtained from these analyses using the three-factor experiment was the similar almayst same for all of respondents.

CONCLUSION

The purpose of this research is the implementation of a Three-factor experimental design as a tool in Applied Statistics. This tool helps:

- To determine whether a factor or a collection of factors, has an effect on the response.

  - In our case, among the three factors the greatest effect was blood pressure after the duration of sleeping, and blood pressure almost had no effect on the response.

- To model the behavior of the response as a function of the factors.

  - In detail was present in part of Results.

- To optimize the response.

  -If students want to reduce their level of stress and anxiety, they have to reduce smoking or give up on smoking and have quality sleep.

References

Beck AT, Epstein N, Brown G, Steer RA: 1988. An inventory for measuring clinicalanxiety: psychometric properties. Journal of Consulting and ClinicalPsychology. 56:893-897

Beyer, W. H. 2002. CRC Standard Mathematical Tables, 31st ed. Boca Raton, FL: CRC Press, pp. 536 and 571.

Creswell, J.W. 2008. Educational Research: Planning, Conducting, and Evaluating, Quantitative and Qualitative Research. Pearson Education, Inc, London, United Kingdom

Douglas C. Montgomery, 2013. Design and Analysis of Experiments, John Wiley and Sons, New York, ISBN 978-1118-14692-7

Gordana Markovik, Vesna Knights, 2022. Model of optimisation of the sustainable diet indicators. Journal of Hygienic Engineering and Design, In print.

Gordana Markovik, Vesna Knights, Daniela Nikolovska Nedelkovska, Dragan Damjanovski, 2020. Statistical analysis of results in patients applying the sustainable diet indicators. Journal of Hygienic Engineering and Design, Vol. 30, pp. 35-39, 2020.

Levine, D. 2014. Even You Can Learn Statistics and Analytics: An Easy to Understand Guide to Statistics and Analytics 3rd Edition. Pearson FT Press

Prof William M.K. Trochim 2022, Research Methods Knowledge Base. Access address: https://conjointly.com/kb/factorial-design-variations; Date of access: 12/05/2022.

Salkind, N. 2016. Statistics for People Who (Think They) Hate Statistics: Using Microsoft Excel 4th Edition.

Vogt, W.P. 2005. Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences. SAGE.

Implementation of Data Mining for the Purpose to Analyze Student's Decisions After High School

Marija Prchkovska[1], Xhenete Aliji[1], Faton Ademi[1], Elham Nuishi[1], Erzan Murati[1], Majlinda Axhiu[1]

[1] Mother Teresa University, Faculty or Computer Science, Informatics, 1000,Skopje, Republic of North Macedonia

*Corresponding author e-mail: marija.prckovska@students.unt.edu.mk

Abstract

This paper presents a student project for the implementation of data mining. For this purpose, has been created questionnaire which intends to collect information from graduate students in high school about their decisions after finishing high school. The questionnaire consists of 13 questions, of which 9 are with completion of offered answers, and 4 with supplementation. We have a total of 723 respondents The survey was conducted in high schools in Skopje, Veles, Prilep, Resen, Bitola, Gostivar, and Tetovo. The data has been processed in the Anaconda package in Jupyter notebooks. The paper is presented dataset creation, Model selection, and algorithm, Data processing, and Making predictions.

Key words: *Data maining, Anaconda, Python, Data processing.*

A Comparative Study of Artificial Neural Network, Gaussian Process Regression, response Surface Methodology Modeling Approaches for Predicting Amount of Corn Bread Phytic Acid

Dilşad Yıldız Kaçar[1], **Hülya Bayrak**[2*]

[1] Turkish Patent and Trademark Office, IT Department, 06560, Ankara, Turkey
[2] Gazi University, Faculty of Science, Department of Statistic, 06500, Ankara, Turkey

*Corresponding author e-mail: hbayrak@gazi.edu.tr

Abstract

This study describes the comparison of the Phytic acid amount (mg/100g) predictive models based on two advance Machine Learning (ML) teqniques, Artificial Neural Network (ANN) and Gaussian process regression (GPR)-based approach, and the Response Surface Methodology (RSM). These three models were validated by comparing model simulations with observed values for unseen data. The models were compared based on performance indices like coefficient of determination, mean square error, rootmean square error (RMSE), model predictive error, mean average deviation. The effect of three variables, wheat flour addition rate (0,15 qnd 30%), Yeast amount (2,4 and 6%) and Fermantation time (60,90 ve 120 minutes) on Phytic acid amount (mg/100g), were summarized by a one replicate experiment in the Central Composite Design (CCD). The results showed that GPR-based models were found to be a better alternative to RSM and ANN.

Key words: *Artificial Neural Network (ANN), Gaussian process regression (GPR), Response Surface Methodology (RSM), Phytic Acid*

Using ROC Analysis with Propensity Score on PEG Feed Patients

Arzu Baygül Eden[1*], **Neslihan Gökmen İnan**[2]

[1] Koç University, School of Medicine, Biostatistics, 34010, İstanbul, Turkey
[2]Koç University, Engineering Faculty, Computer Engineering, 34450, İstanbul, Turkey

*Corresponding author e-mail: abaygul@ku.edu.tr

Abstract

Receiver operating characteristic (ROC) curves are widely used in medicine to evaluate continuous variables as diagnostic biomarkers. ROC analysis is one of the most extensive methods to evaluate diagnostic accuracy by taking into account sensitivity and specificity at all levels. The optimal cut-off value regarding sensitivity and specificity can be also obtained from ROC analysis. In ROC analysis, various methods have been proposed to reduce the effects of covariates. In this study, we proposed a novel approach to obtain higher performance of diagnostics compared to the classical approach. We used the PEG feed patient dataset. To determine the diagnostic performance of Transferrin alone, ROC analysis is performed. Furthermore, we implemented logistic regression analysis with MNA which is another potential risk factor to predict mortality. The results reveal that the AUC (0.729) is higher when the predicted probability is used for ROC analysis. To validate these results, we performed simulations for different mortality and MNA prevalence, and n values. As a result of the simulation study, we obtained higher performance in ROC analysis by using this method. In conclusion, this reliable method can be used to increase the diagnostic performance of biomarkers.

Key words: *ROC analysis, Propensity score, Logistic regression*

On the Propagation of Uncertaininty in Parameter Estimation for Kinetic Models of Metabolism

Emrah Nikerel[1*]

[1] Yeditepe University, Department of Genetics and Bioengineering, 34755, Istanbul, Turkey

*Corresponding author e-mail: emrah.nikerel@yeditepe.edu.tr

Abstract

Kinetic models for biological systems are useful in (among many others): (i) understanding metabolism, (ii) finding genetic engineering targets, (iii) exploratory simulations of alternative process scenarios (iv) functional analysis and (v) identify drug targets. Construction of such models encompasses, constructing mass balances, selecting appropriate kinetic rates and finally parametrizing the model using experimental data. As the data always contains uncertainty, a key challenge in constructing kinetic models is the estimation of parameter uncertainty, propagated from experimental error. This would yield not only better models that are generalizable, but also it will pave the way to the much-desired model-based-experimental design. This estimation is currently performed using Monte Carlo approach, where synthetic, typically white error is added to the available measurements and the parameter estimation is performed. This yields a distribution of estimated parameters, from which the uncertainty can be estimated. An alternative and interesting approach is the direct, possibly analytical, estimation of error propagation. One way to perform this is to use the "Delta method", a first-order Taylor series transformation, approximating the variance of any parameter (γ) as a (potentially non-linear) function of some random variables $(x_1,x_2,\ldots,x_n)$*, for which the measurements are available each with its own estimate of variance as (omitting co-variance):*

$$var(\gamma)=\sum_{i=1}^{n} var(x_i)\cdot\left[\partial f/(\partial x_i)\right]^2$$

The talk will cover key fundamental concepts in parameter estimation for kinetic models of metabolism, in particular in estimating error propagation in these models. Examples will be presented for small scale black-box models of microbial fermentation, where the parameter uncertainty will be estimated using Monte-Carlo approach and the above-described Delta method. Extensions of this direct estimation for genome-scale linear models as well as limitations and potential applications of the direct estimation will be discussed.

Key words: *Error propagation, Dynamic models, Delta Method, Experimental error, Model observability*

Öznitelik Seçiminin Yazılım Maliyet Tahminine Etkisi ve Makine Öğrenmesi Tabanlı Yazılım Maliyet Tahmini Yapan Çalışmaların Karşılaştırmalı Analizi

Şükran Ebren Kara[1*], Rüya Şamlı[2]

[1]Şırnak Üniversitesi, Cizre Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, 73200, Şırnak, Türkiye
[2]İstanbul Üniversitesi-Cerrahpaşa, Bilgisayar Mühendisliği Bölümü, 34315, İstanbul, Türkiye

[*]Sorumlu Yazar e-mail: sukranebren@hotmail.com

Özet

Öznitelik seçimi, veri setine uygulanan modelin performansını olumsuz etkileyen ya da hiç etkisi olmayan özniteliklerin veri setinden çıkarılması işlemidir. Çoğu zaman öznitelik seçimi sonrasında elde edilen yeni veri kümeleri ile doğruluk oranı daha yüksek sınıflandırma ve tahmin sonuçları elde edilmektedir. Yazılım maliyet tahmini, bir geliştiricinin yazılım projesini geliştirmeye başladığı esnada ihtiyaç duyduğu yaklaşık süre ve kaynakların tahminidir. Yazılım maliyet tahmini, yazılım projesinin maliyetini belirlemek ve müşteriyi ikna etmek için yazılım geliştirme sürecindeki en önemli aşamalardan birisidir. Gerçek maliyete en yakın maliyet tahminini yapmak hem yazılım geliştiricileri hem de müşteriler için çok büyük bir önem arz etmektedir. Bu yüzden yazılım maliyet tahmini için çok farklı tahmin yöntemleri geliştirilmiştir. Makine Öğrenmesi tabanlı yapılan tahmin yöntemleri, bu yöntemlerden bazılarıdır. Yazılım maliyet tahmini için kullanılan veri setlerinde öznitelik seçiminin maliyet tahminine olan etkisini belirlemek için bu çalışmada veri setleri üzerinde farklı öznitelik seçim yöntemleri denenmiştir. Elde edilen yeni veri kümelerine Makine Öğrenmesi tahmin algoritmaları uygulanarak yazılım maliyet tahmini gerçekleştirilmiştir. Bu sayede kullanılan tahmin algoritmalarının ve kullanılan öznitelik seçim tekniğinin performansları incelenmiştir. Değerlendirme ölçütü olarak korelasyon katsayısı, hata oranları olarak MAPE (Mean Absolute Percentage Error – Ortalama Mutlak Hata Yüzdesi) ve MAE (Mean Absolute Error – Ortalama Mutlak Hata) kullanılmıştır. Elde edilen bulguların literatürdeki çalışmalar ile karşılaştırılması için Makine Öğrenmesi ile yazılım maliyet tahmini gerçekleştiren çalışmaların karşılaştırmalı bir analizi bu çalışmada sunulmuştur.

Key words: *Öznitelik Seçimi, Yazılım Maliyet Tahmini, Makine Öğrenmesi, MAE, MAPE*

Classification of Brain Tumors on MR Images Using Pixel-Based Segmentation and Machine Learning Methods

Neslihan Gökmen İnan[1]*, Arzu Baygül Eden[2], Ozan Kocadağlı[3]

[1] Koç University, Engineering Faculty, Computer Engineering, 34450, İstanbul, Turkey
[2] Koç University, School of Medicine, Biostatistics, 34010, İstanbul, Turkey
[3] Mimar Sinan Fine Arts University, Faculty of Science and Letters, Statistics, 34380, İstanbul, Turkey

*Corresponding author e-mail: ninan@ku.edu.tr

Abstract

Brain cancer is a life-changing severe disease having high prevalence [1]. One of the primary diagnostic methods in brain tumors is imaging. Image processing is an emerging research area used in diagnosis and treatment in medical applications by medical healthcare specialists. Specifically, image segmentation plays a major role in helping to extract suspicious areas from medical images. This study deals with segmentation of the brain Magnetic Resonance Imaging (MRI) for the detection of tumor localization. 82 tumorous and 72 non-tumorous MRI images are included in the study [2]. In this context, firstly the entropy-based segmentation was handled, and for further steps, shape features such as area, centroid, eccentricity, equiv diameter, major and minor axis length were extracted from binary segmented images and Artificial Neural Networks (ANNs), Decision Trees (DTs) were used to classify tumorous and non-tumorous images. The configurations of the models were selected considering some remarkable performance criteria such as sensitivity, specificity, and accuracy. The analysis results showed that the ANNs and DTs have superior performance to classify the tumorous and non-tumorous images. The test accuracy is found 100.0% with ANN (hidden layer activation function: hyperbolic tangent, output layer activation function: softmax) whereas test accuracy is found 97.7% for CHAID which is a DT technique. Consequently, the classification systems allow radiologists to reach more reliable diagnoses and provide them a decision support system.

Key words: *Image processing, Image Segmentation, Brain tumor, Artificial Intelligence, Classification methods*

Contibution on the Links Between Socio-Economic Inequalities and Inclusive Growth (Evidence from Mena Region)

Nahi Brahim[1*], Ritahi Oussama[1], Echaoui Abdellah[1]

[1] Mohamed V University of Rabat, Faculty of Legal, Economic and Social Sciences-Souissi, Morocco

*Corresponding author e-mail: brahinahi@yahoo.fr

Abstract

For more than a century, the relationship between inequality and economic growth has been the main concern of economists. Yet the nature of this relationship remains elusive. Ideally, if economic growth is assumed to benefit the poor on the same terms as the rich. However, the process of growth and wealth creation is always accompanied by dynamic changes in the distribution of its dividends. The question of wealth distribution is one of the most revisited and discussed issues in the development world today. Over the long term, how has it evolved? In the 19th century, Marx argued that the dynamics of private capital accumulation lead to the high concentration of wealth and power in the hands of a minority. What roles do the so-called balancing forces of growth, competition and technical progress actually play in reducing inequality and stabilising the phases of development as argued by Kuznets in the 20th century? Also, the classical theory approach which defines a positive correlation between inequality and growth (see, for example, Stiglitz, 1969). Such a positive correlation is due to the assumption that the savings rate is higher for the rich than it is for the poor. Thus, increasing inequalities raise aggregate savings, investment and, in turn, promote economic growth. Piketty (2014) believes that unsustainable and arbitrary inequalities are automatically generated by the capitalist system in ways that radically challenge the meritocratic values of current democratic societies. Moreover, some studies have recently shown that the effects of growth are not automatically translated into a fair improvement in living conditions for all parts of society. Its effects have not systematically manifested themselves in terms of paid employment and increased social welfare. The reality and trajectory of some countries, particularly in Asia, over the past decades, show that despite the observation of a remarkable increase in income levels due to the growth rates achieved, socio-economic inequalities have been accentuated at the same time. Several empirical studies have contributed to this policy shift. While Stiglitz (2012) and Piketty (2014) have done most to shed new light on the growth-equity link, the trigger probably came from an unexpected source, an IMF study, which found that "lower net inequality is strongly correlated with faster and sustainable growth" (Ostry et al., 2014). This, however, led to the realisation that what matters is not growth per se but the type and pattern of growth. Thus, the relationship between growth and inequality remains complex and multi-dimensional. This paper examines the complex links between economic growth and inequality, with causation going in both directions. Using a panel data set of Middle East and North African countries between 1990 and 2020, we study the relationship between economic growth and inequality as well as it's relationship with other variables such as human capital and investment. After running stationarity tests and cointegration tests, the study uses dynamic panel framework to establish the relationship between these variables.

Key words: *Economic growth, Inequalities, Panel data, Dynamic panel estimation, Human capital, Inclusive growth, Health, Education.*

INTRODUCTION

The issue of inclusive growth is reviving the debate on economic policies aimed at reducing inequality, fighting poverty and achieving sustainable development. In order to significantly reduce poverty and mitigate inequality, a rapid pace of growth is necessary and continuous over the long term while affecting all sectors of activity, with the inclusion of the entire working population (Ianchovichina & Lundstrom, 2009).

For more than a century, the relationship between inequality and economic growth has been the primary concern of economists. Yet the nature of this relationship remains elusive. Ideally, while it is assumed that economic growth benefits the poor on the same terms as the rich, the process of wealth creation is usually accompanied by dynamic changes in the distribution of its dividends.

By asking the question about the nature of the links that exist between inequalities and inclusive growth, and what impacts can be generated between these variables, this paper examines the complex relationship between economic growth and inequality, with causation going in both directions.

Using a panel data set of Middle East and North African countries between 1990 and 2020, we study the relationship between economic growth and inequality as well as its relationship with other variables such as human capital and investment. After running stationarity tests and cointegration tests, the study uses dynamic panel framework to establish the relationship between these variables.

The remainder of the present paper is organized as follows; point 2 will present an overview of the previous literature. Point 3 will explore research methods, whereas point 4 will be focused on the results and discussion. Point 5 will conclude.

LITERATURE REVIEW

Inequality and growth: a theoretical perspective

Inequality is defined as perceived unfair differences in terms of economic and social resources distribution. Inequality as a social difference [1] translates into unequal access to certain social advantages or disadvantages. Inequalities are multidimensional (M. Galy, 2018) affecting all forms of access to resources and benefits (economic, social or political and symbolic resources). Purely economic inequalities reflect an unequal distribution of wealth and correspond to all differences in income and wealth between individuals or social groups.

The concept of inclusive growth emerged in the economic literature and policy debate following the failure of the Washington consensus in 1990. Although there is no consensus on the definition of the concept of inclusive growth, the literature nevertheless agrees on two approaches to inclusive growth. The first emphasizes participation in the process of wealth creation and the second the distribution of the dividends of this wealth. In the first case, the inclusiveness of growth is based on its capacity to involve the largest possible number of citizens in the productive effort. In the second case, inclusiveness is based on the need for a fair distribution of the benefits of growth among the population.

There is an extensive literature on the relationship between income inequality and economic growth. Analyzing the effects of economic growth and income distribution on living standards has preoccupied the different economic currents. Traditional growth theory establishes a direct relationship between economic growth and living standards. The first thesis, developed by Kuznets in 1955 (Barthélemy, 1995), states that the relationship between GDP per capita and inequality is in the form of an inverted U-shaped curve. Kuznets started from the idea that there is only one relationship between income inequality and GDP growth by proposing a general law that determines the relationship (Henni, 2004). According to his curve, inequality increases concomitantly up to a certain threshold of development where it stabilises and then decreases. In addition, several theoretical models have studied the impact of a non-egalitarian distribution of resources on the development process (Orazio A and Binelli C, 2004). These theories are mainly grouped into two broad categories: the first postulate the existence of a positive relationship between inequality and growth (channels of individual savings and investment incentives), while the second identifies the negative effects of unequal distribution of resources on growth prospects (channels of social and political instability, fiscal and tax policy and imperfect credit markets).

From inequalities to growth

Traditional thinking on the dynamics of development considers that all economic growth can cause an income growth effect (poverty reduction) and an income distribution effect (changes in inequality) (Henni S, 2004). Dubois (1997) argues that the increase in inequality occurs with growth since all countries have relatively egalitarian income structures.

Inequalities slow or even reverse growth

Theories that income inequality reduces growth can be classified into three main categories (Panagiotis E. Petrakis, 2020) : the social and political instability stream (Hibbs, 1973; Venieris and Gupta, 1983 & 1986, ; Gupta, 1990; Alesina and Perotti, 1996), the political economy stream (Bertola, 1993; Alesina and Rodrik, 1994; and Persson and Tabellini, 1994) and the credit and capital market imperfection stream (Banerjee and Newman, 1993; and Aghion and Bolton, 1997).

Inequality and growth: an empirical analysis

The studies that have attempted to assess the overall impact of inequality on growth differ in terms of the structural specification of the data (cross-section or panel), the dimension of inequality examined (income, land ownership, human capital, etc.), the indicator used (Gini, nth quantile share, etc.) and the econometric estimation method. Thanks to the contributions and essays by Persson and Tabellini (1994) and Alesina and Rodrik (1994), the way has been opened for the production of several studies on the links between inequality and growth. Empirically, three patterns of response have been identified in this regard, as follows:

Inequality is good for growth

The standard view, especially before the 1990s, was that inequality is good for growth insofar as it encourages effort and saving. As a result, Adelman and Robinson (1989) argued that "inequality is necessary for accumulation and contains the seeds of eventual growth in individual income". Moreover, Li and Zou (1998) also find a positive relationship between inequality and economic growth, which is supported by Deninger & Squire (1998), who state that land inequality has a significant effect on human capital investment and, therefore on investment and growth. In the same vein, Forbes (2000) has shown that, in the short and medium-term, increasing inequality has a positive effect on economic growth, something that is affirmed by Ostry et al. (2014) who have shown that, for a given level of redistribution, a high level of inequality is robustly correlated with faster and sustainable growth.

Inequality is bad for growth

A new view has emerged from empirical studies in the 1990s. Thus, Persson and Tabellini (1994) showed, on the basis of a redistribution model from the richest to the poorest, that inequality is harmful to growth. On the other hand, Alesina and Rodrik (1994) confirmed that wealth and income inequalities are negatively correlate with growth. In the same vein, **Clarke** (1995) also concludes that inequality is negatively correlated with growth, noting that the magnitude of the relationship is relatively small, and the direction of causality is ambiguous.

Furthermore, Alesina and Perotti (1996) find that inequality causes socio-political instability by reducing investment and growth. Also, Diniger and Squire (1998) find a strong relationship by showing that income inequality reduces income growth for the poor and not for the rich (Amarante, 2014). Finally, Castello (2010) finds a negative effect of income and human capital inequality in low- income countries (Kh. Mdingi, S. Ho, 2021).

No link between inequality and growth

This new view was challenged by another wave of panel data studies in the late 1990s, which showed that there was no stable relationship between growth and inequality. For example, Barro (2000) finds little or no effect of inequality on growth and investment, adding that higher inequality would retard growth in poor countries and encourage it in rich countries. He believes that inequality would have a negative role on growth when GDP per capita is below $2,000 and a positive role if GDP is above that amount. In the same vein, OECD experts (2012) consider that "despite extensive theoretical work on the links between inequality and growth, no consensus has emerged and the empirical evidence is inconclusive (OECD, 2012)".

RESEARCH METHODS:

The data used in this study are extracted mainly from World Bank, IMF, World Inequality Database (WID), Trademap data, The UNESCO Institute for Statistics (UIS) and IKH-Barro database. Gini index was used to measure Income Inquality. GDP per capita at 2010 constant prices was used to measure Economic Growth. Human capital was measured by secondary school completion rate (% of population completing secondary school). The study uses panel data set of 15 countries from the Middle East North Africa (MENA) region between 1990 and 2020.

Before going to the econometric approach, the authors have analyzed the data and used Principal Component Analysis (PCA). The objective of this analysis is to identify the group of countries that have common characteristics.

First, we used Levin, Lin & Chu (LLC) method (Levine et al. 2002); I'm & Shin (IPS) method (Im et al. 2003); ADF-Fisher Chi-square and Hadri to examine the stationarity of our variables. The analysis is intended to see whether the data used in this study has reached the stationary or not. Second, we test for a cointegration test to identify the presence of a long-term relationship between the variables (Engle and Granger, 1987).

The third stage consists of checking the causality analysis between the economic growth and income inequality is performed using panel vector error correction model (VECM).

The general function of the investigation is: The modified function can be demonstrated as panel cointegration

Where: Gini index (GINI), The GDP per capita in US$ (GDP), the gross fixed capital formation in US$ (K), Population growth (POP), Health expenditure in USD (HEALTH), Public education expenditures (EDUC), money supply (MS) and human capital (HK).

RESULTS AND DISCUSSION

PCA Analysis

PCA analysis was conducted to identify the clusters of countries that have common characteristics. The Barlett's test of sphericity (BTS) above shows an approximate Chi-Square value of 87.26 and a p- value of 0.00. This result indicates that the correlation between the variables is significantly different from one, and so it is appropriate to institute factor analysis for the variables in the dataset.

The first three principal components have initial eigenvalues greater than 1. Following Kaiser criterion, we have selected these three components for further analysis. Cumulative extracted variance for these three components is about 73.3%.

Based on the PCA mapping of countries, the authors identified that there are two groups of countries we will take into consideration for econometric analysis.

North African countries: Morocco, Algeria, Mauritania, Tunisia, Egypt

Middle East countries: Iran, Saudia Arabia, Turkey, Bahrain, Oman, Jordan, Liban, United Arab Emirates, Kuwait, and Qatar.

Unit root tests

Before running regression models, panel unit root tests were used for all the variables included in the model using the IPS test, LLC test, ADF Fisher, and Hadri tests.

The data achieve stationary if it has a p-value < 0.05. For North African countries, using the four statistical tests of stationarity. Gini index is not stationary at level since the p-value of LLC, IPS and

ADF-Fisher is greater than 5%, using the first difference, the p-value of these statistical tests is less than 5%, meaning that the Gini index is stationary in at first difference. Similarly, all the other variables are stationary at first difference except population growth rate, as it is stationary at level. For the Middle East countries, similar behavior was revealed across all the variables, as they are stationary in the first difference.

All the results are statistically significant at the level of 1% and 5%. The results allowed the test for panel cointegration between the Gini, GDP, K, HK, POP, EDUC, HEALTH and MS to be established.

Table 1. Extract from the unit root test results

Variables	Statistic Test	North Africa		Middle East	
		At level	First difference	At level	First difference
GINI	LLC	-2.5	-1.79**	2.9	-7.55058***
	IPS	-1.69	-1.77**	3.57	-9.60563***
	ADF-Fisher	15.97	18.44**	8.29	117.5***
	Hadri	7.05***	-0.13	5.03***	-0.18
GDP	LLC	3.25	-3.96181***	-0.24	-14.8427***
	IPS	2.82	-5.88566***	0.93	-12.8901***
	ADF-Fisher	5.19	55.23***	8.97	138.07***
	Hadri	3.38***	0.77	5.66***	2.51***

* significant at the 10% level, ** significant at the 5% level *** significant at the 1% level
Source: author's calculations

Cointegration test

In order to examine the long run relationship between the variables, the authors used Kao test, if the p-value < 0.05 indicates that there is cointegration between the variables. Conversely, if the p-value >

0.05 indicates that there is no cointegration between the variables. The result of Kao's residual panel cointegration test in

Table 2. Cointegration test

		North Africa	Middle East
Kao Residual Cointegration Test	Statistics	-4.562	-2.6491
	P-value	0.0000	0.0040

Source: authors' calculations

Based on the results of Kao panel cointegration test. We can reject the null hypothesis. This indicates that a long-run relationship exists between the variables.

Panel VECM model results

After confirming the long-run relationship between the variables. The panel VECM approach was used to estimate the long-run coefficients and the short-run dynamics between the variables.

For North African countries, the results provide empirical evidence dealing with the long-run relationship between income inequality and economic growth. In the long-run, there is a negative relationship between the two variables, where economic growth can reduce income inequality. In other words, increasing output in the local economy in these countries, in the long-run, encourages the distribution of income in the community. For the Middle East of countries, the relationship between economic growth and inequality is positive. In other words, economic growth can increase income inequality in these countries. The coefficient associated with education expenditure is positive and significant, meaning that it impacts positively economic growth. Also health expenditure affects positively economy for the two groups of countries. We note that the error correction term is negative and statistically significant at 1%.

Regarding the causality relationship between the variables. The results of Granger causality tests show that the bi-directional causality between economic growth and inequality exist in North African countries. Granger causality test is statistically significant at 5%. For the second group –The Middle East –, the results of the tests show that there is no Granger causality between economic growth and inequality of income, Granger causality test is not significant at 5%.

CONCLUSION

The main objective of our study to analyze the relationship between income inequality and economic growth for 15 selected countries in MENA region. The results of Principal Component Analysis confirm that there are two groups of countries to be analyzed separately in econometric modelling. Using cross-section panel data and employing panel vector error correction model, the key conclusions of this study are as follows:

Table 3. Extract from the model results

	North Africa	Middle East
Cointegrating Eq:	CointEq1	CointEq1
LOG_GDP(-1)	1.000000	1.000000
LOG_GINI(-1)	1.100580	-4.142010
	(0.25091)	(1.11933)
	[4.38629]	[-3.70043]
C	-6.418299	-144.8191
Error Correction:	D(LOG_GDP)	
CointEq1	-0.088646	-0.006662
	(0.03319)	(0.00297)
	[-2.67101]	[-2.23973]

(..): standard deviation of estimators & [..]: student's statistic

Source: authors' calculations

The findings proved that for the two groups of countries, the long-run relationship between economic growth, income inequality, health expenditure, education expenditure, population growth, human capital, trade openness and investments. For the first group of countries (North Africa) the relationship between economic growth and income inequality is negative and significant. **However** the relationship between economic growth and income inequality is positive and significant for the second group of countries (Middle East). The error correction term is negative and statistically significant at 1%.

Table 4. Granger causality test for North African countries

Dependent variable: D(LOG_GDP)			
Excluded	Chi-sq	df	Prob.
D(LOG_GINI)	8.165432	3	0.0409

Source: authors' calculations

Table 5. Granger causality test for Middle East countries

Dependent variable: D(LOG_GDP)			
Excluded	Chi-sq	df	Prob.
D(LOG_GINI)	1.344996	2	0.5104

Source: authors' calculations

Regarding causality relationship between economic growth and income inequality, the study found that Granger causality between the two variables is bidirectional and significant in the two directions for the first group of countries (North Africa). **In comparison** the Granger causality between the two variables is not significant for the second group of countries (Middle East).

This paper is one more stone added to the scientific construction examining the relationship between inequality and inclusive growth. It aims to enrich the debate and discussion on the issue by considering the links between these two socio-economic phenomena.

References

Amarante V, 2014. Revisiting inequality and Growth: Evidence for developing Countries, Growth and Change,https://www.researchgate.net/publication/262232026Revisiting_Inequality_and_Growth_Evidence_for_Developing_Countries.

Alesina A. and Perotti R, 1996. Income Distribution, Political Instability and Investment, European Economic Review, 81, 5, 1170-1189.

Alesina et Rodrik, 1994. Distributive Politics and Economic Growth. The Quarterly Journal of Economics, 09, No. 2 pp. 465-490.

Aghion P. and Howitt P, 1997. Endogenous Economic Growth, Cambridge MA, MIT Press. Allison, 1978. Measures of Inequality. American Sociological Review, 43 No. 6, pp 865- 880. Aghion P, Bolton P, 1997. A Theory of Trickle-Down Growth and Development. Review of
Economic Studies, 64(2), 151-172.

Banerjee A., Newman A, 1993. Occupational Choice and the Process of Development. Journal of Political Economy, 101(2), 274-298.

Barthélemy Ph, 1995. L'hypothèse de Kuznets est-elle encore d'actualité ?, Revue Région et Développement n°2-1995

Brueckner M, and Lederman D, 2015. Effects of Income Inequality on Aggregate Output, World Bank Policy Discussion Paper 7317.

Barro, 2000. Inequality, Growth and Investment. National Bureau of Economic Research, No 7038 Choi, I., 2001. Unit root tests for panel data. Journal of International Money and Finance, 20, pp.249-272.

Clarke GRG, 1995. More Evidence on Income Distribution and Growth, Journal of Development Economics, 47(1995), pp. 403-427.

Deininger K, Squire L, 1998. New way of Looking at old Issues: Inequality andGrowth. Journal of Development Economics, 57(2), 159-287.

Forbes, 2000. A Reassessment of the Relationship Between Inequality and Growth, American Economic Review, 90, pp. 869-887.

Ianchovichina, Elena et Susanna Lundstrom, 2009. Inclusive Growth Analytics, Policy Research Working Paper, n° 4851, mars,

Helpman E, 1997. General Purpose Technologies and Economic Growth, Cambridge MA, MIT Press.

Henni S, 2004. Inégalités sociales, croissance et développement durable, Bordeaux IV , Université de Versailles Saint-Quentin En Yvelines C3ED_ henni75@hotmail.com

Im K.S., Pesaran, M.H, and Shin Y, 2003. Testing for Unit Roots in Heterogeneous Panels, Journal of Econometrics, 115(1), pp. 53-74.

Galy M, 2018. How to analyze social structure, from http://eduscol.education.fr

Kao C, 1999. Spurious regression and residual-based tests for cointegration in panel data.Journal of Econometrics, 90(1), pp. 1–44.

Kholeka Mdingi and Sin-Yu Ho, 2021. Literature Review on Income Inequality and Economic Growth, MethodsX, 2021 https://doi.org/10.1016/j.mex.2021.101402Obtenir des droits et du contenu.

Kuznets (1955) Economic Growth and Income Inequality. American Economic Review, 45, No.1, pp. 1-28.

Levin A, Lin C.F, and Chu C.S.J, 2002. Unit Root Test in Panel Data: Asymptotic and Finite Sample Properties, Journal of Econometrics, 108, pp. 1-24.

Li, Squire et Zou, 1998. Explaining International and Intertemporal Variations in Income Inequality, Economic Journal, 108 pp. 26-43.

Mulligan, Gill et Xavier Sala-i-Martin, 2003. Do democracies have different public policies than nondemocracies? National Bureau of Economic Research, Working Paper 10040

Panagiotis E Petrakis, 2020. The New Political Economy of Greece up to 2030, Springer Science and Business Media LLC

Partridge, 1997. Is Inequality Harmful for Growth? Comment. American Economic Review, 87, No. 5 pp. 1019-1032

Persson et Tabellini, 1994. Is Inequality Harmful for Growth? American Economic Review, 84, No. 3 pp. 600-621

Perotti R, 1996. Growth, Income Distribution, and democracy: what the Data say?, Journal of Economic Growth 1(2): 149—187, Département de la politique économique et de la dette, Washington DC, Banque mondiale.

Phillips P.C.B, and Sul D, 2003. Dynamic panel estimation and homogeneity testing under cross section dependence. Econometrics Journal, 6(1), pp. 217-259.

Orazio A and Binelli C, 2004. Inégalités, croissance et politiques redistributives, Afrique contemporaine 2004/3 (n° 211), p. 107-139.

Ostry J D, Berg A, and Tsangarides G D, 2014. Redistribution, Inequality, and Growth, IMF Staff Discussion Note No. SDN/14/02, February.

Sala-i-Matin, Xavier X, 1997. I Just Ran Two Million Regressions. American Economic Review, 87, No. 2, pp. 178-183.

ANNEX

1. Unit Root Test Results

Table 1. Unit Root Test Results

Variables	Statistical test	North Africa countries		Middle East countries	
		At level	1st	At level	1st
GINI	LLC	-	-1.79351**	2.90559	-
	IPS	-	-1.77283**	3.57781	-
	ADF-	15.9710	18.4409**	8.29674	117.514***
	Hadri	7.05345	-0.13375	5.03502***	-0.18738
GDP	LLC	3.25021	-	-0.24078	-
	IPS	2.82831	-	0.93796	-
	ADF-	5.18995	55.2297***	8.97916	138.076***
	Hadri	3.38580	0.77194	5.66443***	2.51512***
K	LLC	-	-	-1.56712*	-
	IPS	-	-	0.77731	-
	ADF-	14.5584	51.2204***	11.6341	106.875***
	Hadri	3.27849	1.20718	9.83005***	-1.43909
POP	LLC	-	-	-	-
	IPS	-	-	-	-
	ADF-	49.0435	32.9316***	62.7253***	55.7652***
	Hadri	5.30153	1.84092**	1.44830*	-0.55056
HEALTH	LLC	1.70315	-	1.56716	-
	IPS	1.19996	-	1.57707	-
	ADF-	4.37036	34.1471***	16.0365	71.0020***
	Hadri	8.26615	-0.62893	8.32751***	1.63150*
EDUC	LLC	0.71393	-	-	-
	IPS	0.26332	-	-	-
	ADF-	11.7802	39.9031***	44.5555***	151.485***
	Hadri	5.13037	0.52795	6.10283***	3.20747***
MS	LLC	-	-	3.97453	-
	IPS	1.01631	-	3.47371	-
	ADF-	7.39126	45.2391***	5.12478	69.7473***
	Hadri	7.07446	1.11794	9.27760***	3.57052***
OPENNESS	LLC	1.56942	-	-1.01251	-
	IPS	0.73910	-	0.23101	-
	ADF-	13.1267	74.7194***	17.1620	61.9978***
	Hadri	6.09436	0.72889	5.31934***	0.65609
HK	LLC	0.42509	2.03960	0.04265	-17.3399
	IPS	3.33510	-	1.35230	-17.0766
	ADF-	0.96630	34.4907***	15.2456	220.871
	Hadri	3.80685	0.48210	8.82079***	3.97627***

Meaning: *significant at the 10% threshold, **at the 5% threshold, ***at the 1% threshold
Source: Author's elaboration

1. *Results of the cointegration test:*

Table: Kao test results

		North Africa countries	Middle East countries
Kao Residual Cointegration Test	Statistics	-4.562164	-2.649128
	P-value	0.0000	0.0040

Source : Author's elaboration

2. *Results of the Panel-VECM model*

Table 2. Model results

	North Africa countries	Middle East countries
Cointegrating Eq:	CointEq1	CointEq1
LOG_GDP(-1)	1.000000	1.000000
LOG_GINI(-1)	1.100580	-4.142010
	(0.25091)	(1.11933)
	[4.38629]	[-3.70043]
LOG_K(-1)	0.143783	-13.84834
	(0.04449)	(2.84771)
	[3.23168]	[-4.86297]
LOG_EDUCATION(-1)	-0.177918	-4.889238
	(0.07740)	(1.48031)
	[-2.29879]	[-3.30284]
LOG_HK(-1)	0.017968	15.34312
	(0.06117)	(3.32240)
	[0.29374]	[4.61808]
LOG_HEALTH(-1)	-0.420197	-3.143197
	(0.07226)	(2.09274)
	[-5.81470]	[-1.50196]
LOG_MONETARY_SUPPLY(-1)	0.221980	1.995703
	(0.07173)	(0.46872)
	[3.09480]	[4.25774]
LOG_OPENNESS(-1)	-0.251951	25.36877
	(0.05551)	(20.1039)
	[-4.53880]	[1.26188]
LOG_POP(-1)	0.072875	0.503365
	(0.08016)	(1.58694)
	[0.90908]	[0.31719]
C	-6.418299	-144.8191
Error Correction:	D(LOG_GDP)	
CointEq1	-0.088646	-0.006662
	(0.03319)	(0.00297)
	[-2.67101]	[-2.23973]

Source: Author's elaboration Meaning value between (..) the standard deviation of the estimators, value between [..] student statistic

3. *Granger causality test results*

Table 3. Granger causality test for North African countries

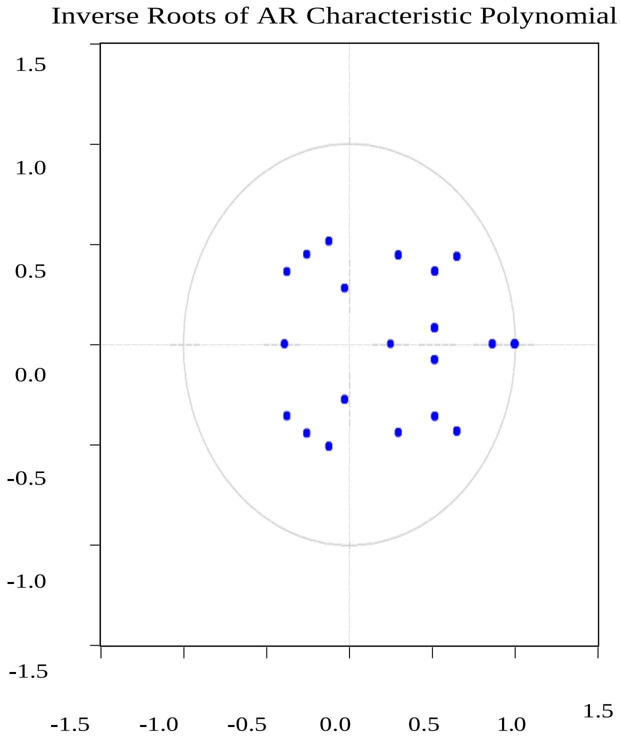
Dependent variable: D(LOG_GDP)			
Excluded	**Chi-sq**	**df**	**Prob.**
D(LOG_GINI)	8.165432	3	0.0409

Table 4. Granger causality test for Middle East countries

Dependent variable: D(LOG_GDP)			
Excluded	**Chi-sq**	**df**	**Prob.**
D(LOG_GINI)	1.344996	2	0.5104

4. *Model stability test*

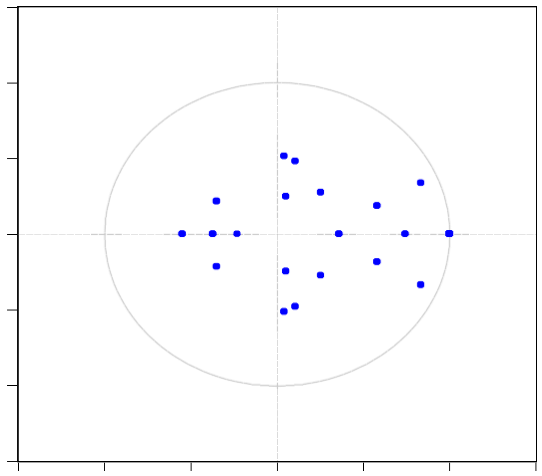
- *Middle East countries*



We notice that all the roots (the cloud of points) are inside the circle of unity, which shows that the model is stable.

- *North Africa countries*

Inverse Roots of AR Characteristic Polynomial



We notice that all the roots (the cloud of points) are inside the circle of unity, which shows that the model is stable.

Comparison of Bank Performances Using Multi-Criteria Decision Making Methods in the Perspective of COVID-19

İsmail Durak [1*], Binnur Çiğse Bal[2]

[1] Duzce University, Faculty of Business Administration, Department of Business Administration, 81000, Duzce, Turkey
[2] Duzce University, Institute of Graduate Education, Department of Business Administration, 81000, Duzce, Turkey
*Corresponding author e-mail: ismaildurak@duzce.edu.tr

Abstract

The purpose of this study is to analyze bank performance before and during the pandemic in developing countries where COVID-19 cases are most common. Four nations and five financial ratio criteria were involved in the analysis. Turkey, Brazil, Argentina, and Russia were the nations chosen based on the number of cases and included in the research. Capital, Liquidity, Profitability, and Income-Expense criteria were determined as a consequence of the literature investigation and included in the research. The entropy approach was utilized to calculate the criteria weights in accordance with the study aim. The rankings of the countries during COVID-19 were then computed using the VIKOR technique, one of the multi-criteria methods, based on the weights of the criteria obtained. According to the entropy results, although the weight of the Profitability criteria was first before COVID-19, the weight of the Liquidity criterion was first after COVID-19. The VIKOR research undertaken in this context revealed that the country rankings before and during the COVID-19 period changed. As a consequence, the causes for this difference were explored, and recommendations were given to industry leaders and policymakers regarding COVID-19's developing performances.

Key words: *Bank performance, COVID-19, Entropy, VIKOR, Multi-Criteria Decision Making*

Examining the Factors Affecting the Adoption of Financial Technologies (Fintechs)

İsmail Durak [1*], Gülçin Çelik[2]

[1] Duzce University, Faculty of Business Administration, Department of Business Administration, 81000, Duzce, Turkey
[2] Duzce University, Institute of Graduate Education, Department of Business Administration, 81000, Duzce, Turkey
*Corresponding author e-mail: ismaildurak@duzce.edu.tr

Abstract

Financial Technologies can be defined as developing new financial products and services and companies providing these products in the financial sector where digitalization is rapidly increasing. With the advancement of Fintech applications, identifying the variables influencing people's acceptance of this innovation has become a crucial study issue. The aim of this research is to examine the effects of variables such as trust, perceived convenience, social impact, and perceived risk in the adoption of Fintech services, with a holistic perspective. For this purpose, a research model was established, and an online survey was conducted with 230 people to collect the data to be used in the research. 220 participants completed the questionnaire and provided relevant data. First, Explanatory Factor Analysis (EFA) was applied to the obtained data in SPSS 24 program. Confirmatory Factor Analysis (CFA) and Path Analysis were then done in the AMOS 24 software to discover the hypotheses and linkages of all latent variables. The model constructed based on the fit index values demonstrated a good degree of fit in the applied Path Analysis. It was revealed that the path coefficients of perceived risk and confidence in the regression relationships tested in the analysis were not statistically significant. It was also seen that social impact and perceived usefulness had a positive effect on the attitude, and the attitude hds a positive effect on the actual use of Fintech.

Key words: *Fintech Adoption, Financial Technology, Attitude, Path Analysis*

G 20 Ülkelerinde Yaşam Kalitesinin CRITIC ve WASPAS Yöntemleri ile Karşılaştırılması

Ayşe Elif Yazgan[1]*

[1] Necmettin Erbakan Üniversitesi, Uygulamalı Bilimler Fakültesi, Finans ve Bankacılık Bölümü, Konya, Türkiye

*Corresponding author e-mail: aeyazgan@erbakan.edu.tr

Özet

Yaşam kalitesi, toplumların giderek artan uygarlığının bir sonucu olarak modern zamanların ortaya çıkan bir terimidir. Mevcut araştırmalar bize, teknolojideki gelişmeler ve gelir düzeylerindeki artış sayesinde refahın artık yaşam kalitesinin yegâne olmadığını, daha birçok faktörün yaşam kalitesi üzerinde önemli derecede etkisinin olduğunu göstermektedir. Bu doğrultuda çalışmada, G 20 ülkelerinin 2021 yılına ait yaşam kalitesi düzeyi Çok Kriterli Karar Verme (ÇKKV) teknikleri içerisinden CRITIC (Criteria Importance Through Intercriteria Correlation) ve WASPAS (Weighted Aggregated Sum Product Assessment) yöntemleri yardımıyla analiz edilerek değerlendirilmiş; ilgili ülkeler yaşam kalitesi düzeyine göre sıralanmıştır. Çalışmada öncelikle yaşam kalitesi düzeyine etki eden sekiz adet kriter, ilgili literatür taranarak tespit edilmiştir. Bu kriterler, satın alma gücü, güvenlik, sağlık hizmeti, iklim, yaşam maliyeti, emlak fiyatları /gelir, trafikte harcanan süre ve kirlilik olarak belirlenmiştir. Yaşam kalitesinin belirlenmesinde kullanılan bu kriterlere ilişkin veriler Numbeo isimli internet sitesinden elde edilmiştir. Daha sonra tespit edilen kriterler CRITIC yöntemi ile ağırlıklandırılmış; bu kriterler arasından en çok önem arz eden kriterin yaşam maliyeti, en az önem arz eden kriterin ise trafikte harcanan süre olduğu belirlenmiştir. Son olarak WASPAS yönteminden yararlanarak ilgili ülkelerin yaşam kalitesi düzeyi analiz edilmiştir. Yapılan analizler sonucunda, yaşam kalitesi düzeyi en yüksek ülkenin Avustralya; en düşük ülkenin ise Endonezya olduğu görülmüştür. Türkiye ise sıralamada, 5. sırada yer almıştır.

Key words: *Yaşam Kalitesi, G 20 Ülkeleri, ÇKKV, CRITIC, WASPAS*

Investigation of Performance of J48 and Reptree Algorithms from Family of Decision Trees

<u>Özlem Bezek Güre</u>[1]*

[1] Batman University, Health Services Vocational School, Medical Documentation and Secretarial Program, Batman, Turkey

*Corresponding author e-mail: obezekgure@gmail.com

Abstract

Decision trees are frequently used among data mining methods. In this study; It is aimed to compare the performances of the J48 and RepTree algorithms from the decision tree family by using the FEV (Forced Expiratory Volume) data set on the Journal of Statistics Education (2017) website.FEV data set; It was created to examine the effects of parents' smoking habits on children's respiratory functions. In the data set in question; There are measurements of 654 children, 318 girls and 336 boys, whose age range is between 3-19 and 46-74 inches in smoking and non-smoking environments. In the study; FEV values were analyzed as dependent variables, gender, age, height and smoking status of parents as independent variables.FEV values below and above the mean value were converted into a two-level categorical form. Analyzes were made using J48 and RepTree algorithms in the Weka program. According to the results of the analysis, the correct classification rate of the J48 algorithm was 82.41% and the error rate was 0.2331, while the correct classification rate of the Random Tree algorithm was 80.23% and the error rate was 0.2297. According to the analysis results, it was concluded that the J48 algorithm performed better than the RepTree algorithm.

Key words: *J48, Random tree, FEV.*

Bütünleşik CRITIC ve ARAS Yöntemleri ile G7 Ülkelerinin Siber Güvenlik Performanslarının Karşılaştırılması

Ayşe Elif Yazgan[1]*

[1] Necmettin Erbakan Üniversitesi, Uygulamalı Bilimler Fakültesi, Finans ve Bankacılık Bölümü, Konya, Türkiye

*Corresponding author e-mail: aeyazgan@erbakan.edu.tr

Özet

Siber güvenlik ve bilişim teknolojisi konuları insan yaşamını kolaylaştırmakla birlikte bilişim güvenliği hususunda tehlikenin artmasına ve kaygılara neden olmaktadır. Siber dünyada meydana gelen suçlar; mağdur durumda kalanlar ile bilişim suçlularını aynı atmosferde olmamalarına rağmen biraraya getirmektedir. Suç örgütleri veya tekil şahıslar siber dünyayı reklam amaçlı kullanmakla birlikte siber dünya içerisinde örgütsel etkinliklerini artırmaktadırlar. Dünya çapında meydana gelen bu teknolojik saldırılardan ötürü siber güvenlik konusunun önemi her geçen gün artmakta olup günümüzün en önemli savunma sistemi haline gelmiştir. Ekonomik açıdan dünya sıralamasında en önemli ülkelerden olan G7 ülkeleri, hem kendi ülkelerinde hem de diğer ülkelerde kaos ortamının oluşmaması, yeni saldırılar ile ülkelerinin zarar görmemesi ve dünya barışının devamlılığının sağlanabilmesi için siber güvenlik hususunu mutlak ve mutlak önemsemelidir. Bu doğrultuda yapılan çalışmada, G7 ülkelerinin 2020 yılına ait siber güvenlik performansları Çok Kriterli Karar Verme (ÇKKV) yöntemleri içerisinden CRITIC (Criteria Importance Through Intercriteria Correlation) ve ARAS (Additive Ratio Assessment) yöntemlerinin birlikte kullanımı ile analiz edilmiş; ilgili ülkelerin siber güvenlik performansları açısından sıralaması yapılmıştır. Ülkelerin siber güvenlik performansını belirleyen kriterler literatürden yararlanarak yasal ölçümler, teknik ölçümler, organizasyonel ölçümler, kapasite gelişimi ölçümleri ve iş birliği ölçümleri olarak tespit edilmiştir. Analize konu olan ülkelerin belirlenen bu kriterlere ait verileri Uluslararası Telekomünikasyon Birliği (International Telecommunication Union – ITU)'nin resmi web sitesinden alınmıştır. Belirlenen kriterlerin ağırlıklarının sıralanmasında CRITIC yöntemi kullanılmış; siber güvenlik performansına ait en fazla önem arz eden kriterin organizasyonel ölçümler, en az önem arz eden kriterin ise yasal ölçümler olduğu görülmüştür. Daha sonra ARAS yöntemini kullanarak G7 ülkelerinin siber güvenlik performansları incelenerek analiz sonucunda siber güvenlik performansı en yüksek üç ülkenin sırasıyla ABD, İngiltere ve Kanada olduğu tespit edilmiştir. İlgili ülkeler arasında siber güvenlik performans düzeyi en düşük ülkenin ise Almanya olduğu görülmüştür.

Key words: *Siber Güvenlik Performansı, G 7 Ülkeleri, ÇKKV, CRITIC, ARAS*

A Hybrid Structure for Classification of Brain Cysts Images

Oznur Ozaltin[1*], **Aynur Yonar**[2], **Orhan Coskun**[3], **Ozgur Yeniay**[1]

[1] Hacettepe University, Institute of Science, Department of Statistics, 6800, Ankara, Turkey
[2] Selcuk University, Faculty of Science, Department of Statistics, 42031, Konya, Turkey
[3] Health Sciences University Gaziosmanpaşa Training and Research Hospital, Pediatric Neurology

*Corresponding author e-mail: oznurozaltin@hacettepe.edu.tr

Abstract

Deep learning algorithms, particularly Convolutional Neural Networks (CNN), are extensively used for medical image classifications, segmentation, and automatic feature extractions. For this purpose, we design a novel CNN architecture, named OzNet, to detect Brain Cysts from Magnetic Resonance Imaging (MRI). In this study, the Brain MRI dataset is collected for the <18 age group from Turkey. First of all, we classify different types of Brain Cysts by using OzNet. Although obtained the experimental results are good, we want to improve the classification results. In the following of this study, OzNet is tasked with feature extraction from the dataset and so 4096 features are obtained for each image. Then, we benefit from a heuristic method, Genetic Algorithm (GA), for dimension reduction of these features. Furthermore, Artificial Neural Networks (ANN) are used for the classification of reduced features. Eventually, we obtain the best hybrid structure, OzNet-GA-ANN, to detect different types of Brain Cysts.

Key words: *ANN, Brain Cysts, Deep Learning, Genetic Algorithm, OzNet*

Kredi Hacmini Etkileyen Faktörlerin Yapay Sinir Ağları ile İncelenmesi

Fatih Çemrek[1], **Özge Demir**[1*]

[1] Eskişehir Osmangazi Universitesi Fen Edebiyat Fakültesiİstatistik Bölümü, 26040, Eskişehir, Turkiye

*Corresponding author e-mail: zdemir94@gmail.com

Özet

Bankacılık sektöründe önemli bir gösterge olan kredi, gerçek veya tüzel kişilere belirli bir vade ve belirli bir faiz oranı ile verilen ve geri ödenmesi talep edilen kredi veya kredi olarak tanımlanmaktadır. Kredi alan kişi ve kurumlar yapacakları iş veya yatırımda finansal işlemlerini bu şekilde gerçekleştireceklerdir. Bankalar kredi verirken kredi verecekleri gerçek veya tüzel kişiler, krediyi ödeyip ödeyemeyecekleri, gelir-gider durumları gibi birçok kriteri dikkate alırlar. Kredinin dört unsuru vardır: zaman, güven, risk ve gelir. Kredi, dolaşım aracı işlevini yerine getirmek, atıl sermaye ve tasarrufları iş alanlarına aktarmak, ekonomik faaliyetleri hızlandırmak ve mal arz ve talep dengesini korumak gibi işlevlere sahiptir. Bu çalışmada Türkiye'de kredi hacmini etkileyen faktörler incelenmiştir. Kredi hacmi Bin TL, döviz kuru (TL/$), aylık faiz oranı (%) ve mevduat hacmi (Bin TL) değişkenleri arasındaki ilişkiler Yapay Sinir Ağları ile analiz edilmiştir. Veriler Türkiye Cumhuriyet Merkez Bankası Elektronik Veri Dağıtım Sistemi'nden (EVDS) alınmıştır. Veriler, 2010:01-2021:12 dönemini kapsayan aylık verilerden oluşmaktadır.

Key words: *Kredi Hacmi, Mevduat, Döviz Kuru, Yapay Sinir Ağları*

Çok Kriterli Karar Verme Modelleri: AHP ve Bulanık AHP Yöntem Karşılaştırması

Mehmet Kenan TERZİOĞLU[1], Aysu YAŞAR[2*],

[1] Trakya Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, Edirne, Turkey
[2] Nişantaşı Üniversitesi, İktisadi İdari ve Sosyal Bilimler Fakültesi, İşletme Bölümü, Edirne, Turkey

*Sorumlu Yazar E-mail: yasar.aysu@nisantasi.edu.tr

Özet

Seçim süreçlerinde alınan yanlış kararlar kuruluşlar için üretkenliği, hassasiyeti, esnekliği ve üretim yeteneklerini olumsuz etkilemektedir. Sektörde artan alternatifler arasında en iyinin seçimi, nicel ve nitel birçok özelliğin varlığında çok kriterli karar verme problemlerini beraberinde getirmektedir. Hem akademik araştırmalarda hem de uygulamalarda çeşitli çok kriterli karar verme problemlerinin değerlendirilmesinde analitik hiyerarşi süreci (AHP) yaygın olarak kullanılmaktadır. Karar vericinin/vericilerin yargılarındaki belirsizlik ve geleneksel AHP'deki net ikili karşılaştırması, karar vericinin/vericilerin doğru yargılarını yakalamak için yetersiz ve belirsiz olduğundan bulanık sayı mantığı geleneksel AHP'ye dahil edilerek belirsizlikler ortadan kaldırılmaya çalışılmaktadır. Çalışma kapsamında, her iki tekniğin de yer aldığı bulanık mantık ve AHP bir araya gelerek bulanık AHP yönteminin uygulama adımlarının ve alanlarının anlaşılması amaçlanmaktadır.

***Anahtar Kelimeler**: Analitik Hiyerarşi Süreci (AHP), Bulanık Mantık, Çok Kriterli Karar Verme.*

Multi-Criteria Decision Making Models: AHP and Fuzzy AHP Method Comparison

Abstract

Wrong decisions taken during selection processes negatively affect productivity, sensitivity, flexibility and production capabilities for organizations. The selection of the best among the increasing alternatives in the sector brings with it multi-criteria decision-making problems in the presence of many quantitative and qualitative features. Analytical hierarchy process (AHP) is widely used in the evaluation of various multi-criteria decision-making problems in both academic research and practice. Since the uncertainty in the judgments of the decision maker(s) and the clear pairwise comparison in traditional AHP are insufficient and uncertain to capture the correct judgments of the decision maker(s), fuzzy number logic is included in the traditional AHP to eliminate the uncertainties. Within the scope of the study, it is aimed to understand the application steps and areas of the fuzzy AHP method by combining fuzzy logic and AHP, which includes both techniques.
***Key words**: Analytic Hierarchy Process (AHP), Fuzzy Logic, Multi-Criteria Decision-Making.*

GİRİŞ

Bilgisayar tabanlı sistemlerin günlük hayatımızdaki rolü arttıkça kullanılan programları değerlendirmek ve güvenli şekilde çalışmasını sağlayarak işlevselliğini artırmak da giderek önemli hale gelmektedir. Bilgisayar tabanlı kaynakların davranışını kontrol etme süreci ve belirsizlikle ilgili problemlerin çözümü başta olmak üzere birçok alanda tercih edilen bulanık mantık, çevresel etkilerin karmaşıklığı ve dil bilimindeki belirsizliğin matematiksel olarak modellenmesi amacıyla geliştirilen klasik küme teorisinin bir genellemesidir. Bulanık mantık, insan düşüncesine ve algısına mümkün olduğunca yakın olmaya çalışan bir mantık olup insanların kesin değişkenler (evet / hayır) yerine bir dizi "bulanık" değerlerle (oldukça iyi, çok iyi, kararsız, kötü vb.) daha iyi çıkarımda bulunduğu varsayımına dayanmaktadır.

Her türlü insan faaliyetiyle ortaya çıkan rekabetçi ortamlarda bulanık mantığa veya kişisel yargılara dayalı çok kriterli karar verme problemleri ile kriterler kategorize edilebilmektedir. Klasik çok kriterli karar verme yöntemleri kesin olmayan bilgiler içerdiğinden ve kriterlere ilişkin performans seviyelerini atamak zor olduğundan bulanık mantıkla esnek hale getirilmektedir. Bulanık mantık ile analitik fonksiyonların bulunmadığı sistem davranışına yaklaşılmakta ve davranışların iyi anlaşılmadığı durumlarda hızlı ve yaklaşık çözümün sağlandığı karmaşık sistemler içeren çözümler elde edilmektedir. Bununla birlikte, bulanık mantık akıl yürütme yeteneklerinin yapay bilgi tabanlı sistemlere uygulanmasını sağlayan çıkarım yapısına sahip olmakta ve çıkarım mekanizması, öğrenme, uyum, paralellik ve genellemede önemli bir rol oynamaktadır.

Uygun değerlendirme yapısının belirlenmesinde uzmanlar, aralarından seçim yapmaları gereken bir dizi alternatifle karşı karşıya kalmaktadır. Alternatif seçmeye yönelik karar sürecinde ise genellikle organizasyonel ihtiyaçlar, hedefler, riskler, faydalar, kaynaklar gibi birçok faktörü dikkate almak zorundadır. Karar verme süreçlerinde tek bir kriter göz önüne alındığında değerlendirmeler sezgisel olurken birden fazla kriter söz konusu olduğunda değerlendirmeler karmaşık hale gelmektedir. Kriterlerin genellikle eşit öneme sahip olmaması ve alternatiflerin çeşitli performanslara sahip olmasından dolayı Analitik Hiyerarşi Prosesi (AHP), İdeal Çözüme Benzerliğe Göre Tercih Sıralaması Tekniği (TOPSIS) ve Veri Zarflama Analizi (DEA) gibi yapılandırılmış çok kriterli karar verme yöntemlerine ihtiyaç duyulmaktadır.

Sayısallaştırılması zor karar unsurlarının karşılaştırılmasını içeren karmaşık kararlar için uygun olan AHP, öznel yargıların karar öğelerini ortak özelliklerine göre kümelemek olduğu varsayımına dayanmaktadır. Karşılaştırma sırasında öznel yargılar belirsiz olabileceğinden bilişsel belirsizlik ve hesaplamalı sinir ağlarında da kullanılan bulanık kümeler, AHP ile birleştirilmekte ve bulanık AHP veya FAHP olarak adlandırılmaktadır. Analitik Hiyerarşi Süreci (AHP), farklı seviyelerdeki karmaşık kriterler yapısı arasında karar vermenin en iyi yollarından biri olurken Bulanık Analitik Hiyerarşi Süreci (FAHP) ise karar vericilerin bulanıklığı dikkate alındığında klasik AHP yönteminin sentetik bir uzantısı olmaktadır.

ANALİTİK HİYERARŞİ PROSESİ

Satty(1980) tarafından geliştiren çok kriterli karar verme yaklaşımı AHP'de, hiyerarşik yapının doğru oluşturulması ve amaç kapsamında değerlendirmelerin doğru bir biçimde yapılması önemli olmaktadır. AHP yönteminde karmaşık problemler hiyerarşik yapı ile çözülürken hem nitel hem de nicel kriterleri kullanılmaktadır.

Adım 1 Hiyerarşik yapının oluşturulması: Karar verilmesi gereken sorunun AHP yaklaşımı ile çözülüp çözülemeyeceğinin belirledikten sonra deneyimli uzman görüşleri doğrultusunda problemin çözülebileceği belirlenerek çözüm algoritmasını oluşturmaktadır. Oluşturulan hiyerarşik yapının en tepe noktasında amaç yer alırken bir alt seviyede amacı etkileyen kriterler ve varsa alt kriterler belirlenerek sonucu belirleyen seçenekler yapısı oluşturulmaktadır.

Adım 2: İkili karşılaştırma matrislerinin oluşturulması: İkili karşılaştırma matrisi, hiyerarşik yapıda bir düzeyde yer alan kriterlerin bir üst faktör kapsamında ikili olarak birbirleriyle karşılaştırılmasıyla elde edilmektedir. Matriste yer alan ham veriler Tablo 1'deki 1-9 ölçeği kullanılarak seçilen nitel kriterlere göre önem kıyası yapılarak oluşturulmaktadır.

Tablo 1. Karşılaştırma Ölçeği

Önem Yoğunluğu	Tanım
1	Eşit Önem
3	Bir Faktörün Diğerine Kıyasla Orta Önemi
5	Bir Faktörün Diğerine Kıyasla Güçlü Önemi
7	Bir Faktörün Diğerine Kıyasla Çok Güçlü Önemi
9	Bir Faktörün Diğerine Kıyasla Aşırı Derecede Önemli Olması
2,4,6,8	Ara Değerler
$1/\boldsymbol{a_{ij}}$	Ters Karşılaştırma Için Karşılıklar

Kaynak: Saaty, T. L. (1980)

İkili karşılaştırma yargısı a_{ij} bir üst düzeydeki faktöre göre i. ve j. kriterlerinin göreceli önemini göstermek üzere, A matrisi,

$$A = A(a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \tag{1}$$

olarak ifade edilmektedir. Karşılaştırma yapılırken ikili karşılaştırma matrisinin tüm değerleri 1 olan köşegenin üstünde kalan elemanlar için i. kriter ile j. kriterin karşılaştırma değerini olmak üzere a_{ij}; a_{ji} değerinin karşıt değerlendirmesi ise $1/a_{ij}$ eşitliğinden elde edilmekte ve İkili karşılaştırma matrisi Tablo 2'de gösterilmektedir (Saaty, 1980).

Tablo 2. Karşılaştırma Matrisi

A	Kriter 1	Kriter 2	Kriter 3	…	Kriter $_n$
Kriter 1	1	a_{12}	a_{13}	…	a_{1n}
Kriter 2	a_{12} $=1/a_{12}$	1	a_{23}	…	a_{2n}
Kriter 3	a_{13} $=1/a_{13}$	$1/a_{23}$	1	…	a_{3n}
…	…	…	…	1	…
Kriter $_n$	a_{n1} $=1/a_{1n}$	a_{n2} $=1/a_{2n}$	a_{n3} $=1/a_{3n}$	…	1

Adım 3: İkili karşılaştırma matrislerinin normalize edilmesi: İkili karşılaştırma marisi oluşturulduktan sonra normalizasyon işlemi

$$a'_{ij} = \frac{a_{ij}}{\sum_{i=1}^{n} a_{ij}} \tag{2}$$

denklemi ile elde edilmektedir.

Adım 4: Öncelik vektörünün hesaplanması: Normalize edilen matriste herbir satır toplamı matrisin boyutuna bölünerek ortalamaları alınmaktadır. Hesaplanan değerler, her bir kritere ait önem ağırlıkları olmakta ve öncelik vektörü olarak adlandırılan önem ağırlıkları w_i,

$$w_i = \left(\frac{1}{n}\right) \sum_{j=1}^{n} a'_{ij} \qquad i, j = 1,2, \dots, n \tag{3}$$

şeklinde hesaplanmaktadır.

Adım 5: Tutarlılık oranının hesaplanması: Kriterler arasında karşılaştırma sonucunu belirleyen değerlerin tutarlı olup olmadığının kontrol edilmesi gerekmektedir. λ_{max} özdeğer olmak üzere tutarlılık endeksi CI;

$$\lambda_{max} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\sum_{j=1}^{n} a_{ij} w_j}{w_i}\right] \tag{4}$$

$$CI = \frac{\lambda_{max} - n}{n-1} \tag{5}$$

olarak hesaplanmaktadır. Hesaplanan tutarlılık endeksinin değerlendirilmesi için "Random Indeks" (RI) değerlerinin bilinmesi gerekmektedir. Her matrisin boyutuna karşılık gelen bir random indeks değeri bulunmakta ve bu değerler Tablo 3'te gösterilmektedir (Kwiesielewicz vd., 2004). CI ve RI değerleri belirlendikten sonra tutarlılık oranı (CR),

$$CR = CI/RI \tag{6}$$

hesaplanmakta ve elde edilen değerin 0.10'dan küçük olması durumunda oluşturulan karar matrisinin tutarlı olduğuna karar verilmektedir.

Table 3. Random Indeksi

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.53	1.56	1.57	1.59

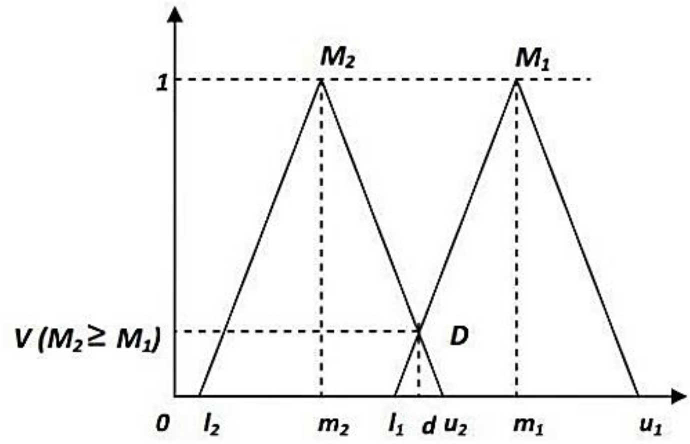
Adım 6: Seçeneklerin ikili karşılaştırması: Seçeneklerin tümünün her kritere göre önemini belirlemek için Eşitlik 1'de formüle edilen ikili karşılaştırma matrisleri tekrardan oluşturulmaktadır. Kriterlerin ağırlığının belirlendiği süreçte Eşitlik (2) ve (3) denklemleri kullanılarak tutarlılık oranları kontrol edilmelidir.

Adım 7: Seçeneklerin Sıralanması: Seçeneklerin amaca göre sıralamasının belirlenmesinde ana kriterler, varsa alt kriterlerin kendi aralarında ve seçeneklerin ise her bir kriterle karşılaştırmaları oluşturularak öncelik vektörlerinin belirlenmesi gerekmektedir. Her bir kritere ait genel ağırlık değerleri ile seçeneklerin o alt kritere göre olan tercih değerleri çarpılarak seçeneklerin ağırlıkları elde edilmektedir. Elde edilen seçeneklerin ağırlık değerleri toplanarak o seçeneğin önem sıralaması bulunmaktadır (Özbek, 2021).

BULANIK ANALİTİK HİYERARŞİ PROSESİ

Karmaşık karar verme sürecini basit karşılaştırmalara bölen AHP yöntemi, bilişsel faktörleri dikkate almamaktadır. Saaty'nin teorisinin bir uzantısı olan ve üçgen bulanık sayıların kullanıldığı bulanık AHP, klasik yöntemlerine göre karar verme sürecinde ağırlıkların daha iyi tanımlanasını sağlamaktadır. AHP, organizasyon hiyerarşisinde birden fazla değerlendiricinin koordinasyonuna ve sentezine yardımcı olan katılım odaklı bir metodolojidir. Yöntemdeki katılımcılar, 1-9 arasında bir ölçek kullanarak değerlendirme yapmaktadır. Nitel değerlendirmelerin nicelleştirilmesiyle kesin karara ulaşmak için üyelik fonksiyonundaki tüm olası değerleri göz önünde bulundurarak eksik veya kusurlu bilgiden kaynaklanan belirsizlikleri ortadan kaldırmak için bulanık mantık kullanımı önemli olmaktadır. (Sharma vd. 2014).

Chang (1992), ikili karşılaştırmanın sentetik kapsam değeri için üçgen bulanık sayılara dayalı ikili karşılaştırma ölçeğini işlemek için yeni bir yaklaşım sunmaktadır. Bu yöntemde ilk adım, Bulanık AHP ölçeği aracılığıyla ikili karşılaştırma için üçgen bulanık sayılar kullanılmakta ve daha sonra sentetik kapsam değerleri kullanılarak öncelik ağırlıklarının elde edilmesi için bir kapsam analizi gerçekleştirilmektedir. Üçgen bulanık sayılar arasındaki kesişimi gösteren Şekil 1, her bir sayı kümesinin üyelik fonksiyonunu temsil etmekte ve şekildeki üyelik fonksiyonları, kümelerin birbiriyle örtüşmesi ile oluşmaktadır



Şekil 1. Üçgen Bulanık Sayılar Arasındaki Kesişim

Üçgen bulanık sayılarda karar vericilere ait değerlendirmeler, M genellikle (l, m, u) olarak temsil edilmekte olup l alt sınır, m orta sınır ve u üst sınır olmak üzere üyelik fonksiyonu,

$$\mu M(x) = \begin{cases} \frac{x}{m-l} - \frac{l}{m-l}, x \epsilon [l, m] \\ \frac{x}{m-u} - \frac{u}{m-u}, x \epsilon [m, u] \\ 0, \; Diğer \end{cases} \tag{6}$$

denklemiyle açıklanmaktadır. Karar vericiler için belirsizlikleri ortadan kaldırmak amacıyla geliştirilmiş olan bulanık AHP ölçeği Tablo 1'de gösterilmektedir.

Tablo 1. Kriterlerin Ağırlıklarını ve Derecelendirmelerin Değerlerini Açıklayan Dilsel Değişkenler

Nitel Ölçek	Üçgen Bulanık Ölçek	Üçgen Bulanık Ölçek Karşıt Değeri
Eşit Önem	$\tilde{1}$	(1, 1, 1)
Orta Önemi	$\tilde{3}$	(1, 3, 5)
Güçlü Önemi	$\tilde{5}$	(3, 5, 7)
Çok Güçlü Önemi	$\tilde{7}$	(5, 7, 9)
Aşırı Derecede Önemli Olması	$\tilde{9}$	(7, 9, 9)
Ters Karşılaştırma Için Karşılıklar	$(\frac{1}{u_i}, \frac{1}{m_i}, \frac{1}{l_i})$	

Kaynak: Kabir & Hasin (2011a)

Üçgen bulanık sayı matrisi $A = \tilde{a}_{ij}'nin\ i$, j $= 1{,}2 \dots \dots$ n olmak üzere karar verici değerlendirmeleri (l_{ij}, m_{ij}, u_{ij}) olmaktadır. A karşılaştırma matrisini oluşturduktan sonra, A matrisinin öncelik vektörü oluşturulmaktadır. A matrisindeki her satır için bulanık sentetik derece değeri $\tilde{S}_i$,

$\sum_{j=1}^{m} \tilde{a}_{ij} = \left(\sum_{j=1}^{m} l_{ij}, \sum_{j=1}^{m} m_{ij}, \sum_{j=1}^{m} u_{ij}\right),\ \forall i = 1,2, \dots, n,$ ve

$$\left[\sum_{i=1}^{n} \sum_{j=1}^{m} \tilde{a}_{ij}\right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{m} u_{ij}}, \frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{m} m_{ij}}, \frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{m} l_{ij}}\right)$$

olmak üzere,

$$\tilde{S}_i = \sum_{j=1}^{m} \tilde{a}_{ij} \otimes \left[\sum_{i=1}^{n} \sum_{j=1}^{m} \tilde{a}_{ij}\right]^{-1} \tag{7}$$

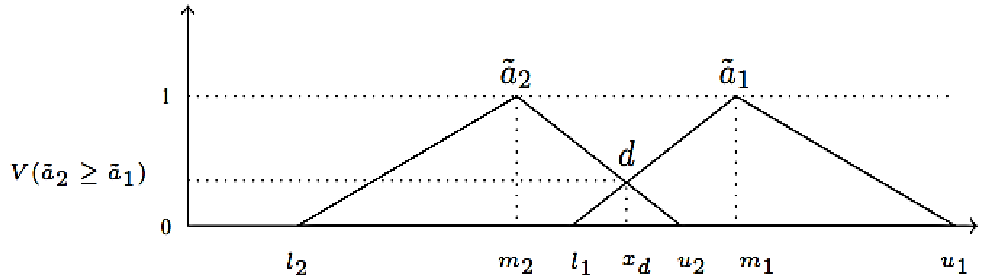
şeklinde hesaplanmaktadır. Sentetik değer belirlendikten sonra elde edilen bulanık sayı/sentetik değerin diğerinden daha büyük olma olasılığının derecesi belirlenmektedir. $\tilde{a}_1 = (l_1, m_1, u_1)$ ve $\tilde{a}_2 = (l_2, m_2, u_2)$ iki üçgen bulanık sayı olmak üzere $\tilde{a}_1$ ve $\tilde{a}_2$ arasındaki olasılık derecesi, $V(\tilde{a}_2 \geq \tilde{a}_1)$,

$$V(\tilde{a}_2 \geq \tilde{a}_1) = \begin{cases} 1 & m_2 \geq m_1 \\ 0 & l_1 \geq u_2 \\ \frac{l_1 - u_2}{m_2 - u_2 + m_1 - l_1} & Diğer \end{cases} \quad (8)$$

elde edilmekte ve dışbükey bulanık sayı için olasılık derecesi:

$$V(\tilde{a}_2 \geq \tilde{a}_1) = hgt(\tilde{a}_1 \cap \tilde{a}_2) = (\frac{l_1 - u_2}{m_2 - u_2 + m_1 - l_1} = d \quad (9)$$

şeklinde hesaplanmaktadır. Burada d; $\tilde{a}_1$ ve $\tilde{a}_2$ arasındaki en yüksek kesişim noktasının ordinatıdır. hgt terimi, $\tilde{a}_1$ ve $\tilde{a}_2$ 'nin kesişimindeki bulanık sayıların yüksekliğini göstermektedir ve Şekil 2'de gösterilmektedir.



Şekil 2. $\tilde{a}_2 \geq \tilde{a}_1$ için Olasılık Derecesi

Son olarak, aşağıdakiler varsayılarak kriterler için ağırlık vektörünü ölçülmektedir;

$$d'(A_i) = \min V(\tilde{S}_i \geq \tilde{S}_j), j = 1,2, \ldots, n, j \neq i \quad (10)$$

burada $A_i (i = 1, 2, \ldots, m)$, m karar alternatifidir ve n kriter sayısı olmak üzere ağırlık vektörü

$$W'(A_i) = (d'(A_1), d'(A_2), \ldots, d'(A_m))^T, A_i (i = 1, 2, \ldots, m), \quad (11)$$

elde edilmekte ve denklem normalleştirilerek W, bulanık olmayan sayı ve alternatiflerin düzenini temsil etmek üzere,

$$W(A_i) = (d(A_1), d(A_2), \ldots, d(A_m))^T \quad (12)$$

ağırlık vektörleri hesaplanmaktadır. Her kriter için normalleştirilmiş ağırlığı, normalleştirilmiş ağırlık vektörlerinin toplamına bölerek n kriter sayısı olmak üzere kriterlerin önem dereceleri W_{c_j},

$$W_{c_j} = \frac{W(A_i)}{\sum_{i=1}^{n} W(A_i)}, j = 1, \ldots, n \quad (13)$$

hesaplanmaktadır.

TARTIŞMA VE SONUÇ

Geleneksel AHP yönteminde, en iyi alternatifin seçilmesine ilişkin her seviye için ikili karşılaştırmalar dokuzlu ölçek kapsamında yapıldığından AHP uygulamasında eksiklikler bulunmaktadır. AHP yöntemindeki tam sayılı ölçek kesin bir yargılama yaratmakta ve karar vericilerin öznel yargıları, seçimleri ve tercihleri hesaplama sonucunda dengesizlik yaratmaktadır. AHP yöntemi, kişinin yargısını tam bir sayıya eşlendiğinden belirsizlikler hesaba katılmamaktadır. Nitel niteliklere ilişkin insan değerlendirmeleri öznel olduğundan kesin yargı oluşturmamaktadır. Bu nedenle geleneksel AHP, karar vericinin gereksinimlerini açık bir şekilde yakalamakta yetersiz görünmektedir (Kabir ve Hasin, 2011b). Bireylerin tercihlerindeki belirsizliği modellemek ve AHP'nin dilsel değişkenleri ele almadaki yetersizliğinin üstesinden gelmek amacıyla AHP yöntemine bulanık kümeler dahil edilerek Bulanık AHP yaklaşımı geliştirilmektedir. Tam bir sayı değerlendirmesi yerine üçgen sayıları kullanan Bulanık AHP yaklaşımı, karar verme sürecinin daha doğru bir şekilde tanımlanmasını sağlamakta ve belirsizliklerin ortadan kaldırılmasına yardımcı olmaktadır.

Kaynakça

Sharma MJ, Yu SJ, 2014. "Fuzzy analytic hierarchy process–based decision tree in identifying priority attributes for supply chain coordination", International Journal of Logistics Systems and Management, vol. 17, no. 1, pp. 46-65.

Kabir G, Hasin MAA, 2011a. “Multi-Criteria Inventory Classification through Integration of Fuzzy Analytic Hierarchy Process and Artificial Neural Network”, International Journal of Industrial and System Engineering (IJISE), Article in Press.

Kabir G, Hasin MAA, 2011b. “Evaluation of Customer Oriented Success Factors in Mobile Commerce Using Fuzzy AHP”, Journal of Industrial Engineering and Management, 4(2), 361- 386.

Chang DY, 1992. “Extent analysis and synthetic decision”, Optimization Techniques and Applications, 1, 352-355

Kızılırmak Havzasında Meteorolojik Kuraklık Analizi

Nurcan Menevşe[1*], **Kadir Yürekli**[1]

[1] Tokat Gaziosmanpaşa Üniversitesi, Ziraat Fakültesi, Biyosistem Mühendisliği Bölümü, 60250, Tokat, Turkey

*Corresponding author e-mail: menevsenurcan@gmail.com

Özet

Kuraklık, su kaynaklarını besleyen yağışlardaki önemli düşmelerin olduğu dönemler olarak tanımlanmaktadır. Kuraklık, taşkın gibi ani olarak meydana gelmemesine rağmen, dünyadaki doğa olayları içinde maliyeti en fazla olan ve küresel anlamda yıllık olarak ortalama 6-8 milyar dolar zarara neden olan doğa olayıdır, bunun yanında diğer doğa olayları ile karşılaştırıldığında insanlığı en fazla tehdit edendir (Wilhite, 2000). Bu doğrultuda bu çalışmada Kızılırmak havzası içinde bulunan 6 yağış istasyonunda (Kastomonu, Çankırı, Sivas, Kırıkkale, Yozgat ve Kırşehir) 1950-2019 yılları arasında kaydedilen aylık toplam yağmur miktarları kullanılarak McKee et al. (1993) tarafından geliştirilmiş Normalleştirilmiş Yağış İndeksi (SPI) yöntemine dayanarak 3 ve 12 aylık zaman dilimleri için kuraklık analizi gerçekleştirilmiştir. Kuraklık analizini gerçekleştirmeden önce verilerin homojen olup olmadıkları Mann-Whitney U testi ile kontrol edilmiştir. Çankırı ve Sivas istasyonlarının 10. aya ait verileri ile Yozgat ilinin 11. ayına ait aylık yağmur verilerinde homojenliğin bozulduğu tespit edilmiştir. Bahsedilen ayların aylık yağmur verilerinin homojenliği çift birikimli eğri yöntemi ile sağlanmıştır. İki ayrı zaman periyodu (3 ve 12) için yağış istasyonlarının SPI indisleri iki parametreli Gama dağılımına bağlı olarak gözlem süresi için elde edilmiştir. Havzada göz önüne alınan yağış istasyonlarının SPI-3 ve SPI-12 için elde edilen kuraklık indisleri değerleri çoğunlukla normal kuraklık sınıfındadır. Diğer taraftan ekstrem kuraklıklar, ekstrem ıslak dönemlerden daha fazla tecrübe edilmiştir. Üç aylık zaman periyodunda ekstrem kuraklıklar oniki aylık zaman periyodunkinden daha fazla meydana gelmiştir. Ekstrem yaş kuraklık kategorisinde oniki aylık zaman periyodunda üç aylık döneme göre daha fazla tekrarlanmıştır.

Key words: *Kuraklık, Standart Yağış İndisi, Kızılırmak Havzası*

Tokat Yöresi Topraklarının Bazı Özellikleri ile Aşınıma Duyarlılıklarının Karşılaştırılması

Rümeysa Armağan[1*], **İrfan Oğuz**[1*]

[1] Tokat Gaziosmanpaşa Üniversitesi, Ziraat Fakültesi, Toprak Bilimi ve Bitki Besleme Bölümü, 60250, Tokat, Türkiye

*Corresponding author e-mail: rrumeysaderin@gmail.com

Özet

Bu çalışmada Tokat Merkez ve Artova ilçelerinde yayılım gösteren 20 farklı lokasyondan alınan yüzey toprak örneklerine ait pH, EC, organik madde, kireç, tekstür, çok ince kum, agregat stabilitesi, hidrolik geçirgenlik, kil oranı I, kil oranı II, kil oranı III, dispersiyon oranı, süspansiyon yüzdesi ve toprak aşınıma duyarlılık faktör olmak üzere bazı fiziksel ve kimyasal özellikleri belirlenmiştir. Toprak aşınıma duyarlılık parametresi (USLE-K) kullanılmıştır. Aşınıma duyarlılık değerleri 0.07- 0.21 t ha-1 ha MJ-1h mm-1 arasında değişmiş, toprakların aşınabilirlik faktörleri birbirinden farklılık göstermiştir. Genel olarak yöre topraklarının erozyona karşı oldukça hassas olduğu sonucu elde edilmiştir. Toprakların aşınıma dayanıklılık durumunu gösteren K faktör, dispersiyon oranı, süspansiyon oranı ve kil oranlarının birbirleriyle uyumlu sonuçlar vermediği görülmüştür. Çalışma alanı toprakları özelliklerinin bireysel olarak K faktörü ile ilişkilendirilmesi uygun görülmüştür.

Key words: *Toprak özellikleri, Toprak aşınımı duyarlılığı, USLE-K, Tokat*

GİRİŞ

Toprak erozyonu, şiddetli yağışlar, toprak bozulması ve yoğun tarımsal faaliyetlerin sonucunda ciddi çevresel bir problem olarak karşımıza çıkmaktadır. Dünya üzerindeki toplam kara alanlarının yaklaşık %10,95'i toprak erozyonuna maruz kalmaktadır (Han, 2016). Muhtemelen gelecekte hidrolojik döngüyü yakından etkileyen iklim değişikliğinden dolayı dünyanın pek çok bölgesi toprak erozyonuna daha duyarlı bir hale gelecektir (Amore, 2004).

Türkiye'de, erozyon riski oldukça düşük olan araziler sadece %13,86 iken, şiddetli ve çok şiddetli erozyonun etkisinin görüldüğü arazilerin oranı ise yaklaşık %58,74'tür. Su erozyonuna maruz 57,15 milyon ha araziye karşılık, 506.309 ha alanda farklı düzeylerde rüzgâr erozyonu riski bulunmaktadır. İşlemeli tarım yapılan 27,7 milyon ha arazinin 16,4 milyon ha kısmında ana sorun olarak erozyondur. (Anonim, 1987 ve 1998). Erozyon sorunu yanında 1,08 milyon ha arazi taşlılık ve 5,40 milyon ha 'da çeşitli uygunsuz toprak şartlarının sorun olduğu görülmektedir.

Toprak erozyonu topraktaki bitki besin elementlerinin içeriklerini etkilemektedir. Verimli üst katmanın uzaklaşmasıyla organik karbon içeriği azalmakta ve toprak verimliliğinde azalmalar meydana gelmektedir. Sonuçta aşınıma bağlı olarak fiziksel ve kimyasal toprak özelliklerinde birtakım değişiklikler meydana gelmektedir. Benzer şekilde toprakların oluşumlarından kaynaklanan genetik özellikleri, toprakların erozyona karşı farklı direnç göstermesine yol açmaktadır.

Bu çalışmada, Tokat yöresi topraklarını temsil eden farklı yerlerden alınan yüzey toprak örneklerinin bazı fiziksel ve kimyasal özellikleri ile aşınıma duyarlılıkları dikkate alınarak karşılaştırılmıştır. Yöre topraklarının su erozyonuna direnç durumları ele alınmaya çalışılmıştır.

MATERYAL VE METOD

Çalışma, Tokat Merkez ve Artova ilçelerinde yayılım gösteren topraklarda yürütülmüştür. Karadeniz Bölgesi'nde yer alan Tokat İli, kuzeyinde Samsun kuzeydoğusunda Ordu, güney ve güneydoğusunda Sivas, güneybatısında Yozgat, batısında Amasya iliyle çevrilidir. Coğrafi koordinatları, 39° 51' – 40° 55' kuzey enlemleri ile 35° 27'- 37° 39' Doğu boylamları arasındadır. Toplam yüzölçümü 10.044 km^2'dir. Ortalama yüksekliği 1050 m olan Tokat ilinde karasal iklim hakimdir (Altunsu, 2019). Toprak ve Su Kaynakları Tokat Araştırma Enstitüsü meteoroloji istasyonunun uzun yıllar (1966-2006) yağış verilerine göre, en yüksek yağış 616.3 mm ile 1996 yılında, en düşük yağış ise 271.1 mm ile 2001 yılında gözlemlenmiştir. (Oğuz, 2008)

Toprak örnekleri Tokat yöresinde erozyona maruz 20 farklı lokasyonda, bozulmuş ve bozulmamış yüzey toprak örneklerinin (0-20 cm) alınarak laboratuvarda analizleri sonucunda gerçekleştirilmiştir. Arazi çalışmaları ile farklı lokasyonlardan alınan yüzey toprak örnekleri önce kurutulmuş ve daha sonra 2 mm'lik elekten elenerek analize hazır hale getirilmiştir. Bozulmamış toprak örneklerinde toprakların hidrolik iletkenlikleri, sabit ve değişken su seviyeli hidrolik geçirgenlik setiyle (Klute ve Dirksen, 1986); organik madde, Walkley-Black yöntemiyle (Kacar, 1994); pH 1:1 oranında hazırlanan toprak-su süspansiyonunda cam elektrotlu pH metre ile (Soil Survey Laboratory, 1992); kireç, Scheibler Kalsimetresi ile (Kacar, 1994); elektriksel iletkenlik 1:1 oranında hazırlanan toprak-su süspansiyonunda elektriksel kondaktivite aleti ile (Richards, 1954); toprak strüktürü arazide doğrudan belirlenmiş; tekstür hidrometre yöntemi ile (Bouyoucous, 1951); kil Oranı I, mekanik analizle belirlenen kum+silt % değerinin, kil % değerine oranlanmasıyla (Chandra,1978), kil oranı II, kil % değerinin kum+silt % değerine oranlanmasıyla (Römkens,1985); kil Oranı III, kum % değerinin kil+silt % değerine oranlanması (Ngatunga et al, 1984) ile; dispersiyon oranı, süspansiyonda dispers edilmeden ölçülen silt+kil % değerinin, mekanik analizde ölçülen silt+kil % değerine oranlanmasıyla (Lal,1988); agregat Stabilitesi ıslak eleme yöntemiyle (Kemper ve Rosenau, 1986) belirlenmiştir. Aşınıma duyarlılık faktörü (K), her bir noktasal toprak örneği için olmak üzere aşağıdaki ampirik eşitlik ile belirlenmiştir (Wischmeier ve Smith, 1978).

$$\mathbf{100 = ((2.7x10{-}4)x(M1{,}14)x(12-a) + 3.25x(b-2) + 2.5x(c-3))x0.1317} \qquad \text{(Eşitlik 1)}$$

Eşitlikte; K: Toprak aşınım faktörü; M: Zerre irilik parametresi, a: Organik madde içeriği, %; b: Strüktür tipi kodu; c:Su geçirgenliği kodu. Eşitlikte yer alan zerre irilik (M) parametresi aşağıdaki eşitlik yardımıyla belirlenmiştir.

$$\mathbf{M = (Çok\ ince\ kum + Silt)x(100 - Kil)} \qquad \text{(Eşitlik 2)}$$

BULGULAR VE TARTIŞMA

Toprak örneklerinde yapılan analiz sonuçları ve hesaplanan çeşitli aşınıma duyarlılık parametreleri Tablo 1 ve Tablo 2'de verilmiştir.

Araştırma yeri topraklarının pH değeri 7.0-7.9 arasında değişmiş, hafif alkali reaksiyona sahiptir. Toprakların EC değeri 0.06-0.21 dSm^{-1} arasında değişmiş, tuz içeriği tuzsuzdur. Organik madde kapsamları %0.28-6.36 değerleri arasında değişim göstermiştir. Toprak örneklerinin 2 adeti çok az (% 0-1), 7 adeti az (% 1-2), 6 adeti orta (% 2-3) ve 5 adeti çok yüksek (% 4<) organik madde sınıfında yer almıştır. Toprakların kireç içerikleri %3.84-83.77 arasında değişmiş olup, 1 adet az (% 1.0-5.0), 3 adet orta (% 5.0-15.0), 5 adet fazla (% 15.0-25.0) ve 9 adet toprak örneği yüksek kireç (>% 25) içeriğine sahiptir (Tablo 1).

Tablo 1. Çalışma yeri toprakları kimyasal özellikleri

Örnek No	pH	EC (dSm^{-1})	Organik Madde (%)	Kireç (%)
1	7.1	0.09	6.39	16.49
2	7.3	0.12	4.26	15.45
3	7.2	0.18	4.20	19.66
4	7.3	0.09	1.84	9.80
5	7.3	0.09	1.57	33.52
6	7.2	0.11	2.93	12.97
7	7.5	0.07	2.66	7.44
8	7.6	0.06	0.28	13.34
9	7.5	0.08	1.03	34.36
10	7.5	0.04	1.62	3.84
11	7.1	0.19	6.71	21.37
12	7.6	0.10	1.76	15.26
13	7.3	0.05	1.71	5.70
14	7.2	0.11	2.34	25.26
15	7.4	0.11	1.54	37.75
16	7.5	0.09	2.87	57.69
17	7.1	0.21	4.51	57.80
18	7.6	0.09	2.77	70.06
19	7.5	0.09	0.75	83.77
20	7.6	0.10	2.75	69.57

Araştırma yeri topraklarının bazı fiziksel özellikleri ve aşınıma duyarlılık eğilimleri incelenmiştir (Tablo 2). Toprakların kil, silt, kum içerikleri sırasıyla % 12.56-46.56, % 10-28, % 31.44-77.44 arasında değişmiştir. Toprakların tekstürel fraksiyonu olan çok ince kum değerleri ise % 0.97-3.84 arasında değişim göstermiştir. Agregat stabilite değerleri % 23.15- 82.91 değerleri arasında değişmiştir. Toprak örneklerinin bir kısmı oldukça düşük agregat stabilitesi göstermiştir. Toprakların hidrolik geçirgenlik değerleri 0.00-4.13 cm h^{-1} arasında değişim göstermiş olup çok yavaş, yavaş, orta yavaş ve orta hidrolik kondaktivite değerlerine sahiptir. Bu nedenle yöre toprakları yüksek yüzey akış eğilimine sahip olduğu söylenebilir.

Çalışma kapsamında değerlendirilen toprakların erozyon eğilimleri Kil Oranı I, II, III, Dispersiyon Oranı, Süspansiyon Yüzdesi ve K Faktör indisleri kullanılarak araştırılmıştır. Kil oranı I, toprakların kum ve silt içeriklerinin kil içeriklerine oranı olup, çalışma yeri topraklarında bu oran 1.15-5.04 arasında değişmiştir. Kil oranı II, toprakların kil içeriklerinin kum ve silt içeriklerine oranlanmasıyla belirlenmiş ve çalışma yeri topraklarında 0.14-0.74 arasında değişmiştir. Kil oranı III değeri ise, kum içeriklerinin kil ve silt içeriklerine oranlanmasıyla belirlenmiş olup bu değer çalışma yeri toprakları için 16.2-29.4 arasında değişmiştir. Artan kil oranı I değeri toprakların aşınıma duyarlılığını artırırken, kil oranı II ve kil oranı III değerleri arttıkça toprakların aşınıma duyarlılığı azalmaktadır (Altunsu ve ark., 2019).

Erozyon eğilimini tanımlamak için toprakların dispersiyon oranları belirlenmiş ve bu değerin %15'ten büyük toprakların erozyona dayanıksız, küçük olanların ise dayanıklı olduğu bildirilmektedir (Sönmez 1994). Çalışma yeri topraklarının dispersiyon oranı değerleri 12.48-64.54 arasında değişmiş ve çoğunlukla erozyona dayanıksız topraklar olduğu belirlenmiştir. Süspansiyon yüzdesi su ile dispers edilmiş toprağın süspansiyona geçebilme özelliğini tanımlar. Bu değerin büyümesi toprakların erozyonla daha kolay taşınabileceği anlamına gelmekte olup, toprakların süspansiyon yüzde değerleri 3.28-12.28 arasında değişmiştir. Oğuz ve ark. (2020) Sivas Ulaş rüzgar erozyon sahasında yaptıkları çalışmada, süspansiyon yüzdesi değerlerinin % 24.2-37.2 olarak belirlemiştir.

Tablo 2. Çalışma yeri toprakları fiziksel ve aşınıma duyarlılık özellikleri

Örnek No	Kil,%	Silt,%	Kum,%	ÇİK	AS,%	HG, cm h^{-1}	KOI	KOII	KOIII	DO,%	SY,%	K Faktör
1	32.56	22.00	45.44	1.74	82.91	0.11	2.07	0.48	23.40	45.01	12.28	0.16
2	24.56	22.00	53.44	3.04	40.07	0.50	3.07	0.33	24.20	44.16	10.28	0.07
3	26.56	25.00	48.44	3.12	28.93	1.74	2.77	0.36	26.80	36.00	9.28	0.09
4	38.56	26.00	35.44	1.98	40.43	0.01	1.59	0.63	26.90	31.85	10.28	0.21
5	28.56	24.00	47.44	2.84	24.48	0.68	2.50	0.40	25.70	39.12	10.28	0.18
6	28.56	24.00	47.44	2.25	40.33	3.64	2.50	0.40	25.70	27.70	7.28	0.10
7	30.56	24.00	45.44	3.84	23.15	1.39	2.27	0.44	25.50	19.35	5.28	0.17
8	46.56	22.00	31.44	2.23	26.40	0.02	1.15	0.87	22.70	18.32	6.28	0.19
9	30.56	28.00	41.44	1.84	42.06	0.54	2.27	0.44	29.40	18.03	5.28	0.21
10	12.56	10.00	77.44	0.97	45.14	0.54	6.96	0.14	16.20	64.54	7.28	0.14
11	28.56	24.00	47.44	2.40	46.71	1.09	2.50	0.40	25.70	12.48	3.28	0.09
12	34.56	19.00	46.44	1.52	30.83	1.70	1.89	0.53	20.30	30.92	8.28	0.14
13	16.56	16.00	67.44	2.58	32.75	1.48	5.04	0.20	20.10	50.86	8.28	0.15
14	40.56	22.00	37.44	2.54	26.47	0.00	1.47	0.68	22.90	29.67	9.28	0.19
15	42.56	22.00	35.44	1.52	45.22	0.13	1.35	0.74	22.80	34.94	11.28	0.16
16	42.56	26.00	31.44	1.79	52.26	4.13	1.35	0.74	26.70	29.99	10.28	0.15
17	34.56	20.00	45.44	2.14	38.20	0.24	1.89	0.53	21.30	34.02	9.28	0.11
18	22.56	14.00	63.44	2.10	27.50	0.45	3.43	0.29	16.80	50.77	9.28	0.15
19	40.56	18.00	41.44	1.47	43.68	0.53	1.47	0.68	19.00	24.86	7.28	0.16
20	42.56	24.00	33.44	1.35	25.91	0.00	1.35	0.74	24.80	24.88	8.28	0.19

Kısaltmalar:ÇİK, Çok İnce Kum; AS,Agregat Stabilitesi; HG,Hidrolik Geçirgenlik; KOI, Kil Oranı I; KOII, Kil Oranı II; KOIII, Kil Oranı III; DO, Dispersiyon Oranı; SY, Süspansiyon Yüzdesi; K,Aşınıma Duyarlılık Faktörü.

Yüzey toprak örneklerinin bazı tanımlayıcı istatistikleri Tablo 3'de verilmiştir. Araştırma yeri topraklarının kimyasal özelliklerinden ortalama olarak pH % 7.37, EC değeri % 0.17, organik madde % 2,72 ve kireç içeriği % 30,56 olmuştur. Toprak fiziksel özelliklerinden ortalama olarak kil %32.26, silt % 21.60, kum % 46.14, çok ince kum % 2.18, agregat stabilitesi % 38.17 ve hidrolik geçirgenlik 0,95 cm h^{-1} olmuştur. Araştırma yeri topraklarının ortalama olarak kil oranı I % 2.44, Kil oranı II % 0.50, Kil oranı III % 23.34, dispersiyon oranı % 33.37, süspansiyon yüzdesi % 8.43 ve K Faktör değeri % 0.15 olarak belirlenmiştir.

Toprak örneklerinin çeşitli özellikleri arasındaki farklılıklar standart sapma, varyans ve değişim katsayısı kriterleriyle karşılaştırılmıştır. Toprakların çeşitli özelliklerine ait değişim katsayısı değerleri, düşük (<%15), orta (%15-%35) ve yüksek (>%35) olarak sınıflandırılmıştır (Mallants et al. 1996). Buna göre, pH değerleri düşük, kil, silt, kum, kil oranı III, süspansiyon yüzdesi, çok ince kum, K Faktör değerleri orta derecede ve kil oranı I, kil oranı II, organik madde, agregat stabilitesi, hidrolik geçirgenlik, dispersiyon oranı, kireç ve EC değerleri yüksek derece değişkenlik göstermiştir.

Normal dağılımda simetrikliğin bozulma derecesi olan çarpıklık (skewness) (Saygın,2018) değerleri incelenmiştir. Çarpıklık değeri normal dağılımda 0 olarak varsayılır. Bir dağılımın çarpıklık katsayısı nagatif ise dağılım sağa, pozitif ise sola çarpıktır. ± 2 saçılımı çoğu istatistikçi tarafından normal dağılım olarak varsayılmaktadır. Tablo 3 incelendiğinde çarpıklık katsayısı kil, silt, kum, kil oranı II, kil oranı III, organik madde, agregat stabilitesi, hidrolik geçirgenlik, dispersiyon oranı, süspansiyon yüzdesi, çok ince kum, kireç, pH, EC ve K Faktör değerleri için ± 2 saçılımı içerisine dahil olduğu için normal dağılım gösterirken, diğer değerler normal dağılımdan uzaktır. Kil, silt, kil oranı III, süspansiyon yüzdesi, pH ve K Faktör değerleri sağa çarpık, diğer toprak özellikleri ise sola çarpıktır.

Normal dağılım eğrisinin sivrilik veya yuvarlaklık derecesine basıklık (kurtosis) denir (Yıldız ve ark. 1998). Basıklık değeri <0 ise dağılım normalden basık, yayvan ve verilerin heterojen olduğuna işaret ederken, sıfıra eşit olduğu değerler dağılımın basıklığı normal dağılım kadar olduğu ve >0 değerler ise dağılım normalden sivridir ve verilerin homojen olduğunu göstermekte olup ± 2 saçılımını normal olarak kabul edilmektedir. Basıklık değerlerine göre kil, kil oranı II, kil oranı III, kireç, pH ve K Faktör değerlerinin normalden basık, yayvan ve hetorojen olduğu görülmektedir (Tablo 3). Kil oranı I, agregat stabilitesi, hidrolik geçirgenlik değerleri ± 2 saçılım sınırında kalmakta ve önemsiz düzeyde normal dağılıma uygun olacak şekilde heterojenlik göstermiştir. Basıklık değeri ± 2 değerini aşan toprak özellikleri kil oranı I, agregat stabilitesi, hidrolik geçirgenlik olmuştur.

Tablo 3. Çeşitli toprak özelliklerine ait bazı tanımlayıcı istatistikler

Özellik	Örn.	Minimum	Maximum	Ortalama	Standart Sapma	Varyans	Değişim Katsayısı	çarpıklık	basıklık
Kil	20	13	47	32.26	9.11	83.063	28.24	-0.43	-0.272
Silt	20	10	28	21.60	4.39	19.305	20.32	-1.17	1.379
Kum	20	31	77	46.14	12	144.221	26.00	1.16	1.359
KOI	20	1	7	2.44	1.40	1.964	57.38	2.16	5.297
KOII	20	0.14	0.87	0.50	0.20	0.039	40	0.07	-0.732
KOIII	20	16	29	23.34	3.51	12.331	15.04	-0.54	-0.259
OM	20	0.28	6.71	2.72	1.73	2.984	63.60	1.03	0.680
AS	20	23	83	38.17	13.67	186.922	35.82	1.84	5.201
HG	20	0	4	0.95	1.16	1.342	122	1.80	2.907
DO	20	12	65	33.37	12.93	167.361	38.74	0.64	0.354
SY	20	3	12	8.43	2.23	4.976	26.45	-0.56	0.141
CIK	20	1	4	2.18	0.68	0.465	31.19	0.64	0.527
KIREÇ	20	4	84	30.56	24.34	592.707	79.67	0.95	-0.335
pH	20	7	8	7.37	0.18	0.033	2.45	-0.15	-1.438
EC	20	40	210	0.17	0.07	0.005	41.17	0.59	1.605
K	20	0.07	0.21	0.15	0.04	0.002	26	-0.25	-0.431

Kısaltmalar:KOI, Kil Oranı I; KOII, Kil Oranı II; KOIII, Kil Oranı III, OM; Organik Madde, AS; Agregat Stabilitesi, HG;Hidrolik Geçirgenlik ,DO;Dispersiyon Oranı, SY; Süspansiyon Yüzdesi, ÇİK; Çok İnce Kum, K;Aşınıma Duyarlılık Faktörü.

Toprakların çeşitli özellikleri arasındaki korelasyonlar araştırılmış ve Tablo 4'te verilmiştir. EC ve diğer toprak özellikleri arasında korelasyon belirlenememiştir. Araştırma yeri topraklarının organik madde içerikleri ile çeşitli toprak özellikleri arasındaki korelasyonlar Tablo 4'te verilmiştir. Organik maddede azalmanın agregat stabilitesinde azalmaya yol açtığı bildirilmektedir (Castro Filho ve ark., 1998).

Benzeri olarak çalışma yeri topraklarının organik madde kapsamları ile agregat stabilitesi arasında (0,482*) önemli pozitif bir korelasyon elde edilmiştir (Şekil 1). Toprağın organik madde düzeyindeki artış organik maddenin toprak zerreleri bağlayıcı etkisi agregatların suya daha fazla dayanıklı olmasına neden olmuştur. Organik madde kapsamları ile toprağın aşınıma duyarlılık faktörü arasındaki negatif ilişki ise organik maddenin agregat stabilitesi üzerinde göstermiş olduğu olumlu etkiye benzer olduğu değerlendirilmiştir. Organik madde toprakların su erozyonuna karşı daha stabil agregat oluşturmasını sağlayarak K faktör değerlerinin düşmesine ve erozyona karşı daha dirençli olmasına neden olmuştur. Kanarya adalarında yer alan Andosollerde doğal orman vejetasyonu altında yürütülen çalışma bulgusuna göre organik maddenin aşınıma duyarlığı (K faktör) azalttığı ve agregat stabilitesini arttırdığı rapor edilmiştir (Rodriguez ve ark., 2006). Organik madde kapsamı ile toprak pH değerleri arasında belirlenen negatif çok önemli korelasyon organik maddenin ayrışmasıyla açığa çıkan organik asitlerin toprak reaksiyonunu etkilemesinden kaynaklandığı düşünülmektedir. Bu etki alkali reaksiyonlu yöre topraklarının organik madde düzeylerinde iyileştirilmeler yapılması durumunda toprakların pH değerlerinde azalma eğilimi sağlayacağını göstermektedir. Birçok çalışmada artan organik madde kapsamının toprakların pH değerlerinde düşüşe yol açtığına dair bulgular rapor edilmiştir (Kütük ve ark. 1999; Sharif ve ark. 2004; Doğan 2000). Erzurum Doğu Anadolu Tarımsal Araştırma Enstitüsü Toprak ve Su Kaynakları arazisinde yürütülen bir başka araştırmada ise, yarasa gübresinin geleneksel yöntem olan çiftlik gübresi ile karşılaştırılmış ve bu gübrelerin toprakların bazı makroelement içeriği, pH ve katyon değişim kapasitesine etkisi araştırılmıştır. Her iki organik materyal, ağırlık esasına göre artan dozlarda 0 (kontrol); 500; 1000; 1500; 2000 kg.da-1 olmak üzere 5 uygulama seviyesinde uygulanmış ve test bitkisi olarak buğday kullanılmıştır. Araştırmanın yarasa ve çiftlik gübresi sonuçları kontrol konusu ile karşılaştırıldığında, en yüksek yarasa gübresi dozu toprakların azot içeriğini (N) % 32, fosfor içeriğini (P) % 407, potasyum içeriğini (K) % 61, pH % 3 ve katyon değişim kapasitesini (KDK) ise %17 arttırmıştır (Karagöz Sezer ve Hanay, 2020). Hatay-Kırıkhan'da yürütülen araştırmada ise organik madde ile pH arasında istatistiksel ilişki bulunmamıştır (Yeter ve Yalçın, 2020).

Kireç içeriklerinin çok yüksek olması toprakların su tutma kapasitesi ve diğer bazı özelliklerine olumsuz etkisi olmakla birlikte organik maddesi düşük topraklara kireç ilavesinin strüktürü düzelterek, agregasyonu ve erozyona dayanıklılığı artırdığını belirtilmiştir Russell (1973). Çalışmamızda toprakların kireç içerikleri ile diğer toprak özellikleri arasında anlamlı korelasyon belirlenmemiştir.

Toprakların kil içeriği ile silt ve kil oranı II arasında pozitif, kum, kil oranı I ve dispersiyon oranı arasında negatif korelasyon bulunmuştur. Ele alınan parametrelerden diğerleri ile toprakların kil içeriği arasında bir ilişki bulunamamıştır. Toprakların tekstürel fraksiyonları kum, kil ve silt olup bu değerlerin toplamı 100 olduğu için aralarında pozitif ve negatif ilişkilerin olması normaldir. Artan kil içeriği ile kil oranı I ve kil oranı II arasındaki istatiksel korelasyon, söz konusu indekslerin hesaplanmasında, toprakların kil fraksiyonun bir parametre olarak yer almasından kaynaklanmaktadır.

Toprakların çok ince kum içeriği ile aşınıma duyarlılığı arasında ilişkiler bulunmaktadır. Toprakların çok ince kum içeriği zerre iriliğinin silt büyüklüğüne en yakın tekstürel büyüklük olması nedeniyle hem yeterince agregatlaşmada önemli rol oynamaması ve hem de zerre iriliğinin diğer kum fraksiyonlarına göre daha düşük olması nedeniyle inflitrasyon özelliklerinin daha düşük olması gibi nedenlerle yüzey akış ve erozyon riskini artırmaktadır. Bununla birlikte değerlendirmeye almış olduğumuz toprakların çok ince kum içerikleri ile diğer toprak özellikleri arasında korelasyon belirlenememiştir.

Toprak aşınıma duyarlılık özellikleri Eşitlik 1 yardımıyla belirlenmiş ancak toprak örneklerinde ele alınan toprak özellikleri ile K Faktör değerleri arasında korelasyon bulunmamıştır. Yarı kurak bir alanda sulama ve arazi kullanımındaki değişikliklerin toprak özelliklerine ve toprak aşınabilirliğine etkisinin araştırıldığı bir çalışmada tarım, otlak, yonca ve zeytin bahçesi arazi kullanımları incelenmiştir. Çalışma sonucunda yoğun toprak işlemenin toprak organik maddesi, kil ve azot içeriğini önemli ölçüde azalttığı, K faktörünün ise yoğun ekim ile arttığı rapor edilmiştir (Ferreira ve ark., 2015).

Toprakların agregat dağılımları ve stabilite ölçümleri bir kalite göstergesi olarak Kabul edilmektedir (Six ve ark. 2000). Agregat stabilitesi ölçümleri toprak agregatlarının bozulmaya neden olan çevresel etmenlere karşı direncinin belirlenmesinde önemli bir parametre olarak kabul edilmektedir. (Hillel 1982). Toprakların agregat stabilitelerinin yüksek olması istenilen durumdur. Strüktürel bakımdan sağlam yapılı topraklarda suyun dispers edici etkisine karşı agregatların stabiliteleri de yüksek olacağından bu topraklarda erozyona karşı duyarlılık derecesi de düşüktür (Karaman ve ark., 2007). Çalışmamızda toprakların agregat stabiliteleri ile ele alınan diğer toprak özellikleri arasında korelasyon bulunmamıştır.

Toprakların dispersiyon oranı ile kil içeriği arasında çok önemli negatif bir korelasyon bulunmuştur. Artan kil içeriği toprakların dispersiyon oranlarının istatiksel olarak azalmasına yol açmıştır. Kil içeriğindeki artışlar toprakların daha sağlam agragat oluşturması yönünde katkı sağlamış ve sonuçta dispersleşmeye karşı daha dayanıklı bir durum gerçekleştirmiştir. Kahramanmaraş ili sınırları içerisindeki yer alan Çemrengeç yağış havzasında, farklı anakayalar üzerinde oluşan toprakların aşınım özelliklerinin araştırıldığı çalışmada, dispersiyon oranı ile kil içeriği arasında mevcut çalışmamızla benzer olarak negatif ilişki bulmuşlardır.

Silt içeriği ile dispersiyon oranı arasında çok önemli negatif bir korelasyon bulunmuştur. Toprakların silt içeriğinin yüksek olması durumunda toprakların yapısal dayanıklılığını olumsuz etkilediği, inflitrasyonu azaltarak yüzey akışı artırdığı ve zerre iriliğinin düşük olması nedeniyle kolaylıkla taşındığı bilinmektedir. Silt içeriği yüksek topraklar erozyona oldukça hassastırlar. Ancak çalışmamızda analiz edilen toprakların silt içerikleri % 10-28 arasında değişmiş ve tekstürel fraksiyon içerisinde dengeli ve olumlu bir dağılım göstermiştir. Bu durum kum, kil ve silt içeriğinin dengeli dağılımının bir neticesi olarak toprakların yapısal dayanıklılığının artmasına ve dolayısıyla dispersiyon oranı ile arasında negatif çok önemli ilişkiye neden olduğu düşünülmektedir.

Kum içeriği ile kil oranı I arasında pozitif, kil oranı II arasında negatif bir ilişki bulunmuştur. Bu durum kil oranı parametrelerinin toprak tekstürel fraksiyonlarından hesaplanmasının doğal bir sonucu olarak değerlendirilmiştir. Topraklarda artan kum içeriği, dispersiyon oranı değerlerini arttırmıştır. Toprakların kum fraksiyonu agregatlaşma üzerinde çok fazla etkili olmadığı için bu durum beklenilen bir sonuçtur. Volkanik Süphan Dağı ile Muş-Bulanık ilçesi arasında kalan tarım topraklarının bazı fiziksel özellikleri ile aşınabilirlik göstergeleri arasındaki ilişkilerin araştırıldığı bir çalışmada, alt toprakların dispersiyon oranı ile % kum içeriği arasında negatif bir ilişki tespit edilmiştir.

Kil oranı I ile kil oranı II ve kil oranı III arasında negatif ilişki bulunmuştur. Bu sonuç her üç oran değerlerinin de toprakların tekstürel dağılıma bağlı olması nedeniyle beklenen bir sonuç olarak değerlendirilmektedir. Kil oranı I ile dispersiyon oranı arasında pozitif bir ilişki bulunmaktadır. Bu pozitif ilişki kil oranı I değerindeki artışa kil miktarının azalmasının yol açması ve kil miktarındaki azalmanın dispersiyon oranı değerlerinde artışa yol açması ile açıklanabilir. Kil oranı II ve kil oranı III ile dispersiyon oranı arasında çok önemli negatif korelasyon bulunması, kil oranı II değerindeki artışın % kil miktarının artmasına bağlı olması ve sonuçta bu durumun dispersiyon oranını azaltmasına yol açmasından kaynaklandığı görülmektedir.

Süspansiyon yüzdesi ile dispersiyon oranı arasında pozitif bir korelasyon bulunmuştur. Dispersiyon oranı 15’ten büyük olan topraklar erozyona karşı dayanıksız sayılmaktadırlar (Bryan,1968; Lal, 1988).

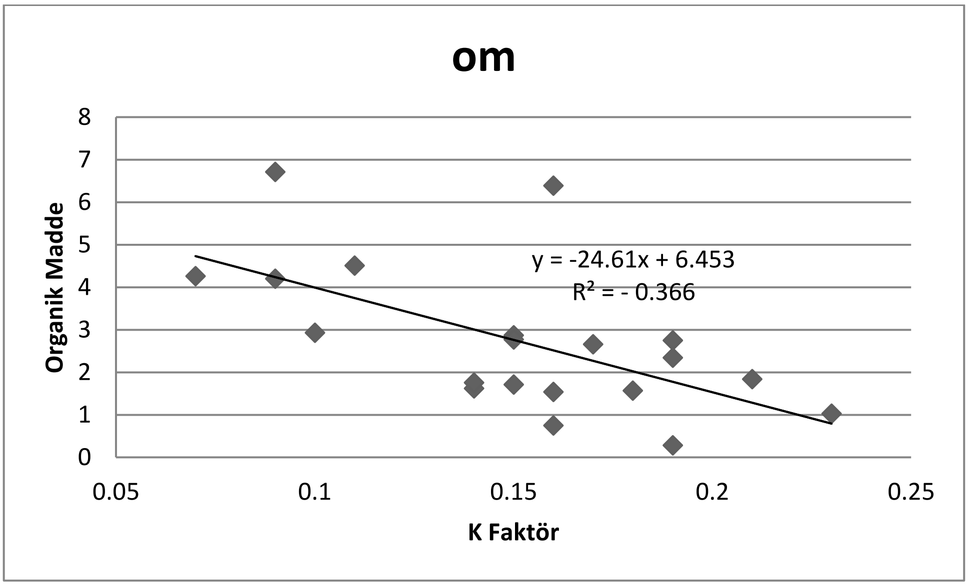
Yüksek dispersiyon oranı gösteren, strüktürel olarak zayıf topraklar 40 saniye okumalarında süspansiyonda kalan zerrelerinde oransal olarak fazla olmasına yol açmıştır. Erozyona dayanıklı düşük süspansiyon yüzdesine sahip topraklar iyi agregatlaştıkları için süspansiyon yüzdesi belirlenmelerinde, hızla Bouycos silindirinde çökelerek süspansiyonda daha az kalma eğilimi göstermektedirler. Araştırmamızda ele alınan toprak özelliklerinden süspansiyon yüzdesi ile EC değeri arasında pozitif korelasyon bulunmuştur. Artan EC ve süspansiyon yüzdesi arasındaki pozitif ilişki toprakların artan tuz ve elektrolit içeriklerinin strüktürel özelliklerine olan olası olumsuz etkisinden kaynaklandığı değerlendirilmiştir.

Toprakların aşınıma duyarlılıkları ile toprak reaksiyonu arasında önemli pozitif bir ilişki bulunmuştur. Toprak killerinin dispersiyonu, düşük EC ve yüksek sodyum adsorpsiyon oranı (SAR) ve pH değerleri ile artmakta olduğu bildirilmektedir (Reinks ve diğerleri 2000).

Tablo 4. Çeşitli toprak özellikleri arasındaki korelasyon katsayıları

Parametreler	Kil	Silt	Kum	KOI	KOII	KOIII	OM	AS	HG	DO%	SY%	ÇİK	Kireç	K	pH	EC
Kil	1	0.523*	-0.950**	-0.917**	0.990**	0.32	-0.2	0.001	-0.11	-0.652**	0.137	-0.18	0.387	0.415	0.18	0.021
Silt		1	-0.762**	-0.693**	0.448*	0.968**	0.163	0.03	0.202	-0.712**	-0.03	0.356	-0.033	0.198	-0.228	0.074
Kum			1	0.949**	-0.915**	-0.597**	0.09	-0.01	0.01	0.755**	-0.09	0.006	-0.282	-0.387	-0.053	-0.043
KOI				1	-0.857**	-0.491*	-0.003	0.01	0.04	0.742**	-0.13	-0.06	-0.379	-0.282	-0.009	-0.101
KOII					1	0.252	-0.26	-0.02	-0.13	-0.601**	0.131	-0.23	0.374	0.435	0.241	-0.002
KOIII						1	0.19	0.042	0.255	-0.605**	-0.09	0.4	-0.171	0.151	-0.272	0.049
OM							1	0.482*	0.081	0.012	0.088	0.235	-0.082	-0.605**	-0.686**	0.186
AS								1	0.064	0.176	0.306	-0.44	-0.039	-0.121	-0.375	0.097
HG									1	-0.099	-0.1	0.123	-0.079	-0.376	-0.015	-0.037
DO%										1	0.542*	-0.2	-0.144	-0.217	-0.046	0.286
SY%											1	-0.14	0.13	-0.003	-0.197	0.594**
ÇİK												1	-0.33	-0.239	-0.307	0.001
Kireç													1	0.124	0.286	0.094
K														1	0.445*	-0.072
pH															1	-0.192
EC																1

Kısaltmalar;KOI, Kil OranıI; KOII, Kil OranıII;KOIII, Kil OranıIII,OM, Organik Madde;AS,Agregat Stabilitesi;HG,Hidrolik Geçirgenlik; DO;Dispersiyon Oranı, SY; Süspansiyon Yüzdesi ,ÇİK; Çok İnce Kum, K;Aşınıma Duyarlılık Faktörü, EC; Elektriksel İletkenlik.



Şekil 1. Toprak organik maddesi ile toprak aşınıma duyarlılığı (K faktör) arasındaki ilişki

SONUÇ VE ÖNERİLER

Tokat yöresini temsilen Tokat merkez ve Artova ilçelerine ait farklı 20 adet lokasyonda alınan yüzey toprak örneklerinde yapılan bazı fiziksel ve kimyasal analizlerin sonuçları değerlendirilmiştir. Genel olarak yöre topraklarının erozyona karşı oldukça hassas olduğu görülmüştür. Toprakların aşınıma dayanıklılık durumunu gösteren K faktör, dispersiyon oranı, süspansiyon oranı ve kil oranlarının birbirleriyle uyumlu sonuçlar vermediği görülmüştür. Yöre topraklarının su erozyonuna dirençlerini artırmak için toprakların daha stabil agregat oluşturmasını sağlayacak toprak yönetimi uygulamalarının gerçekleştirilmesi yararlı olacaktır. Yöre toprakların erozyon duyarlılıklarını azaltmak ve K faktörü değerini düşürmek için toprak koruma yöntemi olarak bitkisel yöntemlere ağırlık verilmelidir. Bitki ekim nöbeti uygulanmalı, toprağa organik madde ilave edilmelidir. İmkanlarla orantılı olarak, toprağı çabuk örten ve uzun süre toprak üzerinde kalan bitki türleri seçilmeye teşvik edilmelidir.

Referanslar

Castro Filho C, Muzilli O, Podanoschi AL, 1998. Estabilidade de agregados e sua relac, ao com o teor de carbono organico num Latossolo Roxo distro´fico, em funca de sistemas de plantio, rotac, oes de culturas e me´todos de preparo das amostras. Rev. Bras. Ci. Solo, 22, 527–538.

Erpul G, Deviren Saygın S, 2012. Ülkemizdeki toprak erozyonu sorunu üzerine: Ne yapmalı?. Toprak Bilimi ve Bitki Besleme Dergisi. 1(1), 26 - 32.

Ferreira V, Panagopoulos T, Andrade R, Guerrero C, Loures L, 2015. Spatial variability of soil properties and soil erodibility in the alqueva reservoir watershed, Solid Earth, 6, 383–392.

Kara Z, Yakupoğlu T, Ballı E, Saltalı K, 2018. Relationship between some soil properties and erodibility of agricultural lands close the suphan mountain, 1. International GAP Agriculture & Livestock Congress, 25-27 April 2018 – Şanlıurfa/TURKEY.

Karagöz K, Hanay A, 2020. Effect of bat guano on cation exchange capacity, some macro nutrient content and reaction of soil (pH), Atatürk Üniversitesi Ziraat Fakültesi Dergisi, 51 (2): 140-144

Kütük C, Çaycı G, Baran A, Baksan O, 1999. Effect of humic acid on some soil properties. Soil Science Department, Agricultural Faculty, Ankara University, Ankara, s. 161.

OĞUZ İ, DEMİR S, AKDOĞAN Y, 2020. Sivas-Ulaş rüzgâr erozyon sahasında buğday ve nadas koşullarında rüzgârla hareket eden sediment miktarı ile sediment yüksekliğinin zamansal değişiminin araştırılması. Gaziosmanpaşa Bilimsel Araştırma Dergisi (GBAD), 9:2, 99-114.

Reinks SM, Botha GA, Hughes JC. 2000. Some physical and chemical properties of sediments exposed in a gully (donga) in northern Kwazulu-Natal, South Africa and their relationship to the erodibility of the colluvial layers. Catena.

Reis M, Tat S, Abız B, 2021. Anakayaların toprak özellikleri ve toprak erodibilitesi üzerine etkileri: Çemrengeç deresi yağış havzası örneği. Turkish Journal of Forestry, 22(4): 353-361.

Russell E.W. 1973. Soil structure, tilth and mechanical behaviour. Russell's soil conditions and plant growth. 10th. ed. Essex: Longman Scientific and Technical, pp 479-519.

Rodriguez R.R., Arbelo C.D., Guerra J.A., Natario M.J.S., Armas C.M., 2006. Organic carbon stocks and soil erodibility ın Canary Islands Andosols. Catena, 66: 228-235. doi: 10.1016/j.catena.

Saygın F, Dengiz O, İç S, İmamoğlu A, 2019. Bazı fiziko-kimyasal toprak özellikleri ile bazı erodobilite parametreleri arasındaki ilişkilerin mikro havza ölçeğinde değerlendirilmesi. Artvin Çoruh Üniversitesi Orman Fakültesi Dergisi 20 (1): 82-91.

Sharif M, Ahmad M, Sarir MS, Khattak RA (2004) Effect of organic and inorganic fertilizers on the yield and yield components of maize. Pakistan Journal of Agriculture, Agricultural Engineering, Veterinary Sciences 20(1): 11-16.

Yeter K, Yalçın M, 2020. Hatay ili Kırıkhan-Kumlu bölgesi topraklarının pH, kireç ve organik madde içeriklerinin belirlenmesi, ISPEC Tarım Bilimleri Dergisi, 4(2):285-293.

Tez:

Altunsu M., 2019. Farklı eğimli ve kurak koşullarda bir arazi topluluştırma sahasında çölleşme potansiyelinin araştırılması, YükseK Lisans Tezi. Tokat Gaziosmanpaşa Üniversitesi, Fen Bilimleri Enstitüsü, Tokat, Türkiye.

Doğan K (2000) Antakya şehir çöplerinden elde edilen kompostun toprakların bazı fiziksel ve kimyasal özellikleri ile domateste verime etkisi. Doktora Tezi, Mustafa Kemal Üniversitesi Fen Bilimleri Enstitüsü, Hatay.

Working Covariance Selection Methods for Negative Binomial Marginal Model

Hatice Tul Kubra Akdur[1]

[1] Gazi University, Faculty of Science, Department of Statistics, 06500, Ankara, Turkey

*Corresponding author e-mail: haticesenol@gazi.edu.tr

Abstract

Negative binomial marginal model is used to analyze correlated count datasets when the Poisson marginal model is not suitable due to over-dispersion in the presence of the greater variability in the data. Assessment of adequacy of both correlation and variance structures for longitudinal count data is important for efficiency of parameter estimations in GEE-based marginal models. In this study, Gaussian pseudolikelihood and Rotnitzky and Jewell criteria for working correlation selection under the negative binomial marginal model are adapted (Carey and Wang, 2011; Rotnitzky and Jewell, 1990). The performances of proportion of selecting the true working correlation structure for the selection criteria are investigated through Monte Carlo simulation study under various scenarios. Parameter estimations of the marginal model based on GEEs are obtained by geeM R package with the user-defined function (McDaniel et al., 2013). For illustration, the method is applied to a real data set which includes overdispersed longitudinal count data.

Key words: *Negative Binomial Marginal models, GEE, Overdispersion, Working covariance selection*

Statistical Evidence of Crime

Afrim Osmani[1*]

[1] University of Tetova, Skopje, 1000, North Macedonia

*Corresponding author e-mail: afrim.osmani@unite.edu.mk

Abstract

No country in the world is immune to crime. But every country in the world fights it with all the mechanisms and opportunities at its disposal. The fight against crime must be continuous, uninterrupted, institutional, but also international. For a more genuine approach to crime, we need to have accurate crime statistics, which are mainly kept by the most relevant institutions, namely: the police, the judiciary and the prosecution, at the national level. Also important are the statistics kept at the international level. Crime statistics often face problems, which hinders the real picture of crime. These problems are: the dark number, the gray number and the excessive number on crime. In the paper below, all three groups of numbers on crime will be presented in more detail.

Key words: *Crime, Statistical Evidence, Dark Number, Gray Number, Excess Number, International Statistics, National Statistics.*

3*2 Lokus Modeli için Bağlantı Dengesizliği Katsayısının Örnekleme Dağılımı Üzerine Bir Simülasyon Çalışması

Selma Kökçü[1*], Orhan Kavuncu[2], Yasemin Gedik[3], Özge Şahin[4]

[1] Ziraat Yüksek Mühendisi, 60500, Tokat, Türkiye
[2] Kastamonu Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Genetik ve Biyomühendislik Bölümü, 37150, Kastamonu, Türkiye
[3] Eskişehir Osmangazi Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, 26160, Eskişehir, Türkiye
[4] Münster Üniversitesi, Biyoloji Fakültesi, Evrim ve Biyoçeşitlilik Enstitüsü, 48143, Münster, Almanya

*Corresponding author e-mail: selma14110396@gmail.com

Özet

Bu çalışmanın amacı çeşitli bağlantı dengesizliği katsayılarına ve iki lokusta üç allel sayısına sahip sanal populasyonlardan simülasyon yöntemiyle çekilen farklı genişlikteki örneklerden hesaplanan bağlantı dengesizliğiyle ilgili kuadratik istatistiklerinin nasıl bir dağılım göstereceğinin belirlenmesidir. $\mu D=0$ ve $\mu D\neq 0$ bağlantı dengesizliği katsayıları için, iki lokus üç allel sayılarına sahip populasyonlardan simülasyon yöntemiyle çekilen n=100 genişlikteki örneklerden hesaplanan birisi metinde (6) numaralı kuadratik bir form olan istatistik olmak üzere üç ayrı kuadratik istatistiğin dağılımı çalışılmıştır. Sonuçta bağlantı dengesizliği katsayısının sıfır olduğu populasyondan çekilen örneklerden hesaplanan bu (6 numaralı istatistik) kuadratik formun 2 serbestlik dereceli (central) merkezi Khi-Kare dağılımı gösterdiği bulunmuştur. Dengede olmayan populasyonlardan çekilen örneklerden hesaplanan istatistikler için denenenler arasında uygun bir dağılım bulunamamış olup başka dağılımlara uygunluğunun test edilmesi gerekmektedir

Key words: *Bağlantı Dengesizliği Katsayısı, Khi-Kare Dağılımı, Noncentral Khi-Kare Dağılımı*

Harvest Efficiency of Chlorella sorokiniana Using a Vacuum Pump Chamber designed for Laboratory Conditions

Çetin Yağcılar[1*], Mehmet Yardımcı[2]

[1] Tekirdağ Namık Kemal University, Faculty of Veterinary Medicine, 59030, Tekirdağ, Türkiye
[2] Tekirdağ Namık Kemal University, Faculty of Veterinary Medicine, 59030, Tekirdağ, Türkiye

*Corresponding author e-mail: cyagcilar@nku.edu.tr

Abstract

Microalgae are microscopic and photosynthetic organisms used in many fields. Recently it has been used as a biofuel or as an alternative food product for humans and animals. The only process limiting the use of microalgae is harvesting, which accounts for about 20-30% of the cost. Microalgae harvesting is briefly the separation of solid particles from the liquid medium. In general, many methods are used either alone or in combination. In general, the harvesting methods used in microalgae are grouped into four main groups as physical, chemical, biological and electrochemical methods. Firstly, physical based methods include precipitation, centrifugation, filtration and flotation. Biological harvesting methods are auto-flocculation and bio-flocculation. Chemical flocculation is considered as a harvesting method that effectively harvests large volumes of microalgae culture and can be applied to many microalgae species. In this method, it is based on the collection of solid particles by providing precipitation in the product with the help of chemicals added to the microalgae culture. Chlorella sorokiniana used in our study is a freshwater microalgae belonging to the chlorophyta group. This microalga has a characteristic emerald green color and a pleasant grass smell. In our research, the differences in the harvesting of microalgae grown in f2 ambient conditions using a vacuum pump with and without a precipitator were investigated. During the process, firstly, cell density was measured by direct counting of cells during the development stages at 2-day intervals. Secondly, the precipitation amount of Chlorella over time was calculated based on the formed sediment thickness. Finally, the harvesting efficiency of chlorella with or without aluminum sulfate, using a vacuum pump chamber or vacuumless was calculated and compared. The calculation was made according to the formula below:

Harvest efficiency = 100 Sample dry weight (g) / Sample wet weight (g)*

The results showed that using a vacuum pomp provided a great advantage in the filtration process of the Chlorella sorokiniana in terms of time and efficiency.

Key words: *Aluminum sulfate, Chlorella, Microalgae, Vacuum pump*

INTRODUCTION

Microalgae are eukaryotic microorganisms with different shapes and sizes (2-50 μm) (Elisabeth et al., 2021). Microalgae, which has gained great popularity around the world today, has started to be used in many areas such as edible energy, food, agriculture, technology and functional medicines (Khan et al., 2018). In addition, microalgae produce potential products with many sustainable biological properties (Ruiz et al., 2016). Microalgae which are among the fastest growing plants in the world (Randrianarison and Ashraf, 2017), started to be used in every sector due to the high protein, oil and other valuable contents.

Members of the Chlorophyta group are mostly spherical or elliptical spherical organisms with a diameter of 2-10 µm in the vegetative stage without flagella (Bingül et al., 2021, Oliveira et al., 2021). Chlorella, which belongs to this group, is a microalgae that can grow rapidly and has an uncomplicated life cycle, contains a high amount of protein, and also has a rich content of carotenoids, vitamins and minerals (Nabavi and Silva, 2019). *Chlorella sorokiniana* species, which is in the Chlorella group, is a microalgae containing high content of protein, antioxidants, fats, vitamins and minerals that can be used as functional food supplements in animals and humans (Jamshidi et al., 2018).

Harvesting of microalgae constitutes the main element of microalgae cultivation (Tan et al., 2020) and covers approximately 20-30% of operating costs (Barros et al., 2015). Harvesting processes of microalgae can be done by physical methods such as centrifugation, precipitation, filtration, flotation, ultrasound, chemical, electrical and biological methods such as coagulation, flocculation (Kumar et al., 2022). In this study, the effect of *Chlorella sorokiniana* microalgae on the harvesting efficiency of the vacuum pump chamber (Figure 1) and the use of aluminum sulfate as a precipitator or not was investigated.

Figure 1. Vacuum pump chamber schematization and orginal view

MATERIAL AND METHODS

Material

Growth Conditions and Growth Medium of the Microalgae

Seed culture of *Chlorella sorokiniana* was provided by a private institution and used for the cultivation under laboratory conditions (Figure 2).

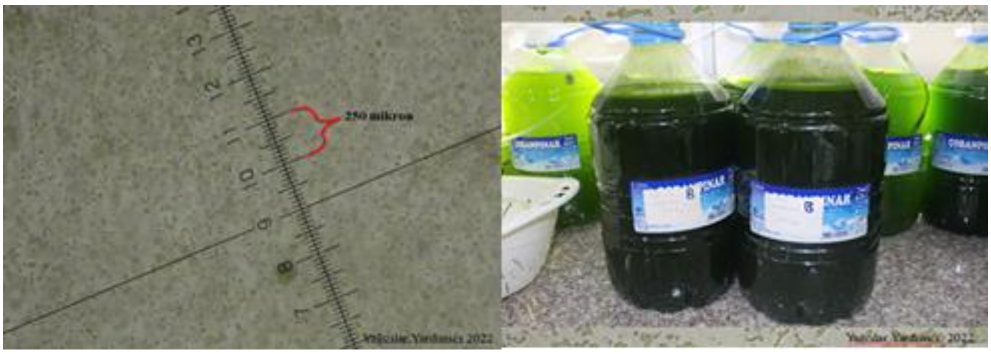


Figure 2. Chlorella sorokiniana microscope view and mode of production

Microalgae cells were cultivated in a modified f/2 medium (Bat et al., 2008). For each liter of *Chlorella sorokiniana* culture, 1 ml of these nutrients is added. (Table 1). Growth of *Chlorella sorokiniana* was carried out in bottles made of plastic material with a volume of 5 liters. They were grown at 32.74±0.811

temperature in 14 hours of light and 10 hours of darkness with white light having an average intensity of 22911±2367.86 lux along the midline of the bottle. Microalgae held in plastic bottles were mixed continuously. After the first seeding of the culture, the control of the growth of the cells was measured at two-day intervals under the microscope with the Fuchs-Rosenthal counting chamber (Figure 3).

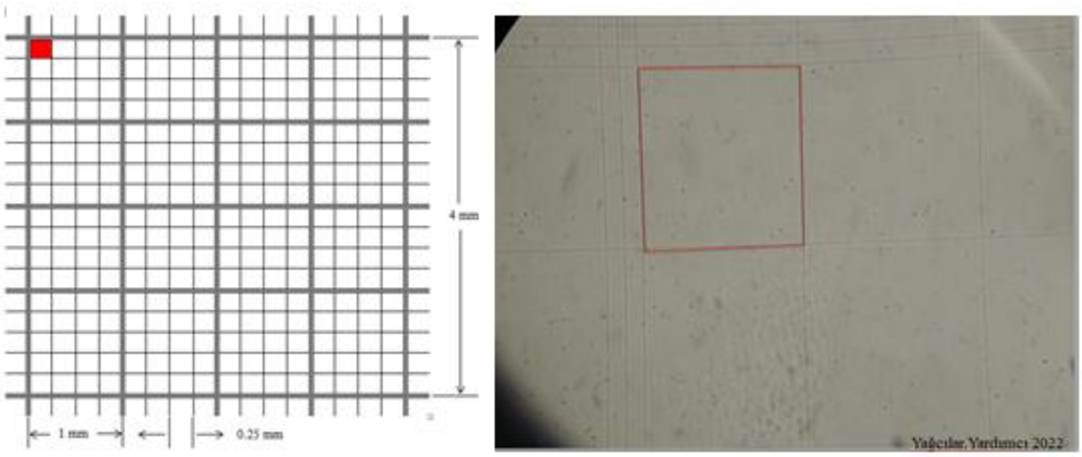


Figure 3. Fuchs-Rosenthal counting chamber

Table 1. Composition of modified f/2 medium

Component	**Chemical compositions (g/L)**
(1) Salt Solutions	
$NaNo_3$	300
KH_2PO_4	30
NH_4Cl	20

* The mixture should be dissolved in 1 liter of dH_2O and kept in an autoclave at 120 °C for 30 minutes and then cooled.

<table>
<tr><th></th><th></th><th>Stock Solutions (ml/L)</th></tr>
<tr><td colspan="3">(2) Trace Metal Solutions</td></tr>
<tr><td>Solutions A</td><td></td><td rowspan="5">10 ml</td></tr>
<tr><td>$ZnSO_4H_2O$</td><td>30</td></tr>
<tr><td>$CuSO_45H_2O$</td><td>25</td></tr>
<tr><td>$CuSO_47H_2O$</td><td>30</td></tr>
<tr><td>$MnSO_4H_2O$</td><td>20</td></tr>
<tr><td>Solutions B</td><td></td><td rowspan="2">10 ml</td></tr>
<tr><td>$FeCl_36H_2O$</td><td>50</td></tr>
</table>

Solutions C		10 ml
$Na_2M_0O_42H_2O$	25	
Solutions D		100 ml
$Na_2EDTA2H_2O$	50	
* Ouantity used should be dissolved in 800 ml of dH_2O and kept in an autoclave at 120 °C for 30 minutes and then cooled.		
(3) Vitamins Solutions		
Biotin (vitamin H)	0,1	10 ml
Cyanocobalamin (vitamin B_{12})	0,1	10 ml
Thiamine (vitamin B_1)	0,01	10 ml
* A stock vitamin solution is prepared by adding the used amounts to a container containing 1 liter of dH_2O.		

Experimental Design

The measurement of the harvest efficiency of *Chlorella sorokiniana* was based on 1) using aluminum sulfate as a precipitation product with or without a vacuum pump (Figure 4), 2) just using a vacuum pump or vacuumless filtration in this study (Figure 5).

Figure 4. Filtration using a vacuum pump chamber or vacuumless with aluminum sulfate

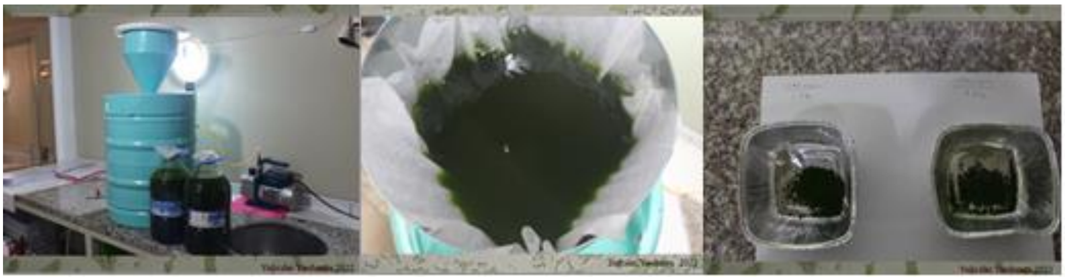


Figure 5. Filtration using a vacuum pump chamber or vacuumless without aluminum sulfate

Methods

The Collection of the Data

The samples were developed in 5 L water-filled plastic bottles. The time of reaching the harvest maturity of the samples was determined according to the apparent color change. During the development period of a two weeks time, cell density was measured by direct counting of cells at 2-day intervals. Secondly, the precipitation amount of chlorella over time was calculated based on the formed sediment thickness. Finally, the harvesting efficiency of chlorella with or without aluminum sulfate, using a vacuum pump chamber or vacuumless was calculated and compared.

Statistical Analysis

Increase rate in the cell numbers was calculated on the basis of the ratio of the next measurement value to the previous measurement value. Due to the high number of cells, logarithmic values of the data were taken as the basis of the graphical representation. Precipitation times were evaluated on the amount of aluminum sulfate. Harvest efficiency was calculated according to the formula below:

Harvest efficiency = 100* Sample dry weight (g) / Sample wet weight (g)

RESULTS AND DISCUSSION

Depending on the time, the increase in the number of cells occurred more rapidly in the first five days compared to the following days. It is thought that this situation was due to the effect of the nutrient element added at first. It was expected that there would be a decrease in the rate of increase since no nutritional supplement was made again until the harvest time (Table 2).

Table 2. Cell number of *Chlorella sorokiniana* in 5 L water sample

Day	Cell count	Increase rate
1	176227916	-
3	280631050	1,6
5	540114750	1,9
7	872185300	1,6
9	1143100000	1,3
11	1321995150	1,2
13	1625945312	1,2
15	2077227031	1,3

The Fuchs-Rosenthal Counting Chamber consists of 16 chambers of one square millimeter areas orientated by triple lines. It is generally recommended to count these 16 areas, preferably 8 in each chamber. In our study, 8 chambers were counted and cell numbers were calculated according to the formula below:

$$\text{Number of cells in the rearing system} = \frac{\text{Number of cells counted} \times \text{Dilution ratio}}{3{,}2} \times \text{amount of culture grown}(cm^3)$$

Considering the increase in the cell number in the log10 base, the following quadratic growth curve was obtained (Figure 6).

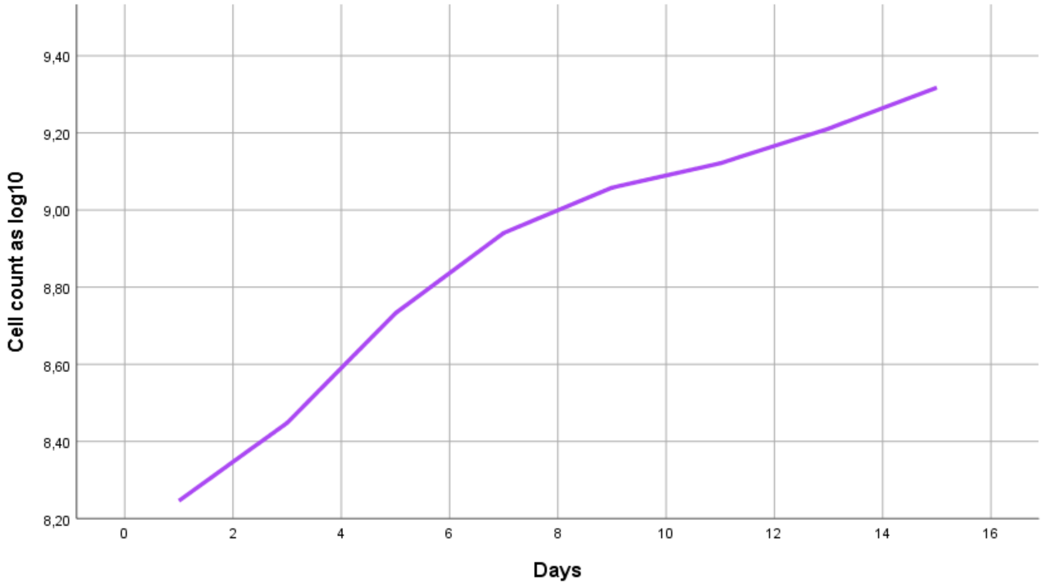


Figure 6. Increase in the number of cells by time

The use of 0.5 g L^{-1} aluminum sulfate caused more than 80% precipitation of microalgae after 15 minutes. At the same time, the use of higher concentration of aluminum sulfate will cause *Chlorella sorokiniana* to precipitate faster (Lira et al., 2022). In another study, the use of 2.5 g/L^{-1} aluminum sulfate for the harvest of *Chlorella vulgaris* provided the highest precipitation rate (Zhu et al., 2018).

Regarding the precipitation time, faster precipitation was formed in samples containing a higher amount of aluminum sulfate (Figure 7-8).

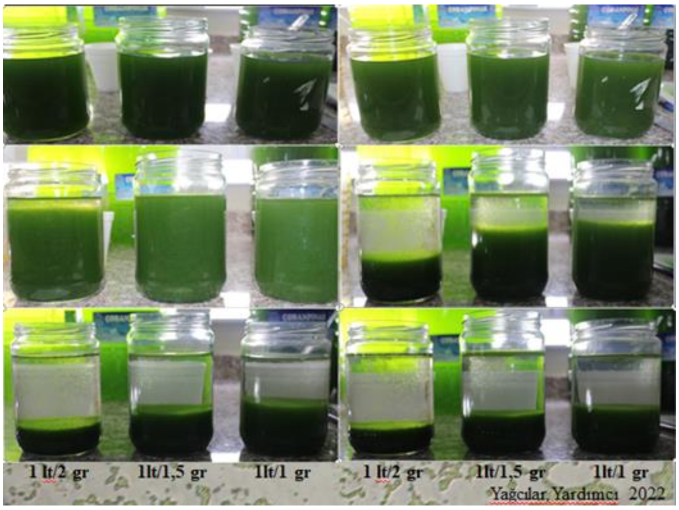


Figure 7. The precipitation amount of Chlorella over time was calculated based on the formed sediment thickness

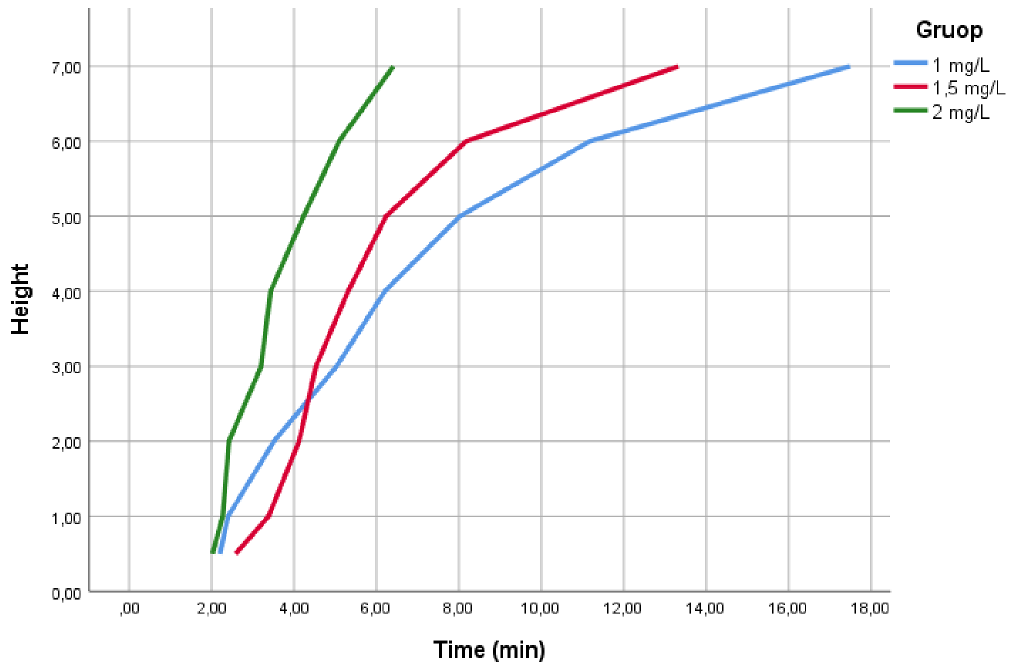


Figure 8. Precipitation of the cells per minute by the measurements of the height of the precipitate

Numerical values indicated that the harvesting efficiency was higher when aluminum sulfate was used. However, the main thing that draws attention was the efficiency achieved over time. Filtration performed by using a vacuum without applying any precipitation process provided a gain of 18 hours compared to the process without vacuum. On the other hand, filtration using a vacuum with aluminum sulfate provided a gain of 1.11 hours compared to the non-vacuum process (Table 3).

Table 3. Harvest efficiency = 100* Sample dry weight (g) / Sample wet weight (g)

	Harvest efficiency	Time (hr)	Time difference (hr)
Aluminium sulphate used Vacuum used	5,96	0,42	1,11
Aluminium sulphate used Vacuum not used	5,84	1,53	
Aluminium sulphate not used Vacuum used	2,91	7,37	18,13
Aluminium sulphate not used Vacuum not used	3,03	25,50	

Considering the findings of this study, the fact that Chlorella sorokiniana reached harvest maturity in a week can be considered a sign of large-scale production in a short time. On the other hand, the use of aluminum sulfate during the harvesting process can be preferred since it is a quick process and easy to apply.

CONCLUSION

The results showed that using a vacuum pomp provided a great advantage in the filtration process of the *Chlorella sorokiniana* in terms of time and efficiency. Moreover, using aluminum sulfate as a precipitator accelerates the process.

References

Bat L, Satılmış HH, Şahin F, Üstün F, Özdemir ZB, Ersanlı E, 2008. Plankton Bilgisi ve Kültürü. Nobel Yayın Dağıtım LTD. ŞTİ, 1287,148-150, Ankara.

Khan MI, Shin JH, Kim JD, 2018. The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. Microbial Cell Factories, 17 (36): 2-21. https://doi.org/10.1186/s12934-018-0879-x

Ruiz J, Olivieri G, Vree JD, Bosma R, Willems P, Reith JH, Eppink MHM, Kleinegris DMM, Wijffels RN, Barbosa MJ, 2016. Towards industrial products from microalgae. Energy Environmental Science, 9: 3036—3043. https://doi.org/10.1039/c6ee01493c

Elisabeth B, Rayen F, Behnam T, 2012. Microalgae culture quality indicators: a review. Critical reviews in biotechnology, 41 (4): 457-473. https://doi.org/10.1080/07388551.2020.1854672

Tan JS, Leec SY, Chewd KW, Lame MK, Lim JW, Ho SH, Show PL, 2020. A review on microalgae cultivation and harvesting, and their biomass extraction processing using ionic liquids. Bıoengıneered, 11 (1): 116-129. https://doi.org/10.1080/21655979.2020.1711626

Barros AI, Gonçalves AL, Simões M, Pires JCM, 2015. Harvesting techniques applied to microalgae: a review. Access address: https://repositorio-aberto.up.pt/bitstream/10216/103426/3/180519.1.pdf ; Date of access: 19.06.2022.

Randrianarison G, Ashraf MA 2017. Microalgae: a potential plant for energy production. Geology, Ecology and Landscapes. 1 (2): 104-120. https://doi.org/10.1080/24749508.2017.1332853

Bingül DY, Demirel Z, Dalay MC, 2021. *Chlorella sorokiniana*' nın izolasyonu, moleküler tanılanması, fototrofik, miksotrofik ve heterotrofik üretimi. Lımnofısh-Journal of limnology and freshwater fisheries research, 7(2): 128-137. https:// doi.org/10.17216/limnofish.703234.

Nabavi SM, Silva AS, 2019. Nonvitamin and nonmineral nutritional supplements. Elsevier, United Kingdom.

Oliveira GA, Teixeira VM, Cunha JND, Santos MRD, Paiva VM, Rezende MJC, Valle AFD, Elia ED, 2021. Biomass of microalgae chlorella sorokiniana as green corrosion ınhibitor for mild steel in hcl solution. International Journal Of Electrochemical Science, 16:2-22. https://doi.org/10.20964/2021.02.42

Jamshidi A, Ebrahimi MA, Rajabian T, Bakhshikhaniki GR, Mozaffari S, 2018. Food enrichment of chlorella sorokiniana (shihira & r.w. krauss). Journal of Experimental Zoology, 25(103):112. https://doi.org/10.30473/EAB.2018.5186

Kumar N, Banerjee C, Negi S, Shukla P, 2022.Microalgae harvesting techniques: updates and recent technological interventions. Critical Reviews in Biotechnology,1-27. https://doi.org /10.1080/07388551.2022.2031089

Lira RDA, Corrêdo LDP, Soares J, Rocha MM, Matos ATD, Coimbra JSDR, Martins MA, 2022. Harvesting of Chlorella sorokiniana BR001 cultivated in a low-nitrogen medium using different techniques. Ciência Rural, Santa Maria, 52(1). https://doi.org/10.1590/0103-8478cr20200293

Zhu L, Li Z, Hiltunen E, 2018. Microalgae Chlorella vulgaris biomass harvesting by natural flocculant: effects on biomass sedimentation, spent medium recycling and lipid extraction. Biotechnol Biofuels, 11(183). https://doi.org/10.1186/s13068-018-1183-z

Conflict of Interest

We declare that there is no conflict of interest.

Birliktelik Kuralları Algoritmaları ile Süt Sığırlarında Güç Doğumun İncelenmesi

Elif Çelik[1*], **Aytaç Akçay**[2], **Murat Abay**[3], **Güven Güngör**[1]

[1]Erciyes Üniversitesi, Veteriner Fakültesi, Biyometri Anabilim Dalı, Kayseri, Turkiye
[2]Ankara Üniversitesi, Veteriner Fakültesi, Biyoistatistik Anabilim Dalı, Ankara, Türkiye
[3]Erciyes Üniversitesi, Veteriner Fakültesi, Doğum ve Jinekoloji Anabilim Dalı, Kayseri, Turkiye

*Corresponding author e-mail: elifcelik149@gmail.com

Özet

Veri madenciliği, bir veri setindeki geçerli ve anlaşılır kalıpları belirleme sürecidir. Büyük veri kümelerinden değerli bilgilerin çıkarılmasına yardımcı olur. Veri madenciliğinin en önemli konularından biri birliktelik kurallarıdır. Birliktelik kuralları yöntemi, ilişkisel davranışları belirleyerek bazı çıkarımlar elde edilmesini sağlayan bir yaklaşımdır. Birliktelik kuralı algoritmaları arasında en yaygın olarak kullanılan algoritmalar Apriori ve Eclat algoritmaları olup bu algoritmalar büyük öğe kümelerini belirlemek için kullanılır. Apriori algoritması, sık görülen alt kümelerin her iterasyonda bir öğe olarak genişletildiği ve aday kümelerin verilere karşı test edildiği "bottom-up" yaklaşımına dayanır ve Breadth First Search şeklinde arama gerçekleştirir. Eclat algoritması ise öğe kümesi madenciliği yapmak için kullanılır ve özyinelemeli olarak tanımlanır. Eclat algoritması, tüm sık kümeleri bulmak için Depth First Search şeklinde bir arama gerçekleştirir. Bu çalışmanın amacı, süt sığırlarında güç doğumu etkileyen faktörleri birliktelik kuralları algoritmaları ile incelemek ve Apriori ve Eclat algoritmalarını karşılaştırmaktır. Çalışma materyali, 108 süt sığırına ait güç doğum, ayak hastalığı, metritis, mastitis, buzağılama mevsimi, buzağılama yaşı ve buzağı doğum ağırlığı verilerinden oluşmaktadır. Analizler R 4.2.0 (www.r-project.org) yazılımındaki "arules" paketi ile yapılmıştır. Apriori algoritmasına göre 9-10 yaş süt sığırlarında mastitis ve güç doğumun birlikte görülme oranı %11; buzağı doğum ağırlığı normal olan ve buzağılama mevsimi yaz olan süt sığırlarında ayak hastalığı ve güç doğumun birlikte görülme oranı %12 olarak bulundu. Eclat algoritması ile elde edilen sonuçlara göre yaz aylarında buzağılama ve güç doğumun birlikte bulunma oranı %23'tür. 9-10 yaş süt sığırlarında güç doğum oranı %17 olarak bulundu. 9-10 yaşlı süt sığırlarında mastitis ve güç doğumun birlikte görülme oranı ise %11'dir. Sonuç olarak Apriori ve Eclat algoritmaları ile elde edilen sonuçların benzer olduğu görülmüştür. Ayrıca birliktelik kurallarının sıklıkla kullanılan istatistiksel yöntemlerle ortaya çıkarılamayan ilişkilerin elde edilmesinde veteriner hekimliği alanında faydalı bir yöntem olduğu görülmüştür.

Key words: *Veri madenciliği, birliktelik kuralları, apriori, eclat, güç doğum*

Determination of Industrial Shelf System Selection Criteria Weights by SWARA Method

Abdurrahman YILDIZ[1*]

[1] Kütahya Dumlupınar University, Engineering Faculty, Department of Industrial Engineering, Kütahya, Türkiye

*Corresponding author e-mail: abdurrahman.yildiz@dpu.edu.tr

Abstract

Warehouse Management is an important business function that is carried out in order to stock the raw materials or products of the enterprises and to reach them effectively when necessary. The most important factor that increases the efficiency of Warehouse Management is a well-designed racking system. Industrial shelving systems are rack systems that are used for stacking raw materials and products with high weights and different properties, and that, unlike standard rack systems, must have some high qualifications. In this study, an evaluation was made with the SWARA method in order to determine the criterion weights to be used in the industrial shelf selection problem for a food wholesale distributor business warehouse. The SWARA Method is a method used to determine criterion weights in multi-criteria decision-making problems. The method basically consists of ranking criteria, pairwise comparison, ratio analysis and determination of weights. For the evaluation, 7 criteria were determined as cost, structural height (maximum height), modularity, shelf and total carrying capacities, shelf material and economic life. As a result of the implementation of the SWARA method steps, criterion weights were determined. The determined criterion weights can be used as an input to the industrial rack selection problem.

Key words: *Industrial shelving systems, SWARA method*

Evaluation of Inventory Management Data Storage Platform Selection Criteria with Analytical Network Process

Abdurrahman YILDIZ[1]

[1] Kütahya Dumlupınar University, Engineering Faculty, Department of Industrial Engineering, Kütahya, Türkiye

*Corresponding author e-mail: abdurrahman.yildiz@dpu.edu.tr

Abstract

Inventory management is one of the issues that businesses work on the most due to increasing costs. In order to control and reduce the increasing costs, an effective inventory management system should be established and operated. One of the most important elements of Inventory Management is the storage of data belonging to the inventory system. Data storage, is the most important information management process for receiving, keeping and evaluating data. For this reason, the selection of the right platform to be used for data storage is one of the main factors that will ensure the effective management of the process. In this study, the weights of the criteria that are considered to be effective in deciding on the data storage platform for the inventory management database to be used in a business were determined. Cost, Security, Capacity, Technological Infrastructure and Access Speed criteria have been determined to be used in the evaluation. Analytical Network Process (ANP) approach was used to determine the criteria weights, since there is an interaction between the determined criteria. Paired comparisons were checked for consistency and the consistency of all comparison matrices was ensured. After the pairwise comparisons, ANP steps were applied and Super Matrix and Limit Super Matrices were obtained. As a result of the ANP method, the weights of the criteria to be used in the evaluation were determined. The determined criterion weights can be used as an input to the data storage platform multi-criteria selection problem, which can be solved by a determined method.

Key words: *Inventory Management, Data Storage, Analytical Network Process*

Covid-19 Salgını Sonrası Sosyal İlişkilerde Sosyal Mesafeye Uyma Sosyal Çevreye Katılma ve Kamusal Alanda Bulunma Niyeti Üzerine Bir Araştırma

Taner Tunç[1*], Murat Şahin[2], Samet Hasan Abacı[3]

[1] Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 55200, Samsun, Türkiye
[2] Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, Sosyoloji Bölümü, 55200, Samsun, Türkiye
[3] Ondokuz Mayıs Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, 55200, Samsun, Türkiye

*Corresponding author e-mail: ttunc@omu.edu.tr

Özet

Yaşlı dünyamız tarih boyu birçok salgın ile karşı karşıya kalmış ve salgınlar süresince de birçok değişime sahne olmuştur. Bu değişimler her alanda olduğu gibi sosyal ilişkilerde de kendini göstermiştir. Salgın zamanları olağanüstü bir hal olarak algılanmış ve yaşam stilleri bu duruma uygun olarak mecburen değişmiştir. Bugünkü salgın da hayatımızda birçok davranış biçiminin ve yaşam stilinin değişmesine sebeptir. Hayatımızı kaplayan tüm alanlarda gerçekleşen bu değişimin ölçülmesine olan ihtiyaç gün geçtikçe artmaktadır. Dünyada meydana gelen birçok salgından sonra belirli değişimler olmuştur. Son olarak ortaya çıkan Covid-19 salgınının da tüm alanlarda belirli değişimlere sebep olacağı açıktır. Bu değişikliklerin ölçülmesi, salgın sonrası yeni normalin sosyal hayata etkilerini tahmin etmede çok önemli bir rol oynayacaktır. Bu bağlamda çalışmanın amacı, salgın sonrası değişmesi beklenen sosyal ilişkilerin nasıl olacağı hakkında verilere dayalı olarak bir çıkarımda bulunabilmektir. Bunun için 2022 yılında Şahin ve Tunç tarafından geliştirilen Salgın Sonrası Sosyal İlişkiler ölçeği kullanılmıştır. nbsp;Çalışma sonuçlarına göre salgın sonrasında; erkeklerin kadınlara göre sosyal mesafeye uyma niyetinin daha fazla olduğu (p=0,017<0,050) ve erkeklerin kadınlara göre kamusal alanda bulunma niyetinin daha yüksek olduğu (p=0,000<0,050) tespit edilmiştir. Cinsiyetin sosyal çevreye katılma niyeti açısından istatistiksel olarak anlamlı bir etkisi yoktur. Virüse çevresinde yakalanan olması durumu sosyal çevreye katılma niyeti üzerinde istatistiksel olarak anlamlı bir etki yapmaktadır. Çevresinde covid-19 a yakalananların sosyal çevreye katılma niyeti daha yüksektir (p=0,001<0,050). Ayrıca virüse çevresinde yakalanan olması durumu ne sosyal mesafeye uyma niyetini ne de kamusal alana katıma niyetini etkilememektedir (p>0,05). Medeni durum, ekonomik durum algısı, virüse kendi yakalanma ve virüsten yakınını kaybetmenin sosyal mesafeye uyma niyeti, kamusal alanda bulunma niyeti sosyal çevreye katılma niyeti üzerinde istatistiksel olarak anlamlı bir etkisi yoktur (p>0,05). Salgının zor geçip geçmeme durumu sosyal çevreye katılma niyetini istatistiksel olarak önemli bir şekilde etkilememektedir (p>0,050). Bunun yanı sıra; salgının zor geçip geçmeme durumu sosyal mesafeye uyma niyetini (p=0,007<0,050) ve kamusal alana katılma niyetini (p=0,002<0,050) istatistiksel olarak önemli bir şekilde etkilemektedir. Salgının zor geçmediğini düşünenlerin sosyal mesafeye daha çok uyma niyetinde oldukları ve kamusal alana daha çok katılma niyetinde oldukları görülmüştür. Eğitim durumu sosyal mesafeye uyma niyetini istatistiksel olarak önemli bir şekilde etkilememektedir (p=0,141>0,050). Bunun yanı sıra; eğitim düzeyi sosyal çevreye katılma niyetini (p=0,00<0,05) ve kamusal alana katılma niyetini (p=0,029<0,050) istatistiksel olarak önemli bir şekilde etkilemektedir. Sosyal çevreye ve kamusal alana katılma niyeti en çok lisansüstü mezunlarında iken en düşük ilköğretim mezunlarındadır. Eğitim düzeyi artarken sosyal çevreye ve kamusal alana katılma niyeti de artmaktadır.

Key words: *Salgın, Sosyal İlişki, Sosyal Mesafe, Kamusal Alan, Yakın Çevre*

Koronavirüs Korkusu Sağlık Algısı ve Yaşam Doyumu Bağlamında Demografik Bir İnceleme

Taner Tunç[1*], **Hasan Fehmi Demirci**[2], **Murat Şahin**[3]

[1] Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 55200, Samsun, Türkiye
[2] Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, Sosyoloji Bölümü, 55200, Samsun, Türkiye
[3] Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, Sosyoloji Bölümü, 55200, Samsun, Türkiye

*Corresponding author e-mail: ttunc@omu.edu.tr

Özet

Ortaya çıkan Covid-19 pandemisinin toplumlar üzerinde olumsuz birçok etkisi olmuştur. İnsanlar kendi alışkanlıklarını bir kenara bırakıp, salgından korunmak için yeni bir hayat tarzı benimser hale gelmiştir. Covid-19 pandemisinden duyulan korku insanların sağlığı algılamaları üzerinde çeşitli olumsuz etkilerde bulunmuş ve bu durum yaşam doyumunu azaltan bir durum olarak karşımıza çıkmıştır. Sağlıklı olmak ve bu sağlıklı olma halini sürdürmek kişinin en temel haklarından biri olduğu gibi bu sürdürülebilirliğin sağlanmasında hem sağlık politikalarının hem de kişinin sorumluluğu büyüktür. Kişinin sağlığı geliştirici davranışlar göstermesinde kendi sağlığını nasıl algıladığı oldukça etkilidir. Kişinin sağlığını geliştirmesi, hastalıklardan korunması kendi sağlık durumunu nasıl algıladığı ile ilgili bir durumdur. Sağlık algısı üzerinde etkisi bulunan birçok faktör bulunmaktadır. Bireyin yaşamdan duyduğu haz, memnuniyet ve yaşam kalitesi sağlık algısı ile doğrudan ilişkilidir. Dünya Sağlık Örgütü yaşam kalitesini, bireylerin ait oldukları kültür ve değer sistemleri, hedefleri, beklentileri, standartları ve ilgileri ile ilişkili olarak yaşamdaki durumlarını algılamaları şeklinde tanımlamıştır. Yaşam kalitesi bireylerin yaşam doyumlarının belirleyicisi durumundadır. Yaşam doyumu, iyi oluşun kilit bir göstergesi olarak kavramsallaştırılmıştır. Bir kişinin hayatının hemen hemen tüm alanlarından (örneğin sağlık, servet, evlilik, eğitim vb.) memnun olması ya da olmaması yaşam doyumu hakkındaki genel yargısını etkileyebilir. Yaşam doyumu öznel ve subjektif iyilik kavramlarıyla ilişkili olarak değerlendirilmektedir. Bu çalışmanın amacı Covid-19 pandemisi sürecinde ortaya çıkan endişe, stres ve anksiyete durumlarının bireylerin sağlık algısı ve yaşam doyumu üzerinde etkisini belirlemektir. Çalışma kapsamında Sağlık Algısı Ölçeği, Yaşam Doyumu Ölçeği ve Koronavirüs Anksiyete Ölçeği kullanılmıştır. Bu bağlamda. Sağlık Algısı Ölçeği'nin tüm alt boyutları, Kontrol Merkezi Puanı (KMP), Sağlık Önem Puanı (SÖP), Kesinlik Puanı (KP), Özfarkındalık Puanı (ÖP), Yaşam Doyum Puanı (YDP) ve Koronavirüs Ankisiyete Puanı (KAP) açısından demografik bir değerlendirme yapılmıştır. Kadınlar ve erkeklerin korona virüs ankiseyete puan ortalamaları arasında istatistiksel olarak anlamlı bir farklılık vardır (p=0,001<0,05). Kadınların ankisyete puanı erkeklerinkinden daha büyüktür. Yaş grupları arasında; Kontrol Merkezi Puanı (KMP), Sağlık Önem Puanı (SÖP) ve Koronavirüs Ankisiyete Puanları (KAP) açısından istatistiksel olarak anlamlı bir farklılık bulunmuştur (p<0,05). Kesinlik Puanı (KP), Özfarkındalık Puanı (ÖP), Yaşam Doyum Puanı (YDP) üzerinde cinsiyetin istatistiksel olarak önemli bir etkisi yoktur (p>0,05). Kontrol merkezi puanı en yüksek 54-67 yaş grubu, en düşük 30-41 yaş grubudur. Sağlık Önem Puanı en yüksek 54-67 yaş grubu, en düşük 42-53 yaş grubudur. Koronavirüs Anksiyete Puanı Puanı en düşük 54-67 yaş grubu, en yüksek 30-41 yaş grubudur. Evlilerin YDP bekarlardan daha yüksektir (p<0,05). Medeni durum KMP, SÖP, KAP, KP, ÖP üzerinde istatistiksel olarak önemli herhangi bir farklılık oluşturmamıştır (p>0,05). Eğitim düzeyi; KMP, KP, YDP, KAP, SAP üzerinde istatistiksel olarak anlamlı bir etkiye sahiptir (p<0,05). Eğitim düzeyi

*arttıkça KMP, KP, YDP, SAP artarken; KAP düşmektedir. Eğitim düzeyi yüksek bireyler sağlıklarını koruma noktasında kendilerine güvenmekte, sağlıklarını korumak için ne yapılması gerektiğini bilmekte ve bunlara paralel olarak yaşam doyumları ve sağlık algıları da yükselmektedir. Bunların tersine yüksek eğitim düzeyine sahip bireyler korona virüs hakkında bilimsel bilgiye rahatlıkla ulaşabildiklerinden hastalık korkuları da düşüktür. Gelir düzeyi KP ve YDP üzerinde istatistiksel olarak anlamlı bir etkiye sahip olduğu görülmektedir (p=0,014<0,050; p=0,000<0,050). Gelir yükseldikçe bireyin sağlığını korumak için yapması gerekenler hakkında daha fazla kesin bilgiye sahip olduğu ve yaşam doyumunun da yükseldiği aşikardır (p=0,000<0,050). Egzersiz yapma alışkanlığı KP, SÖP, ÖP, YDP ve SAP üzerinde istatistiksel olarak anlamlı bir etkiye sahip olmasına rağmen (*p<0,05), KAP üzerinde önemli etkiye sahip değildir (p=0,945>0,050). Egzersiz yapmayanların KP, SÖP, ÖP, YDP ve SAP değerleri egzersiz yapanlardan daha düşüktür. Egzersiz yapan bireyler egzersiz yapmayanlara göre, sağlıklarını koruma noktasında kendilerine daha fazla güvenmekte, sağlıklarını korumak için yapılması gerekenler hakkında daha kesin bilgiye sahip, sağlığının kendi elinde olduğuna yönelik algısı daha yüksek seviyede, daha yüksek yaşam doyumu ve sağlık algısına sahiptirler. Kronik hastalık varlığı, YDP ve SAP üzerinde istatistiksel olarak anlamlı bir etkiye sahip olmasına rağmen (*p<0,050); KP, SÖP, ÖP KAP üzerinde önemli bir etkiye sahip değildir (p=0,945>0,050). Kronik hastalığı olmayan bireylerin, sağlık algısı ve yaşam doyumları kronik hastalığı olanlara göre daha yüksektir. Hastane Başvuru Sıklığı, KP, YDP, KAP ve SAP üzerinde istatistiksel olarak anlamlı bir etkiye sahip olmasına rağmen (*p<0,050); SÖP, ÖP üzerinde önemli bir etkiye sahip değildir (p>0,050). Hastane başvuru sıklığı arttıkça, bireyeler sağlıklarını korumak için yapılması gerekenler hakkında daha az kesin bilgiye, daha az yaşam doyumuna, daha fazla korona virüs yakalanma korkusuna ve daha düşük sağlık algısı puanına sahiptirler. Sağlıklarını korumak için en çok kesin bilgiye sahip olanlar, en yüksek yaşam doyumuna sahip olanlar, en düşük korona virüs korkusuna sahip olanlar ve sağlık algı puanı en yüksek olanlar hastaneye çok nadir başvuran bireylerdir.*

Key words: *Covid 19, Anksiyete, Sağlık Algısı, Yaşam Doyumu*

Test Day-Based Machine Learning Techniques for The Estimation of Norduz Sheep Live Weight

Aslı Akıllı[1*], Suna Akkol[2]

[1] Kırşehir Ahi Evran University, Faculty of Agriculture, Department of Agricultural Economics, 40100, Kırşehir, Turkey
[2] Van Yüzüncü Yıl University, Faculty of Agriculture, Department of Animal Science, 65080, Van, Turkey

*Corresponding author e-mail: asliakilli@ahievran.edu.tr

Abstract

In this study, it was aimed to model live weight in Norduz sheep by using the morphological measurements of age, height of withers, body length, chest width behind paddles, chest depth, chest girth, and thigh circumference. For this purpose, artificial neural networks and least squares support vector machine methods, which are machine learning methods, are discussed and compared with the results of multiple linear regression analysis. The analyses were carried out with data consisting of body measurements and live weight values of 120 Norduz sheep recorded on four test days. Different training algorithms and activation functions were optimized with varying learning parameters in the analyses performed with artificial neural networks. In the least-squares support vector machine analysis, different kernel functions and different parameters of each function are discussed. In the artificial neural network analyses structured with the Bayesian Regulation algorithm, the mean square error results of the live weight estimations performed on four different test- days were obtained as 0.0040, 0.0044, 0.0028, and 0.0096, respectively. The mean square error values for the least squares support vector machines method operated with the Radial Basis Function kernel function were calculated as 0.0181, 0.0172, 0.0102, and 0.0197, respectively. The mean square error results of the multiple linear regression analysis were found to be 0.00421, 0.00402, 0.00379, and 0.0117 for the four test days, respectively. The values obtained as a result of the analyzes show that artificial neural networks are more successful than other methods and can be used as an alternative method in live weight estimation studies.

Key words: *Least Square Support Vector Machine, Neural Networks, Regression Analysis, Sheep Live Weight.*

Using The Random Effect Negative Binomial Model and The Zero Inflated Negative Binomial Model to Examine Sheep Counts in Saray, Van

Koray Ceberut[1], Suna Akkol[1*], Aslı Akıllı[2]

[1]Van Yüzüncü Yıl University, Faculty of Agriculture, Department of Animal Science, 65080, Van, Turkey
[2] Kırşehir Ahi Evran University, Faculty of Agriculture, Department of Agricultural Economics, 40100, Kırşehir, Turkey

*Corresponding author e-mail: sgakkol@yyu.edu.tr

Abstract

In this study, it was aimed to estimate the current sheep presence in the Saray district of Van province. For this purpose, each sheep breeder's goat and cattle assets were included as independent variables in the regression model. Neighborhoods were also included in the regression model to examine the change from neighborhood to neighborhood in terms of sheep presence. Sheep numbers in the Count data form were first analyzed using the Poisson Regression (PR) model. Deviance value, which is an indicator of overdispersion, was found to be 52.47 with PR. Then, Negative binomial regression (NBR) was applied to the same data set, and it was observed that the deviance value decreased up to 1.23. The fact that the Deviance value decreased and took a value very close to 1 revealed that the over-dispersion with NBR was modeled appropriately and was no longer a problem. The random effect NBR model and ZINB regression model were used, respectively, to overcome the issue of the data set having a large number of zero values. The model that best explains the change in sheep numbers was the ZINB regression model with the smallest AIC=20980.13 and BIC=21299.83 values.

Key words: *Count Data, Negative Binomial Model, Poisson model, Sheep.*

Using Content Analysis to Investigate Companies' Activities and Targets

Erk Hacıhasanoğlu[1], Fatma Selen Madenoğlu [1*], Ömer Faruk Ünlüsoy[1]

[1]Abdullah Gul University, Faculty of Managerial Science, Business Administration, 38080, Kayseri, Turkey

*Corresponding author e-mail: selen.madenoglu@agu.edu.tr

Abstract

According to the United Nations 2030 Agenda, the world is expected to be economically wealthy, socially inclusive, environmentally sustainable, and well-governed. Though companies are facing many challenges that limit their potential to grow, such as poverty, inequality, climate change, environmental degradation, injustice, and war issues, mostly environmental, social, and economic issues have been addressed by companies until today. Business growth is generally linked to achieving both macro and micro target. In order to succeed at a micro level, companies must conduce how they can contribute to achieving the targets in a way that motivates their own. Companies are responsible to indicate their activities in reports by categorizing which activities are related to their targets. The aim of this study is to apply content analysis to the reports of companies in Turkey. The study's findings will reveal what the sector's priorities are in terms of targets, internal regulations, and internationalization.

Key words: *2030 Agenda, Interlinkages, Policy Coherence, Content Analysis*

Sales Forecast with Historical Data in a Shrinking Market

Erk Hacıhasanoğlu[1], Fatma Selen Madenoğlu [1*], Ömer Faruk Ünlüsoy[1]

[1] Abdullah Gul University, Faculty of Managerial Science, Business Administration, 38080, Kayseri, Turkey

*Corresponding author e-mail: selen.madenoglu@agu.edu.tr

Abstract

Traditional media tools appear to have a slim chance of surviving in a digital environment. By adapting to current techniques, conventional media tools can provide chances for people. While radio and television are being replaced by mobile applications and subscription-based broadcasting services, news track is being replaced by mobile applications and social media platforms. COVID-19 has expedited digitization, posing a severe threat to the survival of traditional print media. This decline appears to be affecting relevant industries. The study's purpose is to forecast sales by using historical data from a Turkish printing firm based on machine learning techniques. The goal of the study is to make sector-specific predictions using local and global industry indicators, projections, and trends.

Key words: *Sales Forecasting, Sales, Machine Learning*

Some Remarks on BLUPs under Constrained Linear Models

Melek Eriş Büyükkaya[1], **Nesrin Güler**[2*]

[1] Karadeniz Technical University, Faculty of Science, Department of Statistics and Computer Sciences, 61080, Trabzon, Turkey
[2] Sakarya University, Faculty of Political Sciences, Department of Econometrics, 54187, Sakarya, Turkey

*Corresponding author e-mail: melekeris@ktu.edu.tr

Abstract

The best linear unbiased predictors (BLUPs) are defined from the requirement of a minimum covariance matrix of unbiased predictors of unknown vectors under linear regression models. BLUPs are often considered the main object of statistical studies since they have optimality properties. Covariance matrices of BLUPs are usually used as a comparison criterion to determine optimal predictors among other types of unbiased predictors since they have minimum covariance requirements in the Löwner partial ordering. In this study, we consider a general linear model with parameter restrictions, known as a constrained linear model. We derive a group of computational formulas on BLUPs of all unknown vectors under constrained linear models by using quadratics matrix optimization method. This method is related to minimization problems on covariance matrices of predictors. Further, we establish many basic properties of the BLUPs.

Key words: *BLUP, Constrained linear model, Covariance matrix, Inertia, Rank*

Some Remarks on BLUPs and BLUEs under Constrained Linear Models

Melek Eriş Büyükkaya[1*], **Nesrin Güler**[2]

[1] Department of Statistics and Computer Sciences, Faculty of Science, Karadeniz Technical University, 61080, Trabzon, Turkey
[2] Department of Econometrics, Faculty of Political Sciences, Sakarya University, 54187, Sakarya, Turkey

*Corresponding author e-mail: melekeris@ktu.edu.tr

Abstract

The best linear unbiased predictors (BLUPs) and the best linear unbiased estimators (BLUEs) are defined from the requirement of a minimum covariance matrix of unbiased predictors and estimators, respectively, of unknown vectors under linear regression models. BLUPs and BLUEs are often considered the main object of statistical studies since they have optimality properties. Covariance matrices of BLUPs and BLUEs are usually used as a comparison criteria to determine optimal predictors and estimators among other types of unbiased predictors and estimators, respectively, since they have minimum covariance requirements in the Löwner partial ordering. In this study, we consider a general linear model with parameter restrictions, known as a constrained linear model. We derive a group of computational formulas on BLUPs and BLUEs of all unknown vectors under the constrained linear models by using the quadratic matrix optimization method including the rank and inertia of block matrices. This method is related to minimization problems on covariance matrices of predictors. Further, we establish many basic properties of the BLUPs and BLUEs.

Key words: *BLUE, BLUP, Constrained linear model, Covariance matrix, Inertia, Rank*

INTRODUCTION

There are various predictors and estimators of unknown vectors which are possible candidates for optimal predictors and estimators, respectively, to make statistical inference concerning general linear regression models. These predictors and estimators have different properties. The unbiasedness is one of the important properties of predictors and estimators of unknown vectors in general linear regression models. The unbiased predictors and estimators are not necessarily unique. The well-known method for determining the best predictor and estimator for unknown vectors in general linear regression models is to find unbiased predictors and estimators having the smallest covariance matrix among all other unbiased predictors and estimators. In this case, it can be said that the prediction problems under general linear regression models and optimization problems are closely linked with each other in many ways since the best predictors and estimators of unknown vectors in general linear regression models are determined by various optimization methods.

The best linear unbiased predictors (BLUPs) and the best linear unbiased estimators (BLUEs), based on both unbiasedness and minimum covariance matrix criteria, are well-known predictors and estimators, respectively, in statistical theory and its applications. Furthermore, they are usually preferred as predictors and estimators because of their simple and optimal properties in the investigation of the general linear regression models and their statistical applications. Covariance matrices of BLUPs and BLUEs are usually used as a criterion to compare optimality with other types of unbiased predictors and estimators. To derive BLUPs and BLUEs of all unknown vectors under general linear regression models correspond to solving a quadratic matrix optimization problem because of definitions of BLUPs and BLUEs. (Dong et al., 2014) established results for calculating BLUEs and their properties by applying a closed-form solution of a constrained quadratic matrix-valued function optimization problem in the Löwner partial ordering. (Tian, 2015) obtained exact solutions of some constrained quadratic matrix-valued function optimization problems in the Löwner partial ordering and applied them in solving simultaneous estimation/prediction problems of general linear models. The fundamental results on BLUEs were also proved in some other ways; see, e.g., (Drygas, 1970), (Puntanen et al., 2011), (Rao, 1973). As an extension of the problem considered in (Dong et al., 2014), a new method is given by (Güler and Büyükkaya, 2021) for calculating BLUPs and their properties by using a quadratic matrix optimization problem given by (Tian, 2012)

In statistical theory and its applications, there often exist certain restrictions on unknown parameters in linear regression models. These kinds of restrictions occur in many situations such as the linear hypothesis testing on parameters. In this study, we consider a general linear model with parameter restrictions, known as a constrained linear model. By using the approach of (Güler and Büyükkaya, 2021), we derive analytical formulas for calculating the BLUP of unknown vectors and present a series of fundamental properties of the BLUPs under the general assumptions in constrained linear models by solving certain constrained quadratic matrix-valued function optimization problems in Löwner partial ordering. We also give some fundamental results on BLUEs.

Throughout this paper we use the following notations. Let $R^{m\times n}$ stand for the set of all $m\times n$ real matrices. A', $r(A)$, $\mathcal{C}(A)$ and A^{+} denote the transpose, the rank, the column space, and the Moore-Penrose generalized inverse of $A\in R^{m\times n}$, respectively. I_m denotes the identity matrix of order m. $A^{\perp}=I_m-AA^{+}$ stands for the orthogonal projector. $i_{+}(A)$ denotes the positive inertia of symmetric matrix A. The inequalities $A_1-A_2\preccurlyeq 0$ $(A_1\preccurlyeq A_2)$ and $A_1-A_2\succcurlyeq 0$ $(A_1\succcurlyeq A_2)$ mean that the difference A_1-A_2 is negative semi-definite matrix and positive semi-definite matrix in the Löwner partial ordering for the symmetric matrices A_1 and A_2 of same size, respectively.

PRELIMINARY RESULTS

Consider a constrained linear model

$$M\colon y=X\alpha+\varepsilon,\ A\alpha=b \tag{2.1}$$

$$E(\varepsilon)=0\ \text{ and } cov(\varepsilon,\varepsilon)=D(\varepsilon)=\sigma^2\Sigma, \tag{2.2}$$

where $y\in R^{n\times 1}$ is a vector of observable response variables, $X\in R^{n\times k}$ is a known matrix of arbitrary rank, $\alpha\in R^{k\times 1}$ is a vector of fixed but unknown parameters, $\varepsilon\in R^{n\times 1}$ is an unobservable vector of random errors, $A\in R^{m\times k}$ is a known matrix of arbitrary rank, $b\in R^{m\times 1}$ is a known vector, σ^2 is a positive unknown parameter, and $\Sigma\in R^{n\times n}$ is a known positive semi-definite matrix of arbitrary rank. The linear restriction equation $A\alpha=b$ in (2.1) is consistent. The approach of merging of linear regression models is a well-known method in statistical theory and its applications, especially, it is often used when equality restrictions exists in constrained linear regression models. We can merge two given equation parts in (2.1) into the following combined form of vectors

$$\widehat{M}\colon \begin{bmatrix} y\\ b\end{bmatrix}=\begin{bmatrix} X\\ A\end{bmatrix}\alpha+\begin{bmatrix}\varepsilon\\ 0\end{bmatrix} \tag{2.3}$$

and according to the expectation and covariance matrix assumptions in (2.2)

$$E\begin{bmatrix} y\\ b\end{bmatrix}=\begin{bmatrix} X\\ A\end{bmatrix}\alpha \text{ and } D\begin{bmatrix} y\\ b\end{bmatrix}=D\begin{bmatrix}\varepsilon\\ 0\end{bmatrix}=\sigma^2\begin{bmatrix}\varepsilon & 0\\ 0 & 0\end{bmatrix} \tag{2.4}$$

are obtained. Further, we assume that the model in (2.3) is consistent since we make statistical inferences from this model. The model in (2.3) is said to be consistent $\Leftrightarrow \begin{bmatrix} y \\ b \end{bmatrix} \in \mathcal{C}\begin{bmatrix} X & \Sigma \\ A & 0 \end{bmatrix}$ holds with probability (wp) 1; see, e.g., (Rao, 1973).

To estimate the unknown parameter vector α and to predict random error vector ε in (2.3) jointly, we can construct a general vector containing the both unknown vectors as follows

$$\phi = K\alpha + H\varepsilon = K\alpha + [H \quad F]\begin{bmatrix} \varepsilon \\ 0 \end{bmatrix} \tag{2.5}$$

for given matrices $K \in R^{s\times k}$, $H \in R^{s\times n}$, and $F \in R^{s\times m}$.

It can be seen from the expectation and covariance matrix assumptions

$$E(\phi) = K,\ D(\phi) = \sigma^2\begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix},\ cov\left(\phi, \begin{bmatrix} y \\ b \end{bmatrix}\right) = \sigma^2\begin{bmatrix} H\Sigma & 0 \\ 0 & 0 \end{bmatrix}. \tag{2.6}$$

Lemma 2.1. The linear matrix equation $AX = B$ is consistent if and only if $r[A, \quad B] = r(A)$, or equivalently, $AA^{+}B = B$. In this case, the general solution of $AX = B$ can be written in the following form $X = A^{+}B + (I - A^{+}A)U$ where U is an arbitrary matrix.

Lemma 2.2. Let $A \in R^{n\times p}$, $B \in R^{m\times p}$ be given matrices, and $P \in R^{n\times n}$ symmetric positive semi-definite matrix. Also assume that there exists $X_0 \in R^{m\times n}$ such that $X_0A = B$. Then the maximal positive inertia of $X_0PX_0' - XPX'$ subject to all solutions of $XA = B$ is

$$\max_{XA=B} i_{+}(X_0PX_0' - XPX') = r\begin{bmatrix} X_0P \\ A' \end{bmatrix} - r(A) = r(X_0PA^{\perp})$$

Hence there exists solution X_0 of $X_0A = B$ such that $X_0PX_0' \preccurlyeq XPX'$ holds for all solutions of $XA = B$ if and only if X_0 satisfies both $X_0A = B$ and $X_0PA^{\perp} = 0$.

BLUPs and BLUEs' Computations

In this section, we give a group of computational formulas on the BLUPs and the BLUEs of all unknown vectors under the models in (2.3) by using quadratic matrix optimization methods given as in Lemma 2.2, and establish many basic properties of BLUPs/BLUEs. Under our considerations, firstly, we review the predictability/estimability requirement of ϕ and its special cases under the model (2.3) before defining the BLUPs/BLUEs.

Definition 3.1. ([1]) The vector ϕ in (2.5) is predictable in (2.3), i.e., $E\left(L\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) = 0$ holds for some L, if and only if $\mathcal{C}(K') \subseteq \mathcal{C}\left(\begin{bmatrix} X \\ A \end{bmatrix}'\right)$. This requirement also corresponds to the estimability of $K\alpha$ under (2.3). $\begin{bmatrix} X \\ A \end{bmatrix}$ α and Xα are always estimable, and ε is always predictable under (2.3).

Definition 3.2. ([4],[7]) Let ϕ be predictable by $\begin{bmatrix} y \\ b \end{bmatrix}$ in (2.3). If there exists $L\begin{bmatrix} y \\ b \end{bmatrix}$ such that

$$D\left(L\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) = min \text{ s.t. } E\left(L\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) = 0 \tag{3.1}$$

holds in the Löwner partial ordering, the linear statistic $L\begin{bmatrix} y \\ b \end{bmatrix}$ is defined to be the BLUP of ϕ and is denoted by $L\begin{bmatrix} y \\ b \end{bmatrix} = BLUP_{\widehat{M}}(\phi) = BLUP_{\widehat{M}}(K\alpha + H\varepsilon)$. If $H = 0$ in ϕ or $K = 0$ in ϕ, $L\begin{bmatrix} y \\ b \end{bmatrix}$ corresponds the BLUE of $K\alpha$, denoted by $BLUE_{\widehat{M}}(K\alpha)$ and BLUP of $H\varepsilon$, denoted by $BLUP_{\widehat{M}}(H\varepsilon)$, under (2.3).

In order to solve the matrix minimization problem in (3.1), we use (2.7). The following remarkable result concerning analytical solutions of a constrained quadratic matrix-valued function minimization problem includes fundamental equations of BLUPs and BLUEs and their many basic properties.

Teorem 3.1. Let ϕ be predictable by $\begin{bmatrix} y \\ b \end{bmatrix}$ in (2.3). Then the following results hold.

$$BLUP_{\widehat{M}}(\phi) = L\begin{bmatrix} y \\ b \end{bmatrix} = \left(\left[K, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right]W^{+} + PW^{\perp}\right)\begin{bmatrix} y \\ b \end{bmatrix} \tag{3.2}$$

where $P \in R^{s\times(n+m)}$ is an arbitrary matrix and $W = \left[\begin{bmatrix} X \\ A \end{bmatrix}, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right]$. In particular,

(a) L is unique $\Leftrightarrow r(W) = n + m$.

(b) $BLUP_{\widehat{M}}(\phi)$ is unique wp 1 $\Leftrightarrow \widehat{M}$ is consistent.

(c) $r(W) = r\begin{bmatrix} X & H\Sigma H' \\ A & 0 \end{bmatrix} = r\left[\begin{bmatrix} X \\ A \end{bmatrix}, \quad \begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix}\right]$.

(d) Further, the following dispersion matrix equalities hold.

$$\begin{aligned} D[BLUP_{\widehat{M}}(\phi)] = \sigma^2 &\left(\left[K, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right]W^{+}\right)\begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix} \\ &\times\left(\left[K, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right]W^{+}\right)' \end{aligned} \tag{3.3}$$

$$\begin{aligned} D[\phi - BLUP_{\widehat{M}}(\phi)] = \sigma^2 &\left(\left[K, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right]W^{+} - [H, \quad F]\right)\begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix} \\ &\times\left(\left[K, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right]W^{+} - [H, \quad F]\right)' \end{aligned} \tag{3.4}$$

(e) In particular,

$$BLUE_{\widehat{M}}(K\alpha) = ([K, \quad 0]W^{+} + P_1W^{\perp})\begin{bmatrix} y \\ b \end{bmatrix}, \tag{3.5}$$

$$BLUE_{\widehat{M}}(X\alpha) = ([X, \quad 0]W^{+} + P_2W^{\perp})\begin{bmatrix} y \\ b \end{bmatrix}, \tag{3.6}$$

$$BLUP_{\widehat{M}}(H\varepsilon) = \left(\left[0, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right]W^{+} + P_3W^{\perp}\right)\begin{bmatrix} y \\ b \end{bmatrix} \tag{3.7}$$

$$BLUP_{\widehat{M}}(\varepsilon) = \left(\left[0, \quad \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right]W^{+} + P_4W^{\perp}\right)\begin{bmatrix} y \\ b \end{bmatrix} \tag{3.8}$$

where $P_i \in R^{s\times(n+m)}$ is arbitrary matrix, $i = 1, \ldots, 4$.

Proof. Let $L\begin{bmatrix} y \\ b \end{bmatrix}$ and $T\begin{bmatrix} y \\ b \end{bmatrix}$ be two unbiased linear predictors for ϕ under the model in (2.3). Then, we can write the following expressions

$$E\left(L\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) = 0 \Leftrightarrow L\begin{bmatrix} X \\ A \end{bmatrix} = K, \text{ i.e., } [L, \quad -I_s]\begin{bmatrix} X \\ A \\ K \end{bmatrix} = 0, \tag{3.9}$$

$$E\left(T\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) = 0 \Leftrightarrow T\begin{bmatrix} X \\ A \end{bmatrix} = K, \text{ i.e., } [T, \quad -I_s]\begin{bmatrix} X \\ A \\ K \end{bmatrix} = 0, \tag{3.10}$$

$$D\left(L\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) = \sigma^2\begin{bmatrix} (L-H)\Sigma(L-H)' & 0 \\ 0 & 0 \end{bmatrix} \tag{3.11}$$

$$= \sigma^2 [L, \quad -I_s] \begin{bmatrix} I_{n+m} \\ H \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_{n+m} \\ H \end{bmatrix}' [L, \quad -I_s]'$$

$$D\left(T\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) = \sigma^2 \begin{bmatrix} (T-H)\Sigma(T-H)' & 0 \\ 0 & 0 \end{bmatrix}$$

$$= \sigma^2 [T, \quad -I_s] \begin{bmatrix} I_{n+m} \\ H \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_{n+m} \\ H \end{bmatrix}' [T, \quad -I_s]' \tag{3.12}$$

Then the matrix minimization problem for finding the BLUP under the model in (2.3) characterized in Definition 3.2 can be accordingly expressed as to find solution L of the consistent linear matrix equation

$$L\begin{bmatrix} X \\ A \end{bmatrix} = K \text{ such that } D\left(L\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) \preccurlyeq D\left(T\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) \text{ s.t. } T\begin{bmatrix} X \\ A \end{bmatrix} = K$$

i.e.,

$$\begin{aligned} &[L, \quad -I_s] \begin{bmatrix} I_{n+m} \\ H \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_{n+m} \\ H \end{bmatrix}' [L, \quad -I_s]' \\ &\quad \preccurlyeq [T, \quad -I_s] \begin{bmatrix} I_{n+m} \\ H \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_{n+m} \\ H \end{bmatrix}' [T, \quad -I_s]' \end{aligned} \tag{3.13}$$

(3.13) is a standard constrained quadratic matrix-valued function optimization problem in the Löwner partial ordering as given in Lemma 2.2. Applying (2.7) to (3.13), the maximal positive inertia of $D\left(L\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) - D\left(T\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right)$ subject to $T\begin{bmatrix} X \\ A \end{bmatrix} = K$ is obtained as follows:

$$\max_{E\left(T\begin{bmatrix} X \\ A \end{bmatrix} - \phi\right) = 0} i_+\left(D\left(L\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) - D\left(T\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right)\right)$$

$$= r\begin{bmatrix} [L, \quad -I_s] \begin{bmatrix} I_{n+m} \\ H \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_{n+m} \\ H \end{bmatrix}' [L, \quad -I_s]' \\ \begin{bmatrix} X \\ A \\ K \end{bmatrix}' \end{bmatrix} - r\begin{bmatrix} X \\ A \\ K \end{bmatrix}$$

$$= r\left([L, \quad -I_s] \begin{bmatrix} I_{n+m} \\ H \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_{n+m} \\ H \end{bmatrix}' \begin{bmatrix} X \\ A \\ K \end{bmatrix}^{\perp}\right) \tag{3.14}$$

Combining (3.9) with (3.14), we conclude that $D\left(L\begin{bmatrix} y \\ b \end{bmatrix} - \phi\right) = min \Leftrightarrow$ there exists L satisfying both

$$L\begin{bmatrix} X \\ A \end{bmatrix} = K \text{ and } [L, \quad -I_s] \begin{bmatrix} I_{n+m} \\ H \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I_{n+m} \\ H \end{bmatrix}' \begin{bmatrix} X \\ A \\ K \end{bmatrix}^{\perp} = 0$$

i.e.,

$$L\begin{bmatrix} y \\ b \end{bmatrix} = BLUP_{\widehat{M}}(\phi) \Leftrightarrow L\left[\begin{bmatrix} X \\ A \end{bmatrix}, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right] = \left[K, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right].$$

This matrix equation is consistent. By using Lemma 2.1, we obtain (3.2). (a) and (b) follow directly from (3.2). For (c), we refer (Tian, 2013 [Lemma 2.1(a)]). (3.3) is seen from (3.2) and the assumptions in (2.2). By using (2.4) and (2.6),

$$\begin{aligned} cov\{BLUP_{\widehat{M}}(\phi), \phi\} &= \left[K, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right] \\ &\quad \times \left[\begin{bmatrix} X \\ A \end{bmatrix}, \quad \begin{bmatrix} H\Sigma H' & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X \\ A \end{bmatrix}^{\perp}\right] \begin{bmatrix} \Sigma H' & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{3.15}$$

is obtained. (3.4) is seen from (3.3) and (3.15). (3.5)-(3.8) are directly seen from (3.2).

■

Now, consider the constrained linear model M in (2.1) with its partitioned form

$$M_p: y = X\alpha + \varepsilon = [X_1, \quad X_2]\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} + \varepsilon = X_1\alpha_1 + X_2\alpha_2 + \varepsilon, A_1\alpha_1 = b_1, \tag{3.16}$$

where $X_i \in R^{n\times k_i}$, $\alpha_i \in R^{k_i\times 1}$, $A_1 \in R^{m\times k_1}$ is a known matrix of arbitrary rank, and $b_1 \in R^{m\times 1}$ is a known vector, $i = 1,2$, $k_1 + k_2 = k$. The linear restriction equation $A_1\alpha_1 = b_1$ in (3.16) is consistent.

Merging two given equation parts in (3.16), we can write the following model:

$$\widehat{M_p}: \begin{bmatrix}y\\ b_1\end{bmatrix} = \begin{bmatrix}X_1 & X_2\\ A_1 & 0\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} + \begin{bmatrix}\varepsilon\\ 0\end{bmatrix} = \begin{bmatrix}X_1\\ A_1\end{bmatrix}\alpha_1 + \begin{bmatrix}X_2\\ 0\end{bmatrix}\alpha_2 + \begin{bmatrix}\varepsilon\\ 0\end{bmatrix}, \tag{3.17}$$

and a general vector containing the both unknown vectors can be considered as follows

$$\phi_1 = K_1\alpha_1 + H\varepsilon = [K_1, \quad 0]\alpha + [H, \quad F]\begin{bmatrix}\varepsilon\\ 0\end{bmatrix} \tag{3.18}$$

for given matrices $K_1 \in R^{s\times k_1}$, $H \in R^{s\times n}$, and $F \in R^{s\times m}$, where

$$E(\phi_1) = K_1\alpha_1, D(\phi_1) = \sigma^2\begin{bmatrix}H\Sigma H' & 0\\ 0 & 0\end{bmatrix}, cov\left(\phi_1, \begin{bmatrix}y\\ b_1\end{bmatrix}\right) = \sigma^2\begin{bmatrix}H\Sigma & 0\\ 0 & 0\end{bmatrix}. \tag{3.19}$$

Let ϕ_1 be predictable by $\begin{bmatrix}y\\ b_1\end{bmatrix}$ in (2.3). According to Theorem (3.1), following results hold.

$$\mathrm{BLUP}_{\widehat{M_p}}(\phi_1) = \mathrm{L}\begin{bmatrix}y\\ b_1\end{bmatrix} = \left(\left[K_1, \quad \begin{bmatrix}H\Sigma H' & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}X_1 & X_2\\ A_1 & 0\end{bmatrix}^{\perp}\right]W_n^{+} + PW_n^{\perp}\right)\begin{bmatrix}y\\ b_1\end{bmatrix}, \tag{3.20}$$

where $P \in R^{s\times(n+m)}$ is an arbitrary matrix and $W_n = \left[\begin{bmatrix}X_1 & X_2\\ A_1 & 0\end{bmatrix}, \quad \begin{bmatrix}H\Sigma H' & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}X_1 & X_2\\ A_1 & 0\end{bmatrix}^{\perp}\right]$, in particular,

$$\begin{aligned} D\left[BLUP_{\widehat{M_p}}(\phi_1)\right] = \sigma^2\left[K_1, \quad \begin{bmatrix}H\Sigma H' & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}X_1 & X_2\\ A_1 & 0\end{bmatrix}^{\perp}\right]W_n^{+}\begin{bmatrix}H\Sigma H' & 0\\ 0 & 0\end{bmatrix} \\ \times\left(\left[K_1, \quad \begin{bmatrix}H\Sigma H' & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}X_1 & X_2\\ A_1 & 0\end{bmatrix}^{\perp}\right]W_n^{+}\right)' \end{aligned} \tag{3.21}$$

$$\begin{aligned} D\left[\phi_1 - BLUP_{\widehat{M_p}}(\phi_1)\right] = \sigma^2\left(\left[K_1, \quad \begin{bmatrix}H\Sigma H' & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}X_1 & X_2\\ A_1 & 0\end{bmatrix}^{\perp}\right]W_n^{+} - [H, \quad F]\right) \\ \times\begin{bmatrix}H\Sigma H' & 0\\ 0 & 0\end{bmatrix}\left(\left[K_1, \quad \begin{bmatrix}H\Sigma H' & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}X_1 & X_2\\ A_1 & 0\end{bmatrix}^{\perp}\right]W_n^{+} - [H, \quad F]\right)', \end{aligned} \tag{3.22}$$

$$BLUE_{\widehat{M_p}}(K_1\alpha_1) = \left([K_1, \quad 0]W_n^{+} + P_1W_n^{\perp}\right)\begin{bmatrix}y\\ b_1\end{bmatrix}, \tag{3.23}$$

$$BLUP_{\widehat{M_p}}(H\varepsilon) = \left(\left[0, \quad \begin{bmatrix}H\Sigma H' & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}X_1 & X_2\\ A_1 & 0\end{bmatrix}^{\perp}\right]W_n^{+} + P_2W_n^{\perp}\right)\begin{bmatrix}y\\ b_1\end{bmatrix}, \tag{3.24}$$

where P_1 and $P_2 \in R^{s\times(n+m)}$ are arbitrary matrices.

References

Alalouf IS, Styan GPH, 1979. Characterizations of estimability in the general linear model. Ann. Stat., 7: 194 - 200.

Dong B, Guo W, Tian Y, 2014. On relations between BLUEs under two transformed linear models. J. Multivariate Anal., 131: 279 - 292.

Drygas H., 1970. The coordinate-free approach to Gauss–Markov estimation. Springer, Heidelberg.

Goldberger AS, 1962. Best linear unbiased prediction in the generalized linear regression model. J. Amer. Statist. Assoc., 57: 369 - 375.

Güler N, Büyükkaya ME, 2021. Inertia and rank approach in transformed linear mixed models for comparison of BLUPs. Commun. Statist. Theory and Methods, DOI: 10.1080/03610926.2021.1967397.

Penrose R, 1955. A generalized inverse for matrices. Proc. Cambridge Philos. Soc., 51: 406 - 413.

Puntanen S, Styan GPH, Isotalo J, 2011. Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty, Springer, Heidelberg.

Rao CR, 1973. Representations of best linear unbiased estimators in the Gauss-Markoff model with a singular dispersion matrix. J. Multivariate Anal., 3: 276 - 292.

Tian Y, 2012. Solving optimization problems on ranks and inertias of some constrained nonlinear matrix functions via an algebraic linearization method. Nonlinear Analysis, 75: 717 - 734.

Tian Y, 2013. On properties of BLUEs under general linear regression models,J. Statist. Plann. Inference, 143: 771 - 782.

Tian Y, 2015. A new derivation of BLUPs under random-effects model. Metrika, 78: 905 - 918.

Acknowledgment

In this section, acknowledgments can be made to the persons or institutions who supported the study.

Conflict of Interest

The authors have declared that there is no conflict of interest.

Author Contributions

Contributions of the authors to the study can be explained.

Climate Anxiety in Young People and Their Beliefs About Government Responses to Climate Change: A Scale Adaptation Study

Muhammet Atalay[1*], **Ayşenur Gercik**[2]

[1] Kırklareli University, Faculty of Economics and Administrative Sciences, Department of Business Administration, 39100, Kırklareli, Turkey
[2] Kırklareli University, Social Sciences Institute, Department of Business Administration, 39100, Kırklareli, Turkey

*Corresponding author e-mail: atalay@klu.edu.tr

Abstract

In this study, the Turkish adaptation of the Climate Concern and Climate Change Policy Outlook in Children and Young People scale developed by Hickman et al. (2021) was made and the perspectives of young people between the ages of 18-24 on climate change concerns and the work done by governments were measured. In the study, first of all, the definition of climate and climate change and its possible effects are explained. In the research part, the process, method and findings of the research are explained. The scale, which was adapted, was answered by 303 people in the electronic media. First of all, the answers were analyzed by frequency analysis. Then, the factor structure of the scale was revealed by explanatory factor analysis. Finally, the item reliability of the factors was investigated. The factor structure of the scale was found to have one factor in one item and two factors in three items. Item reliability of the factors is sufficient. The findings of the study showed that young people are worried about climate change, they are not very pessimistic even though it affects their daily lives, and they do not have enough information about climate change and its effects.

Key words: *Climate Change, Climate Anxiety, Factor Analysis, Reliability, Scale Adaptation*

Research on Workaholism and Burnout Syndrome in Textile Business Employees

Bihter Çakır[1], Muhammet Atalay[2*]

[1] Kırklareli University, Social Sciences Institute, Department of Business Administration, 39100, Kırklareli, Turkey

[2] Kırklareli University, Faculty of Economics and Administrative Sciences, Department of Business Administration, 39100, Kırklareli, Turkey

*Corresponding author e-mail: atalay@klu.edu.tr

Abstract

The aim of this study is to determine the relationship between workaholism and burnout levels in blue-collar textile production enterprise employees and to measure their effects. For this purpose, a company operating in Çerkezköy Organized Industrial Zone was reached by convenience sampling method and scales were applied randomly to 513 of these company's 800 employees. The incomplete or erroneous surveys were removed and the remaining 379 questionnaires were evaluated. According to the findings of the study, there is a significant and positive relationship between workaholism and burnout and it was concluded that the level of burnout increased as workaholism increased. In addition, age, experience in the institution, weekly working time, livelihood status, annual leave use status, time allocation to family and social activities variables had a significant effect on the dimensions of workaholism and burnout.

Key words: *Workaholism, Burnout Syndrome, Factor Analysis, Difference Tests, Correlation*

This study is based on the master's thesis written by the first author under the supervision of the second author.

Comparison of Environmental Performances of Cities in Central Anatolian Region Using Fucom and Grey Relational Analysis Methods

Rahmi Baki[1]

[1] Aksaray University, Faculty of Economics and Administrative Sciences, Department of Management Information Systems, 68100, Aksaray, Turkey

*Corresponding author e-mail: rahmibaki@aksaray.edu.tr

Abstract

Comparing various cities or countries in terms of environmental indicators is an important step towards achieving environmental goals. Given the diverse conditions in different regions, the baseline must be accurately measured to improve environmental performances and formulate relevant strategies more effectively. In the current study, it is aimed to compare the provinces in the Central Anatolia region with each other in terms of environmental performance with an integrated approach, in which fuzzy Full Consistency Method (FUCOM) and Grey Relational Analysis (GRA) techniques are used together. In practice, four experts evaluated the criteria with fuzzy FUCOM technique. Then, alternatives are listed through the GRA technique. According to the criteria weights, the average of the particulate matter station values (0.314), the proportion of the population provided with waste service (0.279), the forest area per 〖KM〗^2 (0.147), the satisfaction rate from cleaning services (0.137), and the rate of those experiencing noise problems from the street (0.123) are listed as. The provinces with the highest environmental performance are Eskişehir (0.842), Konya (0.689) and Kırıkkale (0.67), while the lowest ones are Aksaray (0.402), Niğde (0.457) and Nevşehir (0.486).

Key words: *Environmental Performance, Fuzzy Logic, FUCOM, GRA, Multi-Criteria Decision Making*

INTRODUCTION

Today, modern cities are complex systems that contain large numbers of citizens, businesses, different modes of transportation, communication networks and public services (Neirotti et al., 2014). Cities around the world are experiencing a significant urban population growth and urbanization (Carli, Dotoli, & Pellegrino, 2018). Uncontrolled growth in the city population and environmental pollution make the sustainability of cities increasingly difficult. Rapid economic growth and accelerating urbanization cause devastating effects on the environment. For this reason, planners and policy makers in developed and developing countries are working to keep the country's resources available for future generations (Suganthi, 2020).

Environmental performance evaluation is a useful tool that will contribute to decision makers' strategy formation and future decisions. These assessments can be applied to many different contexts at all scales, including villages, cities, countries and international organizations (Zuo et al., 2017). Environmental benchmarks and performance measurements are critical steps for environmental development planning. For an effective understanding of sustainability, environmental and ecological indicators should be considered as a whole. In order to evaluate environmental performances of cities, it is important to compare them with other cities (Eğilmez, Gümüs, & Küçükvar, 2015). For this reason, it is necessary to determine appropriate scientific approaches. In general, different methods have been adopted for environmental performance assessment (García-Sánchez, Almeida, & Camara, 2015). Although many different approaches have been tried in performance evaluation, effective and consistent solutions have been reached through multi-criteria decision-making techniques, the applications of which have been increasing recently (Ayçin & Çakın, 2019).

In Turkey Environmental Problems and Prioritization Report, environmental problems of the provinces in Turkey were evaluated and compared in terms of geographical regions (MEUCC, 2019). The results showed that especially the provinces in the Central Anatolia region have environmental problems in terms of different indicators. In the region where there are thirteen provinces in total, water pollution is the primary environmental problem in six provinces, namely Aksaray, Ankara, Çankırı, Kırşehir, Nevşehir and Yozgat. However, air pollution, waste, noise pollution and erosion are other environmental problems faced by the provinces in the region.

Multi-criteria evaluation methods play a central role in the multidimensional evaluation process (Boggia & Cortina, 2010). For these reasons, multi-criteria methodologies have been widely used in environmental assessments. In the current study, an integrated decision-making framework was used to analyze and compare environmental performances of thirteen provinces in the Central Anatolian Region of Turkey. For this, an integrated approach using fuzzy Full Consistency Method (FUCOM) for determining criterion weights and Grey Relational Analysis (GRA) techniques for ranking environmental performances of provinces has been proposed.

In the second part of this study, studies in the literature that analyze environmental and sustainability performances are included. In the third part, the method adopted in the application is introduced, and in the fourth part, the steps of the application are presented. Finally, in the Discussion and Conclusion section, the findings obtained in the study are explained and suggestions for future studies are made..

LITERATURE REVIEW

Today, it is seen that there is a great urban population growth and urbanization all over the world (Carli, Dotoli, & Pellegrino, 2018). Rapid economic growth and accelerating urbanization cause devastating effects on the environment. In addition, rapid developments in the industrial sector have made it difficult to plan resources (Suganthi, 2020). Comparing different cities or countries in terms of environmental indicators is a serious step towards the realization of environmental targets. In many studies in the literature, different regions have been evaluated in terms of environmental and sustainability performances, considering various criteria. In this part of the study, studies on the subject in the literature are included.

Eğilmez, Gümüş, and Küçükvar (2015) evaluated the environmental sustainability performances of twenty-seven metropolises in the USA and Canada. In the study, firstly, sixteen sustainability indicators were determined and the list was finalized as a result of expert evaluations. Then, sustainable performance scores were obtained through the proposed intuitive decision-making model and expert evaluation. Finally, the sustainability scores and rankings of twenty-seven metropolises are presented. Arı, Özköse and Gencer (2016) compared cities and geographical regions in Turkey according to their air quality for each season separately. With this ranking made using the SMAA-2 technique, the conditions of cities and geographical regions in terms of air quality are revealed. Carli, Dotoli, and Pellegrino (2018) carried out an application in the smart metropolitan city perspective, using the AHP technique, to analyze the sustainability development of energy, water and environmental systems. Thirty-five indicators were used in the study and compared to four metropolitan municipalities in Bari, Italy.

Ayçin and Çakın (2019) proposed a decision-making model that evaluates the environmental performance of selected countries based on Entropy, GRA and MOORA techniques. Şepit and Paksoy

(2019) compared the sustainability performance of nine cities in Turkey by considering eleven criteria. Suganthi (2020) proposes a decision model using Data Envelopment Analysis and AHP techniques to measure sustainable energy development efficiency. Forty-eight developing countries were compared in the study.

Arsu and Ayçin (2021) analyzed OECD countries by dividing them into two clusters. In the study, the weights of the criteria were determined by the CRITIC technique and the alternatives in each cluster were ranked using the MARCOS method. It has been seen that the most important criteria for the first cluster are the inflation rate and unemployment rate, while the most important criteria for the second cluster are the life satisfaction index and ecological footprint. Yang and Zhang (2021) proposed a decision-making framework with eight indicators for evaluating the performance of strategies developed for sustainable urban drainage system development. In the present study, thirteen provinces in the Central Anatolia region were evaluated in terms of their environmental performance based on five indicators.

METHODOLOGY

One of the most important steps in solving decision-making problems is to determine the weights of criteria that express the importance of the criteria and their effects on the evaluation results. There are many techniques used to obtain criterion weights in the literature. FUCOM is a current model proposed for the determination of criterion weights, based on pairwise comparisons of criteria and validation of results along a deviation from maximum consistency (Pamučar et al., 2018). The method has advantages such as having few pairwise comparisons, identifying deviations from maximum consistency, and evaluating transitivity during pairwise comparisons of criteria (Pamučar, Stević, & Sremac, 2018). In solving real decision problems, decision makers prefer to evaluate attributes using linguistic variables instead of exact values. Information on qualifications obtained from decision makers may be unclear or incomplete. Fuzzy set theory is one of the mathematical tools used in solving such problems (Pamučar and Ecer, 2020). Decision-making problems involving uncertain information can be modeled using fuzzy set theory. In the current study, the weights of the criteria were obtained using the fuzzy FUCOM technique.

GRA is a decision-making approach based on measuring the approximation degree between sequences using the grey relationship degree (Tzeng et al., 2009). In order to measure the relationship between two sequences, this relationship needs to be calculated numerically, and the calculated relationship degree is called the grey relationship degree (Feng & Wang, 2000). GRA can provide evaluation by considering many criteria together and can produce results with a small amount of data. Moreover, under conditions where the distribution is unknown or not normal, ranking can be made according to the degree of relationship (Rajesh & Ravi, 2015). The method is used to solve complex relationships between multiple performance characteristics through the optimization of grey relational degrees (Zeng et al., 2007).

In the current study, a ten-step method in which fuzzy FUCOM and GRA techniques are used together is proposed. In the first four steps of the proposed method, weights of criteria are determined with the fuzzy FUCOM technique, while the alternatives are compared with the GRA technique in the other steps. With the developed hybrid approach, it is aimed to benefit from the advantages of the two methods. The proposed model has been tested by comparing the development levels of thirteen provinces in the Central Anatolian Region in terms of environmental indicators.

Step 1. The criteria are ranked according to their importance in line with the opinions of the decision makers.

Step 2. The criteria are compared by means of a fuzzy pairwise comparison scale (Table 1) and using Equation (1), the comparative preferences of the criteria are obtained.

Table 1. Fuzzy pairwise comparison scale

Linguistic Variables	l_{ij}
Equally Important (EQI)	(1; 1; 1)
Less Important (LI)	(0.667; 1; 1.5)
Moderately Important (MI)	(1.5; 2; 2.5)
Quite Important (QI)	(2.5; 3; 3.5)
Extremely Important (EXI)	(3.5; 4; 4.5)

Step 3. The fuzzy weights of the criteria are calculated. At this stage, the conditions specified in Equation (2) and Equation (3) must be met. To determine the optimal fuzzy values of the criterion weights, the mathematical model given in Equation (4) is solved. In order to ensure the highest consistency in the model, $\frac{w_{j(k)}}{w_{j(k+1)}} - \Phi_{k/(k+1)} = 0$ and $\frac{w_{j(k)}}{w_{j(k+2)}} - \Phi_{k/(k+1)} \cdot \Phi_{(k+1)/(k+2)} = 0$ conditions must be met. Here the expressions are given in $\widetilde{w}_j = (w_j^l, w_j^m, w_j^u)$ ve $\widetilde{\Phi}_{k/(k+1)} = (\widetilde{\Phi}_{k/(k+1)}^l, \widetilde{\Phi}_{k/(k+1)}^m, \widetilde{\Phi}_{k/(k+1)}^u)$ format.

$$\frac{w_k}{w_{k+1}} = \Phi_{k/(k+1)} \tag{2}$$

$$\frac{w_k}{w_{k+2}} = \Phi_{k/(k+1)} \cdot \Phi_{(k+1)/(k+2)} \tag{3}$$

$$\begin{aligned}
&\min \chi \\
&\text{Constraints} \\
&\left|w_k - w_{k+1} \otimes \Phi_{k/(k+1)}\right| \leq \chi, \forall \text{j} \\
&\left|w_k - w_{k+2} \otimes \Phi_{k/(k+1)} \otimes \Phi_{(k+1)/(k+2)}\right| \leq \chi, \forall \text{j} \\
&\textstyle\sum_{j=1}^{n} \widetilde{w} = 1, \forall \text{j} \\
&w_j^l \leq w_j^m \leq w_j^u \\
&w_j^l \geq 0, \forall \text{j} \\
&\text{j} = 1, 2, \ldots, \text{n}
\end{aligned} \tag{4}$$

Step 4. The outputs obtained as a result of the solution of the mathematical model give the triangular fuzzy coefficient values (l_j, m_j, u_j) of the criteria. The geometric mean of the criteria weights obtained for all decision makers is taken. Then the fuzzy values are converted to net weights using Equation (5). Each net value is normalized by dividing by the sum of the criteria weights. The criteria weights are displayed in the format w_1, w_2, w_3, … ,w_n. The criterion with the highest value has the highest degree of importance.

$$\text{R}(a_j) = (l_j + 4m_j + u_j)/6 \tag{5}$$

Step 5. The alternative factor series to be compared is determined as shown in Equation (6).

$$x_i = (x_i^{(j)}, \ldots, (x_i^{(n)}) \tag{6}$$

The expression x_i in Equation (6) represents the decision alternatives, and the expression $x_i^{(j)}$ is i[th] decision alternative j[th] indicates the value it receives for the criterion. At this stage, the decision matrix (X) is obtained (Equation (7)).

$$X = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(n)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(n)} \\ \vdots & \vdots & \cdots & \vdots \\ x_m^{(1)} & x_m^{(2)} & \cdots & x_m^{(n)} \end{bmatrix} \tag{7}$$

Step 6. In order to compare the factors in the decision problem, a reference series is created as given in Equation (8). The $x_0^{(j)}$ value in Equation (8) indicates the best value of the j[th] criterion among the normalized values to be obtained. The reference series is created by determining the ideal value for each value in a decision alternative.

$$\mathrm{x}_0 = (\mathrm{x}_0^{(\mathrm{j})}) \;\; \mathrm{j} = 1,2, \dots, \mathrm{n} \tag{8}$$

Step 7. In this step, the values in the decision matrix are normalized. The normalization process takes place in three different ways according to benefit, cost and optimal situations. Normalization is done by using Equation (9) for benefit-based criteria, Equation (10) for cost-based criteria, and Equation (11) for optimal-based criteria. Thus, the normalized decision matrix is obtained.

$$x_i^* = \frac{x_i^{(j)} - \min_j x_i^{(j)}}{\max_j x_i^{(j)} - \min_j x_i^{(j)}} \tag{9}$$

$$x_i^* = \frac{\max_j x_i^{(j)} - x_i^{(j)}}{\max_j x_i^{(j)} - \min_j x_i^{(j)}} \tag{10}$$

$$x_i^* = \frac{x_i^{(j)} - x_{0b}^{(j)}}{\max_j x_i^{(j)} - x_{0b}^{(j)}} \tag{11}$$

Step 8. Via Equation (12), the difference of values between the values of the reference series and the normalized decision matrix is taken. Thus, the absolute value matrix (Δ_{0i}) is formed as stated in Equation (13).

$$\Delta_{0i} = x_0^{(j)} - x_i^{(j)} \tag{12}$$

$$\Delta_{0i} = \begin{bmatrix} \Delta_{01}^{(1)} & \Delta_{01}^{(2)} & \cdots & \Delta_{01}^{(n)} \\ \Delta_{02}^{(1)} & \Delta_{02}^{(2)} & \cdots & \Delta_{02}^{(n)} \\ \vdots & \vdots & \cdots & \vdots \\ \Delta_{0m}^{(1)} & \Delta_{0m}^{(2)} & \cdots & \Delta_{0m}^{(n)} \end{bmatrix} \tag{13}$$

Step 9. Using Equation (14-15-16), grey relational coefficient matrix is obtained. The expression ζ in equation (14) takes a value in the range of [0,1] and is called the discriminating coefficient.

$$\gamma_{0i}^{(j)} = \frac{\Delta_{min} + \zeta\Delta_{maks}}{\Delta_{0i}^{(j)} + \zeta\Delta_{maks}} \tag{14}$$

$$\Delta_{max} = \max_i \max_j \Delta_{0i}^{(j)} \tag{15}$$

$$\Delta_{min} = \min_i \min_j \Delta_{0i}^{(j)} \tag{16}$$

Step 10. The grey relational degree (Γ_{0i}) is obtained through Equation (17). The grey relational degree is a measure of the geometric similarity between an alternative and a reference series. The large grey relational degree of the alternative indicates that there is a strong relationship with the reference series. The $w_i^{(j)}$ value in Equation (17) indicates the criterion weight of the j^th^ criterion.

$$\Gamma_{0i} = \sum_{j=1}^{n} [w_i^{(j)} . \gamma_{0i}^{(j)}] \tag{17}$$

APPLICATION

In the current study, 13 provinces in the Central Anatolia Region were compared in terms of environmental indicators. These provinces are listed as Aksaray (A_1), Ankara (A_2), Çankırı (A_3), Eskişehir (A_4), Karaman (A_5), Kayseri (A_6), Kırıkkale (A_7), Kırşehir (A_8), Konya (A_9), Nevşehir (A_{10}), Niğde (A_{11}), Sivas (A_{12}) and Yozgat (A_{13}). Provinces were evaluated considering average of the particulate matter station values (C_1), the forest area per KM^2 (C_2), the proportion of the population provided with waste service (C_3), the rate of those experiencing noise problems from the street (C_4) and the satisfaction rate from cleaning services (C_5). Information on the said criteria was obtained from the statistical data published by TSI (TSI, 2021).

In practice, first of all, four decision makers evaluated the criteria according to their importance. In this process, fuzzy FUCOM technique was applied. Then, the cities were listed using the GRA technique. A ten-step process was adopted in the study. While the criteria weights were determined in the first four steps, the cities were compared in the next steps. The steps followed in the application are presented below.

Step 1. The results of ranking the criteria in order of importance by the four decision makers are given in Table 2.

Table 2. Ranking of Criteria

DM_j	**Criteria**
DM_1	C_1>C_3>C_2>C_5>C_4
DM_2	C_3>C_1>C_4>C_2>C_5
DM_3	C_1>C_5>C_3>C_2>C_4
DM_4	C_1>C_3>C_5>C_4>C_2

Step 2. Pairwise comparisons of the criteria were made using the scale given in Table 1 (Table 3). Pairwise comparisons of the criteria were calculated using Equation (1).

Table 3. Pairwise comparisons of criteria

DM_j	**Pairwise Comparisons**			
DM_1	C_1-C_3 LI	C_3-C_2 MI	C_2-C_5 QI	C_5-C_4 LI
DM_2	C_3-C_1 EQI	C_1-C_4 MI	C_4-C_2 EQI	C_2-C_5 LI
DM_3	C_1-C_5 LI	C_5-C_3 EQI	C_3-C_2 EQI	C_2-C_4 LI
DM_4	C_1-C_3 LI	C_3-C_5 QI	C_5-C_4 LI	C_4-C_2 EQI

Comparative preferences for DM_1 were calculated as given below.

$$\widetilde{\varphi}_{1/3} = \widetilde{w}_{C_1}/\widetilde{w}_{C_3} = (0.667;\, 1;\, 1.5)\ /\ (1;\, 1;\, 1) = (\tfrac{0.667}{1};\tfrac{1}{1};\tfrac{1.5}{1}) = (0.667;\, 1;\, 1.5)$$
$$\widetilde{\varphi}_{3/2} = \widetilde{w}_{C_3}/\widetilde{w}_{C_2} = (1.5;\, 2;\, 2.5)\ /\ (0.667;\, 1;\, 1.5) = (\tfrac{1.5}{1.5};\tfrac{2}{1};\tfrac{2.5}{0.667}) = (1;\, 2;\, 3.748)$$
$$\widetilde{\varphi}_{2/5} = \widetilde{w}_{C_2}/\widetilde{w}_{C_5} = (2.5;\, 3;\, 3.5)\ /\ (1.5;\, 2;\, 2.5) = (\tfrac{2.5}{2.5};\tfrac{3}{2};\tfrac{3.5}{1.5}) = (1;\, 1.5;\, 2.333)$$
$$\widetilde{\varphi}_{5/4} = \widetilde{w}_{C_5}/\widetilde{w}_{C_4} = (0.667;\, 1;\, 1.5)\ /\ (2.5;\, 3;\, 3.5) = (\tfrac{0.667}{3.5};\tfrac{1}{3};\tfrac{1.5}{2.5}) = (0.191;\, 0.333;\, 0.6)$$

The preference vectors for all decision makers are given below.

$$\widetilde{\Phi}_{DM_1} = ((0.667;\, 1;\, 1.5)),\, (1;\, 2;\, 3.748),\, (1;\, 1.5;\, 2.333),\, (0.191;\, 0.333;\, 0.6))$$
$$\widetilde{\Phi}_{DM_2} = ((1;\, 1;\, 1)),\, (1.5;\, 2;\, 2.5),\, (0.4;\, 0.5;\, 0.667),\, (0.667;\, 1;\, 1.5))$$
$$\widetilde{\Phi}_{DM_3} = ((0.667;\, 1;\, 1.5)),\, (0.667;\, 1;\, 1.5),\, (1;\, 1;\, 1),\, (0.667;\, 1;\, 1.5))$$
$$\widetilde{\Phi}_{DM_4} = ((0.667;\, 1;\, 1.5)),\, (1.667;\, 3;\, 5.247),\, (0.191;\, 0.333;\, 0.6),\, (0.667;\, 1;\, 1.5))$$

Step 3. Using Equation (2) and Equation (3), the mathematical model is established. For DM_1, the transitivity criteria obtained using Equation (3) are given below.

$$\widetilde{w}_{C_1}/\widetilde{w}_{C_2} = \widetilde{w}_{C_1}/\widetilde{w}_{C_3}\ .\ \widetilde{w}_{C_3}/\widetilde{w}_{C_2} = (0.667;\, 1;\, 1.5),\, (1;\, 2;\, 3.748) = (0.667;\, 2;\, 5.622)$$
$$\widetilde{w}_{C_3}/\widetilde{w}_{C_5} = \widetilde{w}_{C_3}/\widetilde{w}_{C_2}\ .\ \widetilde{w}_{C_2}/\widetilde{w}_{C_5} = (1;\, 2;\, 3.748),\, (1;\, 1.5;\, 2.333) = (1;\, 3;\, 8.744)$$
$$\widetilde{w}_{C_2}/\widetilde{w}_{C_4} = \widetilde{w}_{C_2}/\widetilde{w}_{C_5}\ .\ \widetilde{w}_{C_5}/\widetilde{w}_{C_4} = (1;\, 1.5;\, 2.333),\, (0.191;\, 0.333;\, 0.6) = (0.191;\, 0.5;\, 1.4)$$

In order to determine the fuzzy optimal value of the weight coefficients, a model was created for all decision makers by means of Equation (4). The model created for DM_1 is presented below.

$$\min \chi$$
Constraints:
$$\left(w_1^l - 0.667w_3^u\right) \le \chi \quad (w_2^m - 1.5w_5^m) \le \chi \quad \left(w_1^u - 5.622w_2^l\right) \le \chi$$
$$\left(w_1^l - 0.667w_3^u\right) \ge -\chi \quad (w_2^m - 1.5w_5^m) \ge -\chi \quad \left(w_1^u - 5.622w_2^l\right) \ge -\chi$$
$$(w_1^m - w_3^m) \le \chi \quad \left(w_2^u - 2.333w_5^l\right) \le \chi \quad \left(w_3^l - w_5^u\right) \le \chi$$
$$(w_1^m - w_3^m) \ge -\chi \quad \left(w_2^u - 2.333w_5^l\right) \ge -\chi \quad \left(w_3^l - w_5^u\right) \ge -\chi$$
$$\left(w_1^u - 1.5w_3^l\right) \le \chi \quad \left(w_5^l - 0.191w_4^u\right) \le \chi \quad (w_3^m - 3w_5^m) \le \chi$$
$$\left(w_1^u - 1.5w_3^l\right) \ge -\chi \quad \left(w_5^l - 0.191w_4^u\right) \ge -\chi \quad (w_3^m - 3w_5^m) \ge -\chi$$
$$\left(w_3^l - w_2^u\right) \le \chi \quad (w_5^m - 0.333w_4^m) \le \chi \quad \left(w_3^u - 8.744w_5^l\right) \le \chi$$
$$\left(w_3^l - w_2^u\right) \ge -\chi \quad (w_5^m - 0.333w_4^m) \ge -\chi \quad \left(w_3^u - 8.744w_5^l\right) \ge -\chi$$
$$(w_3^m - 2w_2^m) \le \chi \quad \left(w_5^u - 0.6w_4^l\right) \le \chi \quad \left(w_2^l - 0.9w_4^u\right) \le \chi$$
$$(w_3^m - 2w_2^m) \ge -\chi \quad \left(w_5^u - 0.6w_4^l\right) \ge -\chi \quad \left(w_2^l - 0.9w_4^u\right) \ge -\chi$$
$$\left(w_3^u - 3.748w_2^l\right) \ge \chi \quad \left(w_1^l - 0.667w_2^u\right) \ge \chi \quad (w_2^m - 0.9w_4^m) \ge \chi$$
$$\left(w_3^u - 3.748w_2^l\right) \ge -\chi \quad \left(w_1^l - 0.667w_2^u\right) \ge -\chi \quad (w_2^m - 0.9w_4^m) \ge -\chi$$
$$\left(w_2^l - w_5^u\right) \ge \chi \quad (w_1^m - 2w_2^m) \ge \chi \quad \left(w_2^u - 0.9w_4^l\right) \ge \chi$$
$$\left(w_2^l - w_5^u\right) \ge -\chi \quad (w_1^m - 2w_2^m) \ge -\chi \quad \left(w_2^u - 0.9w_4^l\right) \ge -\chi$$
$$\left(w_4^l\right) \ge \chi \quad \left(w_4^l\right) \ge -\chi \quad (w_4^m) \ge \chi \quad (w_4^m) \ge -\chi \quad (w_4^u) \ge \chi \quad (w_4^u) \ge -\chi$$
$$(w_1^l + 4w_1^m + w_1^u)/6 + (w_2^l + 4w_2^m + w_2^u)/6 + (w_3^l + 4w_3^m + w_3^u)/6 + (w_4^l + 4w_4^m + w_4^u)/6$$
$$(w_5^l + 4w_5^m + w_5^u)/6\ (w_6^l + 4w_6^m + w_6^u)/6 = 1$$
$$w_1^l \le w_1^m \le w_1^u;\, w_2^l \le w_2^m \le w_2^u;\, w_3^l \le w_3^m \le w_3^u;\, w_4^l \le w_4^m \le w_4^u;\, w_5^l \le w_5^m \le w_5^u;\ w_6^l \le w_6^m \le w_6^u$$
$$w_1^l, w_1^m, w_1^u, w_2^l, w_2^m, w_2^u, w_3^l, w_3^m, w_3^u, w_4^l, w_4^m, w_4^u, w_5^l, w_5^m, w_5^u, w_6^l, w_6^m, w_6^u, \ge 0$$

Step 4. As a result of the solution of the mathematical model, χ_{DM_1}= 0.086, χ_{DM_2}= 0.09, χ_{DM_3}= 0.089 ve χ_{DM_4}= 0.092 are found. The values determined for the four decision makers are combined by taking their geometric mean. Triangular fuzzy values are converted to net values via Equation (5). All values are normalized by dividing by the sum of the criteria weights. The weights of the criteria; are listed in w_1 = 0.314, w_2 = 0.147, w_3 = 0.279, w_4 = 0.123 and w_5 = 0.137.

Step 5. The decision matrix consisting of the data of the thirteen cities to be compared is obtained (Table 4).

Table 4. Decision matrix

-	C_1	C_2	C_3	C_4	C_5
A_1	61	0.017	0.76	0.168	0.562
A_2	63.375	0.155	0.994	0.27	0.739
A_3	34	0.255	0.697	0.083	0.575
A_4	31	0.252	0.989	0.215	0.81
A_5	79	0.265	0.773	0.099	0.787
A_6	75	0.064	0.997	0.217	0.757
A_7	27	0.097	0.867	0.097	0.556
A_8	35	0.037	0.788	0.116	0.716
A_9	55.5	0.121	0.990	0.155	0.827
A_{10}	47	0.013	0.753	0.117	0.694
A_{11}	68	0.071	0.733	0.071	0.615
A_{12}	30	0.099	0.731	0.132	0.757
A_{13}	42	0.164	0.706	0.142	0.721

Step 6. The reference series was created as specified in Equation (8).

Step 7. In practice, benefit-based criteria C_2, C_3 and C_5 are normalized through Equation (9), and cost-based criteria C_1 and C_4 are normalized through Equation (10). Thus, a normalized decision matrix is created (Table 5).

Table 5. Normalized decision matrix

-	C_1	C_2	C_3	C_4	C_5
Reference Series	1	1	1	1	1
A_1	0.346	0.016	0.21	0.513	0.021
A_2	0.3	0.564	0.991	0	0.674
A_3	0.865	0.96	0	0.944	0.07
A_4	0.923	0.948	0.975	0.274	0.935
A_5	0	1	0.255	0.859	0.853
A_6	0.077	0.201	1	0.266	0.741
A_7	1	0.336	0.568	0.869	0
A_8	0.846	0.097	0.303	0.774	0.59
A_9	0.452	0.428	0.978	0.579	1
A_{10}	0.615	0	0.187	0.771	0.508
A_{11}	0.212	0.232	0.121	1	0.216
A_{12}	0.942	0.341	0.113	0.696	0.74
A_{13}	0.712	0.6	0.031	0.645	0.61

Step 8. Using Equation (12), the absolute value matrix is obtained. The absolute value matrix is given in Table 6.

Table 6. Absolute value matrix

-	C_1	C_2	C_3	C_4	C_5
A_1	0.654	0.984	0.79	0.487	0.979
A_2	0.700	0.436	0.009	1	0.326
A_3	0.135	0.04	1	0.056	0.93
A_4	0.077	0.052	0.025	0.726	0.065
A_5	1	0	0.745	0.141	0.147
A_6	0.923	0.799	0	0.734	0.259
A_7	0	0.664	0.432	0.131	1
A_8	0.154	0.903	0.697	0.226	0.41
A_9	0.548	0.572	0.022	0.421	0
A_{10}	0.385	1	0.813	0.229	0.492
A_{11}	0.788	0.768	0.879	0	0.784
A_{12}	0.058	0.659	0.887	0.304	0.26
A_{13}	0.288	0.4	0.969	0.355	0.39

Step 9. The grey relational coefficient matrix is generated by means of Equation (14-5-16) (Table 7). In this process. the ζ value was accepted as 0.5.

Table 7. Grey relational coefficient matrix

-	C_1	C_2	C_3	C_4	C_5
A_1	0.433	0.337	0.388	0.506	0.338
A_2	0.417	0.534	0.983	0.333	0.605
A_3	0.788	0.926	0.333	0.899	0.350
A_4	0.867	0.906	0.952	0.408	0.885
A_5	0.333	1	0.402	0.780	0.772
A_6	0.351	0.385	1	0.405	0.659
A_7	1	0.430	0.537	0.793	0.333
A_8	0.765	0.356	0.418	0.688	0.549
A_9	0.477	0.466	0.958	0.543	1
A_{10}	0.565	0.333	0.381	0.686	0.504
A_{11}	0.388	0.394	0.362	1	0.389
A_{12}	0.897	0.432	0.361	0.622	0.658
A_{13}	0.634	0.556	0.340	0.585	0.561

Step 10. The grey correlation degrees of the alternatives are calculated using Equation (17). The $w_i^{(j)}$ values used in this step were obtained from the criterion weights obtained in Step 4 by means of the fuzzy FUCOM technique. The grey relational degrees and ranking results obtained by considering the criterion weights of the alternatives are given in Table 8. When Table 8 is examined, it is seen that the provinces with the highest level of development in terms of environmental indicators are Eskişehir (A_4), Konya (A_9) and Kırıkkale (A_7). The provinces with the lowest environmental development level are listed as Aksaray (A_1), Niğde (A_{11}) and Nevşehir (A_{10}).

Table 8. Ranking results

-	C_1	C_2	C_3	C_4	C_5	Γ_{0i}	Ranking
A_1	0.433	0.337	0.388	0.506	0.338	0.402	13
A_2	0.417	0.534	0.983	0.333	0.605	0.608	6
A_3	0.788	0.926	0.333	0.899	0.350	0.635	4
A_4	0.867	0.906	0.952	0.408	0.885	0.842	1
A_5	0.333	1	0.402	0.780	0.772	0.565	9
A_6	0.351	0.385	1	0.405	0.659	0.586	7
A_7	1	0.430	0.537	0.793	0.333	0.670	3
A_8	0.765	0.356	0.418	0.688	0.549	0.569	8
A_9	0.477	0.466	0.958	0.543	1	0.689	2
A_{10}	0.565	0.333	0.381	0.686	0.504	0.486	11
A_{11}	0.388	0.394	0.362	1	0.389	0.457	12
A_{12}	0.897	0.432	0.361	0.622	0.658	0.612	5
A_{13}	0.634	0.556	0.340	0.585	0.561	0.525	10
w_j	0.314	0.147	0.279	0.123	0.137	-	-

DISCUSSION AND CONCLUSION

Given the variety of conditions in different cities, the current situation must be accurately measured in order to improve environmental performances and formulate relevant strategies more effectively. For this reason, evaluating the environmental conditions of regions is a subject of study that attracts the attention of researchers. In the current study, the environmental performance of thirteen provinces in the Central Anatolia Region was compared based on five criteria. In practice, an integrated approach is proposed for regional performance assessment, in which fuzzy FUCOM and GRA techniques are used together.

According to the results of the application, it has been determined that the most effective criterion on the environmental performance of the provinces is the average of the particulate matter station values. The criteria are listed according to their weights as the average of the particulate matter station values (0.314), the proportion of the population provided with waste service (0.279), the forest area per KM^2 (0.147), the satisfaction rate from cleaning services (0.137), and the rate of those experiencing noise problems from the street (0.123). The provinces with the highest environmental performance are Eskişehir (0.842), Konya (0.689) and Kırıkkale (0.67). The provinces with the lowest environmental performance are listed as Aksaray (0.402), Niğde (0.457) and Nevşehir (0.486).

With the integrated methodology proposed in the study, environmental assessment and environmental performance comparison studies have been contributed. Thus, it is aimed to identify the weak and strong points of the regions in terms of environment, to facilitate the development of effective action plans and to serve the stakeholders. In future studies, the proposed approach can also be tested for different regions or country groups. In the current study, alternatives were listed considering only five criteria due to limitations in data availability. Diversifying the application with more criteria is a research topic that will contribute to the literature. In addition, the number of decision makers can be increased with the participation of experts who have experience in different fields of study.

References

Arı E, Özköse H, Gencer C, 2016. Ranking Turkish cities and regions for air quality using a multi-criteria decision-making method. *Polish Journal of Environmental Studies*, 25(5), 1823-1830.

Arsu T, Ayçin E, 2021. Evaluation of OECD countries with multi-criteria decision-making methods in terms of economic, social and environmental aspects. *Operational Research in Engineering Sciences: Theory and Applications*, *4*(2), 55-78.

Ayçin E, Çakın E, 2019. The evaluation of the environmental performance of the countries with multiple criteria decision-making methods and a fuzzy logic based approach. *Eskişehir Osmangazi University Journal of Economics and Administrative Sciences*, *14*(3), 631-656.

Boggia A, Cortina C, 2010. Measuring sustainable development using a multi-criteria model: A case study. *Journal of Environmental Management*, *91*(11), 2301-2306.

Carli R, Dotoli M, Pellegrino R, 2018. Multi-criteria decision-making for sustainable metropolitan cities assessment. *Journal of Environmental Management*, *226*, 46-61.

Eğilmez G, Gümüs S, Küçükvar M, 2015. Environmental sustainability benchmarking of the US and Canada metropoles: An expert judgment-based multi-criteria decision making approach. *Cities*, *42*, 31-41.

Feng CM, Wang RT, 2000. Performance evaluation for airlines including the consideration of financial ratios. *Journal of Air Transport Management*, *6*(3), 133-142.

García-Sánchez, IM, Almeida, TAN, Camara, RPB, 2015. A proposal for a Composite Index of Environmental Performance (CIEP) for countries. *Ecological Indicators*, *48*, 171-188.

Ministry of Environment, Urbanisation and Climate Change (MEUCC), 2019. Turkey Environmental Issues and Priorities Report. Access address: https://ced.csb.gov.tr/turkiye-cevre-sorunlari-ve-oncelikleri-raporu-i-82679; Date of access:27.03.2022.

Neirotti P, De Marco A, Cagliano AC, Mangano G, Scorrano, F, 2014. Current trends in Smart City initiatives: Some stylised facts. *Cities*, *38*, 25-36.

Pamucar D, Ecer F 2020. Prioritizing the weights of the evaluation criteria under fuzziness: The fuzzy full consistency method–FUCOM-F. *Facta Universitatis, Series: Mechanical Engineering*, *18*(3), 419-437.

Pamučar D, Lukovac V, Božanić D, Komazec N, 2018. Multi-criteria FUCOM-MAIRCA model for the evaluation of level crossings: case study in the Republic of Serbia. *Operational Research in Engineering Sciences: Theory and Applications*, *1*(1), 108-129.

Pamučar D, Stević Ž, Sremac S, 2018. A new model for determining weight coefficients of criteria in mcdm models: Full consistency method (FUCOM). *Symmetry*, *10*(9), 393-415.

Rajesh R, Ravi V, 2015. Supplier selection in resilient supply chains: A grey relational analysis approach. *Journal of Cleaner Production*, *86*, 343-359.

Suganthi L, 2020. Sustainability indices for energy utilization using a multi-criteria decision model. *Energy, Sustainability and Society*, *10*(1), 1-31.

Şepit A, Paksoy T, 2019. Evaluation of Sustainability Performance of Cities with A Fuzzy Multi Criteria Decision Making Method. *Selcuk University Journal of Engineering, Science and Technology*, *7*(1), 30-48.

Turkish Statistical Institute (TSI), 2021. Income, Living, Consumption and Poverty. Access address: https://data.tuik.gov.tr/Kategori/GetKategori?p=Gelir,-Yasam,-Tuketim-ve-Yoksulluk-107.; Date of access:26.03.2022.

Tzeng CJ, Lin YH, Yang YK, Jeng MC, 2009. Optimization of turning operations with multiple performance characteristics using the Taguchi method and Grey relational analysis. *Journal of Materials Processing Technology*, *209*(6), 2753-2759.

Yang W, Zhang J, 2021. Assessing the performance of gray and green strategies for sustainable urban drainage system development: A multi-criteria decision-making analysis. *Journal of Cleaner Production*, *293*, 126191.

Zeng G, Jiang R, Huang G, Xu M, Li J, 2007. Optimization of waste water treatment alternative selection by hierarchy grey relational analysis. *Journal of Environmental Management*, *82*(2), 250-259.

Zuo X, Hua H, Dong Z, Hao C, 2017. Environmental performance index at the provincial level for China 2006–2011. *Ecological Indicators*, *75*, 48-56.

Conflict of Interest

The research was prepared by a single author.

Author Contributions

The research was prepared by a single author.

Statistical Indicators of the Physiological Changes in *Daphnia Magna* Exposed to Different Environmental Effects

Mehmet Yardımcı[1*], **Çetin Yağcılar[2]**

[1]Tekirdağ Namık Kemal University, Faculty of Veterinary Medicine, Department of Biostatistics, 59030, Tekirdağ, Türkiye
[2]Tekirdağ Namık Kemal University Faculty of Veterinary Medicine Department of Aquaculture, 59030, Tekirdağ, Türkiye

*Corresponding author e-mail: dr.yardimci@gmail.com

Abstract

Statistics is a branch of science that provides the opportunity to make healthier inferences in research results and it attracts great attention day by day. Today, the level of statistical analysis of the data is considered as a criterion in determining the quality of that scientific research. The subject has come to such a point that the referees and journal editors who evaluate the articles get the first impressions by looking at the sections where the statistical analyzes are located. Applying accurate statistical analyzes to the obtained raw data increases the validity and reliability of the research. In this context, we have observed that the physiological changes caused by various environmental effects only become meaningful with statistical analysis from our research on Daphnia magna. The results of the previous studies conducted by the authors of this presentation were used as the study material. To give a few concrete examples, although the positive effect of music on metabolic activities as an environmental factor has been noted in many other living organisms in the literature, we found that the situation was different in Daphnia magna. Again, while we expected to observe a linear change in heart rate, number of eggs produced and mortalities where we examined the effect of drug residues in water as an environmental factor, we encountered non-linear results in some of our studies. Another determination was on molting. It has been determined that the frequency of molting, which was expected to follow a parallel course with other changes in a certain order from the beginning of the research, varied. In conclusion, statistical analyzes provided serious convenience in the interpretation of the data obtained and provided an objective view in the explanation of unforeseen situations.

Key words: *Daphnia magna, Environmental factor, Linear, Statistics*

INTRODUCTION

Thanks to statistical software, data analysis today is fast, easy and flexible. Every statistical analysis gives researchers a range of potential decisions to take (Caldwell and Cheuvront, 2019).

The term ‘indicator’ in Latin words means to point out, to show or to indicate. It provides information that simplifies reality, for example by extracting data for a specific question or aggregating data on a number of different variables. In this context, a statistical indicator can be an absolute number, ratio, average or other statistics related to a subject or phenomenon and can be defined as a set of observed facts. Indicators can be used either to describe a situation or trend to provide an assessment of progress towards established goals and objectives (Anonymous, 2014). The mean, coefficient of variation, standard deviation, standard error, p value, diagrams and tables are some of the key indicators which summarize and exhibits the features of the data, illustrate patterns, describes the distribution of a variable in the entire population, occurrence of the observations in every class or category of the variable and useful for estimating sample size, effect size and statistical probabilities. An indicator can be in a qualitative or quantitative form which gives a descriptive explanation, a performance measure, a mix of them, impact or response situation obtained directly or indirectly in uni or multi-dimensional base (Petrie and Watson, 2013).

Daphnia magna, the main material of this study, is an invertebrate crustacean aquatic creature known as the water flea. It has been accepted as a suitable biological model organism for scientific studies due to its ability to obtain large numbers of samples for statistical design and analysis, its sensitivity to changes in water quality, pollution and its ability to be produced in a laboratory environment (Yardımcı and Yağcılar, 2021).

Figure 1. Daphnia magna (by Mehmet Yardımcı and Çetin Yağcılar, 2021)

It is possible to encounter unexpected results other than expected responses in studies with Daphnia and other living things. The results, which become understandable when biological knowledge and statistical knowledge are blended, lead to confusion in the absence of one party.

This study was carried out to show how physiological changes under different environmental effects become more understandable with statistical indicators in *Daphnia magna*.

MATERIAL AND METHODS

Material

The results of the previous studies conducted by the authors of this presentation were used as the study material. In the studies, physiological responses to different sound frequencies, possible toxic effects of some chemical substances, fertility and viability were evaluated on Daphnia magna produced under laboratory conditions.

Methods

The Collection of the Data

The results, which were difficult to understand and interpret with a cursory glance, have been evaluated. In both table and graphical representations, the results indicated by the statistical analyzes were emphasized.

Statistical Analysis

No new statistical analysis was made on the results of the researches but the interpretation of the results was emphasized.

RESULTS AND DISCUSSION

The subject of the first study to be referred to was about the effects of different sound frequencies on the heart rate, egg number and survival parameters of *Daphnia magna* (Yağcılar and Yardımcı, 2021). In this study, the research hypothesis was that the sound frequency of 432 Hz would have a positive effect in terms of productivity and vitality in Daphnia, which represents aquatic creatures, as in many other creatures. However, it did not turn out as expected. It has been determined that both sound waves with a frequency of 440 Hz, which is also called the poisonous frequency, or 432 Hz, which is considered as a calming melody were perceived as noise by aquatic creatures (Table 1). These inferences were reached with statistical indicators.

Table 1. Comparison of heart rates of groups by periods

Periods (week)	Groups	n	$\bar{x} \pm S_{\bar{x}}$	p
1	Control	50	338,96±5,45	0,954
	432 Hz	50	340,96±5,18	
	440 Hz	50	338,88±5,62	
2	Control	50	345,28±5,93^{b}	0,000
	432 Hz	50	358,88±4,12^{b}	
	440 Hz	50	295,60±5,39^{a}	
3	Control	50	337,68±4,91^{b}	0,000
	432 Hz	50	362,72±5,46^{c}	
	440 Hz	50	309,36±8,25^{a}	
4	Control	50	329,92±3,81^{b}	0,000
	432 Hz	50	340,80±4,22^{b}	
	440 Hz	50	295,36±4,63^{a}	
5	Control	50	344,32±4,28^{c}	0,000
	432 Hz	50	275,84±6,68^{b}	
	440 Hz	50	178,80±3,29^{a}	

$p < 0.05$; p: ANOVA; Different letters indicate significant difference among treatments by Tukey test

The fact that the data were collected on a weekly basis and subjected to statistical analysis ensured that the first indicator was caught in terms of a healthier evaluation of the answers given by the groups. As a matter of fact, the increase in heart rate for a certain period of time in the 432 Hz group was evaluated as a sign of metabolic resistance to sound as a negative environmental factor in terms of metabolic activities. If inferences were made by measuring only at the beginning and end of the study, it could be concluded that the frequencies of 432 Hz and 440 Hz had similar effects on the groups, and therefore there were small differences between the groups. However, negative effects were observed in a very short time in the 440 Hz group, and significant decreases were detected in heart rate, egg count (Fig 2) and survival rates as statistical indicators.

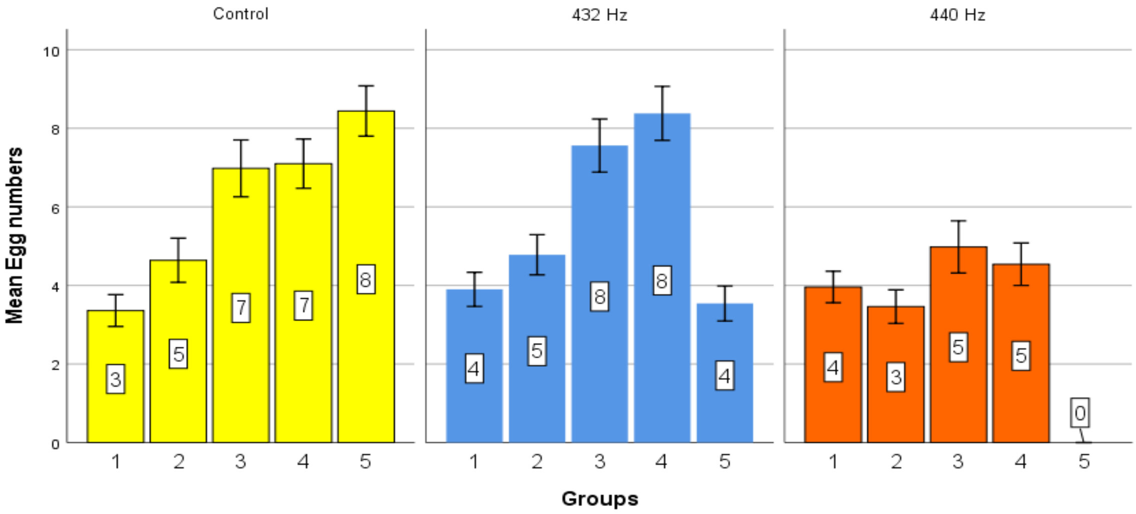


Figure 2. Mean egg numbers in groups by periods

The remarkable results regarding the deaths were observed after the measurements were completed. Namely, after the anticipated trial period was completed, it was determined that no living thing remained in the 440 Hz group within 2 weeks. When we look at the statistical indicators, it can be said that there were signs that this result will be achieved beforehand.

Another study was on the reproductive and physiological responses of daphnia magna exposed to different concentrations of prednisolone (Yağcılar et al., 2022). This research, which examines the effects of drug residues used in human and veterinary medicine and added to wastewater on aquatic organisms was carried out on *Daphnia magna* as a model organism. The effects of a corticosteroid drug, whose trade name is Prednol used at doses of 0.25, 0.75 and 1.25 mg/L. Contrary to increasing Prednisolone doses, a gradual decrease in the number of eggs was observed among the experimental groups ($p<0.05$). This decreasing trend in the number of eggs reached a 10-fold difference between the high dose group and the control group in the last measurement period (Fig 3).

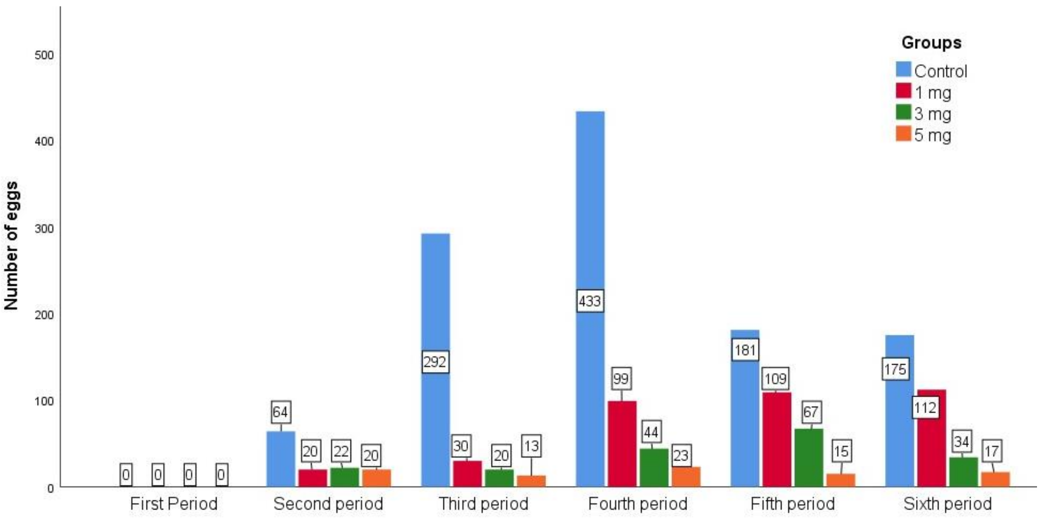


Figure 3. Number of total eggs in groups by periods

In these comparisons, while the change in the number of eggs was a statistical indicator, another striking indicator was the difference in the number of n (Table 2). When both indicators were considered together, it was observed that the high dose had a significant negative effect on both egg number and viability.

Table 2. Comparison of number of eggs between groups by periods

Periods	Groups	n	Mean rank	Kruskal Wallis Test	Pairwise group comparisons		
					Groups	Mann Whitney U	p
1				df=3 H=0.000 p= 1.000	C-1	1800.00	1.000
	Control	60	120.50		C-3	1800.00	1.000
	0,25 g/L	60	120.50		C-5	1800.00	1.000
	0,75 mg/L	60	120.50		1-3	1800.00	1.000
	1,25 mg/L	60	120.50		1-5	1800.00	1.000
					3-5	1800.00	1.000
2				df=3 H=5.976 p=0.113	C-1	1444.50	0.057
	Control	60	122.47		C-3	1423.50	0.112
	0,25 g/L	56	106.86		C-5	1334.00	0.053
	0,75 mg/L	54	109.19		1-3	1478.50	0.737
	1,25 mg/L	52	106.24		1-5	1447.00	0.922
					3-5	1365.50	0.679
3				df=3 H=105.141 p=0.000	C-1	317.50	0.000
	Control	57	156.44[a]		C-3	280.00	0.000
	0,25 g/L	51	88.75[b]		C-5	232.50	0.000
	0,75 mg/L	47	74.40[c]		1-3	995.50	0.048
	1,25 mg/L	44	69.58[c]		1-5	864.50	0.006
					3-5	974.50	0.362
4				df=3 H=92.090 p=0.000	C-1	283.50	0.000
	Control	55	141.09[a]		C-3	155.00	0.000
	0,25 g/L	47	81.60[b]		C-5	108.50	0.000
	0,75 mg/L	40	61.95[c]		1-3	670.00	0.016
	1,25 mg/L	36	51.61[c]		1-5	478.50	0.000
					3-5	607.00	0.177
5				df=3 H=19.794 p=0.000	C-1	854.00	0.152
	Control	53	82.39[a]		C-3	544.00	0.048
	0,25 g/L	39	70.24[ab]		C-5	171.00	0.000
	0,75 mg/L	28	66.45[b]		1-3	516.00	0.699
	1,25 mg/L	18	34.69[c]		1-5	172.50	0.002
					3-5	110.00	0.001
6				df=3 H=21.551 p=0.000	C-1	892.50	0.603
	Control	53	77.52[a]		C-3	312.00	0.001
	0,25 g/L	36	72.13[a]		C-5	199.00	0.000
	0,75 mg/L	22	50.27[b]		1-3	262.00	0.029
	1,25 mg/L	19	37.05[b]		1-5	176.00	0.003
					3-5	139.00	0.056

$p < 0.05$; p: Kruskal-Wallis Test; Different letters indicate significant difference among treatments by Mann Whitnet U Test

Similar results were reported by several researches indicating that exposure to this class of pharmaceuticals caused inhibition of growth population on the freshwater crustacean (Allouche et al., 2022; DellaGrec et al., 2004; Sumiya et al., 2014; Xavier et al., 2017).

We also investigated the chronic effects of Carbamazepine on reproduction and physiological traits of *Daphnia Magna* (in press) using 1,25, 1,75 and 2,25 mg/L dosages (Fig 4). Despite the survival numbers over than 30 in each group (control-59, 1,25 mg/L-52, 1,75 mg/L-50 and 2,25 mg/L-36), the number of molts in 2,25 mg/L group decreased to zero in the last measurement period. This statistical indicator showed that sometimes changes can occur suddenly, surprising the researcher.

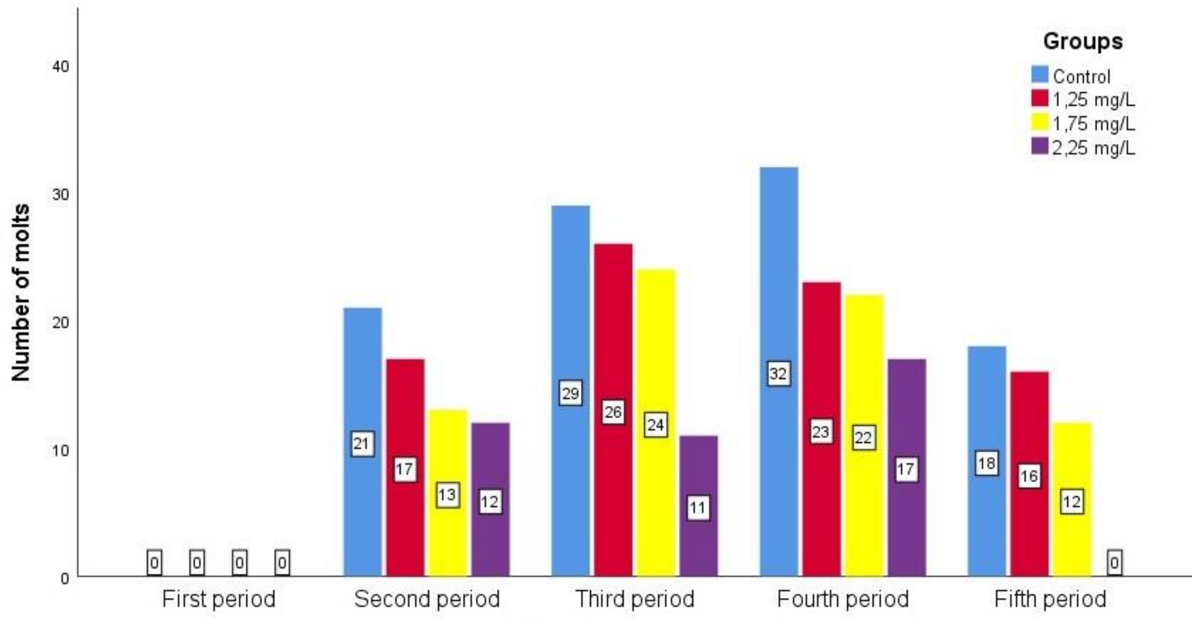


Figure 4. Number of molts by periods

In the literature on the subject, there are many studies reporting that pharmaceuticals adversely affect fertility, molting and number of eggs (Bodar et al., 1990; Chen et al., 2019; Flaherty and Dodson, 2005; İçoğlu 2019).

In this study, we wanted to emphasize that it is inevitable to consider the statistical information and the science of biology together because statistical analysis is a data analysis tool that helps draw meaningful conclusions from raw and unstructured data. In other words, the field of statistics is the science of learning from data.

CONCLUSION

Both clinical and statistical reasoning is essential for advancement in the health sciences. Appropriate interpretation of statistical results allows us to transform the uncertainty and complexity of the facts into measurable parameters during the decision-making processes about interventions.

Statistical significance, a mathematical term, can lead to inaccurate inferences unless evaluated against the actual impact of the results in the field. Therefore, statistical interpretations must have a plausible biological hypothesis behind them. Hence, statistical indicators improve our capacity to understand biological reality and the results produced by research.

In the present study, it was concluded that the results obtained regarding the physiological changes in the studies on Daphnia magna became meaningful with statistical interpretations.

References

Allouche M, Ishak S, Nasri A, Harrath AH, Alwasel S, Beyrem H, Pacioglu O, Boufahja F, 2022. What Is the Impact of Dexamethasone and Prednisolone Glucocorticoids on the Structure of Meiobenthic Nematode Communities? Sustainability, 14, 5344.

Anonymous, 2014. Towards a harmonised methodology for statistical indicators. Part 1: Indicator typologies and terminologies. Publications Office of the European Union, Luxembourg. ISBN 978-92-79-40322-4, doi:10.2785/56118

Bodar CVM, Voogt PA, Zandee DI, 1990. Ecdysteroids in Daphnia magna: their role in moulting and reproduction and their levels upon exposure to cadmium. Aquatic Toxicology, 17(4), 339-350.

Caldwell AR, Cheuvront SN, 2019. Basic statistical considerations for physiology: The journal Temperature toolbox. Temperature, 6(3), 181-210.

Chen H, Gu X, Zeng Q, Mao Z, 2019. Acute and Chronic Toxicity of Carbamazepine on the Release of Chitobiase, Molting, and Reproduction in Daphnia similis. International Journal of Environmental Research and Public Health, 16(2), 209.

DellaGrec M, Fiorentino A, Isidori M, Lavorgna M, Previtera L, Rubino M, Temussi F, 2004. Toxicity of prednisolone, dexamethasone and their photochemical derivatives on aquatic organisms. Chemosphere, 54, 629-637.

Flaherty CM, Dodson SI, 2005. Effects of pharmaceuticals on Daphnia survival, growth, and reproduction. Chemosphere, 61, 200-207.

İçoğlu FA, 2019. Acute and chronic effects of thifluzamide on Daphnia magna. Turkish Journal of Zoology, 43, 554-559.

Petrie A, Watson P, 2013. Statistics for Veterinary and Animal Science. 3rd Ed. London School of Hygiene and Tropical Medicine University of London, London, UK. John Wiley & Sons, Ltd. ISBN: 978-0-470-67075-0

Sumiya E, Ogino Y, Miyakawa H, Hiruta C, Toyota K, Miyagawa S, Iguchi T, 2014. Roles of ecdysteroids for progression of reproductive cycle in the fresh water crustacean Daphnia magna. Frontiers in Zoology, 11, 60.

Xavier CR, López D, Gómez G, Chamorro S, Scholze A, Vidal G, 2017. Sensitivity study comparing Daphnia obtusa (Kurz 1874) and Daphnia magna (Straus 1820) exposure to treated kraft mill effluents, diethylstilbestrol, and androstenedione, BioResources, 12(3), 6558-6567.

Yağcılar Ç, Yardımcı M, 2021. Effects of 432 Hz and 440 Hz Sound Frequencies on the Heart Rate, Egg Number and Survival Parameters in Water Flea (Daphnia magna). Journal of Ecological Engineering, 22(4), 119-125.

Yağcılar Ç, Yardımcı M, Ceylan M, Şensoy E, 2022. Reproductive and Physiological Responses of Daphnia Magna Exposed to Different Concentrations of Prednisolone. Ziraat, Orman ve Su Ürünleri Alanında Uluslararası Araştırmalar II, Editör: Prof. Dr. Hülya Özgenen Özkaya, ISBN: 978-625-8108-74-3.

Yardımcı M, Yağcılar Ç, 2021. The Role of Daphnia Magna as A Model Organism in Scientific Researches. 4th International Nowruz Conference on Scientific Research, March 18-21, 2021, Karabagh, Azerbaijan, Page: 440-447.

Determination of Late Flowering Types in Almonds Using Multivariate Analysis Techniques

Özgür Koşkan[1], Adnan Nurhan Yıldırım[2], Deniz Gülkaya Arıtürk[3*]

[1] Isparta Uygulamalı Bilimler Üniversitesi, Ziraat Fakültesi Zootekni, 32200, Isparta, Turkey
[2] Isparta Uygulamalı Bilimler Üniversitesi, Ziraat Fakültesi Bahçe Bitkileri, 32200, Isparta, Turkey
[3] Isparta Uygulamalı Bilimler Üniversitesi, Bahçe Bitkileri, 32200, Isparta, Turkey
*Corresponding author e-mail: dengulkaya@gmail.com

Abstract

Almond is a fruit with high nutritional value, it also has an important commercial value. One of the most important problems in almond cultivation is late spring frosts, and the selection of late types is important from a commercial point of view. Although Isparta is suitable for almond cultivation, it is known that almond trees are affected by late spring frosts. In this study, the early and late types of almonds were tried to be predicted using statistical methods by using phenological observations, directions and fruit characteristics. The data obtained are evaluated, the difference between the averages of flowering times as a result of the variance analysis of the data obtained in terms of bud length, crown width, shell fruit thickness, inner almond width, inner almond weight characteristics is statistically significant($p<0,05$). The lowest bud length was observed in the 1st flowering period (1,868). It was observed that the lowest average in terms of crown width was 445,4 in the 2nd flowering period. In terms of shell fruit thickness, the first flowering period is higher than other periods with a value of 15,172. In terms of shell fruit weight, the 1st flowering period is higher than the other periods with a value of 4,012, and the 3rd flowering period has the lowest value with 2,868. Considering the averages in terms of inner almond size, it is seen that the highest value is 21,859 and it is observed in the 1st flowering period. According to the directions, the smallest bud length is in the 1st flowering period, it was determined that the highest value of the average in terms of bud length was 7,510 and the "North" value was 1,878 and it belonged to the "Southwest" direction. As a result of the Discriminant analysis made by using all tree and fruit characteristics, the number and degrees of accuracy in the predictions were calculated. All the features are included in the analysis, it is estimated that 84.6% of the trees belong to which direction. In other words, the direction of 66 out of 80 trees was predicted correctly. Flowering times of trees are estimated at 74.4% when only fruit characteristics are included in the analysis. In other words, it was estimated correctly whether 58 of the 80 trees were early or late. All features are included in the analysis, the flowering time of the trees is estimated at 94.9%. In other words, it was correctly estimated that 74 of the 80 trees were early or late. In 2013 according to a study conducted on the detection of young peach fruits with their texture shape and self-qualities in color images taken in garden conditions, it seen that the number of fruits successfully detected by discriminant analysis is 78 %. Similarly outcome of this study of the stepwise disciminant analysis made by utilizing all tree and fruit features, the features included in the model and the number and degrees of accuracy in the predictions are calculated. It is a remarkable result that 87.5% accurate estimation was obtained in separating early and late by making use of only the number of trunk main branches. As a result, with this study, it is aimed to reveal commercially important varieties through predicting late varieties by using of fruit, tree characteristics and various phenological characteristics.

Key words: *prunus amygdalus, discriminant analysis, late flowering types in almonds, cluster analysis*

Longing for Sportive Success Scale: Validity and Reliability Study

Soner Cankaya[1*], Mehmet Derelioglu[2], Seda Sabah[3], Aydan Ermis[4], Samet Hasan Abaci[5]

[1] Ondokuz Mayis University, Yasar Dogu Faculty of Sport Sciences, Department of Sports Management, 55139, Samsun, Türkiye
[2] Avrasya University, Arhavi Vocational School, 61000, Trabzon, Türkiye
[3] Amasya University, Faculty of Education, 05100, Amasya, Türkiye
[4] Ondokuz Mayıs University, School of Foreign Language, 55139, Samsun, Türkiye
[5] Ondokuz Mayis University, Faculty of Agriculture, 55139, Samsun, Türkiye

*Corresponding author e-mail: scankaya@omu.edu.tr

Abstract

The aim of this study is to develop a valid and reliable scale that can be used to determine why sport clubs cannot be successful for a long time by examining how fans are affected by the long term failure of their teams to become champion, and fans' longing for and thoughts about championship. For this purpose, the study was carried out with a cross-sectional survey design in line with quantitative research methods. In this context, the item pool which was prepared based on sport success literature was presented for the views of experts. 36-item draft scale, which was finalized with the suggestions of experts, was applied to 201 individuals chosen with random sampling method among Trabzonspor club fans who could not get a championship for 37 years. In the analysis of the obtained data, exploratory factor analysis was performed for construct validity. Based on the obtained structure, a model was created and confirmatory factor analysis (CFA) was performed to see the fit of the assumed model with the data. Longing for Sportive Success Scale (LSSS) with 17 items and four sub-dimensions was developed as a result of the analyses conducted. These sub-dimensions were called internal factors, knowledge, violence and external factors. It was found that the developed scale explained 56.22% of the total variance and Cronbach's alpha value regarding reliability was 0,809 for the overall scale, 0,551 for internal factors sub-dimension, 0,839 for knowledge sub-dimension, 0,737 for violence sub-dimension and 0,717 for external factors sub-dimension. CFA showed that the fit indices of the four factor model were at "perfect or acceptable" level. Cronbach alpha, Guttman and Spearman-Brown internal consistency coefficients showed that LSSS was highly reliable. Based on these results, it was concluded that LSSS is a valid and reliable data collection tool that can be used to determine how fans are affected from their team's failure to become champion for a long time, to determine fans' longing for championship and their thoughts and the reasons why sports clubs fail to have success for long years.

Key words: *Sportive success, Fan, Scale development*

Principal Components Analysis with Qualitative and Quantitative Variables (PCAmix)

Sıddık Keskin[1], **Nurhan Keskin**[2]

[1] Van Yüzüncü Yıl University, Faculty of Medicine Department of Biostatistics, 65040, Van, Türkiye
[2] Van Yüzüncü Yıl University, Faculty of Agriculture Department of Horticulture, 65040, Van, Türkiye

*Corresponding author e-mail: skeskin973@gmail.com

Abstract

In this study, PCAmix (Principal Components Analysis with Mixed Variables) was explained with its general properties and an application was performed. In most scientific research, the characteristics or variables are either qualitative (categorical or nominal) or quantitative (numerical or continuous). These variables can be linearly or non-linearly related to each other. It is important to accurately determine these relationships using appropriate statistical methods. In data sets containing mixed (qualitative and quantitative) data, PCAmix (Principal Components Analysis with Mixed Variables) can be used to examine the relationships between variables and between variable categories, as well as to make dimension reductions. PCAmix, which can be expressed as a dimension reduction method similar to other multivariate analysis methods such as Factor Analysis and Principal Component Analysis, is an analysis method that visually presents the relations between variable categories as well as the relations between variables in two-dimensional space. PCAmix includes a series of statistical process such as calculating the correlation matrix, eigenvalues and eigenvectors, finding common components in the vector space for qualitative and quantitative variables, determining the ratio of variance and number of principal components that can be explained by each component. In the application, there are thee qualitative (categorical) variables: Berry shape (conical, oval, round), Berry color (blue, purple, yellow and black) and Evaluation type (table or wine). In addion, Cluster weight (g), Berry width (mm), Berry length (mm) and Maturity index are included as quantitative variables. As a result, it has been emphasized that PCAmix can be used to examine both the relationships between variables and between categories of the variable and also to make dimension reductions in data sets containing mixed (qualitative and quantitative) variables.

Key words: *Dimension reduction, Configuration, Mixed data*

Comparison of the Performance of Nominal Logistic Regression and Regression Trees Methods on Distance Education Dataset

Yaren Mert[1*], Berna Yazıcı[1], Ece Özgören[1]

[1] Eskisehir Technical University, Faculty of Science, Statistics, 26050, Eskisehir, Turkey

*Corresponding author e-mail: yarenmert@eskisehir.edu.tr

Abstract

Logistic Regression Analysis, which is frequently used in many different fields such as economy, agriculture, especially in the field of medicine, is a type of analysis that differs according to the problem and is used for classification analysis. In Logistic Regression Analysis, determining the suitable model for the data and the problem is one of the important steps to obtain high performance and accurate results. The purpose of logistic regression is to determine the appropriate model that can explain the relationship between dependent and independent variables, using the least variable and having the best fit. Multiple regression model occurs when independent variables are more than one. The decision trees to be used in the study perform a simple decision-making process in solving the classification and regression problem. It is a method used for both classification and regression, which creates a model in the form of a tree structure, and is very effective in distinguishing important and unimportant data due to the structure of decision trees, which have different definitions by many researchers. Although the formation of decision trees, which is used as a powerful technique for classification and prediction, takes a long time, it is quite easy to understand and apply the rules. The CART (classification algorithm regression trees) technique, which is a very good decision tree algorithm on its own, is one of the most used decision tree algorithms. The CART algorithm is an automated machine learning method. In this study, logistic regression will be used to classify our data set, which consists of 9131 students who graduated from Anadolu University Open Education Faculty, different departments, and Cart application will be used to visualize the analysis and interpret the data more easily. Multinomial Logistic Regression model will be used since the type of occupation in question has more than two categories as the dependent variable. Also, how well the explanatory variables explain the dependent variable of the type of occupation like gender, graduation year, marital status etc. in the data set, will be analyzed.

Key words: *Decision Tree Algorithms, Machine Learning, Regression Tree, Nominal Logistic Regression, Distance Education.*

INTRODUCTION

Relationships between the variable of interest in scientific studies and other variable or variables that are thought to be effective; In general, it can be examined under two headings as linear and nonlinear relationships. In studies where it is desired to determine the direction and degree of these relationships and the effect of other variables on one or some of the variables, regression analysis methods are generally used. These models, which are used to determine the relationships between variables, are generally called regression models (Elasan, 2010).

One dependent variable and one or more independent variables regression analysis methods can be applied in case the relationship between the variables is desired to be examined.

Regression equations or models to be established to determine (linear) relationships between variables; It is called simple regression if it contains one response variable and one explanatory variable, and multiple regression if it contains one response variable and more than one explanatory variable. For this reason, it can be generalized in logistic regression analysis for cases where the outcome variable, that is, the dependent variable, has two or more categories. In the literature, when the response variable has more than two categories, the analysis is called multi-category or multinomial logistic regression analysis. With this analysis, while the response variable has three or more categories, the relationship between the response variable and the explanatory, that is, independent variables, can be determined.

In the multi-category logistic regression analysis, any category of the answer variable is taken as a reference category and other categories are analyzed according to this reference category and other categories are analyzed according to this reference category. For the response variable consisting of M categories, when examining the relationship between the response variable and the explanatory variables, it is necessary to calculate the reference category and M-1 equations in which each category is examined one by one (Elasan, 2010).

In logistic regression, the dependent variable is categorical, while the independent variables can be mixed scale. It is used to determine the cause-effect relationship between the dependent variable and the independent variables without any distribution assumption.

The use of the logistic model dates back to 1845. Logistic regression analysis has been increasingly applied in the fields of meteorology, agriculture, economy, biology, education and health over the last two decades. Although it is applied in different fields, it is seen that the most common usage area is medicine (Yamantaş, 2019).

In order to use multiple regression analysis, some assumptions or prerequisites are required. One of these assumptions is that the variables used in the model are continuous variables. Known standard regression analysis cannot be used when the response variable is binary. Instead, logistic regression analysis must be used. Logistic regression analysis is one of the statistical methods used to determine the relationship between the response variable and the explanatory variables when the response variable is binary or multinomial and the explanatory variables are categorical, sequential or continuous (Yamanataş, 2019; Elesan, 2010 and Bahçe 2017).

MATERIAL AND METHODS

Material

To compare the students' positions before and after graduation, a questionnaire was created. First of all, the demografic situation of the participants were given. The data obtained from the survey were analysed. To do so, multinominal logistic regression, Cart were conducted and the results were summarized in the tables.

Methods

The Collection of the Data

1.Regression Analysis

In the realm of statistical analysis, regression analysis is a widely used data analysis tool. It's a fantastic method that may be used for both descriptive and predictive applications. Simple and multiple linear regression methods are the most well-known of these techniques. It is a numerical data type obtained by measuring the dependant variable in simple and multiple linear regression. When the dependent variable is a categorical/qualitative data type, regression analysis is frequently used. The logistic regression approach is employed when the dependent variable has two or more categories qualitative data types.

The values of predictor variables are used in regression analysis to make predictions about the outcome variable. The regression analysis can be used to determine whether or not there is a link between the variables and, if so, how strong that relationship is. It is important to illustrate the observation values and the impacted events with a mathematical representation, that is, with the help of a function, when completing the regression analysis. The regression model is the name given to this model (Yavuz, 2021).

The structure of the dependent variable affects the regression model. As a result, it's critical to carefully build the dependent variable. The structure of the dependent variable affects the regression model. As a result, it's critical to carefully build the dependent variable. The categorical dependent variable is a binary indicator variable that takes the values 0 and 1 in various regression applications. The responses might be yes or no, or married or single. These outcomes can be classified as a 0 or a 1 (Yavuz, 2021).

1.1 Logistic Regression

Logistic regression can be used to predict a categorical dependent (outcome) variable based on continuous and/or categorical independent (explanatory) variables, as well as to determine the effect size of the independent variables on the dependent variable, rank the relative importance of independent variables, assess interaction effects, and comprehend the impact of covariate control variables. The impact of predictor variables is typically stated in terms of odds ratios, which are the most common effect size metric used in logistic regression. Although it was used in medicine in the past, it is an advanced regression method that has gained popularity in social sciences today (Özer and Aslan, 2019).

After changing the dependent into a logit variable, logistic regression uses maximum likelihood estimation. The natural log of the probabilities of the dependant equaling a given value (typically 1 in binary logistic models or the highest value in multinomial models) is called a logit. The probability of a specific event (value) occurring is estimated using logistic regression. This means that, unlike OLS regression, logistic regression assesses changes in the log odds of the dependant rather than changes in the dependent itself. However, logistic regression requires independent observations and independent variables that are linearly related to the dependent's logit. The classification table, which shows correct and wrong classifications of the binary, ordinal, or polytomous dependent variable, can be used to analyze the predictive performance of logistic regression. Goodness-of-fit tests like as the likelihood ratio test, as well as the Wald statistic to examine the importance of specific independent variables, are available as indicators of model adequacy. Binary logistic regression is extended to accommodate categorical dependent variables with two or more levels in multinomial logistic regression. The categories are not sorted in multinomial logistic regression. Although it is most commonly employed when the dependent variable has three or more classes, it can also be utilized with binary dependent variables because the output tables for multinomial and binary logistic regression processes are different. (Yavuz A, 2021)

While the 'Multinominal Logistic Regression Model' is used when the categorical dependent variable is multi-categorical (ex: marital status; married-single divorced), the 'Ordinal (Ordinal) Logistic Regression Model' is used when the categorical dependent variable is multi-categorical and ordinal (ex: Likert type scales, less-moderate-highly). In some circumstances, the LRA is categorized according to the number of independent variables. When there is just one independent variable, LRA is referred to as 'Univariate Logistic Regression,' and when there are two or more independent variables, it is referred to as 'Multiple Logistic Regression Analysis'. The primary goal of logistic regression is to develop a model that can be used to describe the relationship between independent and dependent variables in a way that assures the best fit with the fewest variables possible.

Any assumptions about the distribution of the independent variables are not required by the LRA. There are, however, several assumptions and conditions that must be met when using LRA.

Losses and extreme values in the data should be investigated, and modifications made if necessary. Although this criterion is not considered an assumption of the relevant analysis, it is said that it should be examined before the analysis because losses and excessive values are thought to influence statistical test findings. The predicted frequency in all cells should be larger than 1 for all pairings of categorical variables, and the number of pores with an estimated frequency less than 5 should not exceed 20%.

There should be no problem with multicollinearity between the variables because the LRA is particularly sensitive to strong correlation between independent variables. When the correlations between variables (r>.90) are large, multicollinearity occurs. If there is an issue with multicollinearity between the variables in the study, it may be possible to remedy the problem by expanding the number of data, although this is not always viable. Another option is to use factor analysis to aggregate these variables and incorporate them in the model as a single variable. It is one of the recommended approaches to eliminate one or more of the variables that create this problem from the model if the solution to the multicollinearity problem could not be obtained with other alternatives. The independent factors and the logit value of the outcome variable are assumed to have a linear relationship in LRA. The significance of the link between the log transform and the independent variable can be used to assess this assumption (Şenel and Alatli, 2014).

When the literature is examined, the headings to be considered in reporting the results of LRA can be listed as follows:

a) General model evaluation comparing the initial model with the intended model,
b) Comments on statistical tests of independent variables (Wald statistics),
c) Comments on goodness-of-fit statistics,
d) Comments on probability or odds ratio,
e) Information on model competencies (competency interpretation from classification tables, examination of residuals, or so-called R^2) (Şenel and Alatli, 2014).

Estimation strategies are addressed in three separate ideas in logistic regression. They're listed below;

1. Maximum Likelihood Method
2. Minimum Logit Chi-Square Method
3. Least Squares Method

1.1.1 Likelihood Function and Maximum Likelihood Estimation Method

If y is encoded as 0 and 1, the expression π(x) returns the conditional probability that y equals 1 for a given value of x. This probability is represented by the symbol π(x) = P (Y =1/ x). Therefore, the expression [1-π(x)] also shows the conditional probability of getting the value 0 for Y. This probability is denoted as [1-π(x}] = P (Y = 0/x). For the (x_i, yi) pair, the contribution to the likelihood function is π(x_i) when x_i=1, and when y_i=0, the contribution to the likelihood function becomes 1-π(x_i).

A reliable way of expressing the contribution of $(x_i, yi$) to the likelihood function is given below:

$$P(Y = y_i) = f_i(y_i) = \pi(x_i)^{y_i}(1 - \pi(x_i))^{1-y_i} \qquad (1.1)$$

i = 1,2,...,n for y_i=1 or $y_i = 0$

The maximum likelihood estimation method is used to calculate the logit coefficients. This method returns values for unknown parameters that maximize the probability of obtaining the observed dataset.

Assuming that the observations are independent of each other, the terms are multiplied to obtain the likelihood function.

$$\mathrm{L}(\beta_0, \beta_1) = g(y_1, y_2, \dots, y_i) = \sum_{i=1}^{n} f_i(y_i)\,) = \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i} \qquad (1.2)$$

Since it will be easier to work with logarithms mathematically, the log-likelihood function is obtained as:

$$\mathrm{L}(\beta_0, \beta_1) = ln \prod_{i=1}^{n} f_i(y_i)\, \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i} \qquad (1.3)$$

$$= \sum_{i=1}^{n} \{ln[\pi(x_i)^{y_i}] + ln[1-\pi(x_i)^{1-y_i}]\}$$

$$= \sum_{i=1}^{n} \{\ln\, [\pi(x_i)^{y_i}(1-\pi(x_i)^{1-y_i}]\}$$

$$= \sum_{i=1}^{n} \{y_i \ln[\pi(x_i)] + (1-y_i)\ln\, [1-\pi(x_i)]\}$$

$$= \sum_{i=1}^{n} \left\{y_i \ln\, [\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} + (1-y_i)\ln\, [1-\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)})]\right\}$$

Therefore (1.3),

$$\mathrm{L}\,(\beta_0, \beta_1) = \sum_{i=1}^{n} \{y_i((\beta_0 + \beta_1 x_i) - \ln\,(1 + (\beta_0 + \beta_1 x_i)\} \qquad (1.4)$$

To find the maximizing values of (1.3), the derivative of $L(\beta_0, \beta_1)$ with respect to β_0 and β_1 is taken and set to zero.

$$\frac{\partial ln(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^{n} \left\{y_i - \frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right\} = 0 \qquad (1.5)$$

(Yavuz, 2021).

1.1.2 Testing the Significance of the Coefficients

To determine whether the independent variable in the model is related to the dependent variable, some statistical hypotheses are determined and these hypotheses are tested. To test the significance of the coefficient of a variable in the model, it is checked whether the model containing this variable provides more information about the response variable than the model without this variable. Three tests are generally used for hypothesis testing in the logistic regression model. These are the likelihood ratio test, the wald test, and the score test (Yavuz, 2021).

1.1.3 Likelihood Ratio Test

In logistic regression, the comparison of observed and expected values satisfies the log probability.

$$\mathrm{D} = -2\log\left[\frac{probability\ of\ the\ initial\ model}{probability\ of\ the\ given\ model}\right] \qquad (2)$$

The expression likelihood ratio given in parentheses in equation (2) is called the "probability ratio". Taking the base (–2ln) is to get a value whose distribution is known mathematically. This value is used for hypothesis testing. Such a test is called the likelihood ratio test.

To decide the importance of an independent variable, the D values in the equation with and without the independent variable are compared. The change in D as it includes the independent variable is as follows.

$$\mathrm{G} = (unvariable\ model) - \mathrm{D}(\ variable\ model) \qquad (2.1)$$

The statistics calculated in equation (2.1) play the same role as the part of the F test used in linear regression. Since the probability of the saturated model is common to both D values to be taken to calculate G, the G statistic takes the following form:

$$\mathrm{G} = -2\ln\,[probability\ without\ a\ variable/\ probability\ with\ variable] \qquad (2.2)$$

Under hypothesis β_1, the G statistic shows the 1-degree-of-freedom x^2 distribution. The validity of the model based on the idea that criteria will be distributed based on the difference of likelihood ratio values for the model estimated by the model including all variables is tested.

1.1.4 Wald Test

The Wald test statistic is obtained by dividing the parameter's maximum likelihood estimate ($\hat{\beta}_i$) by the standard error s ($\hat{\beta}_i$). If the calculated tail probability is less than the determined significance level, the null hypothesis claiming that the coefficient is zero is rejected, thus it is concluded that the coefficient is statistically significant. In logistic regression analysis, the next step after estimating the coefficients is to test whether the model has goodness of fit (Yavuz, 2021; Ergüt et al., 2020).

$$\text{Wald test statistic; } \frac{\hat{\beta}_i}{s(\hat{\beta}_i)} \qquad (3)$$

$$W = \frac{\beta_1}{se(\beta_1)} \qquad (3.1)$$

Under the hypothesis β_1 = 0, W is a standard normal distribution. This test can also be written in an alternative way. Thus, the Wald test statistic is written as:

$$W^2 = \left(\frac{\beta_1}{se(\beta_1)}\right)^2 \qquad (3.2)$$

1.1.5 Score Test

Another test used to test the significance of the parameters is the score test. This test was created according to the conditional distributions of the derivatives of the likelihood equations. In this way, the following formula can be obtained:

$$ST = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sqrt{\bar{y}(1-\bar{y})\sum_{i=1}^{n} x_i(x_i - \bar{x})^2}} \qquad (4)$$

This statistic shows a standard normal distribution for the β_1 = 0 hypothesis showing the slope parameter.

1.1.6 Fitting Multiple Logistic Regression Model

There is more than one independent variable in multivariate logistic regression. Here, the independent variable $\tilde{x} = [1, x_1, x_2, \ldots, xp]$ takes the value of the regression parameters $\tilde{\beta} = [1, \beta_1, \beta_2, \ldots, \beta_p]$ and the dependent variable takes the value 0 or 1. Also, the expected value of errors is $E\{\varepsilon_i\} = 0$. Logistic response function for multivariate case:

$$\pi(x_i) = E(Y = 1|x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \qquad (5)$$

$$\ln\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] = \ln\left[\frac{E(Y = 1|x_i)}{E(Y = 0|x_i)}\right] = \ln\left(e^{\beta_0 + \beta_1 x_i}\right) = \beta_0 + \beta_1 x_i \qquad (5.1)$$

In this way, the model is linearized with the help of general linear models. (Yavuz, 2021)

1.2 Multinomial logistic regression

Multi-category logistic regression analysis is used in cases where the dependent Y variable has more than two values. In multi-category logistic regression models, the dependent variable is measured at the classifier measurement level and must have at least three categories. The multi-category logistic model can be expressed as an extension of the situation where the dependent variable can take two values. The model is similar to the binary logistic model and therefore its assumptions are also similar.

Multi-category logistic regression analysis is used when the dependent variable Y takes more than two values. For example, multi-category logistic regression analysis is used instead of binary logistic regression analysis in cases where the dependent variable has values at three different levels such as good, moderate, and bad. If the number of categories is more than 2, a reference category should be determined for comparison and analysis. Although the selection of the reference category is optional, it is chosen by the researcher. In addition, the change in the reference category causes no change in the structure of the model and causes only the change in parameter estimates and interpretations (Yamantaş, 2019; Elesan, 2010).

1.3 Decision Tree

The most popular decision trees are C4.5 and CART trees. CARTs are the method to be used when regression is used (Harrington, 2012). Classification, decision nodes, and leaf nodes make up the Decision Tree. It's a classification method that uses a tree structure to create a model. The language format for decision trees is "if condition then outcome." In classification difficulties, decision trees are commonly utilized. Other methods are more difficult to design, grasp, and interpret than decision trees. A tree is generated using the data we have in the decision tree approach in order to classify it. The dataset's records are applied to this tree. The records are then classified based on the outcome. To put it another way, when we have data and don't know which class it belongs to in the decision trees we get from the database, the class is guessed using the created rule set. A decision tree is a learning technique that uses simple decision-making procedures to divide vast quantities of data into manageable datasets (Özer and Aslan, 2019).

Advantages and disadvantages of decision trees:

When using decision trees, there are some benefits and drawbacks to consider.
The following are some of the benefits of decision tree-based classification methods:

- It's more practical because it's simple to comprehend and interpret.
- It can be applied to very big datasets with ease.
- Decision trees are capable of handling datasets with missing values.
- Decision trees are self-explanatory and simple to follow when condensed. That is, if the decision tree has a sufficient number of leaves, non-professional users can understand it.
- Any discretevalue classifier can be represented using a decision tree model.

The following are some of disadvantages of decision tree-based classification methods:

- The time and tree complexities fluctuate as the learning set's number of data and variables is large or mixed.
- Some algorithms are only effective with categorical data and may fail to classify continuous variables.
- When the number of inputs is low, the tree may not be able to capture enough information. (Yavuz, 2021; Rokach and Maimon , 2015).

Decision tree algorithms:

Although there are various algorithms based on the logic of "classifying the answer variable according to the explanatory variables" in general terms, each of these approaches serves a different purpose inside the decision tree methods. The AID (automatic interaction detector) algorithm was used to create the first applications that formed the foundation of tree-based approaches, and then further algorithms were developed. The following is a list of the developed algorithms. (Yavuz, 2021).

• CART (classification and regression trees; Breiman et al., 1984)

• CHAID (Chi-squared automatic interaction detector)

• Random Forest (Random Forest; Breiman)

• C4.5 (successor of ID3) decision tree (Quinlan)

• C5 (successor of ID3) decision tree (Quinlan)

• ID3 tree (iterative dichotomiser 3; Quinlan)

• CAL5 decision tree algorithm (Michie et al.)

• MARS (multivariate adaptive regression splies; Friedma)

• SLIQ (supervised learning in QUEST)

•SPRINT (scalable parallelizable induction of decision trees)

• QUEST (quick, unbiased, efficient statistical tree; Loh & Shih)

In this thesis, CART (classification and regression trees) method will be used.

1.3.1 Cart Algorthm

Classification and Regression Trees (CART) is an acronym for Classification and Regression Trees. Breiman et al. (1984) created it, and it is distinguished by the fact that it generates binary trees, with each internal node having exactly two outward edges. The Twoing Criteria are used to determine the splits, and the resulting tree is trimmed using Cost-Complexity Pruning. CART can consider misclassification costs in tree induction if they are provided. Users can additionally give a prior probability distribution. (Rokach and Maimon, 2015).

CART's capacity to create regression trees is an essential feature. The leaves of regression trees predict an actual number rather than a class. CART looks for splits that minimize the prediction squared error in regression (the least-squared deviation). Each leaf's forecast is based on the node's weighted mean.

In research conducted in a range of fields, data is thoroughly evaluated. In the decision-making process, the categorical dependent variable is estimated using classification methods while taking into account numerous variables (Yavuz, 2021).

CART decision tree: It consists of three parts, a branch, a node, and a leaf consist of. In the CART tree structure, each variable is represented by a node and is easy to understand. In other words, tree structure; a root node containing the data,

It consists of inhomogeneous branch nodes and homogeneous leaf nodes. Decision

tree contains branches that are developed according to the yes-no answers applied to the continuing experiment units starting from the root node. A separator to the questions asked at each node and this the process is also called separation.

The most important step in constructing the CART decision tree is the According to which criterion or which branching will be used to classify the variables is to determine what will be done according to the variable. At this stage, the uncertainty rate is the lowest variable is processed and used for testing on the root node (Breiman et al., 1984).

In the CART algorithm, whether the tree growing process is terminated or not it is controlled by "stopping rules". The stopping criteria are enforced in the following conditions, and node splitting is disabled.

- If all observations in a node give identical values for the response variable,
- If the number of nodes and branches is less than the minimum number specified by the user,
- The current tree depth is limited to the maximum tree depth limit set by the user the node is not split if it reaches (Saraç, 2020)

The CART algorithm was developed in 1984 by Breiman et al. suggested by. This algorithm is a continuation of Morgan and Sonquist's decision tree algorithm called AID (Automatic Interaction Detection). CART, within the supervised learning of machine learning, is a classification and regression tree algorithm that uses both categorical and continuous variables. The CART algorithm consists of three steps (Yavuz and Çilengiroğlu ,2020).

a) Creation of the maximum tree

A tree is a structure that divides the cluster of interest into two more homogeneous subsets. The root of the tree includes all risk factors in the data set. Also, this root is thought of as a master node, which is divided into two distinct subnodes at each level. In the next step, each subgroup becomes a main group. Each division includes all risk factors in a subgroup is defined by the value of an explanatory selected to have similar outcome variable values.

b) Tree Pruning

After the maximum tree is built, this tree tends to overfit. That is, the CART algorithm is it grows by dividing continuously without stopping rule. As with other modeling methods, the complexity of the tree and Pruning is required to balance predictive power. In the event that a new division will no longer occur, this time pruning is initiated from the tip to the root. During the pruning process, a series of smaller subtrees derived from the maximum tree are obtained from successive end branches so that different subtrees are compared with the best fit.

c) Selection Of the Optimum Tree

The optimal one has to be selected among the obtained subtrees. This selection process is based on the evaluation of the estimation error is installed. The estimation error is evaluated using the cross-validation test.

RESULTS

With the answers given to the questions addressed in this study; The relationships between age, gender, faculty, place and monthly average personal salary variables were analyzed for each question separately. As a result of this review, the results of the multinomial logistic regression analysis for each question are presented in figures.

Results of Multinomial LRA:

Case Processing Summary

		N	Marginal Percentage
q7.2.jobtype	nojob	542	12,7%
	officer	2413	56,7%
	nonprofitfirm	275	6,5%
	private	824	19,4%
	own	199	4,7%
q12.gender	female	1377	32,4%
	male	2876	67,6%
q1.faculty	openeducation	3215	75,6%
	economy	1038	24,4%
q13.place	village/town	389	9,1%
	city	1238	29,1%
	metropol	2624	61,7%
	5	2	0,0%
Valid		4253	100,0%
Missing		4878	
Total		9131	
Subpopulation		2384[a]	

a. The dependent variable has only one value observed in 2093 (87,8%) subpopulations.

Figure 1. Case Processing Summary

First of all, we get here an overview how many observations in each group there are then the real output starts. In Case Processing Summary, we see the number of participants belonging to categories and variables.

Model Fitting Information Table:

In the image below, the likelihood ratio, deviance, -2LL, model chi-square, or chisquare goodness of fit is found in the "Final" row, "-2 Log Likelihood" column. -2LL will decrease as the quality of fit in subsequent models improves. The likelihood ratio is also the foundation for the likelihood ratio test of difference in fit for nested models, which is addressed further below.

Model Fitting Information

Model	Model Fitting Criteria -2 Log Likelihood	Likelihood Ratio Tests Chi-Square	df	Sig.
Intercept Only	9128,760			
Final	5756,105	3372,655	28	,000

Figure 2. Model Fitting Information

To reach the chi-square result given (9128,760 - 5756,105 = 3372,655), the likelihood ratio test subtracts -2LL for the final model from -2LL for the intercept-only model. At 28 degrees of freedom, 3372,105 is better than the .001 level of the essential chi-square value.

Here, in the model fit information, the part that the model could not explain was 9128,760 it decreased to 5756,105 in the last case, the difference was 3372,655 and our model turned out to be significant.

Goodness of fit tests:

These are alternate tests of the overall model, comparing the researcher's model to the saturated model, which matches the data completely by definition. While there are two types of goodness of fit tests, deviance and Pearson chi-square tests, deviance goodness of fit is the more usually mentioned of the two (Garson, 2014).

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	1,030E+10	9504	,000
Deviance	5101,229	9504	1,000

Figure 3. Goodness- of- Fit

When we look at the model fit here, we see that our model is meaningful. The model we are considering is suitable for us. We are interested in the pearson test in logistic regression. In order for the model to be suitable for the data, the pearson significance value should be less than 0.05 and the continuous significance value should be greater than 0.05.

Pseudo R- Square:

In logistic regression, there is no true counterpart to R-square in OLS regression.

Pseudo R-Square

Cox and Snell	,548
Nagelkerke	,600
McFadden	,324

Figure 4. Pseudo R- Square

It is possible that these values are smaller in the multinomial logistic regression calculated with 3 different analysis methods.

Here, the independent variable explained the dependent variable as our dependent variable, 334 according to Cox and Snell; 0.364 according to Nagelkerke and 0.162 according to McFadden.

Likelihood Ratio Tests:

If the researcher selects "Likelihood ratio tests" in the "Parameters" area under the Statistics button, a table like the one on the following page is produced. These likelihood ratio tests compare the researcher's full model against a simplified model that excludes the row term.

The predictor variables' likelihood ratios are displayed. These likelihood ratio tests compare the whole model of the researcher with a simplified model that excludes a specific effect (Garson, 2014).

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	5756,105[a]	,000	0	.
q14.age	6630,263	874,158	4	,000
q18. monthlyaveragepersonal salary	7184,880	1428,775	4	,000
q12.gender	5907,876	151,771	4	,000
q1.faculty	5765,469	9,364	4	,053
q13.place	5813,274	57,169	12	,000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Figure 5. Likelihood Ratio Tests

In that table, which is likelihood ratio test, where we see the contribution of the variables to the model. We see that our age independent variable had a significant effect on the model. At the same time, the faculty, gender, place variables that we set for the factor also had a significant effect on the model. This is because the significant values are less than .05 (< 0.05).

Parameter Estimator:

In this study, four sets of parameter estimates are output since since the dependent variable, job type has four levels, with the first(nojob) being the reference level.

The odds ratio, which is the key effect size metric for logistic regression, is represented by the "Exp(b)" column, while the odds ratio's confidence limits are represented by the two right-most columns. For reference levels, no coefficients are computed. The obtained coefficients are compared to reference levels and interpreted (Garson, 2014).

When the odds ratio is 1.0, the prediction variable has no effect. So it is not taken into account. The lower the number is below 1.0, the greater the negative impact on the odds. The greater the positive effect, the higher the value above 1.0.

Parameter Estimates

q7.2.jobtype[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound	95% Confidence Interval for Exp (B) Upper Bound
officer	Intercept	-5,089	2,385	4,553	1	,033			
	q14.age	,140	,009	247,395	1	,000	1,150	1,130	1,170
	q18. monthlyaveragepersonal salary	,003	,000	594,249	1	,000	1,003	1,002	1,003
	[q12.gender=1]	-1,002	,133	56,593	1	,000	,367	,283	,477
	[q12.gender=2]	0[b]	.	.	0	.	.	.	.
	[q1.faculty=1]	,144	,156	,854	1	,355	1,155	,851	1,569
	[q1.faculty=2]	0[b]	.	.	0	.	.	.	.
	[q13.place=1]	,806	2,372	,115	1	,734	2,238	,021	233,823
	[q13.place=2]	,555	2,366	,055	1	,814	1,742	,017	180,006
	[q13.place=3]	,321	2,364	,018	1	,892	1,379	,013	141,932
	[q13.place=5]	0[b]	.	.	0	.	.	.	.
nonprofitfirm	Intercept	-13,045	991,603	,000	1	,990			
	q14.age	-,051	,013	15,646	1	,000	,950	,927	,975
	q18. monthlyaveragepersonal salary	,003	,000	472,352	1	,000	1,003	1,002	1,003
	[q12.gender=1]	-1,244	,215	33,471	1	,000	,288	,189	,439
	[q12.gender=2]	0[b]	.	.	0	.	.	.	.
	[q1.faculty=1]	-,147	,236	,388	1	,533	,864	,544	1,370
	[q1.faculty=2]	0[b]	.	.	0	.	.	.	.
	[q13.place=1]	10,417	991,603	,000	1	,992	33435,617	,000	.[c]
	[q13.place=2]	10,885	991,603	,000	1	,991	53372,882	,000	.[c]
	[q13.place=3]	10,552	991,603	,000	1	,992	38245,590	,000	.[c]
	[q13.place=5]	0[b]	.	.	0	.	.	.	.
private	Intercept	-14,557	,302	2325,143	1	,000			
	q14.age	,084	,009	86,079	1	,000	1,087	1,068	1,107
	q18. monthlyaveragepersonal salary	,001	,000	158,515	1	,000	1,001	1,001	1,002
	[q12.gender=1]	-1,325	,130	104,218	1	,000	,266	,206	,343
	[q12.gender=2]	0[b]	.	.	0	.	.	.	.
	[q1.faculty=1]	,362	,156	5,396	1	,020	1,436	1,058	1,948
	[q1.faculty=2]	0[b]	.	.	0	.	.	.	.
	[q13.place=1]	11,871	,232	2620,851	1	,000	143081,541	90824,470	225405,413
	[q13.place=2]	12,330	,151	6703,876	1	,000	226401,828	168537,143	304133,480
	[q13.place=3]	12,464	,000	.	1	.	258770,286	258770,286	258770,286
	[q13.place=5]	0[b]	.	.	0	.	.	.	.
own	Intercept	-15,636	,419	1395,348	1	,000			
	q14.age	,093	,012	57,741	1	,000	1,097	1,071	1,124
	q18. monthlyaveragepersonal salary	,002	,000	134,412	1	,000	1,002	1,001	1,002
	[q12.gender=1]	-2,002	,218	84,068	1	,000	,135	,088	,207
	[q12.gender=2]	0[b]	.	.	0	.	.	.	.
	[q1.faculty=1]	,221	,216	1,050	1	,306	1,247	,817	1,903
	[q1.faculty=2]	0[b]	.	.	0	.	.	.	.
	[q13.place=1]	12,514	,269	2158,835	1	,000	272250,940	160585,228	461565,335
	[q13.place=2]	11,805	,217	2958,511	1	,000	133956,392	87542,485	204978,360
	[q13.place=3]	11,677	,000	.	1	.	117793,892	117793,892	117793,892
	[q13.place=5]	0[b]	.	.	0	.	.	.	.

a. The reference category is: nojob.

b. This parameter is set to zero because it is redundant.

c. Floating point overflow occurred while computing this statistic. Its value is therefore set to system missing.

Figure 6. Parameter Estimator

First of all, all of the the Exp(B) values in the parameter estimator table are between the confidence interval values.

In this study, we see how well the dependent variable subunits (4 categories) are explained by the variables. In the parameter estimate table, there is the gender =1 (female) variable, which is meaningful to us in the first category that is in the officer category. In the parameter estimation table, gender = 1, that is, women's work in the officer job type is lower than men (gender = 2)

Being an officer employee was found to be statistically significant in the model. In addition, our age (independent) variable had a statistically significant contribution to the Office category.

When looking at the second category, which is non-profit firm category was not found to be statistically significant in the model. Also, our age (independent) and gender=1(female) variables contributed statistically significantly to the non-profit form category. In the parameter estimation table, gender = 1, that is, women's work in the non-profit firm job type is lower than men (gender = 2). In addition to this, as the age increases, the probability of working in a non-profit firm decreases.

When looking at the third category that is private category was found to be statistically significant in the model. Also, age(independent), gender = 1(female), faculty = 1 (open education), place = 1(village/town) and place=2 (city) variables contributed statistically significantly to the private category. Women's work in the private job type is lower than men (gender = 2). In Addition to this, those who graduated from the open education faculty prefer the private profession type more than those who graduated from the economics department.In the last category, which is ow job type age, gender= 1(female), place = 1 (village/town) and place=2 (city) variables contributed statistically significantly to the private category. As age increases, people tend to do their own profession. Women's work in the own job type is lower than men (gender = 2). In Addition, those who graduated from the open education faculty prefer the own profession type more than those who graduated from the economics department.

Classification Table:

This table provides an alternative approach for determining the extent of the model effect. The multinomial logistic model has a 71.1 percent accuracy, as seen below.

Classification

	Predicted					
Observed	nojob	officer	nonprofitfirm	private	own	Percent Correct
nojob	345	44	37	116	0	63,7%
officer	74	2202	80	57	0	91,3%
nonprofitfirm	27	32	207	9	0	75,3%
private	94	458	0	272	0	33,0%
own	32	89	0	78	0	0,0%
Overall Percentage	13,4%	66,4%	7,6%	12,5%	0,0%	71,1%

Figure 7. Classification Table for Multinomial Logistic Regression

Results of CART:

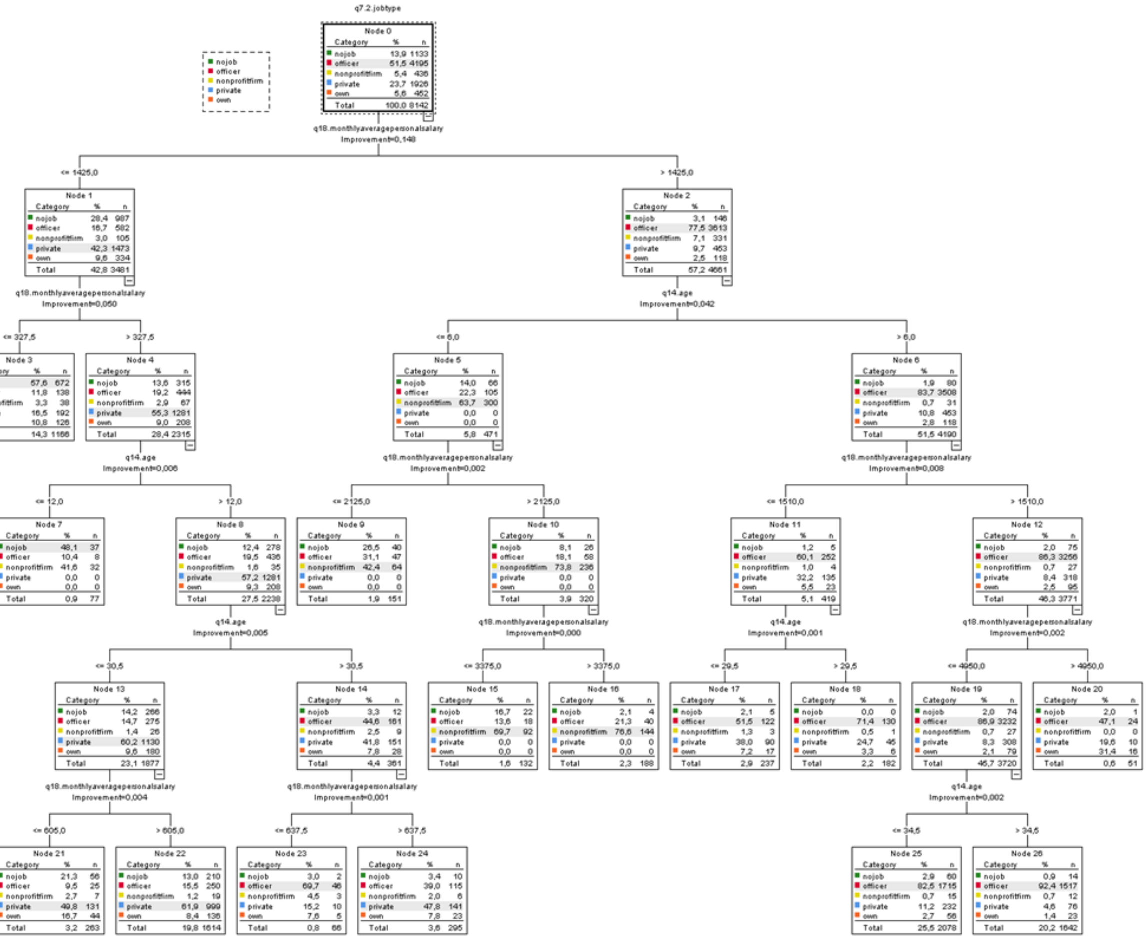


Figure 8. CRT

Classification

Observed	Predicted: nojob	officer	nonprofitfirm	private	own	Percent Correct
nojob	709	82	66	276	0	62,6%
officer	146	3554	105	390	0	84,7%
nonprofitfirm	70	34	300	32	0	68,8%
private	192	463	0	1271	0	66,0%
own	126	123	0	203	0	0,0%
Overall Percentage	15,3%	52,3%	5,8%	26,7%	0,0%	71,7%

Growing Method: CRT
Dependent Variable: q7.2.jobtype

Figure 9. Classification Table for CRT

This table provides an alternative approach for determining the extent of the model effect. The decision tree that is CRT has a 71,7 percent accuracy, as seen above.

DISCUSSION AND CONCLUSION

In the multinomial logistic regression analysis method, any category of the dependent variable is selected as the reference category and other categories are analyzed according to this reference category.

This reference category can be determined at the discretion of the investigator.

When the data obtained in this study are evaluated in general; The effects of other variables on the job type variable were examined. The numbers and percentages of the variables discussed in the study, according to the categories, are given in figure 1. When Figure 4.1 is examined, it is seen that those who participated in the survey; 12.7 percent of the occupation types are unemployed, 56.7 percent are office workers, 6.5 percent are nonprofit forms, 19.4 percent have private occupations and 4.7 percent It is seen that he has his own occupation type.

Examining Figure 2, the likelihood ratio test statistic was found to be significant.

When Figure 5 is examined, it is seen that the faculty, gender, place variables that we set for the factor also had a significant effect on the model. This is because the significant values are less than .05.

In Figure 6, there are regression coefficients in the first column, standard errors of these coefficients in the second column, Wald statistics in the third column, odds ratio in the fourth column and 95% confidence intervals of the odds ratio in the last column. As is known, in logistic regression analysis, as in standard regression analysis, regression coefficients are not directly significant, so these coefficients are not interpreted. Instead, the odds ratios calculated in relation to these coefficients are interpreted.

Figure 1 when examined, it is seen that 8142 people who study distance education: 1133 are no job, 4195 are officer, 436 are non-profit form, 1926 are private and 452 are own professions.

It is seen that 1133 of 8142 people who graduated from distance education faculties are no jobs. Among the most important determinants of no job people are the monthly average personal salary and age variables, respectively.

These determining variables divide the occupation type into 12 different profiles.

Comparison of Logistic Regression and CART Analysis:

Table 1. Analysis Comparison Criteria

	Processing Time	**Classification Succes**
Logistic Regression	13 min. 7 sec.	71,1%
CART	6 min. 20 sec.	71,7%

The short analysis times and the high success of classification and prediction have a very important place in the reliability of the models. However, no matter how short the processing times of the analyzes are, if the classification success is low, these models cannot be trusted, so classification success can be taken as a priority in comparing the models.

As seen in Table 1, the fastest analysis CART in terms of analysis timehas been. For this reason, it can be said that it would be more appropriate to use CART analysis in cases where rapid decision-making is required (Kıran, 2010)

To sum up, there are advantages and disadvantages in both methods, but here it was decided that the performance of the cart application is better in terms of both time and data explainability.

References

Bahçe S, 2017. Multinomiyal Logistik Regresyon Analizi Ve Anket Verilerinde Uygulanabilirliği. Yüksek Lisans Tezi, Yüzüncü Yıl Üniversitesi, Sağlık Bilimleri Enstitüsü, Van.

Breiman L, Friedman J, Stone CJ. Olshen RA, 1984. Classification and Regression Trees. Taylor and Francis, Chapman&Hall/CRC.

Çelik Ö, Aslan A.F, 2019. Gender prediction from social media comments with artificial intelligence, Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi 23(6): 1256-1264.

David Garson, G, 2014. Logistic Regression: Binary & Multinomial. Asheboro, NC 27205 USA: Statistical Publishing Associates.

Elasan, S, 2010. Multinomiyal logistik regresyon analizi ve uygulaması, Yüksek Lisans Tezi, Yüzüncü Yıl Üniversitesi, Sağlık Bilimleri Enstitüsü, Van.

Ergüt Ö, Camkıran C, Çilingirtürk A.M, 2020. Finansal Okuryazarlık Düzeyi ile Sosyal Ağların Etkileşimi Üzerine Ampirik Bir Uygulama, Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi, 18. Özel Sayı: 295-307.

Freund, R. J., & Wilson, W.J, 2006. Regression Analysis: Statistical Modeling of A Response Variable, 2nd ed. Burlington: Academic Press.

Harrington, P, 2012. Machine Learning in Action, Manning Publications Co, United States of America.

Kıran, Z, 2010. Lojistik Regresyon Ve Cart Analizi Teknikleriyle Sosyal Güvenlik Kurumu İlaç Provizyon Sistemi Verileri Üzerinde Bir Uygulama. Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

Rokach, L. & Maimon, O, 2015. Data Mining with Decision Trees: Theory and Applications, 2nd Edition) : 81-85.

Saraç, M. 2020. Van Yüzüncü Yıl Üniversitesi Elektronik Belge Yönetim Sisteminin (Ebys) Kullanım Memnuniyetinin Cart Karar Ağacı İle İncelenmesi, Van Yüzüncü Yıl Üniversitesi Sağlık Bilimleri Enstitüsü, Van.

Şenel S, Alatli B, 2014. Lojistik regresyon analizinin kullanıldığı makaleler üzerine bir inceleme. Journal of Measurement and Evaluation in Education and Psychology. 5(1): 35-52.

Yazıcı B, Kan Kılınç B, Gunsoy B, G, 2020. Perceptions And Opinions Of Graduates About The Effects Of Open And Distance Learning In Turkey. Turkish Online Journal of Distance Education 21(1): 121-132.

Yavuz, A, 2021. Comparison of predictive performances of logistic regression and decision tree algorithms: an application of life satisfaction, Dokuz Eylül Üniversitesi, İzmir.

Yavuz, A, Çilengiroğlu, Ö. V, 2020. Lojistik Regresyon Ve CART Yöntemlerinin Tahmin Edici Performanslarının Yaşam Memnuniyeti Verileri Için Karşılaştırılması. Avrupa Bilim ve Teknoloji Dergisi, 18: 719-727.

Yamantaş, Ö, 2019. Türkiye Sağlık Araştırması Verilerinin Yapay Sinir Ağları Ve Çok Kategorili Lojistik Regresyon Yöntemiyle İncelenmesi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Acknowledgment

This study was supported financially by the Commission of the Scientific Research Projects of Eskişehir Technical University under grant number 22LÖP177.

Conflict of Interest

The authors have declared that there is no conflict of interest.

Author Contributions

Yaren MERT (40 %)

Berna YAZICI (30 %)

Ece ÖZGÖREN (30 %)

Tuz Stresi Altındaki Börülcede Bitki Yaş Ağırlığının CART, CHAID ve Exhausted CHAID Veri Madenciliği Algoritmaları ile Tahmini

Özlem Önal Aşcı[1], Yeliz Kasko Arıcı[2*], Yasin Altay[3], Mualla Altun[1]

[1] Ordu Üniversitesi, Ziraat Fakültesi, Tarla Bitkileri Bölümü, 52200, Ordu, Türkiye
[2] Ordu Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim A.D., 52200, Ordu, Türkiye
[3] Eskişehir Osmangazi Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, 26160, Eskişehir, Türkiye

*Corresponding author e-mail: yelizkasko@gmail.com

Özet

Amaç: Bu çalışmanın amacı, tarla bitkileri alanında ağaç tabanlı CART (Classification and Regression Tree), CHAID (Chi-Squared Automatic Interaction Detector) ve Exhausted CHAID (Exhausted Chi-Squared Automatic Interaction Detector) veri madenciliği algoritmalarının kullanımını göstermek ve algoritmaların tahminleme performanslarını karşılaştırmaktır. Bu amaçla, kaba ve kesif yem olarak hayvan beslemede, yemeklik tane baklagil ve sebze olarak da insan beslenmesinde değerlendirilen tek yıllık baklagil bitkisi olan börülceye (Vigna unguiculata L.) ait veriler kullanılmıştır. Veri madenciliği algoritmaları kullanılarak farklı dozlarda tuz stresi altındaki börülcenin toprak üstü yaş ağırlığı tahmin edilmiştir. Materyal ve Yöntem: Kademeli olarak 9 farklı tuz konsantrasyonu (0, 25, 50, 75, 100, 125, 150, 175 ve 200 mM NaCl) uygulanmış Karagöz ve Ülkem börülce çeşitlerinin morfolojik, fizyolojik ve kimyasal özelliklerine ait toplam yirmidört adet değişkene ait veriler çalışmanın veri setini oluşturmuştur. Toprak üstü yaş ağırlığı bağımlı değişken olarak dikkate alınırken, diğerler değişkenler bağımsız değişken olarak kullanılmıştır. Veri seti SPSS 23 (IBM Corp. Released., 2015) programı kullanılarak analiz edilmiş ve CART, CHAID ve Exhausted CHAID algoritmaları 6 ana düğüm ve 3 alt düğüm kullanılarak çalıştırılmıştır. Çapraz doğrulama işlemi için 5-kat çapraz doğrulama tercih edilmiştir. Algoritmaların tahmin performanslarını karşılaştırmada kullanılan uyum iyiliği ölçütleri R programı "ehaGoF" paketi kullanılarak hesaplanmıştır (R Core Team, 2019). Bulgular: Uyum iyiliği ölçütleri dikkate alındığında, börülce yaş ağırlığını en iyi açıklayabilen veri madenciliği algoritmasının CART olduğu belirlenmiştir. CART'tan düşük olmakla birlikte CHAID ve Exhaustive CHAID algoritmalarının da oldukça iyi tahmin performanslarına sahip oldukları görülmüştür. CART algoritması ile elde edilen ağaç diyagramı 8'i terminal 6'sı iç düğüm olmak üzere toplam 14 adet düğüm içermiştir. Bitki boyu (cm), yaprak alanı (cm2 bitki-1), gövde çapı (mm) ve K (ppm) börülce yaş ağırlığının tahmini için etkili değişkenler olarak belirlenmiştir. Birinci dereceden etkili bağımsız değişken bitki boyu olarak belirlenirken, ikinci dereceden yaprak alanı, üçüncü dereceden gövde çapı, K ve bitki boyu etkili bağımsız değişkenler olarak belirlenmiştir. Börülce yaş ağırlığındaki varyasyonun % 82.5'i CART algoritması ile önemli bulunan bağımsız değişkenler tarafından açıklanmıştır (R2=0.825). Börülce yaş ağırlığının tahminde oluşturulan alt grupların yaş ağırlığı 6.618 gr ile 13.583 gr aralığında değişim göstermiş olup ortalaması 9.389 gr olarak tespit edilmiştir. Tuz stresindeki börülce bitkisinde, en düşük yaş ağırlığa sahip olanların boyu 23.125 cm ve yaprak alanı 123.278 cm2'den küçük veya eşit olan bitkiler olduğu belirlenirken, en yüksek yaş ağırlığa sahip olanların boyu 23.125 cm, yaprak alanı 185.795 cm2 ve K içeriği 1012 ppm'den büyük olan bitkiler olduğu belirlenmiştir. Sonuç: Sonuç olarak, bu çalışmada ağaç tabanlı veri madenciliği algoritmaları kullanılarak tarla bitkileri alanında hasat öncesi isabetli tahminler yapılabileceği ortaya konulmuştur.

Key words: *Veri madenciliği, CART, CHAID, Exhause CHAID, Börülce.*

POSTER PRESENTATIONS

POSTER PRESENTATIONS CONTENTS

Modeling Semi-competing Risks Approach for Predicting Factors of Early Recurrence and Postoperative Death in Patients with Colorectal Cancer

Mohammad Asghari-Jafarabadi [1*], Ramezan Fallah [1], Ghodratollah Roshanaei [2]

[1] Zanjan University, Medical Sciences, 45, Zanjan, Iran
[2] Hamadan University, Medical Sciences, 65781, Hamadan, Iran

*Corresponding author e-mail: m.asghari862@gmail.com

Abstract

Aim: In this study, the aim is to use a semi-competing risks approach with Bayesian methods for modeling Factors of early Recurrence and Postoperative Death in Patients with Colorectal Cancer Methods: This study is a retrospective cohort of 284 patients with colorectal cancer who underwent surgery referred to Imam Khomeini Clinic in Hamadan during the years 1380-1396. The age at the time of diagnosis, sex and clinical and pathological variables, type and date of first treatment, degree of tumor differentiation, disease stage, recurrence and metastasis, types of treatment (surgery, chemotherapy and radiotherapy) were included as the predictors. The patients 'recurrence status was determined from patients' records. The Survival modeling was performed in the presence of semi-competing hazards with Bayesian approach in R 4.1 software to analyze the data. Results: The results showed that for nonterminal event of recurrence Sex, age, stage of disease, , Tumor size, Number of chemotherapy was significantly related. Also, for terminal event of death, age, metastasis to other organs, stage of disease, Grade, Tumor size and Number of chemotherapy was the significant predictors. Also, for terminal event of death conditional on nonterminal event of recurrence, age, metastasis to other organs and Number of chemotherapy was significantly related Conclusion: Suitable and tailored steps could be taken for policy making according to the specific results obtained for terminal and nonterminal events.

Key words: *Semi Competing Risks, Death, Early Recurrence, Colorectal Cancer*

Genomik Araçların Hayvan Yetiştiriciliğinde Uygulanması

Mervan Bayraktar[1], Bahri Devrim Özcan[1*], Gülşah Keklik[1], Elif Dikkaya[1]

[1] Çukurova University, Faculty of Agriculture, Animal Science, 01250, Adana, Turkey

*Corresponding author e-mail: bdozcan@gmail.com

Özet

Genetik bakımından üstün özelliklere sahip bireyleri gelecek generasyonlar için ebeveyn olarak seçip ve bu işlemi hızlı bir şekilde gerçekleştirmek hayvancılığın temel amaçlarından biridir. Seleksiyon başarısının daha isabetli olabilmesi ve generasyon sürenin kısalma yolu ancak genomik seleksiyon tarafından gerçekleştirilebilir. Genomik seleksiyon yeni bir yaklaşım olarak markör destekli seleksiyonun dezavantajlarını gidermek üzere günümüzde sıkça kullanılmaya başlanmıştır. Genomik seleksiyonda bir hayvanın genetik potansiyeli (diğer bir deyişle damızlık değerleri), tüm genom boyunca yayılmış binlerce markör dikkate alınarak tahmin edilir. Yüksek yoğunluklu SNP genotiplemenin ortaya çıkışı, son zamanlarda çiftlik hayvanlarında genomik seleksiyonunun uygulanmasına önemli avantajlar sunmuştur.

Key words: *Genomik seleksiyon, Genetik markörler, SNP*

Strategy for Managing Inputs and Outputs in Agriculture Through Information Technology

Blerta Leka (Moçka)[1*], Daniel Leka[2]

[1] Agricultural University of Tirana, Faculty of Economy and Agrobusiness Albania
, Department of Mathematics and Informatics, 1025, Tirana, Albania
[2] 2Special Court of First Instance for Corruption and Organised Crime , 1016, Tirana, Albania

*Corresponding author e-mail: bmocka@ubt.edu.al

Abstract

Information Technology has advanced greatly by integrating with every possible field including that of Agribusiness. There are many technology systems like sensor and other component that can be used in agricultural systems. We will study this information technology systems that are used tol perform measurements for inputs, factors, different in the field for agricultural production. From the detailed analysis of these factors can be created Strategies to achieve the required outputs. As inputs will be the factors for agricultural production focused on a certain area, like: Cosmic factors, Climatic and terrestrial factors, Technological factors, and Market factors. For each of these factors a technological environment can be developed to help and closely assist agribusiness, focusing concretely on real-time and continuous measurement of key indicators of necessary factors. From the analysis of these indicators, immediate and predictive measures can be taken to increase and maximize agricultural production.

Key words: *Agribusiness, Information Technology, Sensor, And Analyses.*

Affecting Factors During the Anaerobic Digestion Process

Edmond Demollari[1*], **Ndoc Vata**[1], **Ervisa Demollari**[2]

[1] Agricultural University of Tirana, Faculty of Agriculture and Environment, Department of Agronomic Sciences, 1029, Kodër-Kamëz, Tirana, Albania
[2] Central Directorate of the Health-Care Operator, 1002, Tirana, Albania

*Corresponding author e-mail: edemollari@ubt.edu.al

Abstract

Anaerobic digestion is an ideal bioprocess technology in terms of handling and treatment of waste. It has some benefits to be applied and developed in the field of waste management technologies. This is a proven technology for sewage sludge treatment, in which the high-water content sewage sludge can be processed without any pretreatment. The objective of this study was to show the influences that different factors have during the process of biogas production under anaerobic conditions. Anaerobic digestion of organic raw material is a totally complex process involving the decomposition steps of the composite materials. Many factors, including the type and concentration of substrate, temperature, moisture, pH, etc., may affect the performance of the anaerobic digestion process in the bioreactor. Biowaste can be processed into a suspension, containing a high proportion of biodegradable substances by addition of process water and the suspension treated in wet anaerobic digestion (total solid content: TS <15%). From the moisture content of different substrates, two main types of anaerobic digestion processes can be distinguished for organic solid waste treatment generally referred to as wet and dry anaerobic digestion. The activity of anaerobic microorganisms, the AD efficiency (in terms of biogas and digestate yields) and reliability are influenced by many critical factors: type, chemical composition (water and nutrient content, e.g. nitrogen, carbon source, C/N ratio), concentration and pH of raw materials; absence of oxygen; microbial composition; (constant) environmental temperature; presence and amounts of inhibitors (like ammonia); toxic compounds; concentration of intermediate products (like volatile fatty acids). As a result, careful treatment of these factors is considered the most important in the production of biogas under anaerobic conditions.

Key words: *Biowaste, Critical factor, Biodegradable substance*

Süt ve Süt Ürünlerinde Kullanılan İstatiksel Araştırma ve Deneme Metotları

Fatoş Kaplan[1*], **Mustafa Şahin**[2]

[1] Kahramanmaraş Sütçü İmam Üniversitesi, Fen Bilimleri Enstitüsü, 46040, Kahramanmaraş, Türkiye
[2] Kahramanmaraş Sütçü İmam Üniversitesi, Tarımsal Biyoteknoloji Bölümü,46040, Kahramanmaraş, Türkiye

*Corresponding author e-mail: ms66@ksu.edu.tr

Özet

Bu çalışmada, öncelikle süt ve süt ürünlerinde kullanılan araştırma ve deneme metotları hakkında bilgiler verilmiştir. Süt ve süt ürünlerinin öneminden ve Türkiye'deki ve Dünya'daki tüketim istatistiklerinden bahsedilmiştir. Süt ve süt ürünlerindeki deneysel araştırmalar için yaygın olarak kullanılan deneme modelleri sahada kullanılan istatiksel araştırma ve deneme metotları ile uygun olmalıdır. İstatiksel analiz süt ve süt ürünleriyle birlikte tıp, ekonomi, mühendislik gibi deneysel araştırma yapılan pek çok alanda bilimsel çalışmaların doğruluğunun tespiti için gereklidir. Araştırmacılar yapacağı bilimsel çalışmalar için kullanacağı istatiksel deneme modelini belirleyip çalışmaya başlamalıdır.

Key words: *Süt ve Süt Ürünleri, İstatiksel Analiz, Araştırma ve Deneme Metodları*

Estimating Fermentation Capacity of Bacillus Sphaericus MIB Grown Under Different Carbon Sources Using Noisy Measurements

Rümeysa Nur Afşar[1*], **Emrah Nikerel**[1]

[1] Yeditepe University, Department of Genetics and Bioengineering, 34755, Istanbul, Turkey

*Corresponding author e-mail: rumeysanur.afsar@std.yeditepe.edu.tr

Abstract

Bacillus sphaericus offers significant potential as microbial host for various industries including biopesticide (as pest control), bio larvicide (as mosquito control) or production of various enzymes. Typically found in soil, Bacillus sphaericus is a gram-positive bacterium, obligate aerobic and forms spores under extreme cultivation conditions such as high temperature. Large scale cultivation of this bacteria is needed in industrial biotechnology setting, where fermentation conditions, substrate portfolio, growth dynamics need to be crisply estimated. Aim of this study is to monitor the substrate portfolio, the growth rate and yield of Bacillus sphaericus MIB strain in chemical medium with different carbon sources. In doing so, Bacillus sphaericus MIB strain was grown into six different carbon sources including sucrose, glucose, xylose, sorbitol, sodium acetate and succinate, while other conditions were kept the same. Interestingly, this strain grew only in chemical medium with sodium acetate or succinic acid as carbon sources. Individual growth rates were calculated to be 0.185 and 0.141 h-1 for sodium acetate and succinate, with phenotypic differences (changing color) in succinate culture. The noisy measurements are used for state estimation of fermentation: growth, uptake and secretion rates, yields in each growth phase (lag-log-late phases). In doing so, elemental (overall carbon, nitrogen and electron) balances are used in a data reconciliation setting. Significant differences and similarities among substrate portfolio, yields are defined. Overall, this work illustrates the importance of the in-house isolated microorganism as a platform host by crisply measuring several fermentation parameters.

Key words: *Bacillus sphaericus MIB, growth rate, industrial biotechnology, fermentation rates, state estimation.*

Methane Mitigation from Rice Field: An Overview

Dhananjay Gaikwad[1*], **Naleeni Ramawat**[2], **Y. Pravalika**[3], **Dhananjay K. Pandey4**

[1] Amity Institute of Organic Agriculture, Amity University, Delhi NCR, Uttar Pradesh, India
[2] Amity Institute of Organic Agriculture, Amity University, Noida, 201301Uttar Pradesh, India
[3] School of Agriculture, Lovely Professional University, Jalandhar, - 144411 Punjab, India
[4] Amity Institute of Biotechnology, Amity University, 834001 Jharkhand, India

*Corresponding author e-mail: dsgaikwad7@gmail.com

Abstract

The earth's environment is undergoing change because of pollution and greenhouse gas emissions. Each year, the amount of methane (CH4) in the atmosphere increases by 0.3 per cent. Methanogenesis is a geochemical process that occurs in all anaerobic environments and results in the formation of CH4. It is regulated by a number of soil, climatic, and management factors. Rice fields account for about 20% of global methane emissions in the 2000s, global rice cultivated areas produce methane were predicted to range from 18.3 Tg methane per year (under intermittent irrigation) to 33.8 Tg methane per year (under continuous flooding), expressing that the Ch4 emission from the rice cultivated area is majorly dependent on the type of irrigation system provided to the filed. Independently, mitigation methods were considered to significantly address the three basic practices of water-provided rice, rainfed rice and deep-water rice. Rice irrigation is the greatest emitter of methane but is also the best way to reduce this emission by changing crop management. Using mid-season drainage, low emission cultivars, low C: N organic manure, and the direct establishment of rice crops can reduce CH4 emissions from rice fields. In conclusion, field management strategies and soil preparation prior to crop planting are essential for decreasing rice field methane emissions. This review will concentrate on various elements of CH4 emissions from rice fields as well as mitigation techniques.

Key words: *Green house, Methane, Emission, Climate, Rice, Water Management*