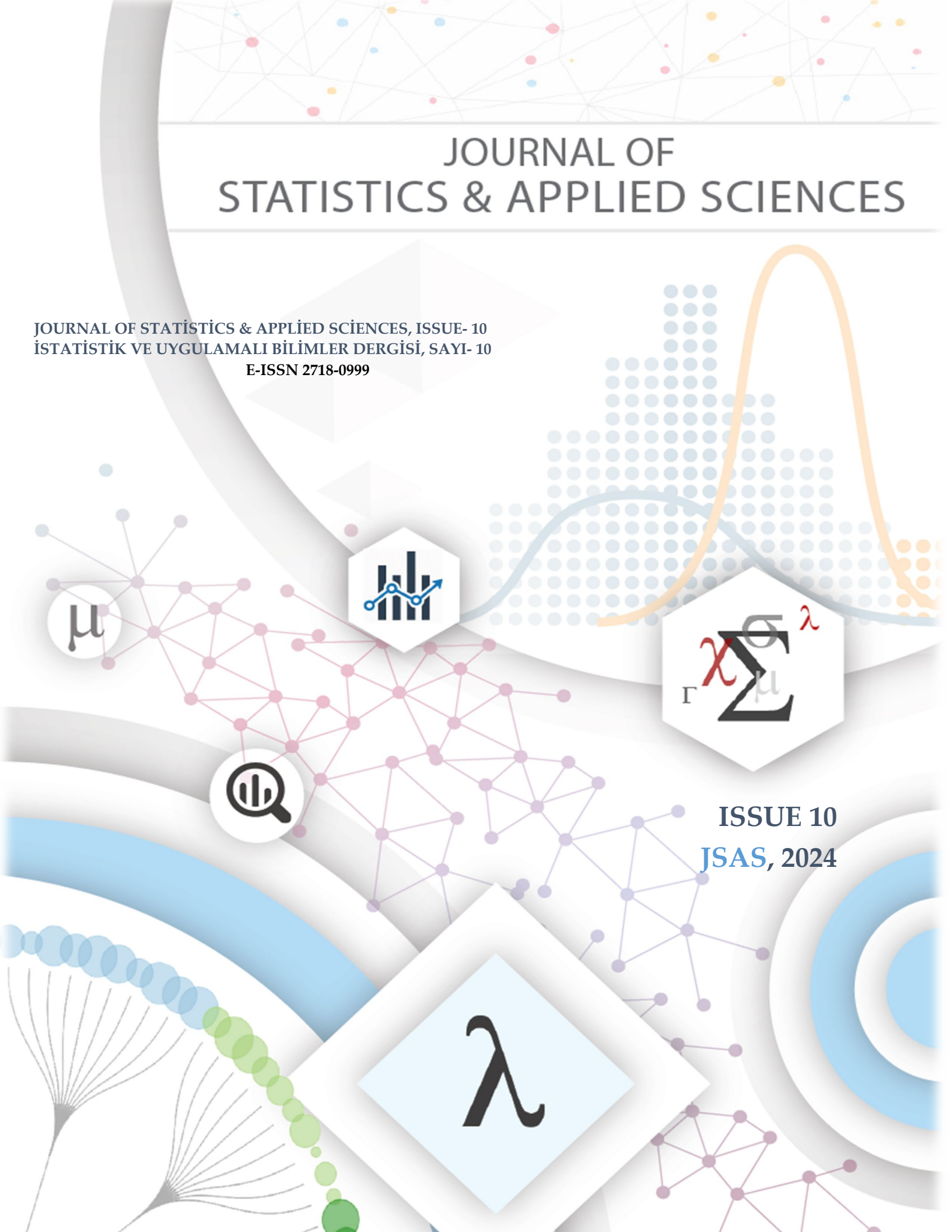
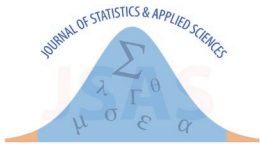


# JOURNAL OF STATISTICS & APPLIED SCIENCES

JOURNAL OF STATISTICS & APPLIED SCIENCES, ISSUE- 10  
İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ, SAYI- 10  
E-ISSN 2718-0999

ISSUE 10  
JSAS, 2024





Journal of Statistics & Applied Sciences, Issue - 10  
ISSN 2718-0999

# -JOURNAL OF STATISTICS & APPLIED SCIENCE- -İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ-

**Aralık 2024**

**Issue: 10**

**Sayı: 10**

**İmtiyaz Sahibi / Owner**

Abdulkadir KESKİN

**Baş Editör / Editor in- Chief**

Dr. Abdulkadir KESKİN

**Yayın Türü:** 6 Aylık, Uluslararası, Hakemli

**Publication Type:** 6 Monthly, International, Refereed

**e-ISSN:** 2718-0999

**Dizinleme Bilgileri/ Abstracting and Indexing Services**


**İletişim / Contact**

E-posta: [jsas.journal@gmail.com](mailto:jsas.journal@gmail.com)

<https://dergipark.org.tr/tr/pub/jsas>

## BİLİM VE DANIŞMA KURULU

Prof. Dr. Serpil AKTAŞ ALTUNAY,	HACETTEPE ÜNİVERSİTESİ
Prof. Dr. Filiz KARAMAN,	YILDIZ TEKNİK ÜNİVERSİTESİ
Prof. Dr. Serhat YÜKSEL,	İSTANBUL MEDİPOL ÜNİVERSİTESİ
Prof. Dr. Ersoy ÖZ,	YILDIZ TEKNİK ÜNİVERSİTESİ
Prof. Dr. Dursun YENER,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Prof. Dr. Serpil KILIÇ DEPREN	YILDIZ TEKNİK ÜNİVERSİTESİ
Doç. Dr. Atif Ahmet EVREN,	YILDIZ TEKNİK ÜNİVERSİTESİ
Doç. Dr. Abdulkadir ATALAN,	ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ
Doç. Dr. Akansel YALÇINKAYA,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Doç. Dr. Ömer BİLEN,	BURSA TEKNİK ÜNİVERSİTESİ
Doç. Dr. Muhammed Fevzi ESEN,	SAĞLIK BİLİMLERİ ÜNİVERSİTESİ
Doç. Dr. Yavuz ÖZDEMİR,	İSTANBUL SAĞLIK VE TEKNOLOJİ ÜNİVERSİTESİ
Doç. Dr. Hasan ŞAHİN,	BURSA TEKNİK ÜNİVERSİTESİ
Doç. Erhan ÇENE,	YILDIZ TEKNİK ÜNİVERSİTESİ
Doç. Kemal Gökhan NALBANT,	BEYKENT ÜNİVERSİTESİ
Doç. Dr. Elif TUNA,	YILDIZ TEKNİK ÜNİVERSİTESİ
Doç. Şahika ÖZDEMİR,	İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ
Dr. Mustafa DEMİRBİLEK,	GAZİANTEP İSLAM BİLİM ve TEKNOLOJİ ÜNİVERSİTESİ
Dr. Sevim ÖZULUKALE DEMİRBİLEK,	YOZGAT BOZOK ÜNİVERSİTESİ
Dr. Adilson SİLVA,	CAPE VERDE ÜNİVERSİTESİ
Dr. Fatma Rabia AKTAŞ ŞENKARDEŞLER	BEYKOZ ÜNİVERSİTESİ
Dr. Abdurrahman KESKİN	BAYBURT ÜNİVERSİTESİ
Dr. Tutku SEÇKİN ÇELİK,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Burak LEBLEBİCİOĞLU,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Batuhan ÖZKAN,	YILDIZ TEKNİK ÜNİVERSİTESİ
Dr. Coşkun PARİM,	YILDIZ TEKNİK ÜNİVERSİTESİ
Dr. Egemen ÖZKAN,	YILDIZ TEKNİK ÜNİVERSİTESİ
Dr. İrfan ERSİN,	İSTANBUL MEDİPOL ÜNİVERSİTESİ
Dr. Enes FİLİZ,	BALIKESİR ÜNİVERSİTESİ
Dr. Hasan Aykut KARABOĞA,	AMASYA ÜNİVERSİTESİ
Dr. Hasan DİLBAŞ,	VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
Dr. Mohamed AHMED,	AL MADİNA HİGHER INSTITUTE FOR MANAGEMENT
Dr. Sinan ATAER,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Yavuz Selim HATİPOĞLU	BAYBURT ÜNİVERSİTESİ
Dr. Müge YETKİN ATAER,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Gözde NALBANT EFE,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. İrem BATIBAY TÜNAYDIN,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Mustafa YAPAR,	RECEP TAYİP ERDOĞAN ÜNİVERSİTESİ
Dr. Burcu KUZUCU YAPAR,	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Tarkan TALAN,	GAZİANTEP İSLAM BİLİM ve TEKNOLOJİ ÜNİVERSİTESİ
Dr. Çağatay TEKE,	BAYBURT ÜNİVERSİTESİ
Dr. Bayram DÜNDAR,	BARTIN ÜNİVERSİTESİ

## EDİTÖR KURULU

Dr. Abdulkadir KESKİN	Baş Editör	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Doç. Dr. Abdulkadir ATALAN	Yardımcı Editör	ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ
Dr. Mehmet Şamil GÜNEŞ	Yardımcı Editör	YILDIZ TEKNİK ÜNİVERSİTESİ
Dr. Enes FİLİZ	Uygulamalı İstatistik Alan Editörü	BALIKESİR ÜNİVERSİTESİ
Doç. Ahmet Atif EVREN	Teorik İstatistik Alan Editörü	YILDIZ TEKNİK ÜNİVERSİTESİ
Dr. Egemen ÖZKAN	Risk Analizi Alan Editörü	YILDIZ TEKNİK ÜNİVERSİTESİ
Dr. Burak LEBLEBİCİOĞLU	Sosyal Bilimler Alan Editörü	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Yasemin Ayaz ATALAN	Mühendislik Alan Editörü	YOZGAT BOZOK ÜNİVERSİTESİ
Dr. Bayram DÜNDAR	Mühendislik Alan Editörü	BARTIN ÜNİVERSİTESİ
Dr. Çağatay TEKE	Mühendislik Alan Editörü	BAYBURT ÜNİVERSİTESİ
Dr. Tarkan TALAN	Eğitim Bilimleri Alan Editörü	GAZİANTEP İSLAM BİLİM ve TEKNOLOJİ ÜNİVERSİTESİ
Dr. Robert RODGERS	Sağlık Bilimleri Alan Editörü	ALLEGHENY HEALTH NETWORK
Dr. Berat KARA	Yayın Editörü	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Abdurrahman KESKİN	Yayın Editörü	BAYBURT ÜNİVERSİTESİ
Dr. Recep Uğurcan ŞAHİN	Mizanpaj Editörü	YALOVA ÜNİVERSİTESİ

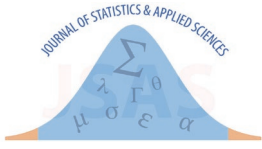
## YABANCI DİL EDİTÖRLERİ

Dr. Batuhan ÖZKAN,	Türkçe Dili Editörü	YILDIZ TEKNİK ÜNİVERSİTESİ
Dr. Tutku SEÇKİN ÇELİK	Yabancı Dil Editörü (İngilizce)	İSTANBUL MEDENİYET ÜNİVERSİTESİ
Dr. Ahmet Furkan EMREHAN	Yabancı Dil Editörü (İngilizce)	YILDIZ TEKNİK ÜNİVERSİTESİ
Umut ÜNAL	Yabancı Dil Editörü (İngilizce)	İSTANBUL MEDENİYET ÜNİVERSİTESİ

## İÇİNDEKİLER (CONTENT)

### ARAŞTIRMA MAKALELERİ (RESEARCH ARTICLES)

<b>Long-run Relationship Between Türkiye and SHANGHAI Cooperation Organization Stock Markets: Cointegration Analysis</b>	<b>1-9</b>
Türkiye ve Şanghay İşbirliği Örgütü Hisse Senedi Piyasaları Arasındaki Uzun Dönem İlişki: Eşbütünleşme Analizi	
<a href="#">Sara Huseynova</a> , <a href="#">Mahammad Charkasov</a>	
<b>A New Approach for the Measurement of Criterion Weights in the Scope of Multi-Criteria Decision Making: Spearman Rank Correlation-based Expanded CRITIC Method (SRCBECM)</b>	<b>10-34</b>
Çok Kriterli Karar Verme Kapsamında Kriter Ağırlıklarının Ölçülmesi için Yeni Bir Yaklaşım: Spearman Sıra Korelasyonuna Dayalı Genişletilmiş CRITIC Yöntemi (SRCBECM)	
<a href="#">Furkan Fahri Altıntaş</a>	
<b>Determination of Critical Factors Affecting The Success of Engineering Students by Structural Equation Modeling</b>	<b>35-49</b>
Mühendislik Öğrencilerinin Başarılarını Etkileyen Kritik Faktörlerin Yapısal Eşitlik Modellemesi ile Belirlenmesi	
<a href="#">Delal Demir</a> , <a href="#">Hasan Şahin</a>	
<b>Kelley's coefficients of skewness using R</b>	<b>50-75</b>
Kelley'nin R ile eğrilik katsayıları	
<a href="#">Jose Moral De La Rubia</a>	
<b>The Impact of Budget Revenue and Expenditure Forecasting Errors on Inflation in Turkey: An Analysis of the 1975-2021 Period</b>	<b>57-66</b>
Türkiye'de Bütçe Gelir ve Harcama Tahmin Hatalarının Enflasyona Etkisi: 1975-2021 Dönemi Analizi	
<a href="#">Gökhan Özdemir</a> , <a href="#">Semih Önüt</a>	
<b>A Robust Initial Basic Subset Selection Method for Outlier Detection Algorithms in Linear Regression</b>	<b>76-85</b>
Doğrusal Regresyonda Aykırı Değer Tespit Algoritmaları için Dayanıklı Bir Temel Küme Seçim Yöntemi	
<a href="#">Mehmet Hakan Satman</a>	
<b>Residual Modelling as a New Approach for Variable Selection</b>	<b>86-95</b>
Değişken Seçiminde Yeni Bir Yaklaşım Olarak Artık-Değer Modellemesi	
<a href="#">Aslı Nurefşan Koçak</a> , <a href="#">Muhammet Furkan Daşdelen</a> , <a href="#">Mehmet Koçak</a>	
<b>Examination of Statistical Methods Used in Group Comparisons in Visual Analogue Scale (VAS) Data in Terms of Type I Error Rate and Power of Test: Monte Carlo simulation study for VAS data</b>	<b>95-104</b>
Görsel Analog Ölçek (VAS) Verilerindeki Grup Karşılaştırmalarında Kullanılan İstatistik Metotların Değerlendirilmesi: VAS verileri için Monte Carlo Simülasyon Çalışması	
<a href="#">Özgür Koskan</a> , <a href="#">Malik Ergin</a>	



---

---

**Prediction of Housing Prices Using Machine Learning Algorithms: The Case of Giresun**

105-111

Makine Öğrenmesi Algoritmalarıyla Konut Fiyatlarının Tahmini: Giresun Örneği

[Mehmet Akif Kara](#)

---

---

**Mühendislik Tasarım Becerileri Algı Testinin Geliştirilmesi (MüTasBAT)**

112-124

Development of Engineering Design Skills Perception Test (EnDeSPT)

[Suat Türkoguz](#) , [Kübranur Sarı](#)

---

---

**Child Suicide News: A Comprehensive Case Study from Turkish Media**

125-132

Çocuk İntiharı Haberleri: Türk Basınından Kapsamlı Bir Vaka Çalışması

[Nazan Çalbayram](#) , [Gamze Balta](#)

---

---

**Derleme Makale (Review Article)**

---

---

**Use Of Renewable Energy Resources In The Service Sector: Literature Review**

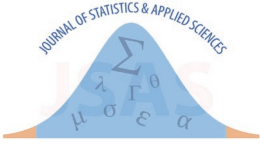
133-145

Yenilenebilir Enerji Kaynaklarının Hizmet Sektöründe Kullanımı: Literatür Çalışması

[Afra Sudenaz Uslu](#) , [Yavuz Özdemir](#) , [Mustafa Yıldırım](#) , [Hakan Kavak](#)

---

---



## ÖNSÖZ

Sayın Okurlar,

Bu yıl, İstatistik ve Uygulamalı Bilimler Dergisi'nin 10. sayısını yayımlamanın mutluluğunu yaşıyoruz. Bu özel sayı, yalnızca geçmişteki başarılarımızı değil, aynı zamanda bilime olan katkılarımızı daha da ileriye taşıma kararlılığımızı da yansıtmaktadır.

Bu sayımızdaki makaleler, istatistiksel yöntemlerin uygulamalı bilimlerle kesişim noktalarını kapsamlı bir şekilde ele almaktadır. Türkiye ve Şanghay İşbirliği Örgütü hisse senedi piyasaları arasındaki ilişkinin incelendiği eşbütünleşme analizinden, mühendislik öğrencilerinin başarılarını etkileyen faktörlerin yapısal eşitlik modellenmesiyle belirlenmesine kadar geniş bir yelpazede derinlemesine araştırmalar sunulmaktadır. Ayrıca, değişken seçiminde yeni yaklaşımlar, makine öğrenmesi algoritmalarıyla konut fiyat tahmini ve yenilenebilir enerji kaynaklarının hizmet sektöründe kullanımına yönelik çalışmalar da bu sayıya değer katmaktadır.

Bu sayı, disiplinler arası araştırmaları teşvik eden ve bilim dünyasına yeni bakış açıları kazandıran bir platform olarak dergimizin amacını bir kez daha vurgulamaktadır. Bu sayının hazırlanmasında emeği geçen yazarlarımıza, hakemlerimize ve editör ekibimize teşekkürlerimizi sunar, burada yer alan çalışmaların bilim dünyasına değerli katkılar sağlayacağına inanıyoruz.

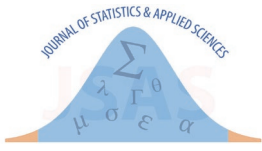
Dergimizin 10. sayısını siz değerli okuyucularımızla buluşturmanın gururunu yaşarken, daha nice sayılarla bilime katkı sağlamayı hedefliyoruz.

Keyifli okumalar dilerim.

Saygılarımızla,

**İstatistik ve Uygulamalı Bilimler Dergisi Editörü**

Dr. Abdulkadir Keskin



## PREFACE

Dear Readers,

This year, we are delighted to publish the 10th issue of the Journal of Statistics and Applied Sciences. This special issue not only reflects our past achievements but also underscores our commitment to advancing academic contributions to science.

The articles in this issue comprehensively explore the intersections of statistical methods with applied sciences. From cointegration analysis examining the relationship between Türkiye and the Shanghai Cooperation Organization stock markets to structural equation modeling identifying factors affecting the success of engineering students, this edition covers a broad spectrum of in-depth research. Additionally, studies addressing new approaches in variable selection, housing price prediction using machine learning algorithms, and the utilization of renewable energy resources in the service sector add significant value to this publication.

This issue reaffirms our mission to promote interdisciplinary research and offer fresh perspectives to the scientific community. We extend our gratitude to the authors, reviewers, and editorial team whose efforts have made this issue possible. We believe that the studies included here will provide valuable contributions to the scientific world.

As we celebrate the release of our 10th issue, we look forward to continuing our journey of contributing to the advancement of science with many more publications.

We hope you enjoy reading this issue.

Sincerely,

**Editor of the Journal of Statistics and Applied Sciences**

Dr. Abdulkadir Keskin

# Long-run Relationship Between Türkiye and SHANGHAI Cooperation Organization Stock Markets: Cointegration Analysis

Sara Huseynova <sup>1\*</sup>, Mahammad Charkasov <sup>2</sup>

<sup>1</sup>Khazar University, School of Economics and Management, Baku, Azerbaijan; sarahuseynova@gmail.com

<sup>2</sup>Istanbul University, Institute of Social Sciences, Istanbul, Türkiye; mahammad.charkasov@ogr.iu.edu.tr

Orcid: 0000-0002-0516-9468<sup>1</sup> Orcid: 0000-0001-7066-7829<sup>2</sup>

\*Correspondence: sarahuseynova@gmail.com

**Abstract:** This study investigates the presence of long-run cointegration relationships between the Turkish Stock Market and the stock markets of the Shanghai Cooperation Organisation (SCO) member countries. In a period of stagnation in the EU membership negotiations, Türkiye's shift towards Eastern markets is noteworthy. Using a dataset covering the period from January 2014 to December 2023, this research applies various econometric analysis techniques such as Augmented Dickey-Fuller test, Johansen Cointegration Test and Granger Causality Test to examine the dynamic relationships between these markets. The findings from these analyses indicate that there is a significant long-run cointegration relationship between Borsa Istanbul and Moscow and Shanghai Stock Exchanges. This highlights the opportunities for strategic investment and portfolio diversification in these regions.  
**Keywords:** Cointegration, Causality, Shanghai Cooperation Organisation, International Equity Indices

## 1. Introduction

The collapse of the USSR and the transformation of the USA into the sole superpower in the world pushed Russia and China to be on the same side and act together. As a result of this unity, firstly, border problems were solved, the Shanghai Five was founded, and then the Shanghai Cooperation Organisation (SCO) was founded.

The five countries of the region - China, Russia, Kazakhstan, Kyrgyzstan and Tajikistan - came together to resolve border issues, held talks on economic, military, cultural and energy issues, and reached common consensus decisions [1].

The name of the Shanghai Five was changed with the invitation of Uzbekistan to the meeting held in Shanghai on 15 June 2001. With the joint declaration of the participating countries, it was announced that the Shanghai Cooperation Organisation (SCO) was established with the inclusion of Uzbekistan as a member of the organisation. The main objective of the SCO is to solve the problems of terrorism, separatist movements and extremism that threaten international and regional peace and security [2].

In the current process, the total number of members of the organisation has increased to 9 with the membership of India and Pakistan in 2017 and Iran in 2023. While Afghanistan, Mongolia and Bulgaria are observer countries in the organization, 14 countries, including Türkiye, have participated in the organization as dialogue partners [3].

With its expansion, the SCO has come to dominate almost 75 per cent of the Eurasian space. Considering the population of the member states, the SCO has approximately 45% of the world population. These two aspects are very important for the organization to have effective power in the international community [4].

Since joining the organization in 2012 as a dialogue partner, Türkiye's negotiations to become a member of the organization have become an agenda item today. At a time when the EU membership process has come to a standstill, Türkiye is expanding its cooperation with the SCO in different fields, especially in the economic field. Türkiye has put the development of relations with Asia, which is the focal point of



---

the new century, at the forefront of its foreign policy agenda due to the shift of capital from the West to the East after the economic crises in the West, and has turned its attention to the East and sought cooperation with the SCO, which it sees as an effective organisation in the region.

In this context, it is of great importance to examine the relations between Türkiye and the financial markets of the member countries of this organisation. Considering the lack of studies in this field in the literature, it is important to examine the long-run relationships between the markets of different countries in order to fill this gap. Investigating the co-movement or dependence between multiple stock markets in different countries is usually based on cointegration and causality analyses. Overall, this research is designed to contribute significantly to the existing literature on international stock market relations by focusing on a relatively less researched area - the linkages between the Turkish Stock Exchange and the stock exchanges of the SCO countries. Through its comprehensive analysis and practical implications, the study aims to be a valuable resource for both academic researchers and practitioners in the field of international finance and investment.

The study consists of four main headings. In the literature review section, a summary of important studies that are similar to the subject is given. In the next section, the data and methodology to be used are mentioned. Finally, the findings are evaluated and the results are discussed.

## **2. Literature Review**

There are various studies that have been published in the academic literature that investigate the interaction between stock markets by employing different methodologies. Korkmaz and İsmail (2008) used the Johansen cointegration test to examine the cointegration relationship between the Istanbul Stock Exchange and the stock exchanges of 12 developed and 22 developing nations. The analytical results indicate that the Turkish stock market is cointegrated with the stock markets of 7 developed countries and 5 developing countries. After identifying the cointegrated stock markets, portfolios were constructed according to different scenarios using the Markowitz mean-variance model. These findings suggest that a better understanding by Turkish portfolio managers of the extent to which developed and emerging stock markets are integrated would be effective in asset allocation and risk minimization [5]. Mukherjee and Bose (2008) examine whether the Indian equity market moves in tandem with other markets in Asia and the United States and whether foreign investors continue to be interested in this market. According to the results of this study using cointegration and causality tests, US indices do not affect the integration of Asian markets, but Japan plays an important role in the integration of Asian markets. In addition, returns on the Indian market significantly impact the returns on major Asian markets [6].

İbicioğlu and Kapusuzoğlu (2011) conducted a study using 1922 daily data from January 2002 to January 2010. The study aimed to analyze the relationship between the Istanbul Stock Exchange (ISE) and the stock exchanges of Mediterranean countries (France, Italy, Spain, Greece, Malta, and Croatia) that are part of the European Union. The Johansen cointegration test, Granger causality analysis, variance decomposition, and impulse-response analyses confirm the existence of a long-run relationship among the stock markets of the examined countries. Specifically, the French Stock Exchange is identified as a driving force that influences other stock exchanges. However, no stock exchange is found to have a causal relationship with ISE [7].

Assidenou (2011) tested the coherence of major stock markets in OECD, Pacific, and East Asian countries by analyzing the daily closing values of market indices from September 2 to August 31, 2009. As a result of the study, he found that Asian major indices are co-integrated and demonstrated that even while certain regional stock markets are not accessible to global investors, they are nonetheless influenced by fluctuations in global markets [8].

Bulut and Özdemir (2012) analysed the relationship between Istanbul Stock Exchange and Dow Jones Industrial using weekly closing prices from 05.01.2001 to 30.12.2010. The causality relationship between the series; Granger method and cointegration analyses were performed using Johansen and Vector Error Correction models. The results of the research indicate that DJI is the three-lag Granger cause of ISE. Based on the findings of the cointegration analysis, it can be concluded that the series exhibit a long-

---

term relationship and are considered to be cointegrated. The error correcting term works in the short run and has a significant effect on the ISE for three periods [9].

Benli (2014) analysed the long-run relationship between the Turkish stock market and the stock markets of developing countries. Emerging markets was categorised according to Morgan Stanley Capital International (MSCI). The long-run relationship between Türkiye and these countries was analyzed with the Johansen cointegration method. An examination of the enduring connection between Türkiye and these markets was conducted using the Johansen cointegration test, covering the time span from December 30, 1994, to September 30, 2013. The results indicate a significant long-run relationship between the stock markets of Türkiye and Colombia, as well as Türkiye and Mexico. However, there is no noteworthy relationship between Türkiye and other emerging market markets. Therefore, there is a opportunity to achieve portfolio diversification and arbitrage by investing in these markets, as they are not closely related to Türkiye in the long term [10].

Gözbaşı (2015) examined the relationship between the Istanbul Stock Exchange (ISE) and the stock markets of seven emerging economies, namely Argentina, Brazil, Mexico, India, Malaysia, Hungary, and Egypt. For this purpose, cointegration and causality analyses were conducted using weekly data for the period 1995 - 2008. The findings indicate a sustained relationship between the ISE and the stock markets of Brazil, India, and Egypt in the long term. Additionally, there is a short-term interaction between the ISE and the stock markets of these three nations, as well as the stock markets of Mexico and Hungary. Accordingly, the ISE is still a stock exchange that is independent from some emerging market stock markets and can offer international diversification opportunities to investors [11].

Ünal and İçigen (2020) investigate the co-integration relationships among the stock indices of the MIST countries (Mexico, Indonesia, South Korea, and Turkey) during the 2000–2016 period using Johansen's co-integration and Vector Error Correction Model (VECM) causality analyses. Their findings reveal a long-term equilibrium relationship among these economies' stock markets. Additionally, they highlight a one-way causality from Indonesia to South Korea and from Turkey to Mexico. These results indicate that the integration level among MIST economies limits portfolio diversification and arbitrage opportunities for investors in the long run. However, the dynamic nature of short-term causality relationships emphasizes the influence of varying timeframes on market interactions, suggesting a need for further exploration of these dynamics in future research [12].

Huseynova and Ganbarli (2024) analysed the trade and economic processes between the stock index indicators of the US, Japan and China based on data from 1995 to 2022. A multivariate regression model was created in the study, and its adequacy was assessed using the F-Fisher test and Student's t test. The stability of the parameters of the regression model was checked using the CUSUM test. As a result, econometrically sound proposals have been developed that allow dynamic analyses to effectively regulate economic processes, integration, international trade and foreign investment operations between the three countries. Subsequently, a VECM model was constructed to depict the long-term equilibrium relationship between the analyzed indicators. During the modelling, all essential statistical procedures were implemented to ascertain and appraise the model's parameters, as well as to verify its sufficiency and the precision of both short-term and long-term forecast values utilizing Eviews 8 tools. The findings demonstrate the significance of government monitoring using the vector error correction model approach for both the states examined in this study and for every country. This monitoring is crucial for ensuring efficient regulation of foreign commerce and for active participation in regional and global integration processes [13].

### **3. Data and Methodology**

This study encompasses the examination of 521 weeks of index values spanning from January 2014 to December 2023. Such a long observation period is chosen to enhance the reliability and applicability of the findings. The data for this research is sourced from investing.com, a widely used platform for financial data. Since all index values are expressed in the local currency of the respective country, all index values are converted into US dollars to harmonise all index values in a common currency. Within the scope of the study, all analyses will be performed via Eviews 13 software [14].

In the methodology section of this scientific article, a three-stage approach will be employed to analyze the data. Initially, the study will utilize the Augmented Dickey-Fuller test, to investigate the stationarity of the variables under consideration. The Augmented Dickey-Fuller test was developed by Said and Dickey (1984) and is a widely used test in time series analysis [15]. In this study, we follow the methods proposed by Gujarati and Porter (2009) for the application of the test [16]. This method is critical as it helps in determining whether the variables exhibit a time-dependent trend or are stable over time, which is essential for ensuring the validity of further statistical analysis.

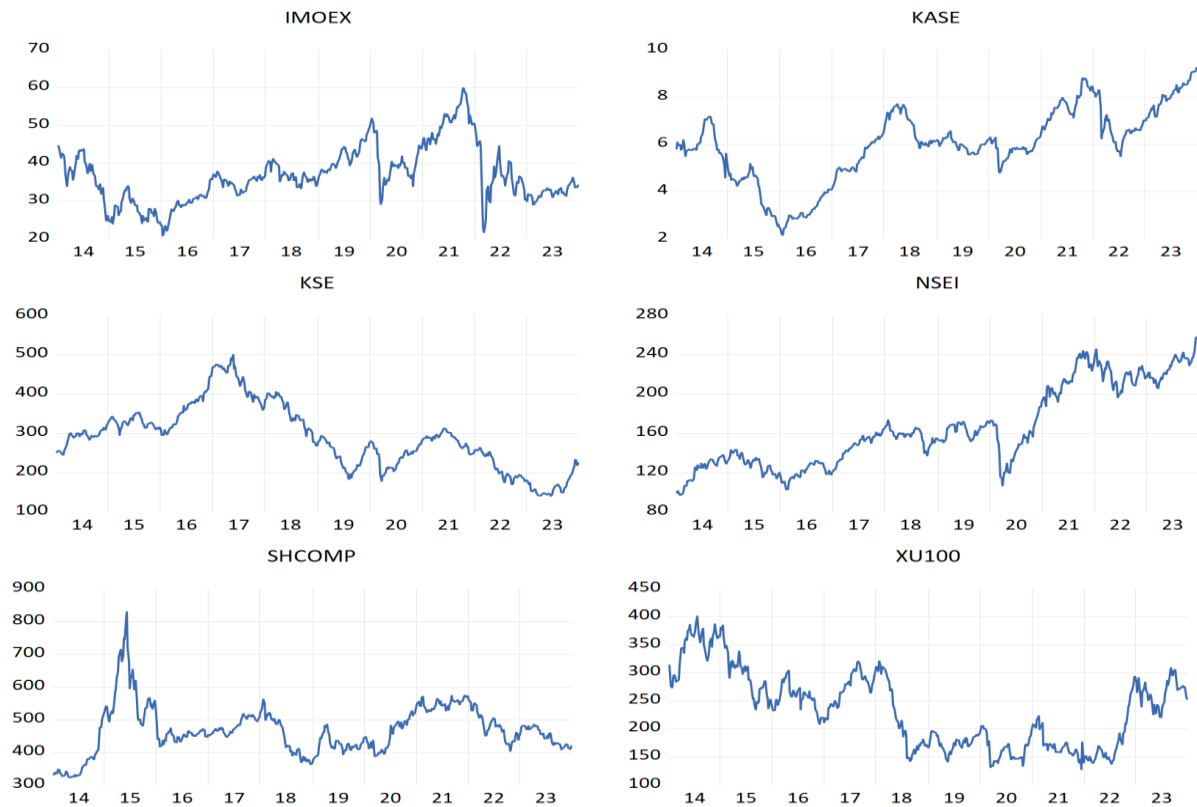
Following the stationarity analysis, the research will proceed to the second stage, where the Johansen Cointegration Test will be applied. This test, developed by Johansen (1988), is a powerful method for analyzing long-run equilibrium relationships [17]. Cointegration analysis is crucial for determining if variables exhibit a shared long-term movement, indicating a common stochastic tendency.

Finally, the third stage of the methodological framework involves the application of the Granger Causality Test [18]. This test is used to detect the direction of relationships between variables. It helps in identifying whether one variable can be used to predict another, thus providing insights into potential causal relationships. Unlike traditional causal analysis, the Granger Causality Test does not imply a cause-effect relationship in the strict sense but rather indicates predictability from one time series to another.

Within the scope of the study, the indices in Table 1 are used to represent stock market indices. And Figure 1 shows the graphs of the price levels of the country indices to be included in the research for the last 10 years.

**Table 1.** Stock Market Indices Used in the Study

Country	Stock Market	Index	Code
China	Shanghai Stock Exchange	Shanghai Composite	SHCOMP
India	National Stock Exchange of India	Nifty 50	NSEI
Kazakhstan	Kazakhstan Stock Exchange	KASE Index	KASE
Pakistan	Karachi Stock Exchange	KSE-100 Index	KSE
Russia	Moscow Stock Exchange	MOEX Russia	IMOEX
Türkiye	Borsa Istanbul	BIST 100	XU100



**Figure 1.** Graphs of Indices at the Level

Firstly, logarithmic series were created by taking the natural logarithms of the weekly closing prices of the stock market values expressed in US dollars and analyses were made over the logarithmic series. Table 2 displays the descriptive statistics for the logarithmic index series.

**Table 2.** Descriptive Statistics at Log Level

Series	LIMOEX	LKASE	LKSE	LNSEI	LSHCOMP	LXU100
Mean	3.584682	1.746857	5.631319	5.079089	6.137844	5.413889
Std. Dev.	0.198601	0.299346	0.292852	0.244732	0.154414	0.293710
Skewness	0.035603	-1.057209	-0.363772	0.175417	0.142087	-0.027498
Kurtosis	2.931438	3.687109	2.609556	1.924960	3.847025	1.771496
Jarque-Bera	0.212116	107.3018	14.80003	27.76050	17.32773	32.82836
Probability	0.899373	0.000000	0.000611	0.000001	0.000173	0.000000

The probability values of the J-B test statistics of the variables in the table give an idea about the normal distribution of the series. LIMOEX series is normally distributed, while other series are not.

**Table 3.** Correlation Coefficients among Series

	LIMOEX	LKASE	LKSE	LNSEI	LSHCOMP	LXU100
LIMOEX	1					
LKASE	0.6169	1				
LKSE	-0.0974	-0.4400	1			
LNSEI	0.4015	0.7502	-0.5375	1		
LSHCOMP	-0.0352	-0.0338	0.1914	0.2780	1	
LXU100	-0.4390	-0.2333	0.2998	-0.3859	-0.0759	1

Table 3 highlights that among the pairs analyzed, the stock markets of India and Kazakhstan exhibit the highest degree of correlation, standing at a significant 75.02%. This strong positive correlation indicates a close relationship between the two markets, suggesting that they often move in tandem.

On the other hand, the correlation table also reveals a markedly different relationship between the Pakistani and Indian stock markets. Here, the correlation is not only lower but distinctly negative, recorded at -53.75%. This negative correlation indicates that these two markets often move in opposite directions.

Mostly stationarity at level values is an exception. The results of analyses on non-stationary series may lead to misleading forecasts. For this reason, stationarity is very important and the series must be stationary before the analysis is performed [19].

Using the ADF test, the stationarity of the modeling variables was examined. The testing results indicate that both the original series and their first differences demonstrate stationarity. These results are presented in Table 4:

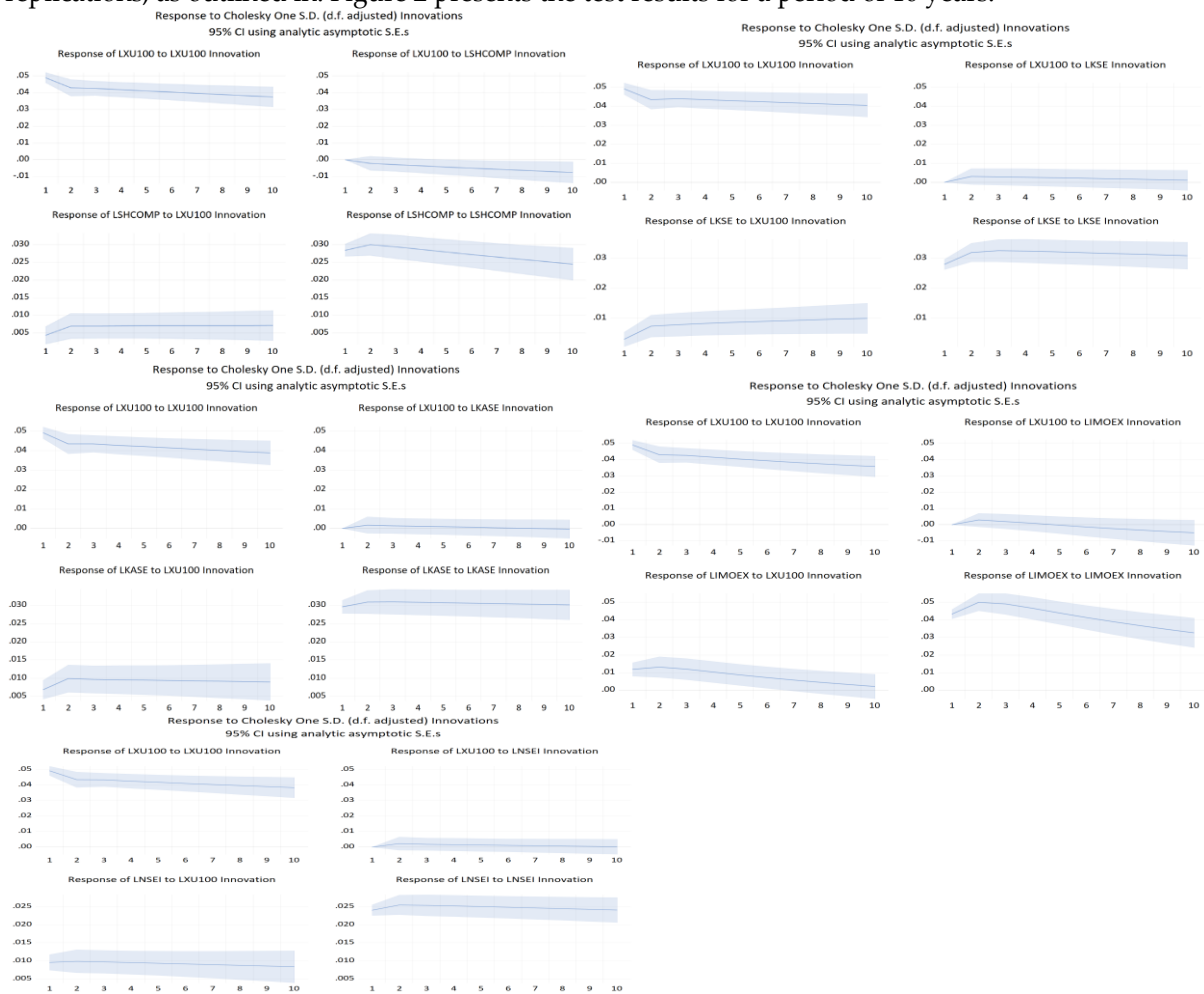
**Table 4.** ADF Test Results

Stock Market Indexes	ADF (Level)		ADF (1st Difference)	
	Intercept	Trend and Intercept	Intercept	Trend and Intercept
LIMOEX	-2.7177	-3.6103	-12.8888	-12.8789
LKASE	-0.8649	-2.0432	-14.3277	-14.3751
LKSE	-1.2348	-2.3457	-19.2398	-19.2410
LNSEI	-1.1774	-2.8472	-21.4096	-21.3890
LSHCOMP	-2.9803	-2.8737	-20.8832	-20.9111
LXU100	-1.8348	-2.2581	-25.3654	-25.3613
<b>Test Critical Values</b>				
%1	-3.4427	-3.9758	-3.4427	-3.9758
%5	-2.8669	-3.4185	-2.8669	-3.4185
%10	-2.5696	-3.1317	-2.5696	-3.1317

It is observed that all the index series applied ADF test contain unit root at level  $I(0)$  both in the model with constant term and in the model with constant term and trend, in other words, they are non-stationary. Non-stationary series become stationary when their first order differences are taken. Since

the stationarity level of the data is I(1), Johansen cointegration analysis can be performed between these data.

In order to provide comprehensive information for the research, it is important to conduct an analysis of the reaction of impulse functions. Impulse response functions, as proposed by Sims (1980), are widely utilized to analyze the dynamic responses of systems to shocks [20]. In this study, the methods recommended by Lütkepohl (2005) have been employed [21]. These functions provide the median estimate, along with a 90% confidence range, of the endogenous variable for each positive shock of one standard deviation of the exogenous variable. They also reflect the amount of time it takes to return to the equilibrium path. The confidence intervals were derived using the bootstrapping method with 100 replications, as outlined in. Figure 2 presents the test results for a period of 10 years.



**Figure 2.** Results of Impulse Response Function

Figure 2 clearly demonstrates that the variables' response to deviations from the main stochastic trend varies. When responding to shocks, the endogenous variable moves partially towards equilibrium.

In order to examine the impact of external factors on an internal factor in the coming decade, the econometric technique of decomposition of forecast error variances was employed. This method assesses the extent to which changes in a specific variable contribute to its own forecast error variance, as well as the variance of other variables. Upon analyzing the control findings of the relevant tests, it is evident that the yearly forecast LXU100, shocks, and LSHCOMP exhibit the most substantial inconsistencies. The error rates for these inconsistencies are 99.88% over a two-year period and 93.14% over a ten-year period, respectively. Yearly forecast LXU100, the largest errors fall on the shocks and LKSE, in the amounts of 99.78% over a two-year horizon and 93.91% over a ten-year horizon; yearly forecast LXU100, the largest errors fall on the shocks and LKASE, in the amounts of 99.93% over a two-year horizon and 91.72% over a ten-year horizon; yearly forecast LXU100, the largest errors fall on the shocks and LIMOEX, in the amounts of 99.81% over a two-year horizon and 95.88% over a ten-year

horizon; yearly forecast LXU100, the largest errors fall on the shocks and LNSEI, in the amounts of 99.87 % over a two-year horizon and 88.87% over a ten-year horizon.

**Table 7.** Results of the Granger Causality Test

Null Hypothesis	F-Statistic	Prob.
DLIMOEX → DLXU100	1.30811	0.2712
DLXU100 → DLIMOEX	0.08349	0.9199
DLKASE → DLXU100	1.81712	0.1635
DLXU100 → DLKASE	2.01294	0.1346
DLKSE → DLXU100	0.89036	0.4111
DLXU100 → DLKSE	5.67884	0.0036*
DLNSEI → DLXU100	0.57593	0.5625
DLXU100 → DLNSEI	1.63526	0.1959
DLSHCOMP → DLXU100	0.92311	0.3979
DLXU100 → DLSHCOMP	2.12942	0.1200

\*indicates the significance at %1

In the Granger causality test, first order differences of logarithmic series were taken. The outputs of the causality test can be found in Table 7. In these tests, which are considered as bivariate model, the null hypothesis is "There is no Granger causality from the independent variable to the dependent variable.". Based on the data in the table, it can be concluded that the null hypothesis between the Borsa Istanbul and other markets, excluding the Karachi Stock Exchange, is rejected. It can be observed that there is a unidirectional causality relationship from Borsa Istanbul to Karachi Stock Exchange.

### 3.1. Cointegration Analysis

In this phase of the research, we will report the findings of the Johansen Cointegration Test regarding the presence of a long-term relationship among Borsa Istanbul and the stock exchanges of Shanghai Cooperation Organization member countries.

Before proceeding with this test, it should be determined with how many week lags the interaction between the series emerges. The optimal lag lengths were determined using VAR analysis. The optimal lag lengths obtained according to the Akaike, Schwarz and Hannan-Quinn information criteria are presented in Table 5.

**Table 5.** Optimal Lag Lengths of the Series

	Lag Length	AIC	SC	HQ
LXU100 - LIMOEX	2	-6.603596*	-6.520940*	-6.571198*
LXU100 - LKASE	2	-7.368301*	-7.285644	-7.335902*
LXU100 - LKSE	2	-7.476937*	-7.394280*	-7.444538*
LXU100 - LNSEI	2	-7.775153*	-7.692497	-7.742754
LXU100 - LSHCOMP	2	-7.448794*	-7.366138	-7.416396*

\* denotes lag order chosen by the relevant criterion

Among the lag lengths, the one that is most validated by the information criteria is selected. According to the results obtained, the two-week lag length is determined as the most appropriate between the XU100 index and all other indices.

The table below presents the results of the cointegration test including the specified lag lengths.

**Table 6.** Results of the Johansen Cointegration Test

Indices	Hypotheses	Trace Stat.	Critical Value	P-value	Max-Eigen Stat.	Critical Value	P-value
LXU100	- r=0	20.50751	15.49471	0.0081	18.47027	14.26460	0.0102
LIMOEX	r≤1	2.037240	3.841465	0.1535	2.037240	3.841465	0.1535
LXU100	- r=0	5.699265	15.49471	0.7307	5.144069	14.26460	0.7235
LKASE	r≤1	0.555196	3.841465	0.4562	0.555196	3.841465	0.4562
LXU100	- r=0	9.013486	15.49471	0.3641	6.000739	14.26460	0.6131
LKSE	r≤1	3.012746	3.841465	0.0826	3.012746	3.841465	0.0826
LXU100	- r=0	6.533325	15.49471	0.6324	4.796692	14.26460	0.7673
LNSEI	r≤1	1.736633	3.841465	0.1876	1.736633	3.841465	0.1876
	r=0	16.56916	15.49471	0.0343	11.13770	14.26460	0.1475

LXU100	-	r≤1	5.431467	3.841465	0.0198	5.431467	3.841465	0.0198
LSHCOMP								

In the literature, it is accepted that there is a cointegration relationship when the critical value probabilities of any of the Trace and Maximum Eigenvalue statistics is below 0.05. The results indicate a significant cointegration relationship between the Borsa Istanbul and Moscow Stock Exchange, as confirmed by both the Trace test statistic and the Max-Eigen value test statistic. A significant cointegration relationship was found between Borsa Istanbul and Shanghai Stock Exchange only according to the Trace test statistics.

#### 4. Conclusions

The primary aim of this research is to examine whether cointegration relationships exist between the Turkish Stock Exchange and the major stock exchanges operating in the member countries of the Shanghai Cooperation Organization (SCO). This study is particularly important given the increasing economic bilateral relations among emerging markets and the growing importance of the SCO in global economic dynamics. By examining these cointegration relationships, the study aims to uncover any long-run equilibrium linkages that may exist between these disparate but economically important markets.

The results of the Johansen Cointegration Test confirm the existence of long-term equilibrium relationships between BIST and the Moscow and Shanghai stock exchanges. The Granger Causality Test results indicate a unidirectional causality relationship between Borsa Istanbul and the Karachi Stock Exchange, suggesting that changes in BIST can predict changes in the Karachi market.

There was no statistically significant long-term relationship observed between the Borsa Istanbul and stock markets in other countries at a 5% probability level. This situation can be considered as a favourable situation for international investors. In particular, international portfolio diversification is possible due to the lack of a long-term relationship among stock exchanges.

These results align with previous studies, such as Benli (2014) and Gözbaşı (2015), which emphasized the diversification opportunities presented by the weak integration of the Turkish stock market with other emerging markets. In this context, the findings contribute to the growing body of literature on international finance by highlighting the potential of Turkey's financial market as an independent component in global investment strategies.

More research should be conducted on Türkiye's financial relationships with other SCO member countries to identify potential long-term cointegration relationships. Policies should be developed to enhance economic cooperation between Türkiye and SCO countries, promoting financial market integration and improving the investment climate. Future studies should use broader datasets and diverse analytical methods to increase the reliability of the findings, with a detailed examination of mutual relationships with other SCO member stock markets.

**Author Contributions:** "Conceptualization, S.H. and M.C.; methodology, S.H. and M.C.; software, S.H.; validation, M.C.; formal analysis, writing - revising and editing, S.H. and M.C.

**Funding:** This research was not financially supported by external sources.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

- [1] A. Melville and T. Shackleina, *Russian foreign policy in transition: Concepts and realities*, Budapest: Central European University Press, 2005.
- [2] A. Frost, "The Collective Security Treaty Organization, the Shanghai Cooperation," *In China & Eurasia Forum Quarterly*, vol. 7, no. 3, pp. 83-102, 2009.
- [3] Shanghai Cooperation Organization, [Online]. Available: <https://eng.sectsco.org/20170109/192193.html>. [Accessed 21 03 2024].
- [4] S. Aydoğan, "Rusya ve Çin'i Şanghay İşbirliği Örgütü'nün Oluşumuna Götüren Sebepler," *Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 22, no. 3, pp. 135-149, 2021.

- 
- [5] T. Korkmaz and E. İ. Çevik, "Türkiye ve uluslararası hisse senedi piyasaları arasındaki eşbütünleşme ilişkisi ve portföy tercihleri," *BDDK Bankacılık ve Finansal Piyasalar Dergisi*, vol. 2, no. 1, pp. 59-84, 2008.
- [6] P. Mukherjee and B. Suchsmita, "Does the stock market in India move with Asia?: A multivariate cointegration-vector autoregression approach," *Emerging Markets Finance and Trade*, vol. 44, no. 5, pp. 5-22, 2008.
- [7] M. Ibicioğlu and K. Ayhan, "An Empirical Analysis on the Integration of the Stock Exchanges of the Ise with those of European Union Mediterranean Countries," *Anadolu University Journal of Social Sciences*, vol. 11, no. 3, pp. 85-102, 2011.
- [8] K. E. Assidenou, "Cointegration of major stock market indices during the 2008 global financial distress," *International Journal of Economics and Finance*, vol. 3, no. 2, pp. 212-222, 2011.
- [9] Ş. Bulut and Ö. Abdullah, "İstanbul Menkul Kıymetler Borsası ve "Dow Jones Industrial" Arasındaki İlişki: Eşbütünleşme Analizi," *Yönetim ve Ekonomi Dergisi*, vol. 19, no. 1, pp. 211-224, 2012.
- [10] Y. K. Benli, "Türkiye Borsasının Gelişmekte Olan Ülkeler Borsaları ile Eşbütünleşme Analizi," *Journal of Management and Economics Research*, vol. 12, no. 23, pp. 18-32, 2014.
- [11] O. Gözbaşı, "İMKB ile Gelişmekte Olan Ülkelerin Hisse Senedi Piyasalarının Etkileşimi: Eşbütünleşme ve Nedensellik Yaklaşımı," *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 35, pp. 99-118, 2015.
- [12] S. Ünal and F. Köse, "MIST ülkelerinin borsa endeksleri arasındaki eşbütünleşme ilişkisi," *Balkan Sosyal Bilimler Dergisi*, vol. 9, no. 17, pp. 137-145, 2020.
- [13] S. Huseynova and U. Qambarli, "Cointegration and causality relationship of the USA stock market with selected world markets," *Business inform*, vol. 12, pp. 128-134, 2024.
- [14] "Eviews 13," [Online]. Available: <https://eviews.com/home.html>. [Accessed 24 04 2024].
- [15] S. E. Said and D. A. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, vol. 71, no. 3, pp. 599-607, 1984.
- [16] D. N. Gujarati and D. C. Porter, *Basic econometrics*, McGraw-hill, 2009.
- [17] S. Johansen, "Statistical analysis of cointegration vectors," *Journal of economic dynamics and control*, vol. 12, no. 2-3, pp. 231-254, 1988.
- [18] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424-438, 1969.
- [19] A. Keskin and B. Kara, "Türkiye'De Ticari Açıklık, dış ticaret vergi Gelirleri Ve GSYH Arasındaki ilişki: 1993-2019 dönemi için ARDL Sınır Testi yaklaşımı," *Turkish Studies-Economics, Finance, Politics*, vol. Volume 16 Issue 4, no. Volume 16 Issue 4, pp. 1297-1315, 2021.
- [20] C. Sims, "Macroeconomics and reality," *Econometrica*, vol. 48, no. 1, pp. 1-48, 1980.
- [21] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Heidelberg: Springer-Verlag, 2005.



Research Article

Received: date:04.06.2024  
Accepted: date:19.09.2024  
Published: date:31.12.2024

# A New Approach for the Measurement of Criterion Weights in the Scope of Multi-Criteria Decision Making: Spearman Rank Correlation-based Expanded CRITIC Method (SRCBECM)

Furkan Fahri Altıntaş<sup>1\*</sup>

<sup>1</sup>Aydın Provincial Gendarmerie Command, Aydın, Türkiye; furkanfahrialtintas@yahoo.com

Orcid: 0000-0002-0161-5862<sup>1</sup>

\*Correspondence: furkanfahrialtintas@yahoo.com

**Abstract:** Advancing the field of multi-criteria decision making (MCDM), this study proposes the Spearman Rank Correlation-based Expanded CRITIC Method (SRCBECM) as a novel and objective method for computing criteria weight coefficients. Leveraging the intricate revised CRITIC method, SRCBECM aims to enrich and contribute the MCDM landscape. Drawing upon criterion values extracted from Freedom in the world (FIW) index assessments for 19 G20 member nations, the study showcases the sensitivity of SRCBECM in objectively deriving criteria weights for diverse contexts. Further bolstering its credibility and reliability, comparative analyses reveal MIEXCF's alignment with established methodologies such as ENTROPY, CRITIC, SD, SVP, LOPCOW, and MEREC. Notably, the simulation analysis underscores SRCBECM's exceptional and stability in discerning criteria weights and its remarkable stability across diverse scenarios. In conclusion, SRCBECM emerges as a robust and objective criterion weighting technique, poised to make significant contributions to the burgeoning field of the broader MCDM corpus.

**Keywords:** MCDM, CRITIC, rho, SRCBECM

## 1. Introduction

Multi-criteria decision-making (MCDM) is a widely utilized methodology in intricate decision-making processes and mathematical modeling, often encompassing various factors. MCDM aims to assess and prioritize alternatives based on the preferences and priorities of decision-makers. To achieve this, it is crucial to establish the weights of the criteria, which reflect the decision-makers' preferences [1].

The field of MCDM provides a diverse range of techniques, including ENTROPY, CRITIC, SD, SVP, MEREC, and LOPCOW, for the computation of criterion weights. These methods leverage two fundamental characteristics of objective criterion weights: 1) the extent of performance contrast among decision alternatives for each criterion, reflecting the range between maximum and minimum values, and the distinctiveness or conflict among criteria. By comprehending and utilizing these inherent data characteristics, decision-makers can obtain valuable insights to steer their decision-making process [2]. Consequently, in the literature on MCDM criterion weighting, the logic of criterion weighting may vary in methods that consider the relationships between criteria (CRITIC, DEMATEL, MEREC).

One of the most notable features of the CRITIC method is that it is fundamentally based on the Pearson correlation coefficient between criteria [2]. The Pearson correlation coefficient is a parametric method, and it can only measure the correlation value between variables that exhibit a normal distribution. Therefore, the correlation values between variables that do not exhibit a normal distribution may not yield accurate results with this method [3]. In contrast, the relationships between variables that do not exhibit a normal distribution can be measured using the non-parametric Spearman Rank Correlation (rho) coefficient [4]. Moreover, the rho coefficient can also be used to calculate relationships between

variables that exhibit a normal distribution [5]. Hence, it can be considered that the rho correlation coefficient is more advantageous compared to the Pearson correlation coefficient. The primary motivation of this study is to propose the use of the rho correlation coefficient instead of the Pearson correlation coefficient for determining the relationships between criteria when calculating the weight values of criteria within the CRITIC method, regardless of whether the criteria values exhibit a normal distribution. This approach ensures a more accurate calculation of the relationships between criteria. The second motivation of this study is to expand the criteria weighting calculation logic of the CRITIC method. The logic of CRITIC method relies on the differentiation or antagonism among criteria. This is because in the CRITIC method, the weight of a criterion increases with the growth of negative relationships among other criteria (along with a decrease in positive relationships) [2]. In contrast, the DEMATEL method considers not only the antagonism among criteria but also takes into account the similarity among them based on the logic of the method [6]. If a criterion influences other criteria both negatively and positively, and is influenced by other criteria, its weight increases in the DEMATEL method [7]. Therefore, in the DEMATEL method, the separation and similarity among criteria are considered on equal terms when calculating the weight of a criterion. Accordingly, the DEMATEL method possesses a more inclusive characteristic in terms of criterion weighting logic compared to the CRITIC method. Therefore, to make the CRITIC method more comprehensive (by considering both the contrasts and similarities between criteria), it is necessary to implement its calculation logic similar to the DEMATEL method. In this context, in the research, a method called Spearman Rank Correlation-based Expanded CRITIC Method (SRCBECM) is proposed, which forms equations that explain relationships among criteria more meaningfully, provide more separation and similarity among criteria, and thus, more effectively determine the characteristic features of criteria compared to the CRITIC method. The SRCBECM method aims to measure the weight coefficients of criteria objectively.

The data set of the study consists of the values of 12 criteria for the Freedom in the world (FIW) of 19 countries in the G20 group. In this sense, the methods of calculating criterion weights in the scope of MCDM, rho and CRITIC techniques are explained in the method section of the study. The implementation plan of the proposed method is shown in Figure 1.

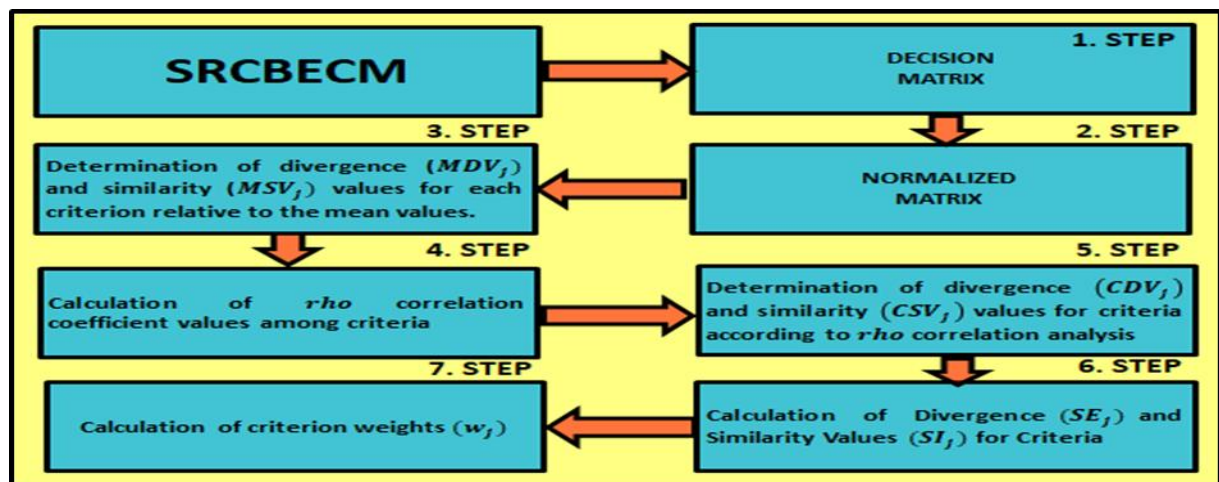


Figure 1. Implementation Plan of Proposed Method

## 2. Materials and Methods

### 2.1. Methods for calculating criterion weights in the scope of MCDM

Navigating the intricacies of decision-making often hinges on identifying the optimal choice amidst a tapestry of alternatives. However, each option may exhibit varying degrees of effectiveness across diverse criteria. This underscores the criticality of pinpointing the relative importance of these criteria to accurately compare the performance of potential solutions and ultimately arrive at the most suitable one [8]. Traditionally, this significance is quantified through the assignment of weight coefficients within the framework of MCDM problems [2].

---

The burgeoning field of Multi-Criteria Decision Making (MCDM) boasts a rich tapestry of objective weighting methods. These include CRITIC (Criteria Importance Through Inter-Criteria Correlation), ENTROPY, CILOS (Criterion Impact Loss Optimization System), IDOCRIW (Integrated Determination of Objective Criteria Weights), SVP (Statistical Variance Procedure), SD (Standard Deviation), MEREC (Method Based on Removal Effects of Criteria), LOPCOW (Logarithmic Percentage Change-driven Objective Weighting), and SECA (Simultaneous Evaluation of Criteria and Alternatives) [2].

The CRITIC method taps into the inherent information within a system by privileging criteria that exhibit greater disorder or distinctiveness compared to others, marking their heightened importance. This approach values the interconnectedness of criteria by meticulously examining their correlations to expose any inconsistencies. These contradictions, meticulously quantified using the standard deviation, inform the determination of criterion weight coefficients. The CRITIC method unfolds by first constructing a decision matrix and then normalizing its values. By analyzing the correlations between these normalized values, the method quantifies the relative weight of each criterion [9-12].

The ENTROPY method adds a valuable tool to the decision-making field. Building on the constructed decision matrix, this method leverages the standardized values and the calculated ENTROPY measure of each criterion to derive their corresponding ENTROPY-based weights [8-13].

The CILOS method prioritizes criteria based on their ability to influence the performance of other criteria relative to their ideal maximum and minimum values. In essence, criteria that cause the least deviation in others receive higher weight coefficients. This approach follows a structured process, beginning with the calculation of the decision matrix, normalization, and square matrix formation. Subsequently, a system of linear equations is solved to uncover the weight coefficients for each criterion [14-15].

The IDOCRIW method bridges the gap between the ENTROPY and CILOS approaches, forging a hybrid path to criterion weight determination. This innovative method delves into the relative impact of a missing index, initially leveraging the decision matrix and both ENTROPY and CILOS methodologies to calculate individual weight coefficients. Subsequently, it seamlessly integrates the ENTROPY and CILOS weights, culminating in the final IDOCRIW weights [14].

The SVP method, a champion of objectivity in weighting criteria, stands tall amidst subjective biases. This method calculates criterion weights with an unwavering impartiality, eliminating the influence of expert opinions or subjective interpretations. SVP dives deep into the variance metrics associated with each criterion, treating them as the sole compass for determining their significance. Following the calculation of individual criterion variances, the method simply divides each value by the total variance across all criteria, yielding their respective weights. In essence, SVP empowers data-driven decision-making, transforming variance into a quantifiable measure of criterion importance [16-18].

The SD method shines a light on the inherent variability of criteria by leveraging their deviation from the average. This straightforward approach first normalizes the values within the decision matrix, ensuring a level playing field. Then, it delves into the standard deviations of each criterion, using these measures as a compass to navigate their relative importance. In essence, SD empowers data-driven decision-making by translating variability into quantifiable weights [19-20].

Like other weighting methods, MEREC lays the groundwork by constructing the decision matrix and its normalized counterpart. Then, it ventures into the realm of performance, calculating the overall effectiveness of each decision alternative through a natural logarithm-powered framework. Building on this foundation, MEREC re-evaluates the performance of each alternative, factoring in the influence of each remaining criterion. This iterative process, driven by natural logarithms, ultimately culminates in the calculation of criterion weight coefficients. These weights reflect the "removal effect" of each criterion, essentially the sum of their absolute influence on the performance of other alternatives. In essence, MEREC recognizes that as a criterion's impact on decision alternatives grows, so too does its weight coefficient [21-22].

The LOPCOW method harnesses the power of multidimensional data, weaving together information from various sources to craft a tapestry of optimal criterion weights. This approach seeks to level the playing field between the most and least influential criteria, while acknowledging the intricate web of connections that bind them. The journey begins with meticulously constructing the decision matrix, followed by a

rigorous normalization process that ensures all voices are heard equally. To overcome the disparities arising from data magnitudes, LOPCOW deploys a potent metric: the average square value expressed as a percentage of the criterion's standard deviation. This measure serves as a compass, guiding us towards the final weight coefficients and ultimately, a balanced and insightful weighting scheme [23].

The SECA method takes a holistic approach, simultaneously unveiling the true potential of decision alternatives and weighing the importance of criteria that shape their performance. This innovative method begins by leveling the playing field through decision matrix standardization. Then, it delves into the realm of disagreement, employing the standard deviation as its tool to quantify discrepancies. These insights, along with standardization values, form the bedrock for determining criterion weights. Finally, SECA leverages the power of multi-objective linear programming, optimizing a model to arrive at the optimal set of weights [24].

The decision matrix in the DEMATEL method is created by obtaining the opinions of experts, and thus, the method is recognized as one of the subjective criterion weighting methods. This is because it involves determining the impact values of variables on each other through the input of experts or their opinions [2]. The relationship structure of variables, including their contributions to the relational structure, influence, relational density, and relational quality (as influencers or influenced), can be determined through the DEMATEL MCDM method [25-26]. Therefore, in the DEMATEL method, variables that possess the quality of being an "influencer" within the relationship structure are considered as causes, while variables that have the quality of being "influenced" are considered as outcomes in said relational structure [23]. The DEMATEL method has demonstrated successful applications in various technical and social problems. In addition to identifying the relationship structure between variables, the DEMATEL method allows for the calculation of the significance values of variables [10]. In the DEMATEL method, for the preparation of the decision matrix, the impact values of variables on each other can be determined subjectively by obtaining the opinion of an expert or the opinions of multiple experts [2]. Apart from this, In the DEMATEL method, the weight of a criterion increases as its impact on other criteria, either positively or negatively, and its susceptibility to the influence of other criteria become more significant. Additionally, one of the most crucial features of the DEMATEL method is that in determining the weights of criteria, the interactive structure among criteria considers both positive and negative influences equally. This is because the weights of criteria can be calculated by taking the square root of the sum of the squares of the total of positive and negative impact values and the square of the difference between positive and negative impact values [27].

## 2.2. Spearman Rank Correlation Coefficient

The Spearman Rank Correlation Coefficient ( $\rho$ ) is a non-parametric correlation coefficient and is referred to as an alternative to the Pearson correlation coefficient. In this sense, the rho correlation coefficient is utilized to measure the linear relationship between two variables that do not exhibit a normal distribution feature [28-29]. Additionally, when variables possess the normal distribution feature, the relationship between the two variables can also be assessed using the rho coefficient. A coefficient value of -1 indicates a perfect negative relationship, whereas a value of +1 signifies a perfect positive relationship. If the coefficient value is 0, it implies no relationship between the two variables. The rho coefficient also possesses a symmetrical property; therefore, the relationship coefficient remains the same even if the roles of the variables change between dependent and independent variables. To measure the rho relationship coefficient between two variables, the following steps are followed [30-39].

**Step 1:** Calculation of the total values of row entries in the contingency table ( $t'$ )

Let  $t$  represent the sum of each row. Therefore, the sum of the row values in the contingency table is shown in Equation 1.

$$t' = \frac{(\sum t^3 - \sum t)}{12} \quad (1)$$

**Step 2:** Calculate the sum of the column values in the contingency table ( $u'$ )

Let  $u$  represent the sum of each column. Therefore, the sum of the column values in the contingency table is shown in Equation 2.

$$u' = \frac{(\sum u^3 - \sum u)}{12} \quad (2)$$

**Step 3:** Calculation of the rho ( $r_{rho}$ )

$$r_{rho} = \frac{n^3 - n - 6 \sum d^2 - 6(t' + u')}{\sqrt{n^3 - n - 12t'} \sqrt{n^3 - n - 12u'}} \quad (3)$$

When the correlation coefficient literature is examined, it is possible to come across many studies that measure the relationships between two variables with the *rho* coefficient. Studies related to the current *rho* correlation coefficient are shown in Table 1.

**Table 1.** *rho* literature

Author(s)	Method(s)	Theme
[40]	rho	Relationship between atmospheric stability and cloud tops temperature of Himawari-8 IR satellite images
[41]	rho	Relationship between communication skills and emotional intelligence among nurses
[42]	rho	Relationships between various cyber threats and their patterns with respect to duration, IP address, target systems/ports
[43]	rho	Relationship between hemoglobin and lactate dehydrogenase
[44]	rho	Relationship between Dysphonia and voice fatigue
[45]	rho	Relationship between aortopulmonary collaterals and common non-invasive clinical variables
[46]	rho	Relationship between bilateral dFIWcit and maximal sprint speed judo test, maximal aerobic speed Judo test and special judo fitness test
[47]	rho	Relationship between performance of magnetic resonance imaging and histopathology
[48]	rho	Relationship between abundance of short-chain fatty acids (SCFAs) and lipopolysaccharide
[49]	rho	Relationship between neutrophil-to-lymphocyte ratio and alzheimer-related biomarkers in cerebrospinal fluid

### 2.3. CRITIC method

The CRITIC (Criteria importance through inter criteria correlation) method is a technique that objectively measures the weight coefficients or importance levels of criteria based on the data of decision alternatives regarding the criteria [19]. The foundation of the method is based on the intensity of the dichotomy within the structure of the decision-making problem [2]. Furthermore, this method provides an analytical approach to revealing all the information inherent within the criteria [50]. The most significant feature that sets the CRITIC method apart from other weight coefficient calculation techniques is that it calculates the weight coefficients of criteria not based on subjective results provided by expert opinions, but by taking into account standard deviation and correlation analysis of the criteria [2]. In this context, the application steps of the method are explained below [50].

$A_i$ : Decision alternative  $i$

$C_j$ :  $J$  - th evaluation criterion

$x_{ij}$ : The value of alternative  $i$  according to evaluation criterion  $j$ .

$x_j^{mak}$ : The maximum value of decision alternatives according to criterion  $j$ .

$x_j^{min}$ : The minimum value of decision alternatives according to criterion  $j$ .

$r_{ij}$ : The value received by alternative  $i$  according to evaluation criterion  $j$ .

$p_{jk}$ : Relationship coefficients between any  $j$  criterion and  $k$  criterion

$\sigma_j$ : Standard deviation value of criterion  $j$  ( $j = 1, 2, \dots, n$ )

$w_j$ : Weight of evaluation criterion ( $j = 1, 2, \dots, n$ ).

**Step 1:** Acquisition of the Decision Matrix ( $X$ )

$$X = \begin{matrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{matrix} \begin{bmatrix} x_{11} & x_{12} & x_{1n} \\ x_{21} & x_{22} & x_{2n} \\ \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & x_{mn} \end{bmatrix} \quad (4)$$

**2. Step:** Normalization Process of the Decision Matrix.

For Benefit-Oriented Criteria

$$r_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \dots \dots \dots j = 1, 2, \dots, n \quad (5)$$

For Cost-Oriented Criteria

$$r_{ij} = \frac{x_j^{\max} - x_{ij}}{x_j^{\max} - x_j^{\min}} \dots \dots \dots j = 1, 2, \dots, n \quad (6)$$

**3. Step:** Creation of the Relationship Coefficient Matrix ( $p_{jk}$ )

$$p_{jk} = \frac{\sum_{i=1}^m (r_{ij} - \bar{r}_j) \cdot (r_{ik} - \bar{r}_k)}{\sqrt{\sum_{i=1}^m (r_{ij} - \bar{r}_j)^2 \cdot (r_{ik} - \bar{r}_k)^2}} \quad j, k = 1, 2, \dots, n \quad (7)$$

**4. Step:** Measurement of  $C_j$  Values

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^m (r_{ij} - \bar{r}_j)^2}{m-1}} \quad (8)$$

$$C_j = \sigma_j \cdot \sum_{k=1}^n (1 - p_{jk}), \quad j = 1, 2, \dots, n \quad (9)$$

**5. Step:** Measurement of Criterion Weights (Importance Degrees) ( $w_j$ )

$$w_j = \frac{C_j}{\sum_{k=1}^n C_k} \quad (10)$$

When the MCDM literature is examined, it is observed that many researchers have benefited from the CRITIC method in calculating the weights of criteria according to decision alternatives. The current research related to the CRITIC method is explained Table 2.

**Table 2.** CRITIC literature

Author(s)	Method(s)	Theme
[51]	CRITIC based MARCOS	Evaluation of zero-carbon measures for sustainable transportation in smart cities
[52]	CRITIC based TOPSIS	Optimum Site Selection for Solar PV Farm
[53]	CRITIC based TOPSIS	Analytically Identify the Air Pollutant's
[54]	CRITIC based GRA	investment portfolio selection
[55]	CRITIC based TOPSIS	Solving the material handling equipment selection problem
[56]	Fuzzy CRITIC based TOPSIS	Smartphone addiction assessment
[57]	CRITIC based EDAS	Geometric aggregation operator
[58]	CRITIC based TOPSIS	Stakeholder assessment in construction projects
[59]	CRITIC based MultiMOORA	Warehouse manager selection
[60]	CRITIC-SD based GRA and TOPSIS	Assessing the energy security of European Union countries

#### 2.4. Proposed method: Spearman Rank Correlation Coefficient-based expanded CRITIC method (SRCBECM)

The Pearson correlation coefficient is a parametric measure, so it can be used to measure the relationships between criteria under the assumption of normal distribution of the data. However, when the relationship between criteria in a dataset that does not follow the normal distribution is measured with the Pearson correlation coefficient, the relationship between the criteria may not reflect reality [4]. In contrast, the rho coefficient is a non-parametric measure, so the normal distribution assumption is not required for measuring the relationship between criteria [35]. In addition, the rho coefficient can be used to obtain real results in detecting nonlinear relationships between variables [61]. In this context, the rho coefficient is considered to be more advantageous than the Pearson correlation coefficient for calculating the weights of criteria due to its aforementioned properties [62]. Therefore, the relationships between criteria can be calculated with the rho coefficient in the CRITIC method.

In the ENTROPY method, as the uncertainty level of a criterion increases relative to other criteria, the weight coefficient of the criterion decreases. Consequently, the criterion with the highest degree of uncertainty attains a more pronounced significance compared to other criteria [2]. In the MEREC method, the weight of a specific criterion increases when the absolute difference between the averages of criteria concerning decision alternatives, either by excluding or considering the criterion, becomes smaller [21]. As a result, the criterion with the highest weight exerts the greatest influence on decision alternatives. In the SD method, the weight of a criterion is maximized when its standard deviation is the highest [19]. Similarly, in the SVP method, the weight of a criterion is maximized when its variance is the highest [20]. In the LOPCOW method, the mean square value of each criterion is calculated as a percentage of the standard deviations, effectively addressing the discrepancy (gap) arising from the dimensionality of the data. Consequently, a decrease in the standard deviation of a criterion diminishes the gap attributed to the data size for that criterion, resulting in an increase in the weight assigned to the criterion [23]. Therefore, in these methods, the weight calculation logic is based on the degree of separation (contrast) of the criteria from each other; the more a criterion is separated or contrasted, the higher its weight value.

As is well known, the increase in the weight coefficient of a criterion in the CRITIC method depends on the intensity of its contrast with other criteria. Therefore, if the contrast intensity of a criterion with other criteria increases, the weight value of the criterion also increases [2]. This is because, according to Equation 9 in the CRITIC method, the importance and weight of a criterion is highest if its negative relationship with other criteria is the greatest and its positive relationship is the least [19]. Therefore, if a criterion is the most distinct from other criteria, the weight of that criterion is greater. In conclusion, the method's logic takes into account the degree of separation of the criteria from each other.

In addition to Equation 9, the CRITIC method also takes into account the standard deviation values of the criteria. In the CRITIC method, the standard deviation only provides information on how far the average value of any criterion's data set is from the mean or how different the criterion's data are from each other. Therefore, when the standard deviation values of any criterion are calculated, the standard deviation values of other criteria are not taken into account. Thus, the standard deviation value in Equation 8 does not support the separation (contrast) logic of the CRITIC method. This is because the standard deviation value shown in Equation 8 shows how far each criterion deviates from the mean based on its own data. In this case, the separation (contrast) between criteria that constitute the logic of the CRITIC method is supported excluded standard deviation only by the process in Equation 9 ( $\sum_{k=1}^n(1 - p_j)$ ).

To increase the contrast of a criterion with another criterion in the CRITIC method, the standard deviation shown in Equation 9 ( $\sigma_j \cdot \sum_{k=1}^n(1 - p_j)$ ) can be replaced with the absolute difference between the arithmetic mean of the normalization values of any criterion and the arithmetic mean value of all criteria (MDV), which shows how much the mean value of the criterion deviates from the arithmetic mean values of all criteria. This will strengthen the degree of separation (contrast) between criteria by multiplying the calculated deviation value (MDV) with the Equation 9 ( $MDV_j \cdot \sum_{k=1}^n(1 - p_j)$ ).

In the DEMATEL method, the criterion weighting logic is more comprehensive than in the CRITIC method. This is because in the DEMATEL method, the weight coefficient of a criterion increases as it affects other criteria both positively and negatively, and is affected by other criteria. Accordingly, in the DEMATEL method, the weight coefficient of a criterion is reflected by the square root of the sum of the squares of the values of the criterion affecting and being affected by other criteria, and the sum of the squares of the difference between the values of the criterion affecting and being affected by other criteria [25]. This situation is illustrated in Equation 11, which shows the criterion weights in the final step of the DEMATEL method [23-24].

In the DEMATEL method, the criterion weighting logic is more comprehensive than in the CRITIC method. This is because in the DEMATEL method, the weight coefficient of a criterion increases as it affects other criteria both positively and negatively, and is affected by other criteria. Accordingly, in the DEMATEL method, the weight coefficient of a criterion is reflected by the square root of the sum of the squares of the values of the criterion affecting and being affected by other criteria, and the sum of the

squares of the difference between the values of the criterion affecting and being affected by other criteria [25]. This situation is illustrated in Equation 11, which shows the criterion weights in the final step of the DEMATEL method [25-26].

$w$  = Weight of criteria.

$d$  = Total influence value of a criterion on other criteria.

$r$  = Total influenced value of a criterion by other criteria.

$$w = \sqrt{(d+r)^2 + (d-r)^2} \quad (11)$$

The similarities and contrasts between criteria shown in Equation 11 are explained in Table 3.

**Table 3.** The similarities and contrasts of a criterion with other criteria according to the DEMATEL method

$d$	$r$	Conditions	$d+r$	$d-r$
$d > 0$	$r > 0$	$d > r$	Similarity	Similarity
$d > 0$	$r > 0$	$d < r$	Similarity	Similarity
$d < 0$	$r < 0$	$ d  >  r $	Contrast	Contrast
$d < 0$	$r < 0$	$ d  <  r $	Contrast	Contrast
$d > 0$	$r < 0$	$d >  r $	Similarity	Similarity
$d > 0$	$r < 0$	$d <  r $	Contrast	Contrast
$d < 0$	$r > 0$	$ d  < r$	Similarity	Similarity
$d < 0$	$r > 0$	$ d  > r$	Contrast	Contrast

According to the eight scenarios explained in Table 13, the similarities and contrasts of a criterion with other criteria in terms of influencing and being influenced by them within the DEMATEL method are demonstrated. When Table 3 is examined, it is evident that the positive or negative nature of a criterion's relationships (influence and being influenced) with other criteria reflects the criterion's similarity or contrast with those criteria. Therefore, in the DEMATEL method, the importance of criterion weight is evaluated not only by the criterion's separation from other criteria (the negative impact of criteria on each other and negative being affected by each other), but also by its similarity (the positive impact of criteria on each other and positive being affected by each other). As a result, in the DEMATEL method, the weight of a criterion is considered equally in terms of the criterion's negative impact on other criteria and being affected by other criteria, as well as the criterion's positive impact on other criteria and being affected by other criteria. Therefore, the DEMATEL method is based on a more comprehensive criterion weighting logic than the CRITIC method. In light of all this information, the CRITIC method can be improved by taking into account the advantages of the rho correlation coefficient over the Pearson correlation coefficient ( $\rho$  is non-parametric and gives better results than Pearson correlation for non-linear relationships) and the advantageous comprehensiveness feature of the DEMATEL method over the CRITIC method.

Taking into consideration the aforementioned points, in the context of the proposed method (SRCBECM), the application steps are outlined as follows: in the first step, decision is formulated using Equation 4, and in the second step, normalized decision matrix is constructed with Equation 5. The third step of the method involves calculating the divergence ( $MDV_j$ ) through Equation 14, based on Equations 12 and 13 and similarity ( $MSV_j$ ) values for each criterion through Equation 14.

**Third Step:** Determination of divergence ( $MDV_j$ ) and similarity ( $MSV_j$ ) values for each criterion relative to the mean values.

Case 1: Calculation of the mean value for each normalized criterion ( $V_j$ )

$$V_j = \frac{\sum_{i=1}^m r_i}{m} \quad (12)$$

Case 2: Calculation of the mean value of the normalized criteria relative to the overall mean value ( $VV$ )

$$VV = \frac{\sum_{j=1}^n V_j}{n} \quad (13)$$

Case 3: Measure the deviation value of each criterion from the mean value ( $MDV_j$ )

For each criterion, the divergence value:

$$MDV_j = |V_j - VV| \quad (14)$$



Case 4: Measure the inverse of the deviation value of each criterion from the mean value ( $MSV_j$ )

For each criterion, the similarity value:

Similarly, within the context of the expansion of the logical foundation of the CRITIC method, the situation of similarity among criteria can be determined by comparing the value of 1 with the  $MDV_j$  value. This is because as the  $MDV_j$  value decreases (increases), the similarity among criteria will increase (decrease).

$$MSV_j = \frac{1}{MDV_j} \quad (15)$$

**Fourth Step:** Calculation of  $\rho$  correlation coefficient values among criteria

In the fourth step of the method, the  $\rho$  relationship matrix among criteria is constructed using Equations 1, 2, and 3. Statistical programs such as SPSS can be utilized for the calculation of these correlation coefficient values.

**Fifth Step:** Determination of divergence ( $CDV_j$ ) and similarity ( $CSV_j$ ) values for criteria according to  $\rho$  correlation analysis.

Case 1: Divergence of criteria according to  $\rho$  correlation analysis ( $CDV_j$ )

In the 4th step of the CRITIC method, the increase in the difference between the Pearson correlation values of criteria, excluding the 1 value in Equation 9, results in an increase in the weight coefficients of criteria for negative relationships (divergence). Therefore, the divergence degree of criteria according to  $\rho$  correlation analysis can be calculated using Equation 16 in the 4th step of the CRITIC method, utilizing Equation 9 excluding standard deviation.

$$CDV_j = \sum_{i=1}^m (1 - \rho_{oi}), \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n \quad (16)$$

Case 2: Identification of the inverse of the divergence value according to  $\rho$  correlation analysis (similarity) ( $CSV_j$ )

Taking into account Equation 16, the degree of similarity among criteria can be calculated as shown in Equation 16, which involves summing the  $\rho$  coefficients between criteria with a value of 1.

$$CSV_j = \sum_{i=1}^m (1 + \rho_{oi}), \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n \quad (17)$$

**Sixth Step:** Calculation of Divergence ( $SE_j$ ) and Similarity Values ( $SI_j$ ) for Criteria

Case 1: Calculation of divergence values for criteria ( $SE_j$ )

The value ( $SE_j$ ) for each criterion is calculated by dividing the product of the deviation from the mean value for criteria  $MDV_j$  and the divergence value according to  $\rho$  correlation analysis  $CDV_j$  by the total divergence values of criteria.

$$SE_j = \frac{MDV_j \cdot CDV_j}{\sum_{j=1}^n MDV_j \cdot CDV_j} \quad (18)$$

Case 2: Calculation of similarity values for criteria ( $SI_j$ )

The similarity values for each criterion are explained by the summation of the product of the inverse of the deviation from the mean value for each criterion ( $MSV_j$ ) and the similarity according to  $\rho$  correlation analysis ( $CSV_j$ ) as described in Equation 19, which represents the total similarity values for criteria.

$$SI_j = \frac{MSV_j \cdot CSV_j}{\sum_{j=1}^n MSV_j \cdot CSV_j} \quad (19)$$

**Seventh Step:** Calculation of criterion weights ( $w_j$ )

Case 1: Taking into account the equal-weight divergence and similarity situations of criteria ( $EW_j$ )

Considering the divergence and similarity situations of criteria with equal weight, the values of criteria ( $EW_j$ ) are determined, as explained in Equation 20.

$$EW_j = \frac{SE_j + SI_j}{2} \quad (20)$$

Case 2: Calculation of criterion weights

$$w_j = \frac{EW_j}{\sum_{j=1}^n EW_j} \quad (21)$$

---

In light of the above, the SRCBECM method has some advantages compared to other methods. Firstly, one of the advantages of the SRCBECM method is its insensitivity to 0 and negative values. In contrast, in the ENTROPY and MEREC methods, when there are 0 or negative values in the decision matrix, calculations can become undefined as these methods allow for logarithmic transformation. Consequently, negative values in the decision matrix can be transformed using Z-score. However, the presence of a 0 value in the decision matrix poses a challenge to the calculation of criterion weights in these methods. Secondly, the foundation of the method is based on the rho correlation coefficient among criteria. Especially in the CRITIC method, the relationships between criteria are measured with the Pearson correlation coefficient, which has a parametric structure. Therefore, the Pearson correlation coefficient may not yield accurate results, especially among criterion data that do not exhibit a normal distribution. In contrast, in the SRCBECM method, the relationships among criteria, without the assumption of a normal distribution, are measured using the non-parametric *rho* correlation coefficient. Thirdly, another advantage is that the SRCBECM method considers both divergence (opposition) and similarity conditions among criteria on equal terms, similar to the DEMATEL method. In contrast, in methods such as ENTROPY, CRITIC, SD, SVP, MEREC, and LOPCOW, the logic of the weights of criteria focuses solely on the divergence (opposition) of criteria from each other. Therefore, the weighting logic of the SRCBECM method is more comprehensive than other methods. Finally, due to the broad focus of the SRCBECM method, the weights of criteria in SRCBECM have more characteristic features compared to other methods. Thus, the characteristic features of criteria with prominent characteristics result in more differences in the weight coefficients of criteria between them and other criteria compared to other methods.

The SRCBECM method offers a more balanced weighting by considering both similarities and differences among criteria, providing a more comprehensive analysis compared to methods that typically focus solely on differences. Its insensitivity to zero and negative values, along with its lack of a normal distribution assumption, makes the method applicable to a wide range of datasets. Consequently, SRCBECM yields reliable results, particularly in scenarios involving heterogeneous and complex datasets. Furthermore, by considering both similarities and differences among criteria, SRCBECM provides decision-makers with more balanced and accurate weights. This contributes to more accurate outcomes in decision-making processes and makes the method applicable across various sectors, including financial decisions, project management, resource allocation, product design, project selection, environmental impact assessment, and sustainable project selection. In practice, the method aids in making more meaningful and consistent decisions, especially in cases involving complex inter-criteria relationships, particularly in big data analysis. It can be particularly beneficial in decision-making processes such as corporate strategy formulation, investment evaluations, and public policy development. Moreover, SRCBECM can simplify complex decision-making processes for individuals or organizations. For instance, when making a multi-criteria investment decision, the method's criterion weighting approach helps achieve clearer and more reliable results. In daily life, SRCBECM enables individuals or organizations working with data to conduct more sound and comprehensive analyses. This facilitates more effective decision-making in workplace performance evaluations, risk analyses, and multi-criteria choices encountered in everyday life. Thanks to its flexible structure that can adapt to different data types and criterion structures, SRCBECM can be applied to various decision-making mechanisms in daily life. This flexibility makes the method more practical and user-friendly.

## **2.5. Data set and analysis of the study**

The research dataset consists of the criteria from the Freedom in the world (FIW) by Freedom House for the year 2022, focusing on 19 countries within the G20 group. The reason for selecting this dataset is to evaluate the discriminatory effectiveness of the model criteria proposed among countries, considering the significant variations in values within this specific dataset. To enhance clarity in the research, Table 4 provides explanations for the abbreviations associated with this dataset.

**Table 4.** Data Set

FIW Criteria	Criteria Abbreviations
Electoral Process	FIW1
Political Pluralism and Participation	FIW2
Functioning of Government	FIW3
Freedom of Expression and Belief	FIW4
Associational and Organizational Rights	FIW5
Rule of Law	FIW6
Personal Autonomy and Individual Rights	FIW7

### 3. Results (The case study)

#### 3.1. Computational analyses

In the research, the first step of the SRCBECM method involves creating the decision matrix using Equation 4. In the second step of the method, since all criteria are benefit-oriented, the decision matrix values were normalized using Equation 5. In this regard, the decision matrix and the normalized decision matrix values are presented in Table 5.

**Table 5.** Decision and normalized matrix

Decision Matrix							
Country	FIW1	FIW2	FIW3	FIW4	FIW5	FIW6	FIW7
Argentina	11	16	8	15	11	10	14
Australia	12	15	11	15	12	15	15
Brazil	10	13	7	13	9	8	12
Canada	12	16	12	15	12	15	16
China	0	0	1	1	2	2	6
France	12	15	11	14	10	13	14
Germany	12	15	12	14	12	14	15
India	12	12	9	9	7	8	9
Indonesia	11	13	6	9	6	5	8
Italy	12	14	10	15	12	13	14
Japan	12	16	12	15	12	15	14
Mexico	9	13	5	12	7	5	9
Saudi Arabia	11	13	9	14	11	12	13
Russia	0	3	2	2	2	2	5
South Africa	12	13	8	15	12	9	10
South Korea	11	13	9	14	11	12	13
Turkey	5	8	3	5	3	3	5
United Kingdom	12	16	11	14	12	13	15
United States	10	14	9	14	11	11	14
Min.	0	0	1	1	2	2	5
Max.	12	16	12	15	12	15	16

Normalized Decision Matrix							
Country	FIW1	FIW2	FIW3	FIW4	FIW5	FIW6	FIW7
Argentina	0,917	1,000	0,636	1,000	0,900	0,615	0,818
Australia	1,000	0,938	0,909	1,000	1,000	1,000	0,909
Brazil	0,833	0,813	0,545	0,857	0,700	0,462	0,636
Canada	1,000	1,000	1,000	1,000	1,000	1,000	1,000
China	0,000	0,000	0,000	0,000	0,000	0,000	0,091
France	1,000	0,938	0,909	0,929	0,800	0,846	0,818
Germany	1,000	0,938	1,000	0,929	1,000	0,923	0,909
India	1,000	0,750	0,727	0,571	0,500	0,462	0,364
Indonesia	0,917	0,813	0,455	0,571	0,400	0,231	0,273
Italy	1,000	0,875	0,818	1,000	1,000	0,846	0,818

Japan	1,000	1,000	1,000	1,000	1,000	1,000	0,818
Mexico	0,750	0,813	0,364	0,786	0,500	0,231	0,364
Saudi Arabia	0,917	0,813	0,727	0,929	0,900	0,769	0,727
Russia	0,000	0,188	0,091	0,071	0,000	0,000	0,000
South Africa	1,000	0,813	0,636	1,000	1,000	0,538	0,455
South Korea	0,917	0,813	0,727	0,929	0,900	0,769	0,727
Turkey	0,417	0,500	0,182	0,286	0,100	0,077	0,000
United Kingdom	1,000	1,000	0,909	0,929	1,000	0,846	0,909
United States	0,833	0,875	0,727	0,929	0,900	0,692	0,818
Mean	0,816	0,783	0,651	0,774	0,716	0,595	0,603
Mean of Mean				0,705			

In the third step of the SRCBECM method, Equation 12 was used to determine the values of  $V_j$  for the criteria, Equation 13 for  $VV$ , Equation 14 for  $MDV_j$ , and Equation 15 for  $MSV_j$ . For illustrative purposes, the values of  $V_j$ ,  $VV$ ,  $MDV_j$ , and  $MSV_j$  for FIW1 relative to the mean values are calculated below, and the  $MDV_j$  and  $MSV_j$  values for the other criteria are presented in Table 6.

Case 1: Calculation of the mean value for each normalized criterion ( $V_{FIW1}$ )

$$V_{FIW1} = \frac{0,917+1+0,8833+1+0+1+1+1+0,917+1+1+0,750+0,917+0+1+0,917+0,417+1+0,833}{19} = 0,816$$

Case 2: Calculation of the mean value of the normalized criteria relative to the overall mean value ( $VV$ )

$$VV = \frac{0,816 + 0,783 + 0,651 + 0,774 + 0,716 + 0,595 + 0,603}{7} = 0,705$$

Case 3: Measure the deviation value of each criterion from the mean value ( $MDV_{FIW1}$ )

For each criterion, the divergence value:

$$MDV_{FIW1} = |0,816 - 0,705| = 0,111$$

Case 4: Measure the inverse of the deviation value of each criterion from the mean value ( $MSV_{FIW1}$ )

For each criterion, the similarity value

$$MSV_{FIW1} = \frac{1}{0,111} = 9,026$$

**Table 6.** Separation (Deviation ( $MDV_j$ )) and Similarity ( $MSV_j$ ) Values of Criteria

Criteria	FIW1	FIW2	FIW3	FIW4	FIW5	FIW6	FIW7
$MDV_j$	0,111	0,078	0,054	0,069	0,011	0,110	0,102
$MSV_j$	9,026	12,838	18,422	14,402	92,683	9,103	9,792

In the fourth step of the method, the  $\rho$  correlation coefficients between the criteria are first calculated using Equations 1, 2, and 3, and the correlation matrix is created. The calculated  $\rho$  correlation values between the criteria are shown in Table 7.

**Table 7.**  $\rho$  correlation matrix

Criteria	FIW1	FIW2	FIW3	FIW4	FIW5	FIW6	FIW7
FIW1	1,000	0,952	0,883	0,912	0,872	0,801	0,780
FIW2	0,952	1,000	-0,227	-0,244	-0,235	-0,183	-0,137
FIW3	0,883	-0,227	1,000	-0,389	-0,421	-0,388	-0,337
FIW4	0,912	-0,244	-0,389	1,000	-0,306	-0,215	-0,199
FIW5	0,872	-0,235	-0,421	-0,306	1,000	-0,296	-0,279
FIW6	0,801	-0,183	-0,388	-0,215	-0,296	1,000	-0,360
FIW7	0,780	-0,137	-0,337	-0,199	-0,279	-0,360	1,000
Mean	0,886	0,132	0,017	0,080	0,048	0,051	0,067

In the fifth step of the method, the separation ( $CDV_j$ ) values according to  $\rho$  correlation analysis were calculated using Equation 16, and the similarity ( $CSV_j$ ) values were calculated using Equation 17 for the criteria. The corresponding ( $CDV_j$ ) and ( $CSV_j$ ) values are presented in Table 8.

**Table 8.** The separation ( $CDV_j$ ) and similarity ( $CSV_j$ ) values of the criteria based on the  $\rho$ 

( $CDV_j$ )	FIW1	FIW2	FIW3	FIW4	FIW5	FIW6	FIW7
FIW1	0,000	0,048	0,117	0,088	0,128	0,199	0,220
FIW2	0,048	0,000	1,227	1,244	1,235	1,183	1,137
FIW3	0,117	1,227	0,000	1,389	1,421	1,388	1,337
FIW4	0,088	1,244	1,389	0,000	1,306	1,215	1,199
FIW5	0,128	1,235	1,421	1,306	0,000	1,296	1,279
FIW6	0,199	1,183	1,388	1,215	1,296	0,000	1,360
FIW7	0,220	1,137	1,337	1,199	1,279	1,360	0,000
<b>Sum (<math>CDV_j</math>)</b>	<b>0,800</b>	<b>6,075</b>	<b>6,878</b>	<b>6,441</b>	<b>6,665</b>	<b>6,641</b>	<b>6,533</b>
( $CSV_j$ )	FIW1	FIW2	FIW3	FIW4	FIW5	FIW6	FIW7
FIW1	2,000	1,952	1,883	1,912	1,872	1,801	1,780
FIW2	1,952	2,000	0,773	0,756	0,765	0,817	0,863
FIW3	1,883	0,773	2,000	0,611	0,579	0,612	0,663
FIW4	1,912	0,756	0,611	2,000	0,694	0,785	0,801
FIW5	1,872	0,765	0,579	0,694	2,000	0,704	0,721
FIW6	1,801	0,817	0,612	0,785	0,704	2,000	0,640
FIW7	1,780	0,863	0,663	0,801	0,721	0,640	2,000
<b>Sum (<math>CSV_j</math>)</b>	<b>13,200</b>	<b>7,925</b>	<b>7,122</b>	<b>7,559</b>	<b>7,335</b>	<b>7,359</b>	<b>7,467</b>

For illustrative purposes, the discrimination  $CDV_j$  and similarity  $CSV_j$  values for FIW1 were calculated based on the  $\rho$  correlation values. The calculated values for  $CDV_j$  and  $CSV_j$  for FIW1, as well as the corresponding situations for other criteria based on  $\rho$  correlation coefficient values, are presented in Table 8.

Case 1: Divergence of criteria according to  $\rho$  correlation analysis ( $CDV_{FIW1}$ )

$$CDV_{FIW1} = 0 + 0,048 + 0,117 + 0,088 + 0,128 + 0,199 + 0,220 = 0,800$$

Case 2: Identification of the inverse of the divergence value according to  $\rho$  correlation analysis (similarity) ( $CSV_{FIW1}$ )

$$CSV_{FIW1} = 2 + 1,952 + 1,883 + 1,912 + 1,872 + 1,801 + 1,780 = 13,200$$

In the 6th step of the method, the separation values ( $SE_j$ ) and similarity values  $SI_j$  for the criteria are calculated using Equations 17 and 18, respectively. The calculated values for  $SE_j$  and  $SI_j$  are presented in Table 9.

**Table 9.** The separation ( $SE_j$ ) and similarity ( $SI_j$ ) values of the criteria

Criteria	FIW1	FIW2	FIW3	FIW4	FIW5	FIW6	FIW7	Sum
( $MDV_j.CDV_j$ )	0,089	0,055	0,031	0,038	0,005	0,073	0,070	0,361
( $SE_j$ )	0,246	0,153	0,086	0,106	0,015	0,202	0,194	
Criteria	FIW1	FIW2	FIW3	FIW4	FIW5	FIW6	FIW7	Sum
( $MSV_j.CSV_j$ )	34,121	48,601	70,032	54,808	352,760	34,189	37,014	631,525
( $SI_j$ )	0,054	0,077	0,111	0,087	0,559	0,054	0,059	

As an illustrative example, the discrimination values ( $SE_j$ ) and similarity values ( $SI_j$ ) for FIW1 have been calculated below.

Case 1: Calculation of divergence values for criteria ( $SE_{FIW1}$ )

$$SE_{FIW1} = \frac{0,111 \cdot 0,800}{0,361} = 0,246$$

Case 2: Calculation of similarity values for criteria ( $SI_{FIW1}$ )

$$SI_{FIW1} = \frac{9,026 \cdot 13,200}{631,525} = 0,054$$

In the final step of the method, the weights of the criteria ( $w_j$ ) are calculated by evaluating the discrimination ( $SE_j$ ) and similarity ( $SI_j$ ) conditions of the criteria with equal importance, as

expressed in Equations 19 and 20. The resulting weight values ( $w_j$ ) for the criteria are presented in Table 10.

**Table 10.** Criteria weights

Criteria	$EW_j$	w	Rank
FIW1	0,150	0,197	2
FIW2	0,066	0,087	4
FIW3	0,071	0,093	3
FIW4	0,062	0,082	7
FIW5	0,282	0,372	1
FIW6	0,063	0,084	6
FIW7	0,064	0,085	5
<b>Sum</b>	0,759		

Upon examining Table 10, the ranking of criteria weight values is as follows: FIW5, FIW1, FIW3, FIW2, FIW7, FIW6, and FIW4. Furthermore, for illustrative purposes, the weight values of the FIW1 criterion have been calculated and are presented below.

Case 1: Taking into account the equal-weight divergence and similarity situations of criteria ( $EW_{FIW1}$ )

$$EW_{FIW1} = \frac{0,246 + 0,054}{2} = 0,150$$

Case 2: Calculation of weight ( $w_{FIW1}$ )

$$w_{FIW1} = \frac{0,150}{0,759} = 0,197$$

### 3.2. Sensitivity analyses

In the scope of this investigation, we performed an evaluation of the SRCBECM approach to scrutinize its methodological sensitivity. Sensitivity analysis, as applied in the context of MCDM, involves the utilization of diverse criteria weighting methods on the same dataset, enabling a comparison of resulting values and rankings. To ascertain the method's sensitivity in calculating weight coefficients, it is expected that the rankings of criteria weights determined using the chosen method for sensitivity analysis will deviate from the weight coefficient rankings obtained through alternative methods [63].

Following this methodology, we employed established objective weighting techniques to compute and organize the weighting coefficients linked to the components of the FIW. These techniques, widely utilized in scholarly literature, encompass ENTROPY, CRITIC, SD, SVP, MEREC, and LOPCOW. The corresponding numerical outcomes are meticulously presented in Table 11.

**Table 11.** Findings derived from different methodologies for computing objective weighting coefficients

Criteria	CRITIC	Ranking	SD	Ranking	SVP	Ranking
FIW1	0,024	7	0,145	3	0,132	4
FIW2	0,153	6	0,131	6	0,170	3
FIW3	0,197	4	0,153	2	0,103	7
FIW4	0,193	5	0,142	5	0,183	1
FIW5	0,221	1	0,145	4	0,115	6
FIW6	0,212	2	0,167	1	0,180	2
FIW7	0,199	3	0,118	7	0,117	5
Criteria	LOPCOW	Ranking	MEREC	Ranking	ENTROPY	Ranking
FIW1	0,156	2	0,110	6	0,157	5
FIW2	0,167	1	0,126	4	0,131	6
FIW3	0,140	4	0,097	7	0,178	2
FIW4	0,151	3	0,121	5	0,167	3
FIW5	0,135	5	0,175	3	0,159	4
FIW6	0,123	7	0,177	2	0,209	1

An analysis of Table 11 indicates notable disparities in the rankings of FIW weight values when employing the SRCBECM method compared to rankings obtained through other methods for determining FIW criterion weights. This outcome implies that the proposed method exhibits sensitivity to the inherent relationships among the criteria.

### 3.3. Comparative analyses

The comparative analysis explores both the congruities and divergences between the proposed method and other techniques for calculating objective weight coefficients. The proposed method is expected to possess credibility, reliability, and consistency in alignment with other methodologies, while concurrently exhibiting a positive and significant correlation with various weight coefficient methods [20]. To substantiate this, Figures 1 and 2 provide a visual examination of the weight coefficients determined using the weight measurement methods outlined in Table 12.

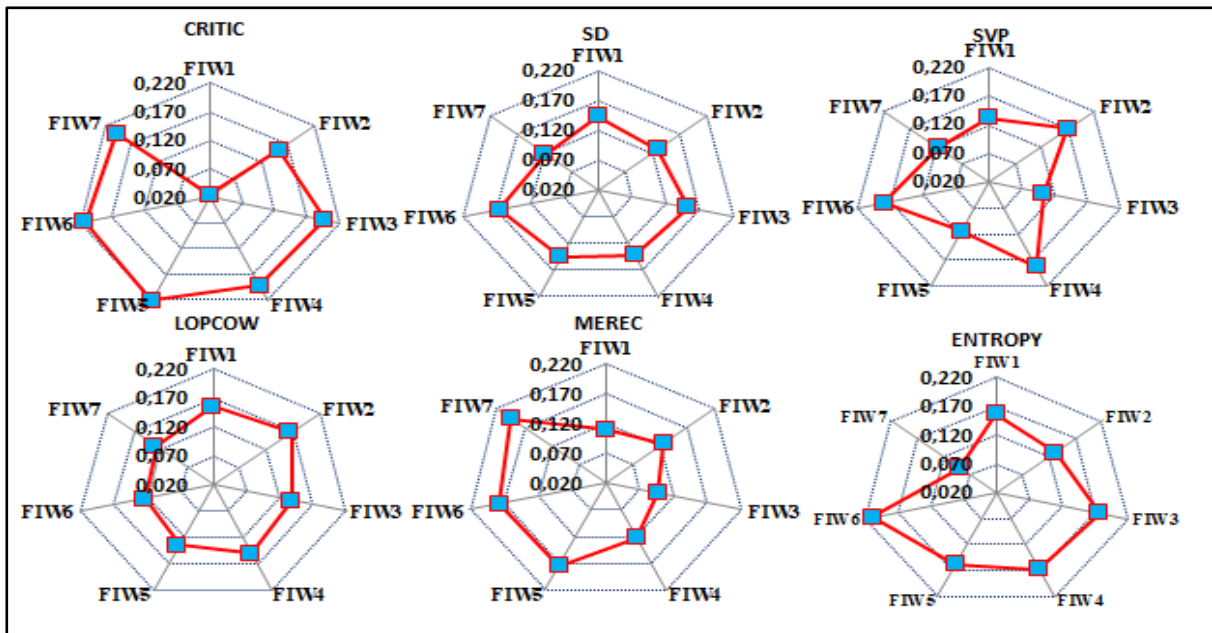


Figure 1. Positions of the ENTROPY, CRITIC, SD, SVP, LOPCOW, and MEREC methods

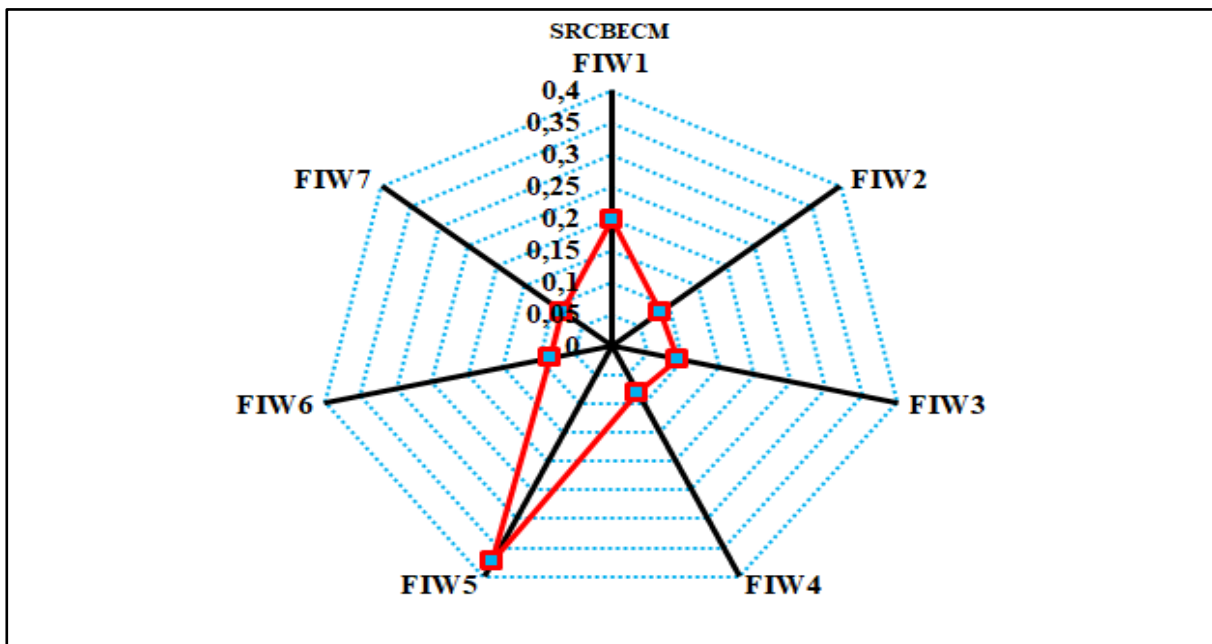


Figure 2. Positions of the SRCBECM method

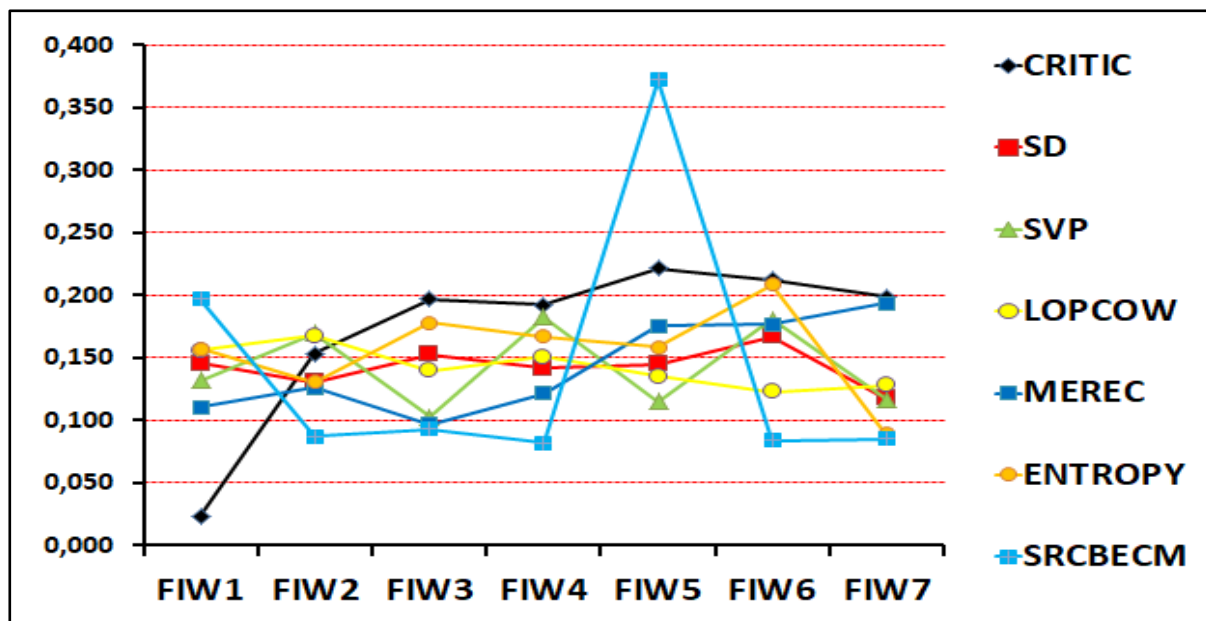


Figure 3. Positions of the Methods 1

When Figure 3 is examined, it is determined that the characteristic properties of FIW1 and FIW5 criteria are reflected more in the SRCBECM method than in other methods. In particular, these criteria are observed to reflect both the separation and similarity of criteria, which constitute the logic of the proposed method, simultaneously. This shows that it explains better which criteria are more important than other methods. In addition, according to Figure 3, it is determined that the differences between the positions of the criterion values of the SRCBECM method are different from other methods. Thus, it is evaluated that the SRCBECM method does not have a positive correlation with other methods. Although it is stated in the MCDM literature that the high level of positive relationship between any objective weighting method and other methods indicates that the credible and reliable level of the proposed method is high, it is the desired result that the SRCBECM method has a low positive and negative correlation with other methods, since the criterion weighting logic in the method (SRCBECM) takes into account both separation and similarity levels (both contrast conditions), unlike other methods. Therefore, in this case, the low positive and negative correlation of the SRCBECM method with other methods shows that the proposed method is credible and reliable. The  $\rho$  correlation coefficient values between the methods are explained in Table 13, since the data of the methods do not show normal distribution.

Table 13.  $\rho$  correlation values of the SRCBECM method with other methods

Criteria	CRITIC	SD	SVP	LOPCOW	MEREC	ENTROPY
SRCBECM	-0,047	0,089	-0,434	-0,071	0,211	0,044

When Table 13 is examined, it is observed that none of the relationships between the SRCBECM method and other objective criterion weighting methods are significant. Based on these quantitative results, it is concluded that the SRCBECM method is different from other methods because the criterion weights are calculated based on both separation and similarity levels (both contrast conditions), and accordingly, the proposed method is reliable and credible.

### 3.4. Simulation Analyses

To assess the robustness of the proposed method, a simulation analysis is undertaken, involving various scenarios created by assigning different values to decision matrices. As the number of scenarios increases, the proposed method is expected to diverge from other methodologies, showcasing its stability. In the subsequent step, the average variance of criterion weights determined by the proposed method across the scenarios should exceed that of one or more alternative objective weighting methods. This signifies the superior discriminative capability of the proposed method in distinguishing between



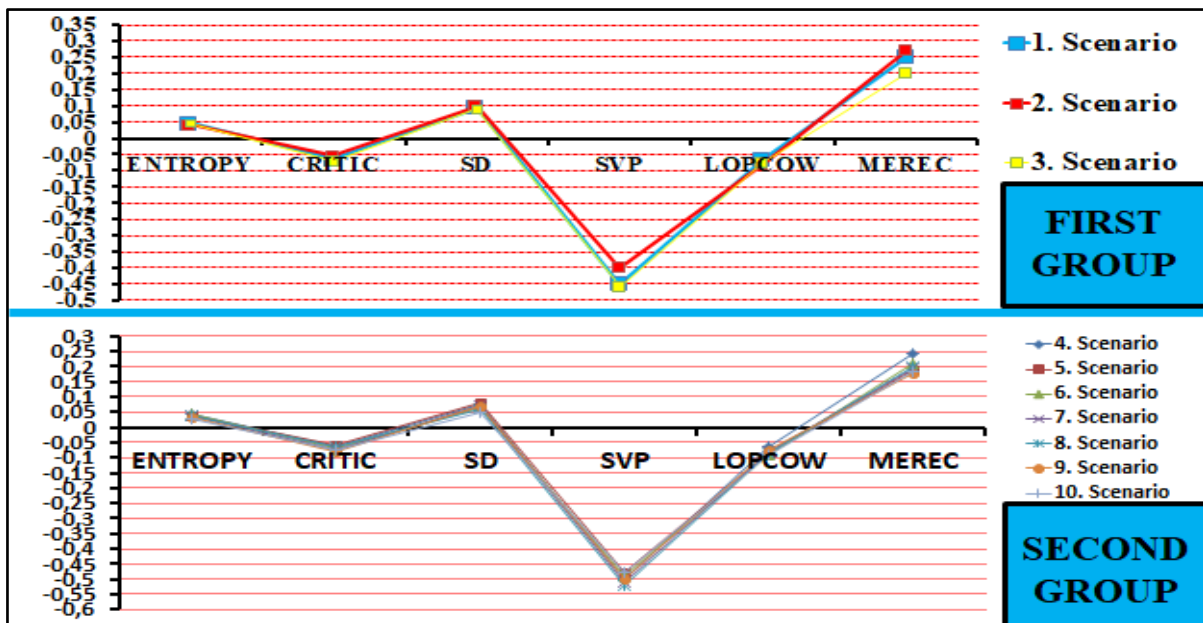
criterion weights. Finally, it validates the consistency of criterion weight variances across methods within the scenarios [20].

To evaluate the coherence of the SRCBECM method with alternative objective weighting methods, a simulation analysis was executed. The simulation involved the generation of 20 distinct scenarios, categorized into 2 groups, each characterized by a unique set of decision matrix values. For each scenario, correlation coefficients between the SDBHA method and other methodologies were computed. The results of the simulation analysis are presented in Tables 14 and 15, and visualized in Figures 4 and 5.

**Table 14.**  $\rho$  correlation coefficients of the SDBHA method with other methodologies across the range of scenarios-A (First Category)

Group	Scenarios	ENTROPY	CRITIC	SD	SVP	LOPCOW	MEREC
First group	1. Scenario	0,048	-0,065	0,095	-0,450	-0,065	0,250
	2. Scenario	0,043	-0,055	0,100	-0,400	-0,080	0,270
	3. Scenario	0,045	-0,070	0,090	-0,460	-0,075	0,200
Second group	4. Scenario	0,044	-0,066	0,075	-0,500*	-0,065	0,240
	5. Scenario	0,040	-0,060	0,080	-0,480	-0,085	0,190
	6. Scenario	0,042	-0,070	0,065	-0,490	-0,090	0,210
	7. Scenario	0,037	-0,075	0,070	-0,510*	-0,080	0,195
	8. Scenario	0,039	-0,065	0,060	-0,520*	-0,085	0,200
	9. Scenario	0,033	-0,080	0,070	-0,500*	-0,075	0,180
	10. Scenario	0,028	-0,077	0,050	-0,480	-0,090	0,185
Mean		0,040	-0,068	0,076	-0,479	-0,079	0,212

$p^* < 0,05$



**Figure 4.** Positions of the Methods 2-A (First Category)

**Table 15.**  $\rho$  correlation coefficients of the SDBHA method with other methodologies across the range of scenarios-B (Second Category)

Group	Scenarios	ENTROPY	CRITIC	SD	SVP	LOPCOW	MEREC
First group	1. Scenario	0,055	-0,078	0,122	-0,5	-0,074	0,324
	2. Scenario	0,059	-0,082	0,132	-0,55	-0,098	0,327
	3. Scenario	0,056	-0,08	0,124	-0,495	-0,087	0,312
Second group	4. Scenario	0,058	-0,077	0,114	-0,455	-0,085	0,308

5. Scenario	0,055	-0,075	0,109	-0,475	-0,077	0,301
6. Scenario	0,048	-0,073	0,102	-0,484	-0,089	0,294
7. Scenario	0,041	-0,067	0,09	-0,501	-0,082	0,299
8. Scenario	0,033	-0,062	0,07	-0,495	-0,078	0,267
9. Scenario	0,031	-0,069	0,06	-0,425	-0,071	0,259
10. Scenario	0,003	-0,066	0,06	-0,412	-0,066	0,239
Mean	0,0440	-0,0730	0,0983	-0,4792	-0,0807	0,293

$p^* < .05$

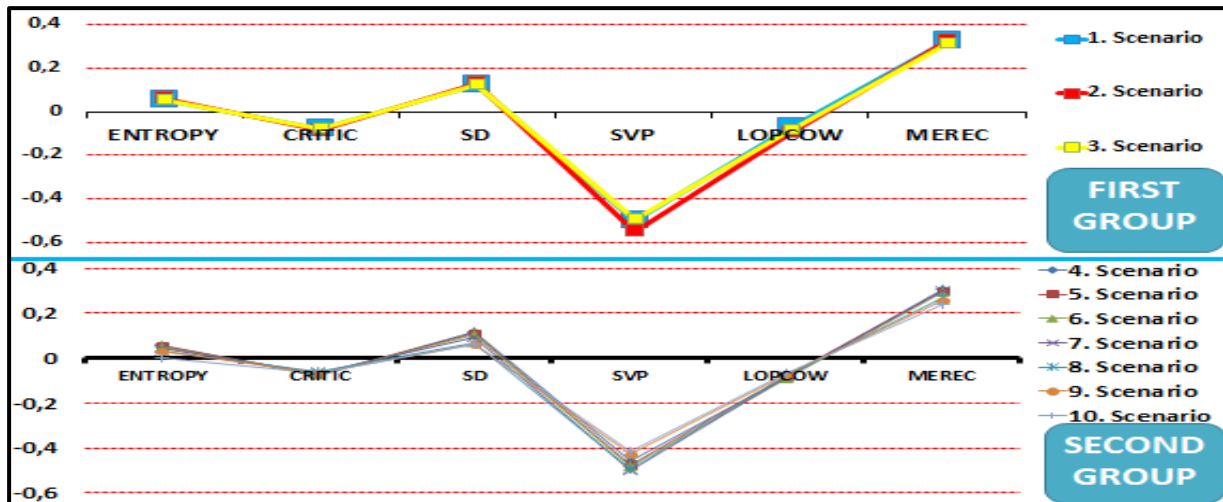


Figure 5. Positions of the Methods 2-B (Second Category)

A simultaneous examination of Tables 14 and 15, and Figures 4 and 5 reveals that the 10 scenarios, categorized into 2 groups, can be further divided into two subgroups. The initial group comprises the first three scenarios, while the subsequent group includes the last seven. When Tables 14 and 15, and Figures 4 and 5 are analyzed together, it is evident that the positive correlation values of the SRCBECM method with the ENTROPY, SD, and MEREC methods decrease as the number of scenarios increases, whereas the correlation values with the CRITIC, SVP, and LOPCOW methods increase negatively. In summary, the SRCBECM method exhibits greater divergence from the other methods as the number of scenarios grows. Consequently, it can be inferred that the distinctive characteristics of the SRCBECM method become more pronounced with an increasing number of scenarios. Additionally, Figures 6 and 7 illustrate the spatial distribution of the relationships between the SRCBECM method and other methodologies within the specified groups across both categories (First Category and Second Category).

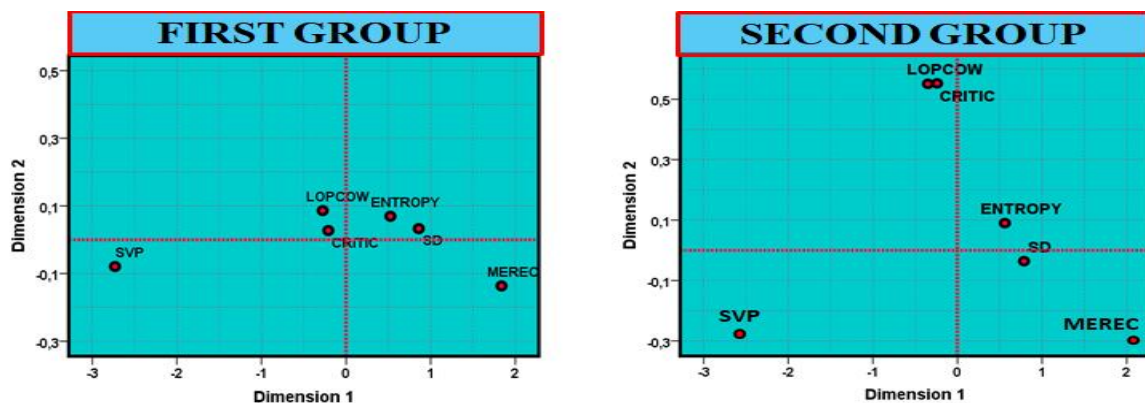
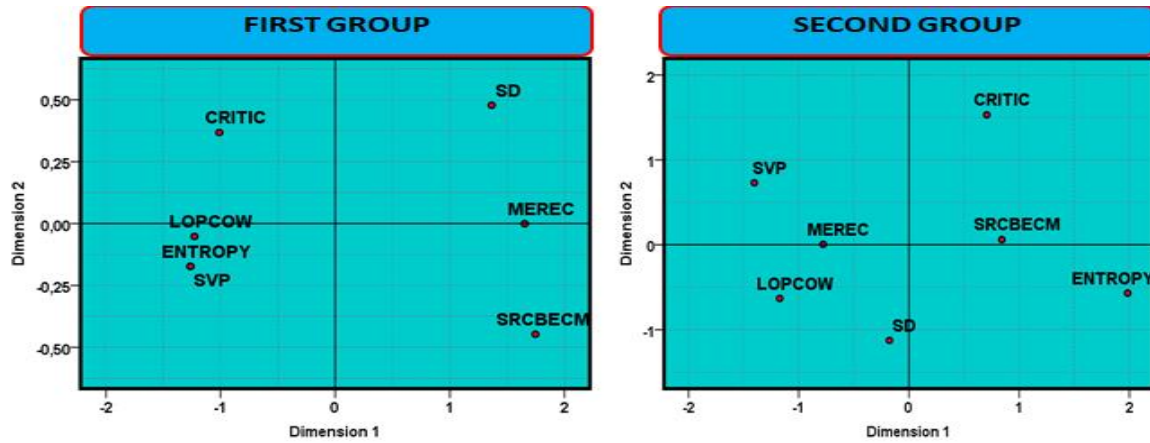


Figure 6. Depicts the discriminant analysis illustrating the rho correlation between the SRCBECM method and alternative methodologies across various scenarios-A (First Category)



**Figure 7.** Depicts the discriminant analysis illustrating the rho correlation between the SRCBECM method and alternative methodologies across various scenarios-B (Second Category)

When Figure 6 and Figure 7 are examined, in scope of first and second categories it is observed that the distribution of the relationships of the SRCBECM method with other objective criterion weighting methods in the 7 scenarios in the 2nd group is more dispersed than the distribution of the relationships of the SRCBECM method with other objective criterion weighting methods in the 3 scenarios in the 1st group. Therefore, it is observed that the *rho* correlation values of the SRCBECM method with other objective criterion weighting methods change as the number of scenarios increases.

The separation power of any objective criterion weighting method to its own criteria depends on the high variance value of the method [20]. Throughout the simulation analysis, variance values for the methods were computed across diverse scenarios for two categories, and the corresponding values are summarized in Tables 16 and 17.

**Table 16.** Variance in method outcomes across different scenarios-A (First category)

Scenario	SRCBECM	CRITIC	SD	SVP	LOPCOW	MEREC	ENTROPY
1. Sce.	0,0125	0,0052	0,0002	0,0011	0,0003	0,0014	0,0014
2. Sce.	0,0125	0,0048	0,0001	0,0014	0,0004	0,0019	0,0017
3. Sce.	0,0138	0,0046	0,0003	0,0013	0,0003	0,0027	0,0013
4. Sce.	0,0091	0,0025	0,0003	0,0015	0,0006	0,0011	0,0019
5. Sce.	0,0145	0,0033	0,0003	0,0009	0,0005	0,0018	0,0021
6. Sce.	0,0084	0,0018	0,0002	0,0015	0,0007	0,0028	0,0015
7. Sce.	0,0132	0,0032	0,0004	0,0014	0,0004	0,0026	0,0013
8. Sce.	0,0117	0,0024	0,0005	0,0015	0,0008	0,0029	0,0012
9. Sce.	0,0113	0,0063	0,0003	0,0013	0,0006	0,0022	0,0012
10. Sce.	0,0124	0,0070	0,0002	0,0012	0,0003	0,0016	0,0016
Mean	0,0119	0,0041	0,0003	0,0013	0,0005	0,0021	0,0015

**Table 17.** Variance in method outcomes across different scenarios-B (Second category)

Scenario	SRCBECM	CRITIC	SD	SVP	LOPCOW	MEREC	ENTROPY
1. Sce.	0,0128	0,0058	0,0003	0,0022	0,0007	0,0024	0,0028
2. Sce.	0,013	0,0051	0,0002	0,0019	0,0006	0,0021	0,0025
3. Sce.	0,0141	0,0049	0,0004	0,0015	0,0005	0,0039	0,0021
4. Sce.	0,0101	0,0028	0,0004	0,0012	0,0007	0,0017	0,0021
5. Sce.	0,0156	0,0036	0,0005	0,001	0,0007	0,0026	0,0023
6. Sce.	0,0099	0,0021	0,0003	0,0018	0,0009	0,0031	0,0017
7. Sce.	0,0149	0,0041	0,0004	0,0016	0,0006	0,0032	0,0019
8. Sce.	0,0131	0,0028	0,0007	0,0017	0,0011	0,0035	0,0016
9. Sce.	0,0128	0,0071	0,0004	0,0017	0,0009	0,0029	0,0015

<b>10. Sce.</b>	0,0131	0,0082	0,0003	0,0014	0,0006	0,0023	0,0018
<b>Mean</b>	0,01294	0,00465	0,00039	0,0016	0,00073	0,00277	0,00203

Based on the data presented in Tables 16 and 17 for both categories, it is evident that the SRCBECM method consistently exhibits higher average variance values across scenarios compared to the other methods. Consequently, it can be inferred that the SRCBECM method demonstrates a superior ability to discriminate criterion weights, as evidenced by its higher average variance value relative to the other methods.

To meticulously examine the homogeneity of variances within the criterion weights integral to the SDBHA method, an ADM (Analysis of Means for Variances, incorporating the Levene modification) was meticulously executed across a diverse array of scenarios. This analytical approach proffers a cogent visual representation of variance uniformity. The graphical manifestation is composed of three cardinal elements: the global average ADM, which serves as the pivotal axis, bordered by the upper and lower decision limits (UDL and LDL). In the event that the standard deviation of an individual cluster transcends these decision boundaries, it intimates a salient deviation from the global average ADM, thereby portending the existence of variance heterogeneity. Antithetically, if the standard deviations of all clusters are ensconced within the confines of the LDL and UDL, this buttresses the notion of variance homogeneity [20]. A clear visual explanation of the ADM analysis is provided in Figures 8 and 9, encompassing both categories (the first category and the second category).

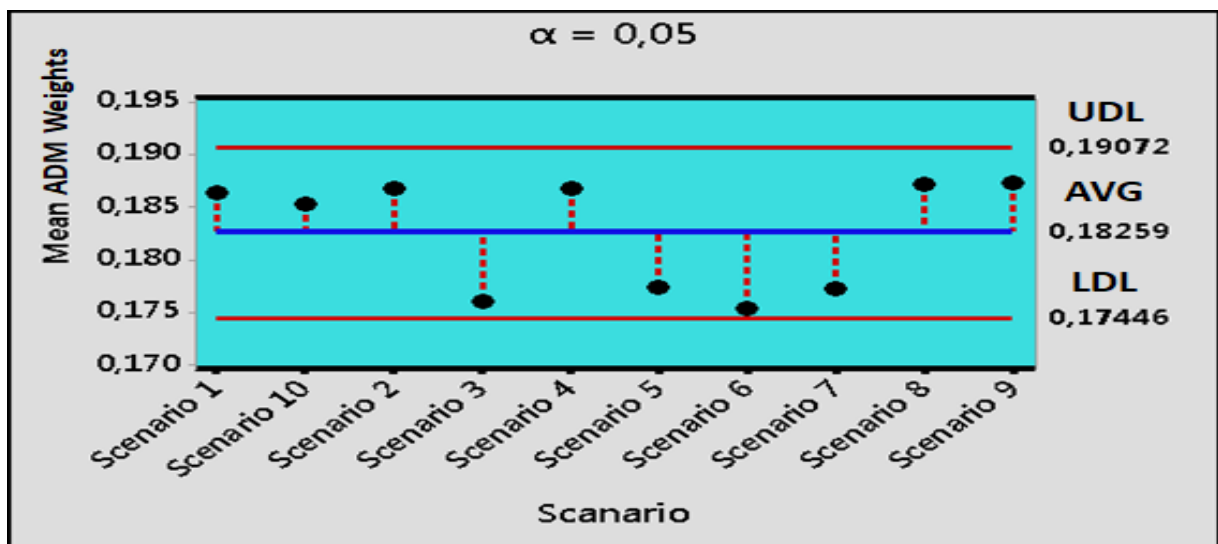


Figure 8. ADM visual-A (First Category)

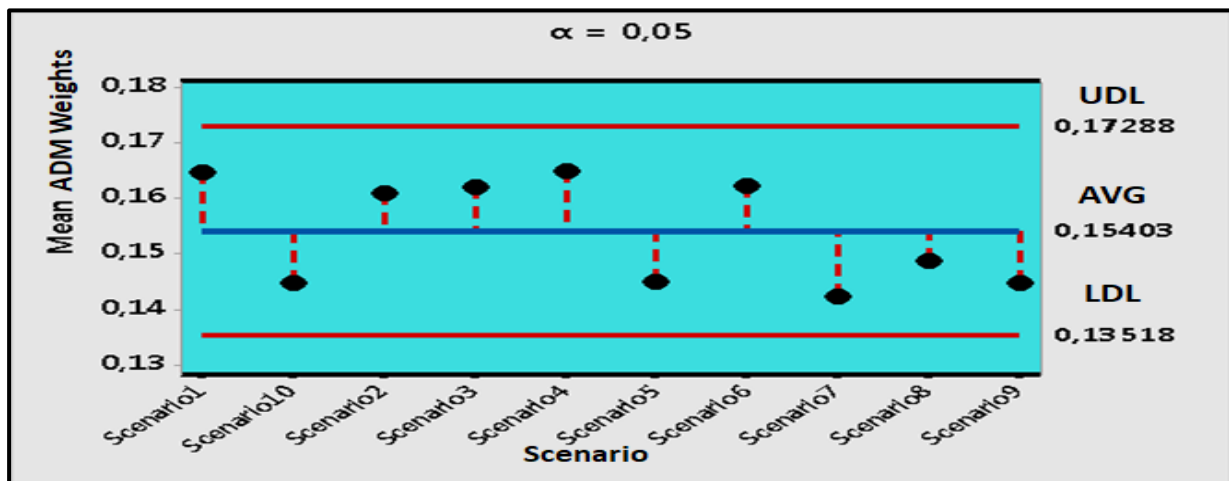


Figure 9. ADM visual-B (Second Category)

Observations from Figures 8 and 9 reveal that for both categories (the first category and the second category), the ADM values for each scenario fall within a range bounded by the UDL and LDL values. This positioning suggests uniformity in the variance of the weights determined for each scenario. This finding is corroborated by the Levene Test, with the essential statistics summarized in Tables 18 and 19 for each category.

**Table 18.** Levene test-A (First Category)

Levene Statistic	df1	df2	<i>p</i>
0,240	2	10	0,130

**p\*\*<.05**

**Table 19.** Levene test-B (Second Category)

Levene Statistic	df1	df2	<i>p</i>
0,298	2	10	0,163

**p\*\*<.05**

Statistical evidence from Tables 18 and 19 indicates that the p-values for both categories exceed the commonly accepted significance threshold of 0.05. This finding signifies that the variances observed in criterion weights across different scenarios exhibit a remarkable degree of homogeneity. In essence, the simulation analysis produces results that affirm the inherent robustness and consistent stability of the SDBHA method.

#### 4. Conclusion

Navigating intricate problems often necessitates a MCDM approach, where different, and often conflicting, factors come into play. Weighing these criteria fairly is paramount, as their relative importance significantly influences the outcome. Recognizing this, researchers have devised various methods for calculating weight coefficients, enriching the MCDM landscape. This study introduces Spearman Rank Correlation-based Expanded CRITIC (SRCBECM) technique.

In the study, the weight coefficients of the FIW criteria were first measured according to the SRCBECM method and the criterion weights were ranked as FIW5, FIW1, FIW3, FIW2, FIW7, FIW6, and FIW4. In the second stage of the study, the weights of the criteria were also calculated using the CRITIC, ENTROPY, SD, SVP, MEREC, and LOPCOW methods as part of the sensitivity analysis and the criteria were ranked. According to the findings, the FIW ranking determined within the scope of the SRCBECM method was completely different from the FIW criterion rankings within the scope of other objective weighting methods. Based on this result, the SRCBECM method was considered to be sensitive.

In the third stage, a comparative analysis was conducted between the results obtained by the SRCBECM method and the results within the scope of other objective weighting methods. When the findings were examined, it was determined that the characteristics of the criteria identified by the SRCBECM method were more pronounced than the other methods. In addition, the *rho* correlation coefficient was calculated between the FIW weights within the scope of the SRCBECM method and the FIW weights calculated by other objective criterion weighting methods. According to the findings, no significant positive correlations were observed between the FIW weights obtained by the SRCBECM method and the FIW weights calculated by other objective weighting methods. Although, in the literature, positive and significant relationships with other objective weighting methods are sought for the proposed criterion weighting method to be reliable and credible, it is thought that the SRCBECM method is reliable and credible in that it considers both separation and similarity conditions, which is different from other methods.

Fourthly, two categories were created with different values assigned to the FIW values of countries, providing 10 scenarios (data sets) for each category: 3 scenarios in the first group and 7 in the second group. Correlations between the SRCBECM method and other objective criterion weighting methods were evaluated based on these scenarios. The findings revealed that as the number of scenarios increased, the SRCBECM method diverged from the other methods. Additionally, the average variance values of the methods were calculated according to these scenarios. The results indicate that the SRCBECM method has a higher separation power than other methods, as its average variance values

---

exceed those of the other methods. Furthermore, the homogeneity test of the SRCBECM method (ADM) was applied across the 10 scenarios in both categories. According to the findings, it is concluded that the SRCBECM method exhibits homogeneous variances, thus demonstrating robustness and stability. The SRCBECM method offers several practical and theoretical benefits based on its unique features. In practical terms, its insensitivity to 0 and negative values enhances its reliability in real-world applications where such values may arise. This avoids the calculation issues seen in methods like ENTROPY or MEREC, resulting in smoother and more reliable computations. Furthermore, the method's flexibility in not assuming a normal distribution of data makes it applicable to a wider range of datasets, particularly in cases where non-parametric relationships are predominant. This allows for better performance across diverse data without the need for complex transformations. Another practical advantage is that the SRCBECM method balances the consideration of both divergence and similarity among criteria, leading to a more holistic assessment. This results in more accurate and meaningful weight assignments in decision-making processes. Additionally, the method emphasizes the characteristic differences between criteria, which helps in distinguishing the most important factors in complex decision-making environments.

## 5. Discussion

From a theoretical perspective, the SRCBECM method provides a robust alternative to parametric correlation measures by utilizing the rho correlation coefficient. This is especially advantageous when the assumption of normal distribution is not met, making the method more adaptable to a variety of data types. Its comprehensive approach to weighting logic, which includes both divergence and similarity among criteria, makes it more nuanced and reflective of complex relationships between variables compared to methods like ENTROPY and CRITIC, which focus solely on divergence. Finally, the SRCBECM method's ability to generate more distinctive weights based on the characteristic differences between criteria enhances its discriminative power, offering a more precise and granular insight into the decision-making process.

This method is considered an important tool for decision-makers in overcoming various challenges encountered in the criterion weighting process. In real-world scenarios, the presence of zero or negative values during the evaluation of criteria is a common issue, which can lead to invalid calculations in some methods. However, the SRCBECM method is insensitive to zero and negative values, helping to overcome such difficulties. Additionally, since this method does not require the assumption of a normal distribution, it offers a wide range of applicability across different datasets, reducing the complexity of the data. This provides decision-makers with a flexible and reliable solution for the various situations they may face. Furthermore, by evaluating both the similarities and differences among criteria in a balanced manner, it enables the determination of more accurate and meaningful weights. Therefore, the SRCBECM method can contribute to more effective outcomes in complex decision-making processes.

It is considered that the criterion weighting logic of the ENTROPY, SD, SVP, LOPCOW, and MEREC methods, which are other objective weighting methods, can be expanded and improved by considering the similarity logic of criteria in addition to the opposition logic, as in the DEMATEL method, in future research. It is considered that the criterion weighting logic of the ENTROPY, SD, SVP, LOPCOW, and MEREC methods, which are other objective weighting methods, can be expanded and improved by considering the similarity logic of criteria in addition to the opposition logic, as in the DEMATEL method, in future research.

**Author Contributions:** Since this is a single-author manuscript, F.A. was solely responsible for the conception, design, data collection, analysis, interpretation, and manuscript preparation.

**Funding:** This research did not receive external funding.

**Conflicts of interest:** The authors declare no conflict of interest.

## References

- [1] Saaty, T, L., "Decision making with the analytic hierarchy process", *International Journal of Services Sciences*, vol.1, no.1,

- pp.83-98, 2008.
- [2] Ecer, F., *Çok Kriterli Karar Verme*, Ankara: Seçkin Yayıncılık, 2020.
- [3] Kalaycı, Ş., *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*, Ankara: Anı Yayın Dağıtım, 2013.
- [4] Kilmen, S., *Eğitim Araştırmacıları İçin SPSS Uygulamalı İstatistik*, Ankara: Edge Akademi, 2015.
- [5] Karagöz, Y., *İlişki Katsayıları*, Ankara: Detay Yayıncılık, 2014.
- [6] Baş, F., *Çok Kriterli Karar Verme Yöntemlerinde Kriter Ağırlıklarının Belirlenmesi*, Ankara: Nobel Bilimsel, 2021.
- [7] Tepe, S., *Örnek Uygulamalarla Çok Kriterli Karar Verme Yöntemleri*, Ankara: Akademisyen Kitapevi, 2021.
- [8] Ayçin, E., *Çok Kriterli Karar Verme*, Ankara: Nobel Yayın, 2019.
- [9] Arslan, R., *CRITIC Yöntemi*, In: *Çok Kriterli Karar Verme Problemlerinde Kriter Ağırlıklandırma Yöntemleri*, Ankara: Nobel Akademik Yayıncılık, 117-134, 2020.
- [10] Dinçer, S. E., *Çok Kriterli Karar Alma*, Ankara: Gece Akademi, 2019.
- [11] Aytekin, A., Korucuk, S., Bedirhanoglu, Ş. B., and Simic, V. (2024). Selecting the ideal sustainable green strategy for logistics companies using a T-spherical fuzzy-based methodology. *Engineering Applications of Artificial Intelligence*, 127, 107347. doi: 10.1016/j.engappai.2023.107347
- [12] Korucuk, S., Demir, E., Karamaşa, Ç., and Stevic, Z. (2020). Determining the dimensions of the innovation ability in logistics sector by using plithogenic-critic method: An application in Sakarya Province. *International review*, 1-2. Doi:
- [13] Zavadskas, E. K. and Podvezko, V., "Integrated determination of objective criteria weights in MCMD", *International Journal of Information Technology & Decision Making*, vol 15, no.2, pp. 267-283, January 2016. doi:10.1142/S0219622016500036
- [14] Sel, A., *IDOCRIW Yöntemi*, In: *Çok Kriterli Karar Verme Problemlerinde Kriter Ağırlıklandırma Yöntemleri*, Ankara: Nobel Akademik Yayıncılık, pp. 37-50, 2020.
- [15] Nasser, A. A., Alkhalaidi, M. N., Hankal, A. M. and Al-olofe, M., "A weighted euclidean distance - statistical variance procedure based approach for improving the healthcare decision making system in Yemen", *Indian Journal of Science and Technology*, vol.12, no.3, pp.1-15, January 2019. doi: 10.17485/ijst/2019/v12i3/140661
- [16] Gülençer, İ. and Türkoğlu, S. P., "Gelişmekte olan Asya ve Avrupa ülkelerinin finansal gelişmişlik performansının istatistiksel varyans prosedürü temelli OCRA yöntemiyle analizi," *Üçüncü Sektör Sosyal Ekonomi Dergisi*, vol.55, no.2, pp. 1330-1344, 2020. doi:10.15659/3.sektor-sosyal-ekonomi.20.06.1375
- [17] Odu, G. O., "Weighting methods for multi-criteria decision making technique", *J. Appl. Sci. Environ. Manage*, vol.28, no.3, pp. 1449-1457, September 2019. doi: 10.4314/jasem.v23i8.7
- [18] Demir, G., Özyalçın, T. and Bircan, H., *Çok Kriterli Karar Verme Yöntemleri ve ÇKKV Yazılımı ile Problem Çözümü*, Ankara: Nobel Yayıncılık, 2021.
- [19] Uludağ, A. S. and Doğan, H., *Üretim Yönetiminde Çok Kriterli Karar Verme*, Ankara: Nobel Yayıncılık 2021.
- [20] Keshavarz-Ghorabae, M., Amiri, M., Zavadskas, E. K., Turskis, Z. and Antucheviciene, J., "Determination of objective weights using a new method based on the removal effects of criteria (MEREC)", *Symmetry*, vol.13, pp. 1-20, March 2021. doi: 10.3390/sym13040525
- [21] Ecer, F. and Pamucar, D., "A novel LOPCOW-DOBI multi-criteria sustainability performance assessment methodology: An application in developing country banking sector", *Omega*, vol.112, pp. 1-17, October 2022. doi: 10.1016/j.omega.2022.102690
- [22] Keshavarz-Ghorabae, M., Amiri, M., E., Zavadskas, E. K., Turskis, Z. and Antucheviciene, J., "Simultaneous evaluation of criteria and alternatives (SECA) for multi-criteria decision-making", *Informatica*, 29(2):265-280, July 2018. doi: 10.15388/Informatica.2018.167
- [23] Keleş, N., *Uygulamalarla Klasik ve Güncel Karar Verme Yöntemleri*, Ankara: Nobel Bilimsel, 2023.
- [24] Paksoy, S., *Çok Kriterli Karar Vermede Güncel Yaklaşımlar*, Ankara: Karahan Kitapevi, 2017.
- [25] Kaya, İ. and Karaşan, A., *Çok Kriterli Karar Verme*, İstanbul: Umuttepe Yayınları, 2020.
- [26] Fontela, E. and Gobus, A., *The DEMATEL Observer*, Geneva: Battelle Geneva Research Center, 1976.
- [27] Güçlü, İ., *Sosyal Bilimlerde Nicel Veri Analizi*, Ankara: Gazi Kitapevi, 2020.
- [28] Hayran, O. and Özbek, H., *Sağlık Bilimlerinde Araştırma ve İstatistik Yöntemler*, İstanbul: Nobel Tıp Kitapevi, 2017.
- [29] Yaratan, H., *Sosyal Bilimler İçin Temel İstatistik*, Ankara: Anı Yayıncılık, 2017.
- [30] Demir, İ., *SPSS İle İstatistik Rehberi*, İstanbul: Efe Akademi Yayınları, 2020.
- [31] Çelikbilek, Y., *Çok Kriterli Karar Verme Yöntemleri*, Ankara: Nobel Akademik Yayıncılık, 2018.
- [32] Öztel, A. and Alp, İ., *Çok Kriterli Karar Verme Seçiminde Yeni Bir Yaklaşım*, İstanbul: Kriter Yayıncılık, 2020.
- [33] Bayram Arlı, N. and Aytaç, M., *Sosyal Bilimlerde Parametrik Olmayan İstatistiksel Yöntemler*, Bursa: Ezgi Kitapevi, 2018.
- [34] Karagöz, Y., *SPSS 21.1 Uygulamalı İstatistik: Tıp, Eczacılık, Diş Hekimliği ve Sağlık Bilimleri İçin*, Ankara: Nobel Akademik Yayıncılık, 2014.
- [35] Karagöz, Y., *SPSS ve AMOS Uygulamalı Nitel-Nicel Karma Bilimsel Araştırma Yöntemleri ve Yayın Etiği*, Ankara: Nobel Akademik Yayıncılık, 2016.

- [36] Karagöz, Y., *SPSS ve AMOS 23 Uygulamalı İstatistiksel Analizler*, Ankara: Nobel Akademik Yayıncılık, 2017.
- [37] Özdamar, Ö., *Paket Programlar ile İstatistiksel Veri Analizi*, Ankara: Nisan Kitapevi, 2013.
- [38] Çoşkun, R., Altunışık, R. and Yıldırım, E., *Sosyal Bilimlerde Araştırma Yöntemleri-SPSS Uygulamalı*, Sakarya: Sakarya Yayıncılık, 2020.
- [39] Purba, L. I., Humaidi, S., Darmawan Y., Zahedi and Nasution, T. K., "Analysis of Spearman rank correlation & linear regression of atmospheric stability and cloud tops temperature of himawari-8 ir satellite images", *Jurnal Pengkajian Ilmu dan*, vol. 11, no.2, pp. 476-484, April 2023, doi: 10.33394/j-ps.v11i2.7818.
- [40] Raeissi, P., Zandian, H., Mirzarahimy T., Delavari, S., Moghadam, T. Z. and G. Rahimi, "Relationship between communication skills and emotional intelligence among nurses", *Nursing Management*, vol.26, no.2, pp. 1-10, March 2019, doi: 10.7748/nm.2019.e1820.
- [41] Pires, S. and Mascarenhas, C., "Cyber threat analysis using pearson and spearman correlation via exploratory data analysis", ICSCCC, Jalandhar. [Third International Conference on Secure Cyber Computing and Communication, p. 257-262, 2023]
- [42] Güler, K., Korelasyon Analizi, In: Psikologlar için spss uygulamaları ve araştırma yöntemleri, Nobel Akademik, Ankara, pp.203-206, 2020.
- [43] Cantarella, G., Alde, M., Consonni, D., Zuccotti, G., Di Berardino, F., Barozzi, S., Bertoli, S., Battezzati, A., Zanetti, D. and Pignataro, L., "Prevalence of dysphonia in non hospitalized patients with covid-19 in Lombardy, the Italian epicenter of the pandemictaged end", *Journal of Voice*, vol.37, no.4 605-509, March 2021, doi: 10.1016/j.jvoice.2021.03.009.
- [44] He, S., Zhang, Q., Wu, F., Chen, J., He, S., Ji, Z., Li, B., Gao, L., Xie, Q. and Zhang, J., "Influence of cigarettes on myocardial injury in healthy population after exposure to high altitude over 5000 m", *Science of The Total Environment*, vol.855, pp. 1-10, September 2023, doi: 10.1016/j.scitotenv.2022.158824.
- [45] Segar, D. E., Pan, A. Y., McLennan, D. I., Kindel, S. J., Handler, S. S., Ginde, S., Woods, R., K., Goot, B. H., and Spearman, A. D., "Clinical variables associated with pre fontan aortopulmonary collateral burden", *Pediatric Cardiology*, vol. 44, pp. 228-236, September 2023, doi: 10.1007/s00246-022-03014-8.
- [46] Kons, R., Ache-Dias, J., Rodrigo, G. G., Bishop, C. V. and Detanico, D., "Bilateral deficit in the countermovement jump and its associations with judo-specific performance", *Research in Sports Medicine*, vol.31, no.5, pp. 638-649, December 2023, doi: 10.1080/15438627.2021.2024542.
- [47] Popița, A. R., Rusu, A., Muntean, V., P. Cadariu, P. A., Irimie, A., Lisencu, C., Pop, B., Resiga, L., Zsolt, F. and Badea, R., "Preoperative MRI accuracy after neoadjuvant chemoradiation for locally advanced rectal cancer", *Medicine Pharmacy Reports*, vol.96, no.3, pp. 258-268, July 2023, doi: 10.15386/mp-2542.
- [48] Jiang, S., Xu, H., Zhao, C., Zhong, F. and Li, D., "Oyster polysaccharides relieve DSS-induced colitis via anti-inflammatory and maintaining the physiological hypoxia", *International Journal of Biological Macromolecules*, vol.238, 2023, doi: 10.1016/j.ijbiomac.2023.124150.
- [49] Wu, X., Chi, F., Wang, B., Liu, S., Wang, F., Wang, J., Tang, X., Bi, X., Lin, X. and Li, J., "Relationship between preoperative neutrophil-to-lymphocyte ratio and postoperative delirium: The PNDABLE and the PNDRFAP cohort studies", *Brain and Behavior*, vol.13, p.1-11, October 2023, doi: org/10.1002/brb3.3281.
- [50] Ulutaş, A. and Topal, A., *Bütünleştirilmiş Çok Kriterli Karar Verme Yöntemlerinin Üretim Sektörü Uygulamaları*, Ankara: Akademisyen Kitapevi, 2020.
- [51] Krishankumar, R., Mishra, A. R., Rani, P., Ecer, F. and Ravichandran, K. S., "Assessment of zero-carbon measures for sustainable transportation in smart cities: A CRITIC-MARCOS frameworkbased on q-rung fuzzy preferences", *IEEE Internet of Things Journal*, vol.10, no.21, pp.18651-18662, November 2023, doi: 10.1109/JIOT.2023.3293513.
- [52] Hassan, I., Alhamrouni, I. and Azhan, N. H., "A CRITIC-TOPSIS multi-criteria decision-making approach for optimum site selection for solar pv farm", *Energies*, vol.16, no.10, pp. 1-26, May 2023, doi: 10.3390/en16104245
- [53] Bhardwaj, R. and Garg, S., "An MCDM approach to analytically identify the air pollutants impact on health", *Atmosphere*, vol.14, no.6, pp. 1-17, May 2023, doi: 10.3390/atmos14060909 .
- [54] Silva, N. F., dos Santos, M., Gomes, C. F. S. and de Andrade, I. P., "An integrated CRITIC and Grey Relational Analysis approach for investment portfolio selection," *Decision Analytics Journal*, vol.8, pp. :1-8, July 2023, doi: 10.1016/j.dajour.2023.100285
- [55] Yazid, N. A., Sabtu, N. I., Azmiral, N. U. S. and Mahad, N. F., "The application of CRITIC-TOPSIS method in solving the material handling equipment selection problem", *Malaysian Journal of Computing*, vol. 8, no.1, pp. 1310-1330, March 2023, doi: 10.24191/mjoc.v8i1.16976.
- [56] Ertemel, A. V., Menekse, A. and Akdag, H. C., "Smartphone addiction assessment using pythagorean fuzzy CRITIC-TOPSIS", *Sustainability*, vol.15, no.5, pp. :1-19, February 2023, doi: 10.3390/su15053955.
- [57] Mallick, R., Pramanik, S. and Giri, B. C., "Neutrosophic magdm based on CRITIC-EDAS strategy using geometric aggregation operator", *Yugoslav Journal of Operations Research*, vol. 33, no.4, pp. 683-698, July 2023, doi: 0.2298/YJOR221017016M.
- [58] Gaur, S., Dosapati, S. and Tawalare, A., "Stakeholder assessment in construction projects using a CRITIC-TOPSIS approach", *Built Environment Project and Asset Management*, vol.13, no.2, pp.217-237, August 2023, doi:



---

10.1108/BEPAM-10-2021-0122.

- [59] Kara, K., Yalçın, G. C. and Edinsel, S., "Warehouse manager selection by CRITIC-MULTIMOORA hybrid method based on single-valued neutrosophic sets", *Deniz Taşımacılığı ve Lojistiği Dergisi*, vol.4, no.1, pp. 48-64, January 2023, doi: 10.52602/mtl.1126012.
- [60] Brodny, J., "Assessing the energy security of European Union countries from two perspectives: A new integrated approach based on MCDM methods", *Applied Energy*, vol.347, pp. 1-26, October 2023, doi: 10.1016/j.apenergy.2023.121443.
- [61] Witold, W. C., "The Possibilities on the use of the Spearman correlation coefficient", *International Electronic Scientific and Practical Journal*, vol.5, no.1, pp. 151-165, July 2022.
- [62] Russell, J. and Coxs, R., *Spearman Rank Correlation*, Norderstedt: Book on Demand GmbH, 2012.
- [63] Gigović, L., Pamučar, D., Bajič, Z. and Milicevič M., "The Combination of expert judgment and GIS-MAIRCA analysis for the selection of sites for ammunition depots", *Sustainability*, vol.8, no.232, pp. 1-30, April 2016, doi: 10.3390/su8040372.

Research Article

Received: date:27.06.2024  
Accepted: date:14.08.2024  
Published: date:31.12.2024

# Mühendislik Öğrencilerinin Başarılarını Etkileyen Kritik Faktörlerin Yapısal Eşitlik Modellemesi ile Belirlenmesi

Delal Demir<sup>1\*</sup>, Hasan Şahin<sup>2</sup>

<sup>1</sup>Bursa Technical University, Department of Industrial Engineering, Bursa Türkiye; delal99sk@gmail.com

<sup>2</sup>Bursa Technical University, Department of Industrial Engineering, Bursa Türkiye; h.sahin@btu.edu.tr

Orcid: 0009-0006-9922-2325<sup>1</sup> Orcid: 0000-0002-8915-000X<sup>2</sup>

\*Correspondence: delal99sk@gmail.com

**Öz:** Mühendislik programlarının, günümüz küresel işgücü piyasasının taleplerini karşılayacak mezunlar yetiştirmedeki zorlukları göz önüne alındığında, bu araştırma, öğrenci başarısını etkileyen kritik faktörleri incelemiştir. Çalışma, Bursa Teknik Üniversitesi'nden 500 mühendislik öğrencisinin katılımıyla anket verilerini toplamış ve Yapısal Eşitlik Modellemesi (YEM) kullanarak analiz etmiştir. Deneyimsel öğrenme, akademik öz-yeterlik, kültürel zekâ ve algılanan istihdam edilebilirliğin, öğrenci başarısı üzerindeki etkisi incelenmiştir. Sonuçlar, bu dört faktörün tümünün öğrenci başarısıyla pozitif bir şekilde ilişkili olduğunu göstermiştir. Deneyimsel öğrenmenin, diğer üç faktör üzerinde özellikle güçlü bir etkisi olduğu bulunmuştur. Bu bulgular, mühendislik programlarının öğrencilere deneyimsel öğrenme fırsatları sunmasını, öz-yeterliklerini ve kültürel zekâlarını geliştirmenin ve istihdam edilebilirliklerini artırmanın önemini vurgulamaktadır. Bu faktörleri eğitim müfredatlarına entegre etmek, mezunların küresel işgücü piyasasında rekabet edebilirliğini artırabilir.

**Anahtar kelimeler:** Mühendislik eğitimi, yapısal eşitlik modellemesi, öz-yeterlik, kültürel zekâ, istihdam edilebilirlik, deneyimsel öğrenme

## Determination of Critical Factors Affecting The Success of Engineering Students by Structural Equation Modeling

**Abstract:** Given the challenges engineering programs face in producing graduates who meet the demands of today's global workforce, this research investigated the critical factors influencing student success. The study utilized a survey methodology, collecting data from 500 engineering students at Bursa Technical University, and employed Structural Equation Modeling (SEM) for analysis. The impact of experiential learning, academic self-efficacy, cultural intelligence, and perceived employability on student success was examined. The results indicated that all four factors are positively associated with student success, with experiential learning demonstrating a particularly strong influence on the other three factors. These findings highlight the importance of engineering programs providing students with experiential learning opportunities to enhance their self-efficacy, cultural intelligence, and ultimately, their employability. Integrating these factors into educational curricula can better prepare graduates to compete in the global workforce.

**Keywords:** Engineering education, structural equation modeling, self-efficacy, cultural intelligence, employability, experiential learning.

### 1. Giriş

Mühendislik eğitimi, küresel işgücü piyasasının artan taleplerine yanıt verebilecek yetkin ve donanımlı mezunlar yetiştirmeyi hedefler [1]. Ancak, bu hedefe ulaşmada çeşitli zorluklarla karşılaşmaktadır. Özellikle, mühendislik öğrencilerinin akademik ve profesyonel başarılarını etkileyen kritik faktörlerin belirlenmesi, eğitim programlarının daha etkili hale getirilmesi açısından büyük önem taşır. Bu bağlamda, deneyimsel

öğrenme, akademik öz-yeterlik, kültürel zekâ ve algılanan istihdam edilebilirlik gibi faktörlerin öğrenci başarısı üzerindeki etkilerini anlamak, mühendislik eğitiminde yenilikçi yaklaşımlar geliştirilmesine yardımcı olabilir. Genel olarak, mühendislik eğitimi sadece teknik bilgi ve becerilere odaklanırken, günümüz iş dünyasının gereksinimleri doğrultusunda öğrencilere kültürel zekâ, iletişim yetenekleri, liderlik becerileri gibi meta-yeterliklerin de kazandırılması önem kazanmaktadır[1]. Deneyimsel öğrenme fırsatlarının artırılması, öğrencilerin öz-yeterlik ve kültürel zekâlarının geliştirilmesi, ve iş bulma becerilerinin güçlendirilmesi, mezunların küresel işgücü piyasasında rekabet edebilirliğini artırmada kritik rol oynayacaktır. Araştırmanın temel odak noktaları deneyimsel öğrenme, akademik öz-yeterlik, kültürel zekâ ve algılanan istihdam edilebilirlik gibi mühendislik öğrencilerinin başarı düzeylerini etkileyen önemli unsurların belirlenmesidir [3]. Bu faktörlerin seçimi, güncel literatür taraması ve uzman görüşleri doğrultusunda yapılmıştır. Belirlenen faktörlerin mühendislik eğitimindeki akademik kalite kültürüne nasıl katkı sağladığı incelenmiş ve öğrencilerin eğitim deneyimlerinin kalitesini artırmada bu faktörlerin rolü değerlendirilmiştir [4]. Günümüz yükseköğretim kurumlarının, öğrencileri teknik bilgilerin yanı sıra genel becerilerle donatma ihtiyacı vurgulanmış ve eğitim programlarının ve öğretim yöntemlerinin yeniden düzenlenmesi için stratejiler önerilmiştir [5]. Bu çalışmada mühendislik öğrencilerinin başarılarını etkileyen kritik faktörleri Yapısal Eşitlik Modellemesi (YEM) kullanarak incelenmiştir [2]. Yapısal Eşitlik Modellemesi (YEM) ve Kısmi En Küçük Kareler (PLS) teknikleri kullanılarak önerilen yapısal modelin geçerliliği test edilebilmektedir [6]. YEM'in tercih edilme nedeni, karmaşık değişkenler arasındaki ilişkileri bütüncül olarak analiz etme kabiliyetidir. Ayrıca çalışmada modelin mühendislik eğitiminde başarıyı artırmak için nasıl kullanılabileceğine dair öneriler de sunulmuştur.

## 2. Literatür Taraması

Akademik başarı, öğrencilerin gelecekteki profesyonel yaşamlarında etkinlik gösterebilmeleri için temel bir öncül olarak değerlendirilmektedir. Bu bağlamda, literatürde başarı kriterleri üzerine yürütülen araştırmalar, öğrencilere kariyer meta-yetenekleri veya Kritik Başarı Faktörleri (KBF) kazandırmanın önemini vurgulamaktadır. Çok sayıda araştırmacı, eğitim deneyimlerinin, öğrencilerin profesyonel alanda başarıya ulaşabilmeleri için gerekli olan kariyer meta-yeteneklerinin edinilmesinde kritik bir işlev üstlendiğini ileri sürmektedir [7], [8], [9].

Bununla birlikte, gerçekleştirilen kapsamlı literatür taraması, akademik camiada öğrenci başarısını tanımlama ve ölçme konusunda bir fikir birliğinin bulunmadığını ortaya koymaktadır. Bu çerçevede, mevcut çalışmanın temel amacı, mühendislik öğrencileri için kariyer meta-yeteneklerinin belirlenmesi hususundaki bu epistemolojik belirsizliği ele almak ve bu alanda bütüncül bir kavrayış geliştirmektir [10]. Aşağıda çalışmada kullanılan faktörlerin tanımları yer almaktadır:

- Deneyimsel Öğrenme: Öğrencilerin teorik bilgileri pratik deneyimlerle birleştirerek öğrenmelerini sağlayan bir öğrenme yöntemidir. Archida & Vigil, öğrencilerin iş dünyasındaki gereksinimlere uygun olarak yetişmelerini sağlamak amacıyla Deneyimsel Öğrenmenin önemini vurgular [33]. Çalışma, öğrencilerin saha deneyimleri ve stajların kariyer gelişimlerine nasıl katkı sağladığını incelemiştir. Öz-Yeterlik: Bireyin belirli bir görevi başarma konusundaki inanç düzeyidir. İbrahim, öğrencilerin Öz-Yeterlik düzeylerinin kariyer başarısı üzerindeki etkisini araştırmıştır [34]. Öğrencilerin öz-yeterlik algılarının, kariyer hedeflerine ulaşmada ve zorluklarla başa çıkmada ne kadar etkili olduğunu analiz etmiştir.
- Kültürel Zekâ: Farklı kültürlerle etkileşimde bulunma ve bu etkileşimleri yönetme yeteneğidir.
- Kendi Algılanan İstihdam Edilebilirlik: Bireyin kendi iş bulunabilirlik düzeyine dair algısıdır.

Bu kavramlar arasındaki ilişkiler, mühendislik öğrencilerinin kariyer başarılarını belirlemede kritik rol oynayan faktörleri anlamamıza katkı sağlamak amacıyla detaylı bir şekilde değerlendirilmiştir.

Araştırma kapsamında, ilgili literatürde belirtilen başarı faktörleri ve kritik başarı faktörleri sistematik bir şekilde incelenmiş ve analiz edilmiştir. Bu analiz, her bir çalışmanın metodolojik yaklaşımını, örneklem karakteristiklerini ve temel bulgularını içeren kapsamlı bir değerlendirmeyi içermektedir. Bu sayede, mühendislik eğitiminde öğrenci başarısını etkileyen faktörlerin çok boyutlu bir perspektiften ele alınması ve bu faktörler arasındaki karmaşık ilişkilerin Yapısal Eşitlik Modellemesi (YEM) gibi ileri analitik teknikler kullanılarak incelenmesi hedeflenmektedir.

## 2.1. Eğitim Deneyimleri ve Kariyer Başarısı

Coetzee ve Harry'nin çalışması, eğitim deneyimlerinin öğrencilere kariyer başarısı için temel bir zemin sağlaması gerektiğini vurgular [11]. Bu araştırma, eğitim kalitesinin önemini genel olarak ele alırken, belirli başarı faktörleri üzerinde derinleşmemektedir. Ancak, çalışma mühendislik eğitiminin kalitesinin, öğrencilerin gelecekteki kariyer başarılarıyla doğrudan ilişkili olduğunu öne sürmektedir.

Coetzee ve Schreuder tarafından yapılan araştırma ise, öğrenci başarısının belirlenmesinde öğretim kalitesi ve öğrenci memnuniyetinin önemini ele alır [12]. Bu çalışma, iyi bir eğitim deneyiminin kritik başarı faktörleri arasında olduğunu belirtirken, özellikle mühendislik eğitiminde öğretim kalitesinin öğrenci memnuniyeti üzerindeki etkisini vurgular. Araştırmacılar, yüksek kaliteli öğretimin, öğrencilerin motivasyonunu artırdığını ve dolayısıyla akademik performanslarını iyileştirdiğini savunmaktadır.

Brown ve arkadaşlarının çalışması, yüksek öğretim kurumlarının performans göstergeleri olarak kullandığı öğrenci deneyimi ve mezun çıktıkları arasındaki ilişkinin sorgulanması gerektiğini savunuyor [7]. Bu araştırma, Avustralya Hükümeti tarafından toplanan veriler kullanılarak, yüksek öğretim kalitesi ile mezunların iş bulma olasılığı arasındaki ilişkiyi araştırıyor. Çalışmada, Öğrenci Deneyimi Anketi (CEQ) ölçeğinin alt ölçekleri olan İyi Öğretim (GTS), Mezun Becerileri (GSS) ve Mezun Nitelikleri (GQS) inceleniyor. Araştırmacılar, GTS'nin mezunların genel ders memnuniyetini tahmin ettiğini ancak GSS ve GQS'nin birleşik bir şekilde mezunların iş bulma olasılıklarına ihmal edilebilir düzeyde etki ettiğini keşfediyorlar. Bu bulgular, yüksek öğretim liderlerinin işe yerleştirme becerileri ile iş bulma arasındaki ilişki hakkındaki söylemi dikkatlice incelemelerini öne çıkarıyor[13].

## 2.2. Bireysel Yetenekler ve Kişisel Hedefler

Papadopoulos ve Armatas'ın çalışması, öğrenci başarısının ölçümünde bireysel yeteneklerin ve kişisel hedeflerin kritik rol oynadığını belirtir [14]. Bu çalışma, öğrencinin kendi hedeflerine ulaşmasına yardımcı olacak faktörler üzerinde durur. Araştırmacılar, mühendislik öğrencilerinin kişisel hedeflerini belirlemelerinin ve bu hedeflere ulaşmak için gerekli yetenekleri geliştirmelerinin önemini vurgular. Çalışma, öğrencilerin kendi güçlü ve zayıf yönlerini anlamalarının, kariyer planlaması ve başarısı açısından kritik olduğunu savunmaktadır.

Miller, Fassett ve Palmer'in çalışması, yüksek öğrenimdeki öğrenci katılımını tahmin etmede başarı hedef yönelimlerinin rolünü araştırmaktadır. Çalışma, bireysel başarı hedef yönelimlerinin, öğrencinin yüksek öğrenimdeki katılım davranışlarını tahmin etmede önemli bir rol oynadığını göstermektedir. Çalışma ayrıca, öğrenci katılımını artırmak için yüksek öğrenimdeki öğretim üyelerinin ve idari kadronun öğrenci hedef yönelimlerine dikkat etmesi gerektiğini savunmaktadır. Bu araştırma, yüksek öğrenimde öğrenci katılımı ile ilgili önemli bir konuyu ele almaktadır ve yüksek öğrenimdeki öğrencilerin başarı hedeflerine ulaşmaları ve hedeflerini belirlemeleri için gerekli kaynaklar sağlayarak akademik gelişimlerini iyileştirmek için pratik öneriler sunmaktadır [15].

## 2.3. İş Bulma ve İşte Kalma Yetenekleri

Literatürde, iş bulma ve işte kalma yetenekleri ile ilgili çeşitli konseptler ve teoriler ele alınmıştır. Örneğin, Campos'un çalışması, öğrencilerin sınıf içinde edindikleri bilgi, beceri ve kariyerle ilgili meta-yetenekleri iş piyasasındaki durumlara nasıl aktarabileceğini değerlendirmiştir[7]. Bu araştırma, mühendislik öğrencilerinin teorik bilgilerini pratik uygulamalara dönüştürme yeteneklerinin önemini vurgular.

Clark'ın çalışması, İngiltere ve Avustralya'daki yükseköğretim kurumlarının, öğrencilerin sınıf içinde edindikleri yetenekleri iş yaşamına nasıl transfer edebileceklerini ele almıştır [16]. Bu araştırma, mühendislik eğitiminde kazanılan becerilerin iş dünyasında nasıl uygulanabileceğine dair önemli içgörüler sunmaktadır.

Yükseköğretimde akademik beceri geliştirme ve iş bulma becerileri arasındaki ilişki üzerine yapılan araştırmalar, öğrencilerin iş hayatına hazırlanmasında üniversitelerin rolüne odaklanmaktadır. Menz'in çalışması[17], öğrencilerin eleştirel düşünme, problem çözme ve etkili iletişim gibi akademik becerilerini geliştirerek işverenlerin beklentilerini karşılamalarına yardımcı olmak için müfredata zorunlu beceri geliştirme modülleri ve öğrenme günlükleri eklenmesini önermektedir . Bu araştırma, akademik becerilerin iş bulma becerileriyle doğrudan bağlantılı olduğunu ve üniversitelerin bu becerileri geliştirmede aktif rol oynamaları gerektiğini vurgular.

## 2.4. Deneyimsel Öğrenme ve Kariyer Gelişimi

Deneyimsel öğrenme, mühendislik öğrencilerinin kariyer gelişiminde önemli bir rol oynamaktadır. Amant'ın çalışması, deneyimsel öğrenmenin öğrenci başarısı üzerindeki etkilerini inceler ve öğrencilerin farklı bakış açılarından deneyimlediği öğrenme süreçlerini analiz eder [18]. Bu araştırma, mühendislik öğrencilerinin pratik deneyimler yoluyla nasıl daha etkili öğrendiklerini ve bu deneyimlerin kariyer başarısına nasıl katkıda bulunduğunu göstermektedir.

Andrew ve Garcia'nın çalışması, deneyimsel öğrenmenin öğrencilerin kariyer gelişimi üzerindeki etkilerini değerlendirir ve öğrencilerin işyeri deneyimlerinin kariyer hazırlıklarındaki rolünü araştırır[19]. Bu çalışma, mühendislik öğrencilerinin staj ve iş deneyimlerinin, onların gelecekteki kariyer başarılarını nasıl etkilediğini ortaya koymaktadır.

Arslan'ın çalışması, deneyimsel öğrenmenin öğrencilerin kişisel gelişimi üzerindeki etkilerini inceler ve öğrencilerin deneyimledikleri olaylardan nasıl öğrendiklerini ele alır [20]. Bu araştırma, mühendislik öğrencilerinin pratik deneyimler yoluyla nasıl kişisel ve profesyonel olarak geliştiklerini göstermektedir.

Bass'ın çalışması, deneyimsel öğrenmenin liderlik becerileri üzerindeki etkilerini analiz eder ve öğrencilerin liderlik deneyimleriyle kazandıkları bilgileri değerlendirir [21]. Bu araştırma, mühendislik öğrencilerinin liderlik becerilerini geliştirmelerinin kariyer başarısı açısından önemini vurgular.

Booth'un çalışması, deneyimsel öğrenmenin öğrencilerin eleştirel düşünme yetenekleri üzerindeki etkilerini inceler ve öğrencilerin deneyimlerinden nasıl eleştirel düşündüklerini araştırır[22]. Bu çalışma, mühendislik öğrencilerinin analitik ve problem çözme becerilerini geliştirmelerinin kariyer başarısı için kritik olduğunu göstermektedir.

Segarra ve Gentry'nin çalışması, deneyimsel öğrenmenin akademik kariyer gelişimi üzerindeki etkilerini incelemekte ve bireylerin kendi profesyonel gelişimlerini yönetmelerinin önemini vurgulamaktadır [23]. Bu araştırma, Amerikan Hücre Biyolojisi Derneği'nin (ASCB) Kariyer Geçişlerini Gerçekleştirme (ACT) Programı'nın deneyimsel öğrenme bileşenini ele almaktadır. Çalışma, doktora sonrası araştırmacıların ve yeni öğretim üyelerinin, özellikle STEM alanlarında yetersiz temsil edilen gruplardan gelenlerin, akademik pozisyonlara geçiş ve kadro alma süreçlerinde ihtiyaç duydukları becerileri geliştirmelerine yardımcı olmak için bireyselleştirilmiş profesyonel gelişim deneyimlerinin (practicum) tasarlanmasının ve uygulanmasının faydalarını ortaya koymaktadır.

## 2.5. Öz Yeterlilik ve Akademik Başarı

Öz yeterlilik, bireyin bir hedefe ulaşma konusundaki kişisel inancını ifade eder. Bandura'nın tanımına göre[24], öz yeterlilik bir davranışı başarıyla gerçekleştirebileceğine dair bir inançtır ve bu inanç, davranış seçimini etkiler. Bandura ise öz yeterliliği, bir kişinin bir hedefe ulaşmak için gerekli olan eylemleri düzenleme ve gerçekleştirme yeteneği olarak tanımlamıştır.

Adediwura, öz yeterliliğinin bir bireyin öğrenme ve performansla başa çıkma kapasitesi üzerinde önemli bir ölçü olduğunu rapor etmiştir [25]. Avustralya'daki birinci sınıf üniversite öğrencileri arasında yapılan çalışmalar, öz yeterliliğinin devamlılık, retansiyon, öğrenci memnuniyeti ve akademik performansın bir göstergesi olarak kabul edildiğini göstermiştir. Bu nedenle öğrenci öz yeterliliğinin eksikliğinin öğrenci memnuniyetini, öğrenmeye devam etmeme ve akademik başarıyı olumsuz etkileyebileceği sonucuna varılmıştır.

Honick ve arkadaşları çalışmalarında[26], Bandura'nın öz-yeterlik ve akademik başarı arasındaki karşılıklı ilişki teorisini incelemektedir. Araştırma, bu ilişkiyi etkileyen faktörleri ve özellikle başlangıç performansı ile görev zorluğunun zaman içindeki etkilerini araştırmaktadır. Çalışma, 261 üniversite öğrencisi üzerinde gerçekleştirilmiş ve şu temel bulguları ortaya koymuştur:

Başarıdan öz-yeterliğe doğru anlamlı pozitif yollar bulunmuştur, ancak öz-yeterlikten performansa doğru anlamlı bir ilişki gözlenmemiştir. Bu durum, karşılıklı etkileşim hipotezini tam olarak desteklememektedir.

Tekrarlı ölçümler ANOVA analizi, başlangıçtaki görev performansının, görev zorluğu ve zaman içindeki başarı arasındaki etkileşim üzerinde düzenleyici bir etkisi olduğunu göstermiştir. Bulgular, özellikle başlangıçta düşük performans gösteren öğrenciler için, akademik performansı kademeli ustalık yoluyla artırmanın önemini vurgulamaktadır. Bu yaklaşım, öğrenilen beceriler için öz-yeterliliği geliştirmede etkili olabilir. Bu çalışma, akademik başarı ve öz-yeterlilik arasındaki ilişkinin karmaşıklığını ortaya koymakta ve eğitimcilere, öğrencilerin performansını ve öz-yeterliliklerini geliştirmek için stratejiler sunmaktadır.

## 2.6. Kültürel Yetkinlik ve Küresel Rekabet Gücü

Gelecekteki mezunları daha iyi hazırlamak amacıyla öğretim ve öğrenme süreçlerine entegre edilecek temel nitelikleri belirlemeye yönelik literatürde çeşitli teoriler ve terimler bulunmaktadır. Hunter, küresel yetkinliğin diğer kültürlerle etkileşimi teşvik eden önemli bir kavram olduğunu belirtmiştir [27]. Downey, mühendislik eğitiminde öğrenmeye küresel yetkinlik eklemenin önemini ele almıştır[28].

Vitto, kültürel yetkinliğin artık sadece güzel bir beceri olmadığını, aynı zamanda ekonomik bir zorunluluk haline geldiğini belirtmiştir. Jesiek, "cross-cultural competence" terimini kullanarak, gelecekteki mühendislerin etkili bir şekilde farklı ülkeler ve kültürler arasında hareket etmeleri gerektiğine dair farkındalık oluşturmanın önemine vurgu yapmıştır[29].

Sternberg, "cross-culturally Intelligence" terimini kullanarak, çevreye etkili bir şekilde uyum sağlama yeteneğini ifade etmiştir. Bu çalışmalar, öğrencilerin sadece teknik becerilerini değil, aynı zamanda kültürel ve global yetkinlikleri de geliştirmelerinin önemine dikkat çekmektedir[30].

Huang araştırmasında[31], kültürlerarası öğrenmenin ergenlerin küresel yetkinliği üzerindeki etkisini ve okul ayrımcılık ikliminin moderatör rolünü incelemektedir. 26 ekonomiden 7,012 okulu temsil eden 211,554 öğrencilik bir örneklem kullanılarak gerçekleştirilen analizler, kültürlerarası öğrenmenin dikkat çekici bir ters U-şeklinde etkiye sahip olduğunu ortaya koymaktadır. Bulgular, öğrencilerin kültürlerarası öğrenme faaliyetlerine katılımı arttıkça, küresel yetkinliklerinin başlangıçta arttığını, ancak kritik bir dönüm noktasına ulaştıktan sonra azaldığını göstermektedir. Ayrıca, çalışma kültürlerarası öğrenmenin eğrisel etkisinin okullardaki mevcut ayrımcılık iklimiyle önemli ölçüde düzenlendiğini ortaya koymaktadır; eğrisel patern, ayrımcılık seviyeleri düşük olduğunda (-1SD) daha belirgin, yüksek okul ayrımcılık iklimi koşullarında (+1SD) ise nispeten daha düz hale gelmektedir. Bu bulgular, kültürlerarası öğrenme girişimlerinin ve okul ikliminin ergenlerin küresel yetkinliğini geliştirmede kritik faktörler olduğuna işaret etmektedir.

## 2.7. İstihdam Edilebilirlik ve Kariyer Başarısı

Günümüzde bireylerin istihdam edilebilirlikleri, Hillard ve Pollard tarafından ortaya konan dört temel unsur etrafında şekillenmektedir[8]. Bu unsurlar; varlıklar (bilgi, beceri ve tutumlar), görevlendirme (kariyer yönetim becerileri, iş arama becerileri ve adaptasyon), sunum (istihdam edilebilirlik varlıklarını pazara erişilebilir bir şekilde sunabilme yeteneği) ve kişisel durumlar ile dış işgücü piyasası olarak belirlenmiştir.

Bireyin sadece sahip olduğu bilgi ve becerilere odaklanmanın yeterli olmadığı; aynı zamanda bu bilgi ve becerileri etkili bir şekilde yönetebilme, iş bulabilme ve değişen koşullara uyum sağlayabilme yeteneğinin de kritik öneme sahip olduğu vurgulanmaktadır.

Rothwell ise istihdam edilebilirliği üç bileşenli bir modelle açıklamaktadır. Bu model, ulusal işgücü düzeyinde istihdam edilebilirlik, insan kaynakları yönetimi ve iş ve kariyer psikolojisi içindeki istihdam edilebilirlik ve bireylerin resmi eğitim içindeki istihdam edilebilirliği içermektedir. Bu anlamda, bir bireyin işgücüne katılımının, ulusal beceri politikaları ve endüstri düzeyindeki istihdam ihtiyaçlarıyla ilişkilendirilebileceği vurgulanmaktadır[32].

Literatür taraması ile mühendislik öğrencilerinin kariyer başarılarını etkileyen önemli faktörler incelemiş olup Deneyimsel Öğrenme, Öz-Yeterlik, Kültürel Zekâ ve İş Bulunabilirlik Algısı gibi unsurların öğrencilerin kariyer gelişiminde kritik roller oynadığı görülmüştür.

## 3. Materyal ve Metot

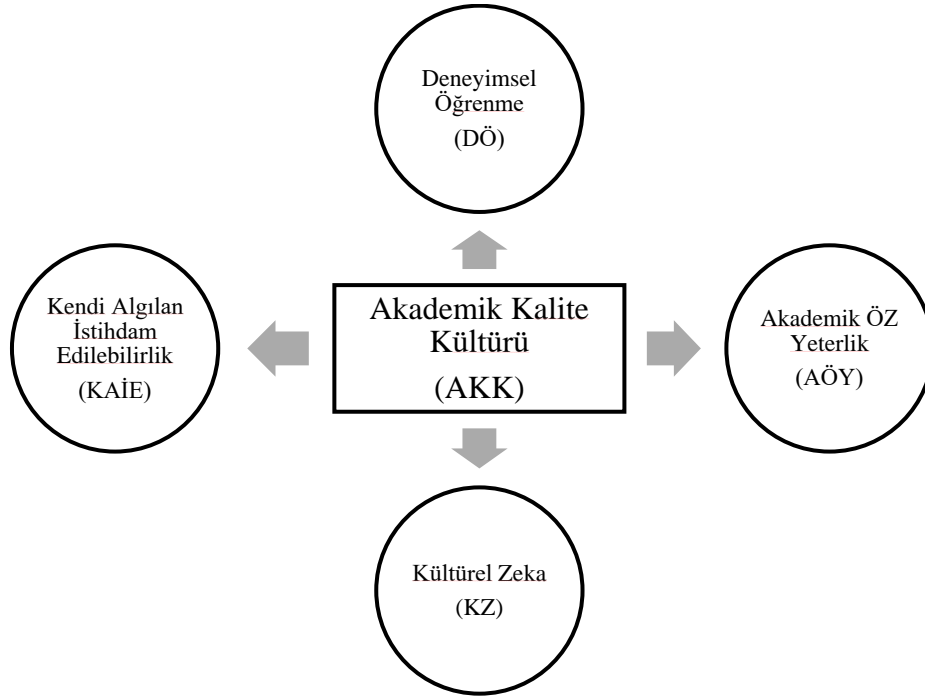
Bu çalışma, mühendislik öğrencilerinin başarılarını etkileyen kritik faktörlerin yapısal eşitlik modellemesi (YEM) ile belirlenmesini amaçlamaktadır. Araştırma, "Hangi kritik faktörler mühendislik öğrencilerinin başarılarını etkiler?" ve "Kurulan modelin doğruluğu ve geçerliliği nasıl belirlenebilir?" sorularına cevap aramaktadır. Bu soruları yanıtlamak amacıyla, Şekil 1'de detaylı olarak gösterilen bir araştırma modeli ve bu modele ait alt değişkenler tanımlanmıştır. Model, öğrenci başarısını etkileyen faktörler arasındaki ilişkileri analiz etmeyi hedeflemektedir. Kısacası, bu çalışma, mühendislik öğrencilerinin akademik başarılarını etkileyen faktörleri belirlemeyi ve bu faktörler arasındaki ilişkileri YEM ile doğrulamayı amaçlamaktadır.

Çalışmada test edilen hipotezler aşağıda sunulmuştur:

**H1:** Akademik öz yeterlik (AÖY) ile kültürel zekâ (KZ) arasında pozitif bir ilişki vardır.

**H2:** Akademik öz yeterlik (AÖY) ile kendi algılanan istihdam edilebilirlik (KAİE) arasında pozitif bir ilişki vardır.

- H3:** Kültürel zekâ (KZ) ile kendi algılanan istihdam edilebilirlik (KAİE) arasında pozitif bir ilişki vardır.  
**H4:** Deneyimsel öğrenme (DÖ) ile akademik öz yeterlik (AÖY) arasında pozitif bir ilişki vardır.  
**H5:** Deneyimsel öğrenme (DÖ) ile kültürel zekâ (KZ) arasında pozitif bir ilişki vardır.  
**H6:** Deneyimsel öğrenme (DÖ) ile kendi algılanan istihdam edilebilirlik (KAİE) arasında pozitif bir ilişki vardır.



**Şekil 1.** Araştırma modeli ve bu modele ait alt değişkenler

Veri Toplama: Veriler, Bursa Teknik Üniversitesi öğrencilerine uygulanan anketler aracılığıyla toplanmıştır. Çeşitli öğrenci gruplarını temsil edecek şekilde özenle seçilen 500 katılımcıdan oluşan örneklem, akademik öz yeterlik, kültürel zekâ, deneyimsel öğrenme ve kendi algılanan istihdam edilebilirlik gibi değişkenleri ölçen anketi yanıtlamıştır.

SmartPLS ile Veri Analizi: Veriler, SmartPLS programı kullanılarak analiz edilmiş ve yapısal model değerlendirilmiştir. Büyük örneklem büyüklüğü (n=500) ve özenle seçilmiş örneklem, istatistiksel tahminlerin doğruluğunu artırarak hipotezlerin güvenilir bir şekilde test edilmesini sağlamaktadır.

**Tablo 1.** Hipotez testinin sonuçlarını

Hipotez	Yol Katsayısı (β)	T Değeri	R <sup>2</sup>	f <sup>2</sup>	Sonuç
H1	0.368	5.24	0.390	0.018	Kabul
H2	0.272	3.91	0.303	0.073	Kabul
H3	0.145	2.57	0.303	0.066	Kabul
H4	0.430	6.78	0.185	0.181	Kabul
H5	0.370	5.56	0.390	0.227	Kabul
H6	0.259	4.12	0.303	0.324	Kabul

Çalışma sonuçları, hipotezlerin doğrulandığını göstermektedir. Akademik öz yeterlik, kültürel zekâ ve deneyimsel öğrenme gibi faktörlerin akademik performans algısı üzerinde pozitif ve anlamlı bir etkisi olduğu bulunmuştur. Bu çalışma, akademik başarıyı etkileyen faktörlerin anlaşılmasına katkı sağlamaktadır. Özellikle öğrencilerin akademik öz yeterliklerinin artırılması, kültürel zekâ ve deneyimsel öğrenme yeteneklerinin desteklenmesi, genel akademik performanslarını iyileştirmede önemli rol oynamaktadır. Anket, literatürdeki bilgilere dayanarak, öğrencileri rekabetçi iş piyasasına hazırlamada önemli faktörler olarak tanımlanan yapılar odaklanmaktadır. Bu faktörler Tablo 2'de detaylı olarak tartışılmaktadır.

**Tablo 2.** Anket bilgileri

<b>Oluşturmak (gizil değişkenler)</b>		<b>Öge/Gösterge (gözlemlenen değişkenler)</b>
<b>Akademik öz-yeterlik</b>	AÖY 1	Kendim için belirlediğim çoğu hedefi başarabiliyorum.
	AÖY 2	Zor görevleri başaracağımdan eminim.
	AÖY 3	Aklıma koyduğum herhangi bir çabada başarılı olabilirim.
	AÖY 4	Hedeflerimi başarmak için azimle çalışabilirim.
	AÖY 5	Yeni görevleri iyi yapabiliyorum.
<b>Kültürel Zekâ</b>	KZ 1	Kariyerim başkalarının kültürel normlarını anlamama yardımcı oldu.
	KZ 2	Çok kültürlü etkileşimlerde kullandığım kültürel bilginin farkındayım.
	KZ 3	Diğer kültürlerin sanat ve zanaatlarını biliyorum.
	KZ 4	Farklı kültürlerden insanlarla etkileşimden keyif alıyorum.
	KZ 5	Çok kültürlü bir durumda gerektiğinde sözsüz iletişim davranışlarımı değiştiririm.
<b>Deneyimsel Öğrenme</b>	DÖ 1	Bu kurs, öğrenme hedefleri bağlamında etkili bir deneyim sunar.
	DÖ 2	İçerik ve öğrenme deneyimi kariyerimle ilgilidir.
	DÖ 3	Bu öğrenme deneyimini gelecekteki iş fırsatlarıyla bağdaştırıyorum.
	DÖ 4	Bu ders bana yeni bilgiler oluşturma fırsatı sağladı.
<b>Kendi Algılanan İstihdam Edilebilirlik</b>	KAİE 1	İşe kolayca çekici olmak için gerekli niteliklere sahibim.
	KAİE 2	Mezuniyetimden sonra iş bulacağıma güveniyorum.
<b>Akademik kalite kültürü</b>	AKK 1	BTÜ, mühendislik alanında önde gelen bir üniversitedir.
	AKK 2	Bildiğim kadarıyla, BTÜ Türkiye’de tanınmış bir üniversitedir.
	AKK 3	BTÜ'nün mükemmellik düzeyinde performans sergilediğine inanıyorum.
	AKK 4	BTÜ'deki eğitim deneyiminden memnunum.

Deneyimsel Öğrenme, Akademik Öz Yeterlilik, Kendi Algılanan İstihdam Edilebilirlik ve Kültürel Zekâ gibi belirlenen kritik başarı faktörleri arasındaki ilişkilerin öğrencilerin eğitim kalitesi algıları üzerindeki etkisini ölçmek için bir anket/soru formu geliştirilmiştir. Anket iki bölümden oluşmaktadır: Birinci bölüm demografik özellikleri (yaş, cinsiyet, bölüm, akademik statü vb.) içerirken, ikinci bölümde öğrencilerden Akademik Öz Yeterlilik, Kültürel Zekâ, Deneyimsel Öğrenme, Kendi Algılanan İstihdam Edilebilirlik ve Akademik Kalite gibi beş gizil değişkeni temsil eden eğitim deneyimlerine yönelik algılarını 5 puanlık Likert tipi ölçekte değerlendirmeleri istenmiştir.

Bu çalışma, nicel bir araştırma yaklaşımı kullanarak gerçekleştirilmiştir. Veriler, Yarı Parametrik En Küçük Kareler Tekniği ile Yapısal Eşitlik Modeli (PLS-SEM) kullanılarak yapılandırılmış bir anket aracılığıyla toplanmış ve modelin doğrulanması için SmartPLS 3.0 yazılımına aktarılmıştır. Yazılım, sağlanan endojen latent değişkenlerin açıklanan varyansını maksimize eden bir algoritma kullanarak, latent değişkenler arasındaki ilişkileri (iç model katsayıları) ve bunların ilgili göstergelerle olan ilişkisini (dış model katsayıları) tahmin etmektedir. Önerilen model, nedenselliğin yapıdan ölçütlere doğru yönlendirildiği yansıtıcı bir modeldir.

#### **4. Analiz Ve Bulgular**

Bu çalışma, SmartPLS3 yazılımı kullanılarak iki ana kısımda gerçekleştirilmiştir.

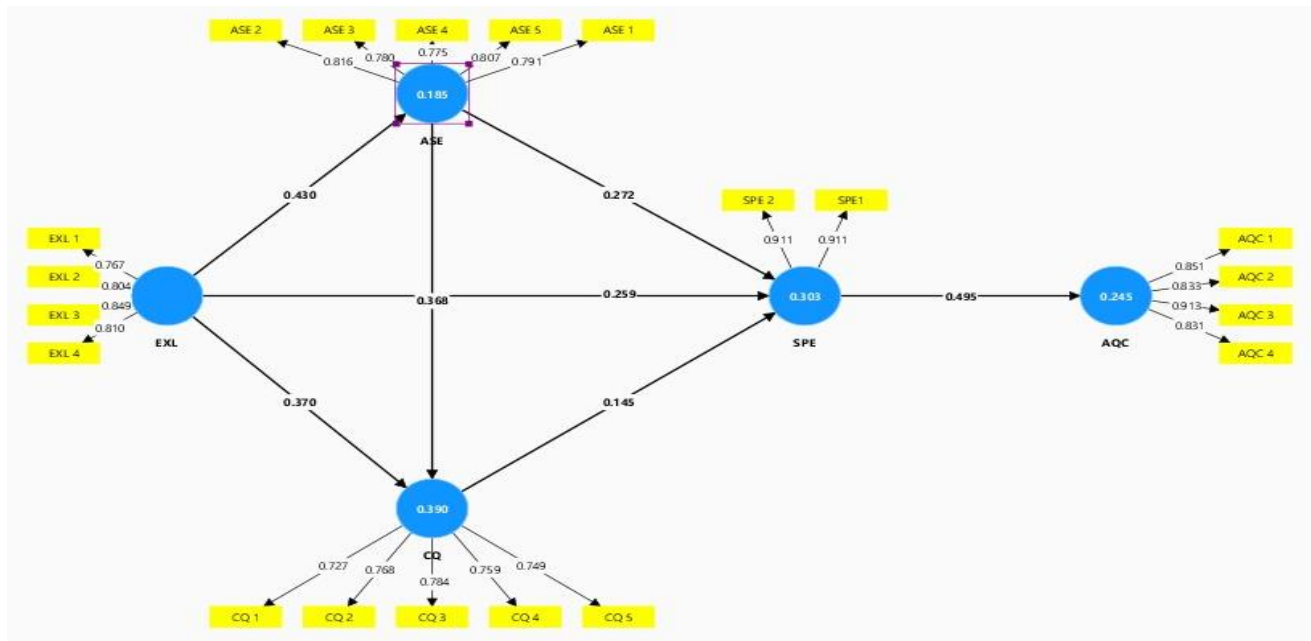
Çalışma, yapısal eşitlik modelinin gözlenen ve gözlemlenmeyen değişkenler arasındaki ilişkileri tanımlayan ölçüm modeli analizi ile başlamaktadır. Bu aşamada, araştırma değişkenlerinin ölçümünün tartışılması ve



göstergelerin geçerliliği ve güvenilirliğine odaklanılmıştır. Ölçüm modeli analizi, çalışmanın temelini oluşturan değişkenlerin doğru bir şekilde ölçüldüğünden emin olmayı amaçlamaktadır.

İkinci kısım ise yapısal model analizidir. Bu aşama, bağımsız ve bağımlı değişkenler arasındaki ilişkinin doğasını ortaya koyar ve bu ilişkileri açıklayan bir içsel model oluşturur. Her bir bağımsız değişkenin bağımlı değişkenler üzerindeki etkisini ve açıklama gücünü araştırır. Yapısal model analizi, çalışma hipotezlerinin sonuçlarını ve ilişkilerin yönünü (pozitif veya negatif) belirler.

Çalışmanın sonuçları, toplanan verilerin ölçüm modeli kriterlerini karşılayıp karşılamadığının değerlendirilmesinin ardından yapısal modelin sonuçlarıyla birlikte sunulmuştur. Bu analizler, çalışmanın amacına ve belirlenen hedeflere ulaşma sürecini tamamlamış ve araştırma sorularına yanıt vermiştir. Sonuç olarak, SmartPLS3 programı kullanılarak yapılan bu iki aşamalı analiz süreci, araştırmanın teorik temelini güçlendirmiş ve hedeflenen sonuçları etkili bir şekilde sunmuştur.



Şekil 2. SmartPLS 3.0'dan çıktı sonuçları

#### 4.1. Ölçüm Modeli analizi (Measurement Model)

**Yakınsak Geçerlilik (Convergent Validity):** Bir ölçeğin farklı ölçümleri arasındaki ilişkiyi değerlendirir. Aynı kavramı ölçen farklı ölçümlerin birbirleriyle tutarlı olup olmadığını gösterir. Yani, ölçek altındaki farklı ölçümler arasında yüksek bir korelasyon olması beklenir. Bu, ölçeğin güvenilirliğini ve geçerliliğini gösterir [35,36].

**Ayırt Edici Geçerlilik (Discriminant Validity):** Bir ölçeğin farklı kavramları ölçme yeteneğini değerlendirir. Farklı kavramları temsil eden ölçekler arasında düşük bir korelasyon olması gerektiğini vurgular. Yani, ölçümler arasında düşük korelasyonlar, farklı kavramları ölçtüğünü doğrular [36].

Bu iki ana bileşen, ölçüm modelinin geçerliliğini ve güvenilirliğini sağlamak için temel olarak kullanılır.

##### 4.1.1. Ölçeklerin yakınsak geçerlilik analizi (Convergent validity)

Yakınsak geçerlilik analizi, bir kavramı ölçmek için kullanılan maddelerin birbiriyle ne derece tutarlı olduğunu değerlendirir. Bu analiz, genellikle Hair Jr, Hult, Ringle ve Sarstedt (2016) tarafından belirtilen üç kriter aracılığıyla gerçekleştirilir:

**İç Tutarlılık (Faktör Doygunlukları):** Bir faktör altında yer alan ölçek maddelerinin birbirleriyle ne kadar tutarlı olduğunu ifade eder. Bu kriter, maddeler arasındaki korelasyonların değerlendirilmesi yoluyla belirlenir.

**Ölçeğin Güvenilirliği (Bileşik Güvenilirlik - CR):** Ölçek altındaki tüm maddelerin bir araya gelerek ölçülen kavramı ne kadar doğru ve güvenilir bir şekilde yansıttığını ifade eder. Bileşik güvenilirlik, ölçeğin iç tutarlılığını değerlendirmek için kullanılır ve genellikle Cronbach'ın alfa katsayısı gibi istatistiksel yöntemlerle hesaplanır.

**Çıkarılan Ortalama Varyans (AVE):** Bir faktör altındaki maddelerin toplam varyansının ne kadarını açıkladığını ifade eder. AVE, bir faktör altındaki maddelerin birbirleriyle ne kadar tutarlı olduğunu değerlendiren bir kriterdir. Genellikle, AVE değeri 0.50'nin üzerinde olmalıdır, bu da maddelerin birbiriyle tutarlı olduğunu gösterir.

Bu üç kriter, ölçek altındaki maddelerin birbiriyle tutarlı olup olmadığını değerlendirmek için kullanılır ve ölçüm modelinin geçerliliğini sağlamak için önemlidir.

**Tablo 3.** Yakınsak Geçerlilik Kriterlerinin Açıklaması.

Standartlar	Onaylanan değerler
İç Tutarlılık- Faktör Doygunlukları	Hair Jr ve diğerleri'ye göre tüm soruların faktör yükü 0,50'den büyük olmalıdır[38].
Ölçeğin güvenilirliği- Bileşik Güvenilirlik (RC)	Hair Jr ve diğerleri'ye göre genel göstergeler arasındaki iç tutarlılığı yansıtabilmek için bileşik güvenilirlik (CR) değerlerinin 0,70'ten büyük olması gerekir[38].
Çıkarılan Ortalama Varyans (AVE)	Hair Jr ve diğerleri'ye, açıklanan varyans (AVE) değerlerinin, tüm faktör sorularının ölçümdeki başarısını yansıması için 0,50'den büyük olması gerekir[38].

**İç Tutarlılık - Doygunluk Faktörü Yükleme:** Araştırma modelinin iç tutarlılığı, soruların (maddelerin) tüm boyutlara ilişkin doyumluklarının test edilmesiyle tahmin edilmektedir. Çalışma modelinin yakınsak geçerliliğini doğrulamak için modelin geçerliliği, çalışma modeli öğelerinin doyumlukları (Faktör yükleme) ölçülerek değerlendirilir; burada Tablo 4'deki sonuçlar, çalışma modelinin doyumluk değerlerini gösterir.

**Tablo 4.** Her Boyut için Soru Doyumluklarının Sonuçları

Outer loadings - Matrix					
	AKK	AÖY	KZ	DÖ	KAİE
AKK 1	0.851				
AKK 2	0.833				
AKK 3	0.913				
AKK 4	0.831				
AÖY 2		0.816			
AÖY 3		0.780			
AÖY 4		0.775			
AÖY 5		0.807			
KZ 1			0.727		
KZ 2			0.768		
KZ 3			0.784		
KZ 4			0.759		
KZ 5			0.749		
DÖ 1				0.767	
DÖ 2				0.804	
DÖ 3				0.849	
DÖ 4				0.810	
KAİE 2					0.911
KAİE 1					0.911
KAİE 1		0.791			

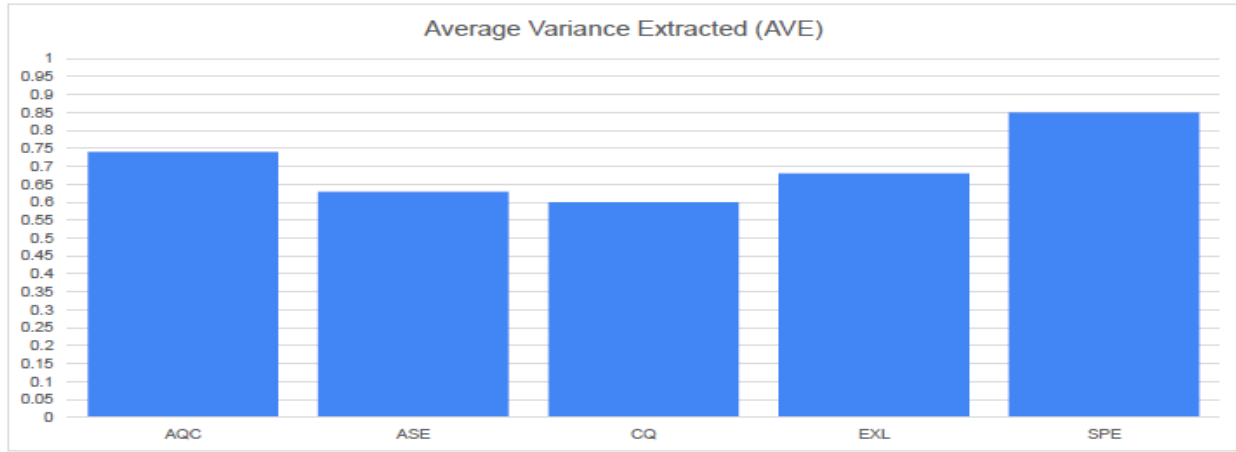
**Bileşik Güvenilirlik Ölçeği (CR):** Bileşik güvenilirlik, faktör yüklerinin toplamına göre gizli değişken faktör yüklerinin toplamı artı hata varyansını ölçer ve Hair Jr ve arkadaşlarına göre [38] önerilen değer 0,7 veya daha yüksek olmalıdır. Bileşik Güvenilirlik (CR) ölçeğinin güvenilirlik sonuçları Tablo 5'te gösterilebilir.

**Tablo 5.** CR ölçeğinin güvenilirlik sonuçları

	Cronbach's alpha	Composite reliability (rho_a)	Composite reliability (rho_c)	Average variance extracted (AVE)
AKK	0.880	0.884	0.918	0.736
AÖY	0.854	0.856	0.895	0.631
KZ	0.814	0.814	0.871	0.574
DÖ	0.822	0.823	0.882	0.653
KAİE	0.795	0.795	0.907	0.830

CR kompozit güvenilirlik ölçümü sonuçları yukarıdaki Tablo 5'te olduğu gibi tüm değerlerin 0,70'ten büyük olduğunu göstermektedir. Hair ve diğerleri 'ya göre [38], çalışma faktörleri yüksek kabul edilmektedir.

**Çıkarılan Ortalama Varyans (AVE):** Açıklanan Ortalama Varyans (AVE) ölçüsü, yapısal model düzeyinde yakınsak geçerliliği kanıtlamak için kullanılan en popüler önlemlerden biridir. Bu ölçü, faktörle ilişkili göstergeler için kare yüklerin genel ortalaması, yani kare yüklerin toplamının gösterge (soru) sayısına bölünmesiyle tanımlanır. Bireysel göstergelerle aynı mantığı kullanarak, 0,50 veya daha yüksek bir AVE değeri, faktörün göstergelerindeki varyansın ortalama yarısından fazlasını açıkladığını gösterir. Buna karşılık, 0,50'den düşük bir AVE değeri, ortalamada hala varyansın olduğunu gösterir.

**Şekil 3.** AVE sonuçları

Çıkarılan Ortalama Varyansın sonuçları, Şekil 3'te gösterildiği gibi, tüm değerlerin önerilen standardı aştığını göstermektedir. Yani ortalama açıklanan varyans kriterinin karşılandığı, yani faktörün soruları arasında uyum olduğu anlamına gelmektedir.

#### 4.1.2. Ayırt Edici Geçerlilik (Discriminant Validity)

Ayırt edici geçerlilik kriteri, kullanılan değişkenlerin mükerrer olmamasını sağlamak amacıyla değişkenlerin birbirinden farklılık derecesini, yani her değişkenin kendisini temsil ettiğini ve diğer değişkenleri temsil etmediğini ifade eder. Fornell ve Larcker [37] kriteri kullanılmış ve çalışma aracının ayırt edici geçerliliğinin olabilmesi için, çalışma değişkenlerinin her biri için Fornell-Larcker kriterinin geri kalan değişkenlerle karşılaştırılabilir en büyük kriter olması gerekmektedir. Yani değişkenin kendisini diğer değişkenlerden daha fazla temsil ettiği ve dolayısıyla çalışma değişkenleri arasında örtüşme olmadığı anlamına gelir. Tablo 6 örtüşmenin olmadığını ve dolayısıyla çalışmanın ayırt edici geçerliliği kriterini göstermektedir.

**Tablo 6.** Ayırt edici geçerlilik sonuçları

	AKK	AÖY	KZ	DÖ	KAİE
AKK	0.858				
AÖY	0.290	0.794			
KZ	0.401	0.528	0.758		
DÖ	0.456	0.430	0.528	0.808	
KAİE	0.495	0.459	0.425	0.452	0.911

Yukarıdakilere dayanarak, sonuçlar hem Yakınsak geçerliliği hem de Ayırıcı geçerliliği olan Ölçüm Modelinin analizi yoluyla değişkenleri ölçmek için kullanılan soruların geçerliliğini göstermektedir. İçsel model olan Yapısal Modelin analiz edilmesiyle devam edilebilir.

#### 4.2.Yapısal Model Analizi

Standart modelin yakınsak ve ayırt edici geçerlilik ölçütlerinin sonuçları kabul edildikten sonra, bir sonraki adım yapısal modelin sonuçlarının değerlendirilmesi ile ilgilidir. Bu aşama, modelin tahmin yeteneklerini ve çalışma değişkenleri arasındaki ilişkileri incelemeyi içerir. Yapısal modelin değerlendirilmesi için kullanılması gereken temel kriterler şunlardır:

**Yorumlama Katsayısı (R<sup>2</sup>):** Bağımlı değişkenlerdeki değişkenliğin ne kadarını bağımsız değişkenlerin açıkladığını ölçer. Yüksek bir R<sup>2</sup> değeri, modelin verileri iyi açıkladığını gösterir [18].

**Etki Boyutu (f<sup>2</sup>):** Bağımsız değişkenlerin bağımlı değişkenler üzerindeki etkisini ölçer. Bu, ilişkinin gücünü ve önemini belirlemeye yardımcı olur [19].

**Hipotez Testi:** Modeldeki ilişkilerin istatistiksel olarak anlamlı olup olmadığını belirler. Bu hipotezlerin kabul edilip edilmediğini gösterir.

Bu kriterler, yapısal modelin başarısını ve ilişkilerin gücünü değerlendirmek için kullanılır.

**Tablo 7.** Yapısal modeli değerlendirme kriterleri

Standartlar	Tanım
Açıklama Faktörü R <sup>2</sup>	Cohen'e göre [40] , R <sup>2</sup> yorumlama faktörünün değeri aşağıdaki gibidir: <ul style="list-style-type: none"> <li>• R<sup>2</sup> &lt;0,02 ise herhangi bir açıklama faktörü yoktur</li> <li>• 0,02 &lt;=R<sup>2</sup>&lt;0,12 ise küçük</li> <li>• 0,12 &lt;=R<sup>2</sup>&lt;0,25 ise orta</li> <li>• 0,25 &lt;=R<sup>2</sup>&lt;1 ise büyük</li> </ul>
Etki Boyutu f <sup>2</sup>	Hair ve arkadaşları'na göre[38] , f <sup>2</sup> etki boyutunun değeri aşağıdaki gibidir: <ul style="list-style-type: none"> <li>• f<sup>2</sup>&lt;0,02 ise herhangi bir etki boyutu yoktur.</li> <li>• 0,02 &lt;= f<sup>2</sup>&lt;0,12 ise küçük</li> <li>• 0,12 &lt;= f<sup>2</sup>&lt;0,25 ise orta</li> <li>• 0,25 &lt;= f<sup>2</sup>&lt;1 ise büyük</li> </ul>
Bootstrapping kullanarak yol katsayısını kontrol etme	Yol katsayısı büyüklük, uyum açısından tahmin edilmelidir ve normalde kullanılan değer aşağıdaki gibidir: <ul style="list-style-type: none"> <li>• 10% anlamlılık düzeyinde 1,65'e eşittir.</li> <li>• 5% anlamlılık düzeyinde 1,96'e eşittir.</li> <li>• Hair ve arkadaşları'na göre[38] , 1% anlamlılık düzeyinde 2,59'e eşittir.</li> </ul>

**Açıklama Katsayısı (R<sup>2</sup>):** Yorumlama katsayısı, yapısal bir modeli değerlendirmek için en yaygın ölçüdür ve R<sup>2</sup> değeri olarak bilinir. Bu katsayı, modelin tahmin gücünün bir ölçüsünü temsil eder ve yapısal modelin gerçek ve tahmin değerleri arasındaki kare korelasyonu olarak hesaplanır.

**Tablo 8.** Yorumlama katsayısı sonuçları

	R <sup>2</sup>	R <sup>2</sup> adjusted
AKK	0.245	0.243
AÖY	0.185	0.183
KZ	0.390	0.387
KAİE	0.303	0.299

Yüksek ve orta bir R<sup>2</sup> değeri, modelin verileri iyi açıkladığını gösterir.

**Etki Boyutu (f<sup>2</sup>):** Etki büyüklüğü değerleri her bir bağımsız değişkenin bağımlı değişkeni açıklama yeteneğini açıklamaktadır.

**Tablo 9.** Etki büyüklüğü sonuçları

f-square-Matrix				
AKK	AÖY	KZ	DÖ	KAİE

AKK			
AÖY		0.181	0.073
KZ			0.018
DÖ	0.227	0.182	0.066
KAİE	0.324		

Tablodaki değerler bağımsız değişkenlerin bağımlı değişkenler üzerindeki etkisinin gücünü ve boyutunu göstermektedir.

**Hipotez Testi:** Bu aşama, araştırmanın ana sorularına ve hipotezlerine cevap verme yeteneğini belirlemeye yöneliktir. Hem yorumlama faktörünün hem de etki büyüklüğünün değerleri doğrulandıktan sonra hipotezlerin sonuçları test edilmiştir. Tablo 10'da yapısal model, çalışmanın hipotezlerinin sonuçlarını göstermektedir.

**Tablo 10.** Hipotez sonuçları

	Path coefficients
AÖY-> KZ	0.368
AÖY->KAİE	0.272
KZ->KAİE	0.145
DÖ->AÖY	0.430
DÖ->KZ	0.370
DÖ->KAİE	0.259

Bursa Teknik Üniversitesi öğrencileri üzerinde yapılan yapısal model analizi sonuçlarına göre, AÖY (Akademik Öz Yeterlik) boyutu ile KZ (Kültürel Zekâ) boyutu arasında pozitif ve anlamlı bir ilişki olduğu bulunmuştur ( $p < 0.05$ ). Bu ilişkiyi ifade eden korelasyon değeri 0.368'dir. Bu, AÖY boyutunun artmasıyla KZ boyutunun %37 arttığı anlamına gelir. T değerine baktığımızda, bu değer 1.96'dan büyük olması, ilişkinin önemli olduğunu gösterir ve bu da birinci hipotezin kabul edildiğini gösterir. Öte yandan, etki boyutu ( $f^2$ ) incelendiğinde, AÖY bağımsız bir faktör olarak KZ üzerindeki etkisinin değeri 0.181'dir. Bu değer, Hair ve diğerlerinin tavsiyesine göre orta bir etki boyutunu temsil eder. Bu sonuçlar, AÖY'nin öğrencilerin kültürel zeka düzeylerini artırma potansiyeline sahip olduğunu göstermektedir. Ancak etkisinin sınırlı olduğunu belirtmek önemlidir.

## 5. Sonuçlar ve Öneriler

Güncel eğilimler, geleceğin mezunlarının akademik hazırlıklarını geliştirmek için kritik başarı faktörleri olarak adlandırılan temel kariyer becerilerinin müfredata ve üniversite deneyimine entegre edilmesi gerektiğini göstermektedir. Bu araştırmanın temel amacı, insan davranışlarını, algılarını ve eğilimlerini etkileyen karmaşık parametreleri incelemek için ikinci nesil çok değişkenli istatistiksel analiz yöntemlerini kullanmaktır. Buradaki motivasyon, yükseköğretimde kaliteyi öznel bileşenlerine göre değerlendirebilmektir.

Bu çalışmada, öğrencilerin deneyimsel öğrenme, akademik öz-yeterlik, kültürel zekâ, öz-algılanan istihdam edilebilirlik ve akademik kalite algıları arasındaki nedensel ilişkileri incelenmiştir. Ayrıca mühendislik mezunlarının akademik başarısını etkileyen kritik faktörleri belirlemek ve bu faktörleri birbirine bağlayan çok değişkenli bir modeli doğrulamak amaçlanmıştır.

- Önerilen kritik faktörlerin akademik kalite kültürü üzerinde belirgin bir etkisi olduğu gözlemlenmiştir.
- Deneyimsel öğrenmenin, öğrencilerin öz-algılanan istihdam edilebilirlikleri üzerinde en güçlü etkiye sahip olduğu bulunmuştur. Bu durum, deneyimsel öğrenmenin kritik düşünmeyi geliştirme, uygulamaya dayalı teoriler oluşturma ve çok kültürlü anlayışı iyileştirme yeteneğiyle ilişkilendirilebilir.
- Deneyimsel öğrenme ile kültürel zekâ arasında istatistiksel olarak anlamlı bir ilişki olduğu tespit edilmiştir. Benzer şekilde, akademik öz-yeterlik ile kültürel zekâ arasında da anlamlı bir ilişki bulunmuştur. Bu sonuçlar, deneyimsel öğrenmenin öğrenci motivasyonunu artırma ve akademik hazırlığı geliştirme potansiyeline sahip olduğunu desteklemektedir.
- Deneyimsel öğrenme (DÖ) ile akademik öz-yeterlik (AÖY) arasında istatistiksel olarak anlamlı bir ilişki bulunamamıştır. Bu, deneyimsel öğrenmenin akademik öz-yeterliliği doğrudan etkilemediği, ancak her ikisinin de kültürel zekayı (KZ) tahmin etmede güçlü göstergeler olduğunu göstermektedir.

Bu araştırma, eğitim programlarının belirlenen kritik faktörleri (deneyimsel öğrenme, akademik öz-yeterlik, kültürel zekâ) içermesinin önemini vurgulamakta ve öğrencilerin başarısını artırmak için bu faktörlerin entegre edilmesini önermektedir. Bu şekilde, öğrencilerin daha motive olmaları ve akademik olarak daha hazırlıklı olmaları sağlanabilir.

Kültürel zekâ (KZ) bulguları, bir kişinin kültürel olarak çeşitli ortamlarda etkili olabilme yeteneğiyle ilgili bir yetenek olduğunu göstermektedir. Bu nedenle, genellikle kültürel olarak etkileşimli durumlardan etkilenen akademik ortamlardaki öz-yeterlilikle doğrudan ilişkilidir. Ayrıca, KZ'nin diğer zekâ türlerini tamamlayan çok boyutlu bir yapı olarak kabul edilmesi, genel zekâ ile de ilişkili olabileceğini düşündürmektedir [37].

Bu çalışma da ayrıca gözlenen ve gizli değişkenleri bir araya getiren bir modelin deneysel olarak doğrulanmasını sağlayarak, yükseköğretim kurumlarının öğrencilerinin hazırlığını artırmak için deneyimsel öğrenim fırsatlarını entegre etmelerinin önemini göstermektedir.

Öneriler arasında, mühendislik eğitim programlarının öğrencilerin Öz-Yeterlik deneyimlerini desteklemesi ve saha deneyimlerini kariyer becerileri olarak geliştirmelerine olanak tanınması bulunmaktadır. Ayrıca, üniversitelerin kariyer danışmanlığı ve öz-yeterlik geliştirme programları sunarak öğrencilerin kariyer hedeflerine güvenle ilerlemelerini sağlamaları önerilmektedir.

- Yükseköğretim kurumları, öğrencilerin deneyimsel öğrenme fırsatlarına erişimini artırmalıdır.
- Müfredat, kültürel zekâyı ve akademik öz-yeterliliği geliştirecek şekilde tasarlanmalıdır.
- Öğrencilerin motivasyonunu ve akademik hazırlığını artırmak için destekleyici bir öğrenme ortamı yaratılmalıdır.
- Bu faktörlerin etkisini daha detaylı anlamak için daha fazla araştırma yapılmalıdır.

Sınırlılıklar:

- Çalışma, belirli bir üniversitedeki mühendislik öğrencileriyle sınırlıdır.
- Bulguların diğer disiplinler ve kültürler için genellenebilirliği sınırlı olabilir.
- Öz-bildirime dayalı veriler kullanılmıştır, bu da yanlılığa neden olabilir.

Gelecek Araştırmalar:

- Farklı disiplinler ve kültürlerdeki öğrencilerle daha fazla araştırma yapılmalıdır.
- Bu faktörlerin uzun vadeli etkilerini incelemek için boyamsal çalışmalar yapılmalıdır.
- Deneyimsel öğrenme, akademik öz-yeterlik ve kültürel zekâyı geliştirmek için etkili stratejiler araştırılmalıdır.

Bu çalışma, mühendislik öğrencilerinin akademik başarısını etkileyen faktörler hakkında değerli bilgiler sunmaktadır. Bulgular, eğitimciler ve politika yapıcılar için öğrenci başarısını artırmak için kullanılacak pratik bilgiler sağlayabilir.

**Yazar Katkıları:** Kavramsallaştırma, H.Ş. ve D.D.; metodoloji, D.D.; yazılım, H.Ş. ve D.D.; doğrulama, H.Ş.; biçimsel analiz, H.Ş.; soruşturma, H.Ş.; kaynaklar, D.D.; veri iyileştirme, A.A.; yazı- orijinal taslak hazırlama, A.A., H.Ş. ve D.D.; yazma- gözden geçirme ve düzenleme, H.Ş. ve D.D.; görselleştirme, H.Ş. ve D.D., Tüm yazarlar makalenin yayınlanan versiyonunu okudu ve kabul etti.

**Finansman:** Bu araştırma dışarıdan fon almadı.

**Çıkar çatışmaları:** Yazarlar çıkar çatışması beyan etmemektedir.

**Bilgilendirme:** Bu çalışma Bursa Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Müdürlüğü Endüstri Mühendisliği Anabilim Dalı Yüksek Lisans öğrencisi Delal Demir'in yüksek lisans tezinden türetilmiştir.

## Kaynaklar

- [1] Y. SARIER, "Türkiye'de Öğrencilerin Akademik Başarısını Etkileyen Faktörler: Bir Meta-Analiz Çalışması," *Hacettepe Üniversitesi - Eğitim Fakültesi Dergisi*, 2016.
- [2] R. E. Keller, E. Johnson, and S. DeShong, "A structural equation model looking at student's participatory behavior and their success in Calculus I," *Int J STEM Educ*, vol. 4, no. 1, pp. 1-11, Dec. 2017, doi: 10.1186/S40594-017-0093-0/TABLES/4.
- [3] X. Wang, M. Dai, and R. Mathis, "The influences of student- and school-level factors on engineering undergraduate student success outcomes: A multi-level multi-school study," *Int J STEM Educ*, vol. 9, no. 1, pp. 1-13, Dec. 2022, doi: 10.1186/S40594-022-00338-Y/TABLES/2.
- [4] A. R. Carberry and D. R. Baker, "The Impact of Culture on Engineering and Engineering Education," pp. 217-239, 2018, doi: 10.1007/978-3-319-66659-4\_10.
- [5] R. Y. Chan, "Understanding The Purpose Of Higher Education: An Analysis Of The Economic And Social Benefits For Completing A College Degree," 2016.

- [6] M. Sarstedt, C. M. Ringle, and J. F. Hair, "Partial Least Squares Structural Equation Modeling," *Handbook of Market Research*, pp. 1–40, 2017, doi: 10.1007/978-3-319-05542-8\_15-1.
- [7] A. C. Campos, J. Da, C. Mendes, J. A. Silva, and P. Oom Do Valle, "Critical success factors for a total quality culture: A structural model," *Tourism & Management Studies*, vol. 10, no. 1, pp. 7–15, 2014, Accessed: Aug. 06, 2024. [Online]. Available: <https://tmstudies.net/index.php/ectms/article/view/641>
- [8] A. Rothwell and J. Arnold, "Self-perceived employability: Development and validation of a scale," *Personnel Review*, vol. 36, no. 1, pp. 23–41, 2007, doi: 10.1108/00483480710716704.
- [9] E. Alyahyan and D. Düşteğör, "Predicting academic success in higher education: literature review and best practices," *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, pp. 1–21, Dec. 2020, doi: 10.1186/S41239-020-0177-7/TABLES/15.
- [10] T. T. York, C. Gibson, and S. Rankin, "Defining and Measuring Academic Success. Practical Assessment, Research & Evaluation, 20(5).," vol. 20, no. 5, 2015.
- [11] M. Coetzee and N. Harry, "Emotional intelligence as a predictor of employees' career adaptability," *J Vocat Behav*, vol. 84, no. 1, pp. 90–97, Feb. 2014, doi: 10.1016/J.JVB.2013.09.001.
- [12] M. Coetzee and D. Schreuder, "The relation between career anchors, emotional intelligence and employability satisfaction among workers in the service industry," *76 Southern African Business Review*, vol. 15, p. 2011.
- [13] J. L. Brown, S. J. Hammer, H. N. Perera, and P. McIlveen, "Relations between graduates' learning experiences and employment outcomes: a cautionary note for institutional performance indicators," *Int J Educ Vocat Guid*, vol. 22, no. 1, pp. 137–156, Apr. 2022, doi: 10.1007/s10775-021-09477-0.
- [14] T. Papadopoulos and C. Armatas, "Evaluation Of An Ict Skills Program: Enhancing Graduate Capabilities And Employability," *International Journal of Electronic Commerce Studies*, vol. 4, no. 1, pp. 93–102, Jul. 2013, doi: 10.7903/IJECS.1115.
- [15] A. L. Miller, K. T. Fassett, and D. L. Palmer, "Achievement goal orientation: A predictor of student engagement in higher education," *Motiv Emot*, vol. 45, no. 3, pp. 327–344, Jun. 2021, doi: 10.1007/S11031-021-09881-7/TABLES/10.
- [16] S. Clark, "Employability and the fine art student," *CEBE Transactions*, vol. 9, no. 2, pp. 7–29, 2012, doi: 10.11120/TRAN.2012.09020007.
- [17] M. Menz, "Integrating academic skills and employability - revisiting the learning journal," *Journal of Perspectives in Applied Academic Practice*, vol. 8, no. 1, pp. 115–120, Sep. 2020, doi: 10.14297/JPAAP.V8I1.393.
- [18] K. S. Amant, "Expanding Internships to Enhance Academic-Industry Relations: A Perspective in Stakeholder Education," <http://dx.doi.org/10.2190/MBJF-PW01-EJ9D-5QT0>, vol. 33, no. 3, pp. 231–241, Jul. 2003, doi: 10.2190/MBJF-PW01-EJ9D-5QT0.
- [19] M. A. Andreu Andrés and M. García Casas, "Perceptions of gaming as experiential learning by engineering students," *International Journal of Engineering Education*, vol. 27, no. 4, pp. 795–804, 2011, Accessed: Aug. 07, 2024. [Online]. Available: <https://riunet.upv.es/handle/10251/57102>
- [20] R. Arslan and S. Di Bono, "A Comparative Study: Challenges and Opportunities for European Union Dual Vocational Training Systems," 2013. [Online]. Available: <https://www.researchgate.net/publication/283796181>
- [21] B. M. Bass, B. J. Avolio, D. I. Jung, and Y. Berson, "Predicting unit performance by assessing transformational and transactional leadership," *J Appl Psychol*, vol. 88, no. 2, pp. 207–218, Apr. 2003, doi: 10.1037/0021-9010.88.2.207.
- [22] J.-M. J. Booth, "Experiential Learning Through Simulation And Prototyping In First Year Engineering Design," 2014.
- [23] V. A. Segarra and W. A. Gentry, "Taking ownership of your career: professional development through experiential learning," *BMC Proc*, vol. 15, no. 2, pp. 1–7, Jun. 2021, doi: 10.1186/S12919-021-00211-W/TABLES/2.
- [24] A. Barnett, "Self-Efficacy," *Encyclopedia of Quality of Life and Well-Being Research*, pp. 6235–6238, 2023, doi: 10.1007/978-3-031-17299-1\_2631.
- [25] A. A. Adediwura, "Effect of peer and self-assessment on male and female students' self-efficacy and self-autonomy in the learning of mathematics," *Gender and Behaviour*, vol. 10, no. 1, pp. 4492–4508, 2012.
- [26] T. Honicke, J. Broadbent, and M. Fuller-Tyszkiewicz, "The self-efficacy and academic performance reciprocal relationship: the influence of task difficulty and baseline achievement on learner trajectory," *Higher Education Research and Development*, vol. 42, no. 8, pp. 1936–1953, 2023, doi: 10.1080/07294360.2023.2197194.
- [27] B. Hunter, G. P. White, and G. C. Godbey, "What Does It Mean to Be Globally Competent?," <http://dx.doi.org/10.1177/1028315306286930>, vol. 10, no. 3, pp. 267–285, Sep. 2006, doi: 10.1177/1028315306286930.
- [28] G. L. Downey et al., "The Globally Competent Engineer: Working Effectively with People Who Define Problems Differently," *Journal of Engineering Education*, vol. 95, no. 2, pp. 107–122, Apr. 2006, doi: 10.1002/J.2168-9830.2006.TB00883.X.
- [29] C. Del Vitto, "Cross-Cultural 'Soft Skills' and the Global Engineer: Corporate Best Practices and Trainer Methodologies Cross-Cultural 'Soft Skills' and the Global Engineer: Corporate 'Best Practices' and Trainer Methodologies," 2008, Accessed: Aug. 07, 2024. [Online]. Available: <http://digitalcommons.uri.edu/ojgeehttp://digitalcommons.uri.edu/ojgee/vol3/iss1/1http://digitalcommons.uri.edu/ojgee/vol3/iss1/1http://digitalcommons.uri.edu/ojgee>
- [30] R. J. Sternberg, "A triarchic view of intelligence in cross-cultural perspective," *Human Abilities in Cultural Context*, pp. 60–85, Feb. 2010, doi: 10.1017/CBO9780511574603.003.
- [31] Q. Huang and A. C. K. Cheung, "More is not Always Better: The Curvilinear Relationship Between Intercultural Learning and Adolescents' Global Competence," *Appl Res Qual Life*, vol. 19, no. 2, pp. 835–857, Apr. 2024, doi: 10.1007/s11482-023-10268-8.
- [32] A. Rothwell and J. Arnold, "Self-perceived employability: Development and validation of a scale," *Personnel Review*, vol. 36, no. 1, pp. 23–41, 2007, doi: 10.1108/00483480710716704.
- [33] M. M. Arshida, T. Abdul, J. Abdullah, and S. Agil, "Critical Success Factors For Total Quality Management Implementation Within The Libyan Iron And Steel Company," 2013.
- [34] M. Zuhdi Ibrahim, M. Nizam Ab Rahman, and R. M. Yasin, "Determining Factors of Students' Satisfaction with Malaysian Skills Training Institutes," *International Education Studies*, vol. 7, no. 6, 2014, doi: 10.5539/ies.v7n6p9.

- 
- [35] J. Henseler, C. M. Ringle, and R. R. Sinkovics, "The use of partial least squares path modeling in international marketing," *Advances in International Marketing*, vol. 20, pp. 277–319, 2009, doi: 10.1108/S1474-7979(2009)0000020014/FULL/XML.
- [36] H. K. Alay, A. Keskin, M. T. Deveciyan, G. Şen, D. Kayaldereden, and Ş. Berber, "The impact of green business ethics and green financing on sustainable business performance of industries in Türkiye: The mediating role of corporate social responsibility," *Sustainability*, vol. 16, no. 17, p. 7868, 2024, doi: 10.3390/su16177868.
- [37] C. Fornell and D. F. Larcker, "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research*, vol. 18, no. 1, p. 39, Feb. 1981, doi: 10.2307/3151312.
- [38] J. F. Hair, G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray, "Evaluation of Reflective Measurement Models," pp. 75–90, 2021, doi: 10.1007/978-3-030-80519-7\_4.
- [39] J. F. Hair, C. M. Ringle, and M. Sarstedt, "Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance," *Long Range Plann*, vol. 46, no. 1–2, pp. 1–12, 2013, doi: 10.1016/J.LRP.2013.01.001.
- [40] J. Cohen, "Statistical Power Analysis for the Behavioral Sciences," *Statistical Power Analysis for the Behavioral Sciences*, May 2013, doi: 10.4324/9780203771587.



Research Article

Received: date:04.07.2024  
Accepted: date:23.09.2024  
Published: date:31.12.2024

# Kelley's coefficients of skewness using R

Jose Moral de la Rubia<sup>1\*</sup>

<sup>1</sup>School of Psychology, Universidad Autónoma de Nuevo León, Monterrey Mexico; jose.morald@uanl.edu.mx

Orcid: 0000-0003-1856-1458<sup>1</sup>

\*Correspondence: jose.morald@uanl.edu.mx

**Abstract:** Kelley developed a robust measure of asymmetry based on quantiles. His proposal was an absolute index which, when divided by the median, results in its relative expression. If the additive complement is standardized with the semi-percentile range, the percentile coefficient of skewness (PCS) is obtained. Additionally, Kelley provided its standard error in case of normal distribution. However, no statistical software currently computes these measures. The aim of this methodological article is to determine their sampling distribution and facilitate their use. Three random samples of 10,000 data points were generated from three symmetric distributions: semicircular (platykurtic), normal (mesokurtic), and logistic (leptokurtic). By bootstrapping, the sampling distribution was obtained for absolute and relative indices, as well as the PCS. The sampling distributions of the absolute index and the PCS conformed to normality, while that of the relative index was leptokurtic with an excessive bootstrap standard error. Furthermore, a script was developed for the R program, adjusted based on these findings, to obtain point and interval estimates of these indices. The script was applied to a random sample as an example. It is concluded that dividing the absolute index by the semi-percentage range is a better standardization option than dividing by the median.

**Keywords:** skewness; quantiles; bootstrap confidence interval; asymptotic standard error; normality.

## 1. Introduction

In 1923, the American psychologist Truman Lee Kelley (1884-1961) proposed a robust measure of skewness based on quantiles, which he had initially proposed in 1921 as a measure of dispersion [1]. The 1923 formulation was an absolute index that, when divided by the median, results in its relative expression, which is not bounded between -1 and 1 [2]. Moreover, his original measure of skewness is interpreted in the opposite way compared to the classical coefficients of Pearson [3-4], Bowley [5], Yule [6], and Fisher [7], so it went somewhat unnoticed, except among researchers in the field of psychology [8]. However, if the additive complement of the absolute index is standardized by dividing it by the semi-percentile range, the percentile coefficient of skewness is obtained, which is bounded between -1 and 1. Additionally, Kelley [2] provided the standard error of the absolute index for a normal distribution, which can be adapted to its two standardized expressions, whether divided by the absolute value of the median or by the semi-percentile range.

It has not been encountered so far, no statistical program computes this measure. The objective of this methodological article is to determine the sampling distribution of the absolute index and the two relative modalities proposed, as well as to facilitate the use of these indices by means of a script developed for the R program, which has been developed by the mathematical community and is freely available [9]. This script allows for the point and interval estimation, providing the information required to choose between the asymptotic or bootstrap confidence interval and, in the latter case, to decide among three estimation methods: normal, percentile, and bias-corrected and accelerated percentile. For this purpose, the assumptions of randomness and normality of the sample are tested. The sample is represented by a histogram with overlaid density and normal curves. In addition, the symmetry, skewness, and normality of the bootstrap sampling distribution are checked.

## 2. Kelley's absolute index of skewness

Kelley's absolute index of skewness is the signed distance from the mean percentile range to the median. It is calculated by subtracting from the median (minuend) the mean percentile range (subtrahend) and can be denoted as *AIS*. See Equation 1, where  $q_p(x)$  represents the  $p$ -order sample quantile. It is interpreted oppositely to the traditional measures of skewness [3-7], as well as more contemporary measures by Bickel [9], Altınay [10], Singh, Gewali, and Khatiwada [11], and Eberl and Klar [12]. Negative values indicate that the right side of the distribution is longer than the left side or the presence of a long tail on the right, taking the median as the axis of symmetry. Conversely, positive values reveal that the left side of the distribution is longer than the right or the presence of a long tail on the left.

$$\widehat{AIS}(x) = d\left(\frac{q_{0.1}(x) + q_{0.90}(x)}{2}, q_{0.5}(x)\right) = q_{0.5}(x) - \frac{q_{0.1}(x) + q_{0.90}(x)}{2} \quad (1)$$

The calculation of sample quantiles can be done by rules 6, 7, 8, or 9 in the R program [14]. The SPSS program calculates quantiles using rule 6, which is the most used method [15]. This rule expresses the order  $p$  of the quantile  $x_{(p)}$  as the average or expected value of the  $i$ -th order statistic in a sample of size  $n$  drawn from a standard continuous uniform distribution  $U[0, 1]$ . This approach is used as a non-informative prior distribution when estimating a probability in Bayesian inference [16]. Rule 7, which is the default in the R program, expresses the order  $p$  of the quantile  $x_{(p)}$  as the mode of the  $i$ -th order statistic in a sample of size  $n$  [17]. Rule 8 is especially recommended when the distribution of the variable is unknown [18-19], as suggested by the simulation study of Hyndman and Fan [20] and the exploratory data analysis of Tukey [21]. This rule expresses the order  $p$  of the quantile  $x_{(p)}$  as the median of the  $i$ -th order statistic in a sample of size  $n$  drawn from a standard continuous uniform distribution. For normal distributions, rule 9 [22] is recommended. In this rule, the quantile order is expressed as  $p = (i - a) / (n + 1 - a - b)$ , where  $a = b = 3/8$ ,  $n$  is the sample size, and  $i$  denotes the unknown order of the quantile among the  $n$  sample data, so  $i = 0.375 + p \times (n + 0.25)$ . If  $i$  is an integer, after sorting the sample in ascending order, the data in that order is the quantile. If  $i$  is a number with decimals, the quantile is obtained by linear interpolation (Equation 2).

$$\begin{aligned} & x = \{x_i\}_{i=1}^n = \{x_1, x_2, \dots, x_n\} \subseteq X \\ & x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(i)} \leq \dots \leq x_{(n)} \\ & \begin{array}{l} \text{Rule 6} \quad a = b = 0 \quad i = p(n + 1) \\ \text{Rule 7} \quad a = b = 1 \quad i = 1 + p(n - 1) \\ \text{Rule 8} \quad a = b = 1/3 \quad i = 1/3 + p(n + 1/3) \\ \text{Rule 9} \quad a = b = 3/8 \quad i = 0.375 + p(n + 0.25) \end{array} \\ p = \frac{i - a}{n + 1 - a - b} \Rightarrow & \\ q_p(x) = & \begin{cases} x_{(i)} & i = [i] = [i] \\ x_{([i])} + (i - [i])(x_{([i]+1)} - x_{([i])}) & i = [i] + (i - [i]) \end{cases} \end{aligned} \quad (2)$$

In addition, Kelley [1-2] provided a formula for the standard error (*se*) or standard deviation in the sampling distribution of the *AIS*, assuming a normal distribution. See Equation 3, where  $PR(x)$  is the sample percentile range,  $q_p(x)$  represents the sample quantile of order  $p$ , and  $n$  denotes the sample size. The constant 0.599143 is a scaling factor derived from theoretical calculations that model the dispersion of the inter-percentile range, divided by the square root of the sample size, in random samples from a standard normal distribution. This factor ensures that the calculated standard error aligns with the asymptotic behavior of percentiles in random samples of varying sizes drawn from a standard normal distribution [1].

$$se[\widehat{AIS}(x)] = 0.599143 \times \frac{PR(x)}{\sqrt{n}} = 0.599143 \times \frac{q_{0.9}(x) - q_{0.1}(x)}{\sqrt{n}} \quad (3)$$

For a random sample drawn from a normal distribution, or a large sample drawn from a variable with finite mean and variance, this formula for the standard error enables the asymptotic estimation of the confidence interval at a confidence level of  $(1 - \alpha) \times 100$ . See Equation 4, where *LL* represents the low limit and *UL* denotes the upper limit of the confidence interval.

$$\begin{aligned} P(LL \leq AIS \leq UL) &= P(AIS(X) \in [LL, UL]) = 1 - \alpha \\ LL &= \widehat{AIS}(x) - z_{1-\frac{\alpha}{2}} \times se[\widehat{AIS}(x)] \end{aligned} \quad (4)$$

$$UL = \widehat{AIS}(x) + z_{1-\frac{\alpha}{2}} \times se[\widehat{AIS}(x)]$$

$$z_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right); \Phi^{-1}\left(1 - \frac{0.05}{2}\right) = \Phi^{-1}(0.975) \approx 1.96$$

In turn, for any type of distribution or a random variable with an unknown distribution, the confidence interval can be calculated using bootstrap with the extraction of 1000 or more samples, applying the bias-corrected and accelerated percentile, percentile, or normal method, unless the bootstrap standard error is too large. The bootstrap standard error should preferably be less than one-fourth of the sample interval. A value greater than half of the sample range indicates a very large error [23]. Additionally, for the method to be accurate, a random sample that is representative of the population and preferably not less than 30 data points is required. If the bootstrap sampling distribution of the skewness statistic follows a normal distribution, the normal method can be used, though the other two methods are not necessarily inadequate. If the distribution is symmetric (tested by the D'Agostino test [24], or another statistical test) with small skewness and acceleration ( $|bias| \leq 0.05$  and  $|a| \leq 0.025$ , respectively), the percentile method can be applied, with the bias-corrected and accelerated method not being inappropriate. Otherwise, the bias-corrected and accelerated percentile method is the best choice [25-26]. When more than one method is applicable, the one that provides the confidence interval with the smallest width can be chosen, as efficiency is a desirable property.

### 3. Kelley's relative index of skewness

The so-called Kelley's relative index of skewness is obtained by dividing the absolute index by the median. This results in a quotient with the difference between the median and the mean percentile range in the numerator and the median in the denominator (Equation 5). Thus, it is standardized in a manner analogous to relative ranges. However, like these measures of variability, the denominator does not convert the quotient into a ratio with a range from -1 to 1, but instead allows it to vary from  $-\infty$  to  $\infty$  [27]. This index can be denoted as *RIS*.

$$\widehat{RIS}(x) = \frac{q_{0.5}(x) - \frac{q_{0.1}(x) + q_{0.9}(x)}{2}}{q_{0.5}(x)} = 1 - \frac{q_{0.1}(x) + q_{0.9}(x)}{2 \times q_{0.5}(x)} \quad (5)$$

If the sampling distribution of the *RIS* were a normal distribution and the median were considered a constant instead of a random variable, the standard error of the *AIS* would be multiplied by the inverse of the absolute value of this constant, resulting in the standard error of the relative index (Equation 6). Consequently, an asymptotic confidence interval could be defined, as shown in Equation 4.

$$se[\widehat{RIS}(x)] = se\left[\frac{\widehat{AIS}(x)}{q_{0.5}(x)}\right] = \frac{1}{|q_{0.5}(x)|} se[\widehat{AIS}(x)] = \frac{0.599143}{\sqrt{n}} \times \frac{PR(x)}{|q_{0.5}(x)|} \quad (6)$$

To maintain the interpretative logic of the skewness coefficients established by Pearson [3-4] and settled by Bowley [5] and Yule [6], the order of the minuend and subtrahend in *AIS* can be reversed (Equation 7). The index value remains the same, with only the sign changing. A positive sign in the additive opposite of *AIS* indicates positive or right-tailed skewness, while a negative sign indicates negative or left-tailed skewness. This reversal can also be applied to the additive opposite of *RIS*, changing only the sign (Equation 8). Additionally, the value of the asymptotic standard error remains unchanged in either index (Equations 9 and 10). These additive opposites can be denoted by *AIS<sub>r</sub>* and *RIS<sub>r</sub>*, where *r* represents the revised or reversed character of the index.

$$1 - AIS(x) = \frac{q_{0.1}(x) + q_{0.9}(x)}{2} - q_{0.5}(x) = \frac{q_{0.1}(x) + q_{0.9}(x) - 2 \times q_{0.5}(x)}{2} = AIS_r(x) \quad (7)$$

$$1 - RIS(x) = \frac{\frac{q_{0.1}(x) + q_{0.9}(x)}{2} - q_{0.5}(x)}{q_{0.5}(x)} = \frac{q_{0.1}(x) + q_{0.9}(x)}{2 \times q_{0.5}(x)} - 1 = RIS_r(x) \quad (8)$$

$$se[1 - AIS(x)] = se[AIS(x)] = 0.599143 \times \frac{PR(x)}{\sqrt{n}} = se[AIS_r(x)] \quad (9)$$

$$se[1 - RIS(x)] = se[RIS(x)] = 0.599143 \times \frac{PR(x)}{\sqrt{n}} = se[RIS_r(x)] \quad (10)$$

In the latter case, if instead of standardizing using the median, it is divided by the semi-percentile range, as in Altinay's measure of skewness [11], it results in the percentile coefficient of skewness as the relative skewness index. This index is a ratio and is bounded between -1 and 1 (Equation 11).

$$RIS_r(x) = \frac{\frac{q_{0.1}(x) + q_{0.9}(x)}{2} - q_{0.5}(x)}{\frac{q_{0.9}(x) - q_{0.1}(x)}{2}} = \frac{q_{0.1}(x) + q_{0.9}(x) - 2 \times q_{0.5}(x)}{q_{0.9}(x) - q_{0.1}(x)} = PCS(x) \quad (11)$$

If the sampling distribution of the percentile coefficient of skewness were normal and the percentile range were considered a constant instead of a random variable, the standard error would be simplified to the expression shown in Equation 12, which would allow the definition of an asymptotic confidence interval as shown in Equation 4.

$$se\left(\frac{1 - AIS(x)}{PR(x)}\right) = \frac{1}{PR(x)} se[AIS(x)] = \frac{0.599143 \times PR(x) / \sqrt{n}}{PR(x) / 2} = \frac{1.19828}{\sqrt{n}} \quad (12)$$

It should be noted that the percentile coefficient of skewness does not have a clear attribution of authorship. In 1975, Hinkley [28] used this coefficient to assess symmetry after applying the Box-Cox power transformation to sample data drawn from non-normal distributions, marking the first published use of this measure of skewness. In the field of economics [29-30] and other disciplines [31-32], it is attributed to Kelly, often without proper citation, likely causing confusion with the psychologist Truman Lee Kelley [33-34].

The percentile coefficient of skewness (*PCS*) deals with one of the infinite possibilities of Bowley's [5] generalization of the quartile coefficient of skewness given by Groeneveld and Meeden [35]. It is similar to the quartile coefficient of skewness except that it uses the 10th ( $q_{0.1}$ ) and 90th ( $q_{0.9}$ ) percentiles instead of the 1st ( $q_{0.25}$ ) and 3rd ( $q_{0.75}$ ) quartiles in the numerator and uses the percentile range (*PR*) instead of the interquartile range (*IQR*) in the denominator. Thus, *PCS* is a way of standardizing the additive opposite of Kelley's *AIS*.

Using quantiles, *AIS* is a robust measure of skewness and can be applied to any type of distribution. The percentile coefficient of skewness is considered a better choice than the quartile coefficient of skewness by Bowley [5] and Yule [6] for platykurtic or thin-tailed distributions, such as the uniform, triangular, or semicircular distribution [29, 36].

#### 4. Materials and Methods

For the first objective of the study, which was to determine the sampling distribution of the three measures of skewness: *AIS<sub>r</sub>*, *RIS<sub>r</sub>*, and *PCS*, three random samples of 10,000 data points were generated using the inverse transform method [37]. Three continuous, symmetric distributions with finite moments were chosen to facilitate compliance with the central limit theorem [17], each with different kurtosis. The Wigner's standard semicircular distribution  $SC(r = 1)$  was selected as the platykurtic distribution. This was generated from a beta distribution by applying the transformation:  $Y = 2rX - r = 2X - 1$ , where the variable *Y* follows a semicircular distribution with unit radius ( $r = 1$ ) and the variable *X* follows a Beta distribution with shape parameters:  $\alpha = 2/3$  and  $\beta = 3/2$  [38]. The standard normal distribution with location parameter 0 and scale 1,  $N(0, 1)$ , was taken as the mesokurtic distribution. The chosen leptokurtic distribution was the standard logistic distribution with location parameter 0 and scale 1,  $Logist(0, 1)$ .

From each of these three source samples of 10,000 data points, 1,000 samples of 10,000 data points were drawn with replacement. In each of these bootstrap samples, Kelley's absolute and relative indices, as well as the percentile coefficient of skewness, were calculated. Thus, three bootstrap sampling distributions were obtained for each of the three measures of skewness.

For each of these nine bootstrap sampling distributions, the presence of outliers was checked using Grubbs' test [39], symmetry by D'Agostino's test [24], mesokurtosis by Anscombe-Glynn test [40], and normality by the Anderson-Darling [41], Shapiro-Francia [42-43], and D'Agostino-Belanger-D'Agostino [44] tests. The difference between asymptotic and bootstrap standard errors was evaluated using the one-sample chi-square test [45].

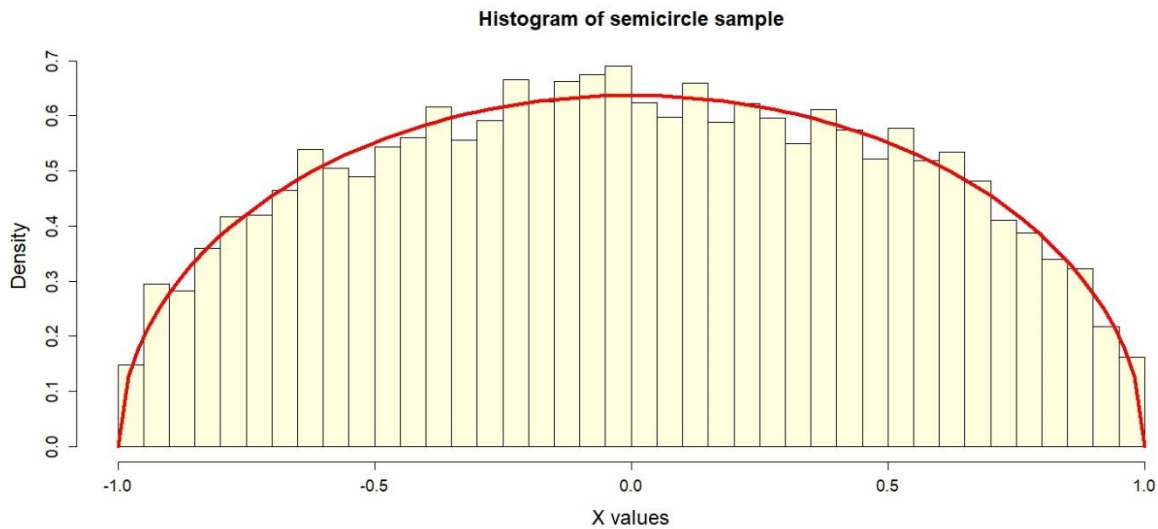
To facilitate the use of these measures of asymmetry, a script was developed for the R program, allowing for both point and interval estimation. The script adjusts the results based on the sampling distributions

of the three indexes studied. Both asymptotic and bootstrap confidence intervals are computed, enabling the testing of the normality of the random sample of the quantitative variable  $X$  to determine which interval to use. Three methods are included to obtain the bootstrap confidence intervals: normal, percentile, and bias-corrected and accelerated percentile. Thus, the symmetry (D'Agostino test [24]), mesokurtosis (Anscombe-Glynn test [40]), and normality (Shapiro-Francia [42-43] and D'Agostino-Belanger-D'Agostino [44] tests) of the bootstrap sampling distribution are tested, and the bias and acceleration ( $a$ ) are reported to select the appropriate method. As an example, the script was applied to a randomly generated sample of 60 data points. To make the example more meaningful, the sample data were generated to correspond to the proportional (monthly) return of stock prices in a volatile stock market, using a Cauchy distribution with a location parameter ( $x_0$ ) of 0.05, representing the population median, and a scale parameter ( $\gamma$ ) of 0.38, constituting the population semi-interquartile range [46-47].

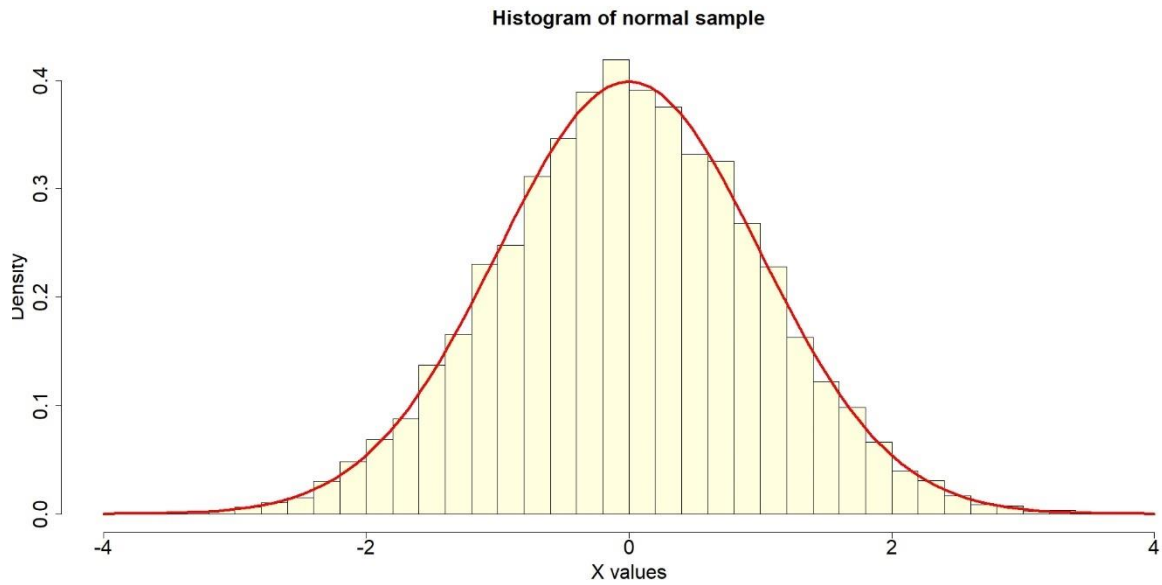
## 5. Results

### 5.1. Sampling distribution of $AISr$ , $RISr$ and $PCS$

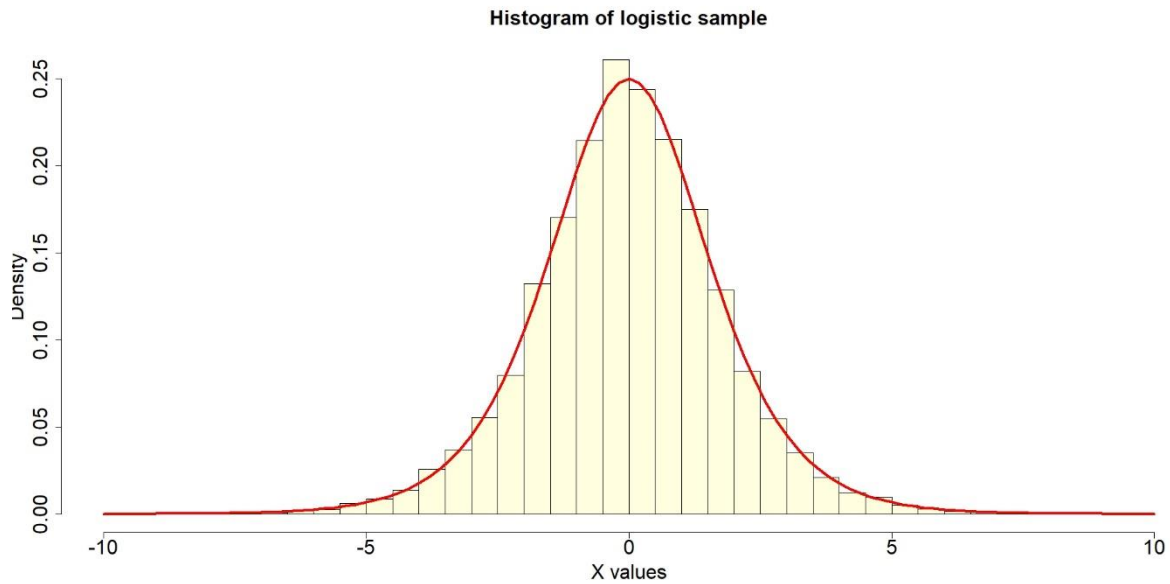
The random samples of 10,000 data points (source samples) are plotted as histograms (50 bins of uniform width) with the curve of the theoretical or generating density function overlaid. Figure 1 shows the sample randomly drawn from the standard Wigner semicircular distribution (unit radius). Figure 2 displays the sample randomly drawn from the standard normal distribution, and Figure 3 shows the sample randomly drawn from the standard logistic distribution. From these three source samples, 1,000 samples of 10,000 data points each were drawn with replacement. In each of these samples, Kelley's absolute and relative reversed indices, as well as the percentile coefficient of skewness, were calculated, thus obtaining the bootstrap sampling distribution of these measures of skewness.



**Figure 1.** Histogram with the overlaid standard semicircular distribution curve. Source: prepared by the author.



**Figure 2.** Histogram with the overlaid standard normal curve. Source: prepared by the author.



**Figure 3.** Histogram with the overlaid standard logistic curve. Source: prepared by the author.

At the 5% significance level, the bootstrap sampling distributions of  $AIS_r$  showed no outliers by Grubb's test in the three samples generated from 10,000 data points from three symmetric distributions with different kurtosis. Their profiles were symmetric by the D'Agostino test and mesokurtic by the Anscombe-Glynn test. They were fitted to a normal distribution by the Anderson-Darling, Shapiro-Francia, and D'Agostino-Pearson tests. Thus, the asymptotic and bootstrap standard errors were very similar, with absolute differences less than or equal to 0.0010, resulting in a non-significant difference in the bootstrap sampling distribution derived from the logistic distribution (Tables 1-3). Bootstrap standard errors were small, ranging from 0.007 to 0.026. In the normal Q-Q plots, an alignment of the points around the center line at 45 degrees can be seen, and in the histogram, the bell-shaped profile of the normal distribution can be observed, whether the bootstrap sampling distribution is derived from semicircular, normal, or logistically distributed data (Figures 4-6).

The same results of good fit to normality were also observed in the bootstrap sampling distribution of  $PCS$ , with absolute differences between bootstrap and asymptotic standard errors less than or equal to 0.0015, resulting in a non-significant difference in the bootstrap sampling distribution derived from the

sample generated from the logistic distribution. Bootstrap standard errors were small, ranging from 0.011 to 0.012. See Tables 1-3 and Figures 7-9.

However, the bootstrap sampling distributions of *RISr* deviated from normality, presenting outliers, skewness (positive with data from the semicircular and logistic distributions and negative with data from the normal distribution), and marked leptokurtosis (Tables 1-3). The profile of the normal Q-Q plot shows central flattening with a concave curvature at the left end and convex at the right end, which is typical of a strongly leptokurtic distribution. This non-normal, pointed profile can be observed in the histogram, whether the bootstrap sampling distribution is obtained from semicircular, normal, or logistically distributed data (Figures 10-11). Bootstrap standard errors were very large. The largest occurred with the mesokurtic distribution (*bse* = 38.998), followed by the leptokurtic (*bse* = 13.294) and platykurtic (*bse* = 4.897) distributions. Consequently, these estimates are not reliable.

**Table 1.** Tests of randomness, symmetry and meso-kurtosis for the bootstrap sampling distribution of three measures of skewness.

Type of distribution	Statistic of skewness	Grubbs' test			D'Agostino's test			Anscombe-Glynn test		
		g	u	<i>p</i> -value	$\sqrt{b_1}$	z	<i>p</i> -value	b <sub>2</sub>	z	<i>p</i> -value
Semi-circle	AISr	3.307	0.989	0.457	-0.060	-0.774	0.439	2.884	-0.687	0.492
	RISr	29.773	0.112	<0.001	26.136	44.543	<0.001	788.660	23.624	<0.001
	PCS	3.302	0.990	0.467	-0.058	-0.748	0.454	2.887	-0.669	0.504
Normal	AISr	2.984	0.991	0.999	0.035	0.456	0.649	2.900	-0.571	0.568
	RISr	29.376	0.135	<0.001	-25.454	-44.260	<0.001	747.730	23.566	<0.001
	PCS	3.027	0.991	0.999	0.0247	0.321	0.749	2.905	-0.538	0.591
Logistic	AISr	3.596	0.987	0.155	-0.040	-0.519	0.604	3.014	0.203	0.839
	RISr	18.725	0.649	<0.001	6.065	28.904	<0.001	168.091	21.445	<0.001
	PCS	3.490	0.988	0.233	-0.032	-0.417	0.677	2.999	0.106	0.915

<sup>1</sup> Note. Statistic: *AISr* = Reversed Absolute Index of Skewness, *RISr* = Reversed Relative Index of Skewness, and *PCS* = Percentile Coefficient of Skewness. Grubbs' test for one outlier: *g* = Grubbs' test statistic, *u* = normalized test statistic, and *p*-value = one-tailed *p*-value (alternative hypothesis: lowest or highest value is an outlier). D'Agostino skewness test:  $\sqrt{b_1}$  = measure of skewness based on the standardized third central moment, *z* = test statistic, and *p*-value = two-tailed *p*-value (alternative hypothesis: data have a skewness towards left or right tail). Anscombe-Glynn kurtosis test: *b*<sub>2</sub> = measure of kurtosis based on the standardized fourth central moment. *z* = test statistic, and *p*-value = two-tailed *p*-value (alternative hypothesis: the value of kurtosis [ $\beta_2$ ] is not equal to 3). Probability values less than 0.05, indicating the statistical test's significance, are highlighted in bold. Source: prepared by the author.

**Table 2.** Tests of normality for the bootstrap sampling distribution of three measures of skewness.

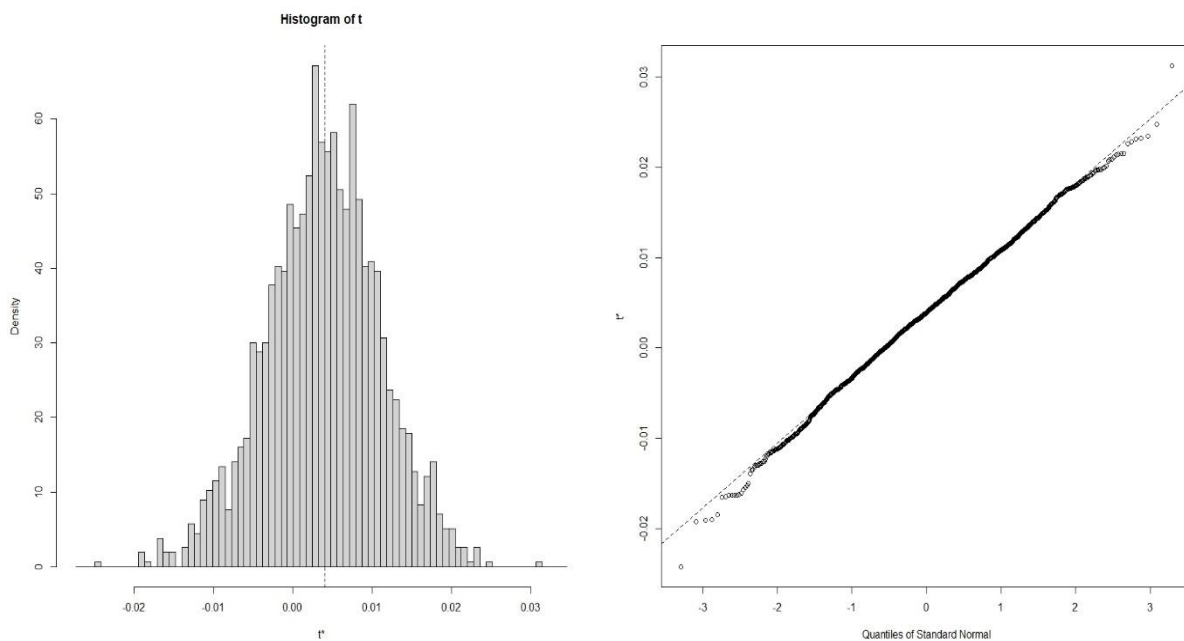
Type of distribution	Statistic of skewness	Anderson-Darling		Shapiro-Francia		D'Agostino et al.	
		AD	<i>p</i> -value	w	<i>p</i> -value	k <sup>2</sup>	<i>p</i> -value
Semi-circle	AISr	0.172	0.930	0.999	0.830	1.072	0.585
	RISr	285.900	<0.001	0.088	<0.001	2542.163	<0.001
	PCS	0.167	0.938	0.999	0.863	1.007	0.604
Normal	AISr	0.312	0.551	0.999	0.631	0.533	0.766
	RISr	267.320	<0.001	0.110	<0.001	2514.260	<0.001
	PCS	0.319	0.534	0.999	0.675	0.392	0.822
Logistic	AISr	0.288	0.617	0.998	0.434	0.311	0.856
	RISr	215.960	<0.001	0.284	<0.001	1295.340	<0.001
	PCS	0.2873	0.620	0.998	0.475	0.185	0.912

<sup>1</sup> Note. Statistic: *AISr* = Reversed Absolute Index of Skewness, *RISr* = Reversed Relative Index of Skewness, and *PCS* = Percentile Coefficient of Skewness. Anderson-Darling normality test: *AD* = test statistic, and *p*-value (alternative hypothesis: sample was not drawn from a normal distribution). Shapiro-Francia normality test: *w* = test statistic, and *p*-value (the same alternative hypothesis). D'Agostino-Berlanger-D'Agostino Normality test: *k*<sup>2</sup> = test statistic, and *p*-value (the same alternative hypothesis). Probability values less than 0.05, indicating the statistical test's significance, are highlighted in bold. Source: prepared by the author.

**Table 3.** Point estimation of skewness measures and bootstrap confidence interval, bias and comparison between bootstrap and asymptotic standard errors.

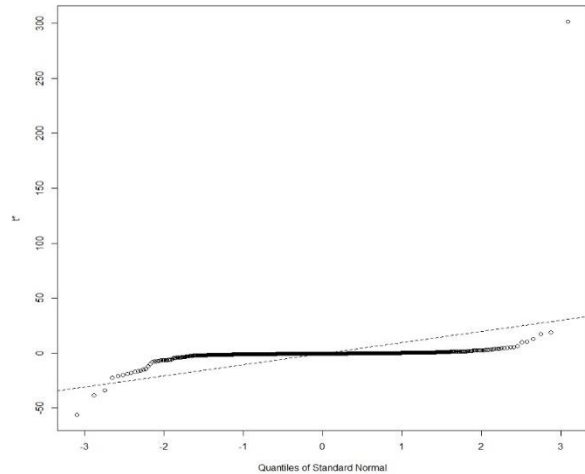
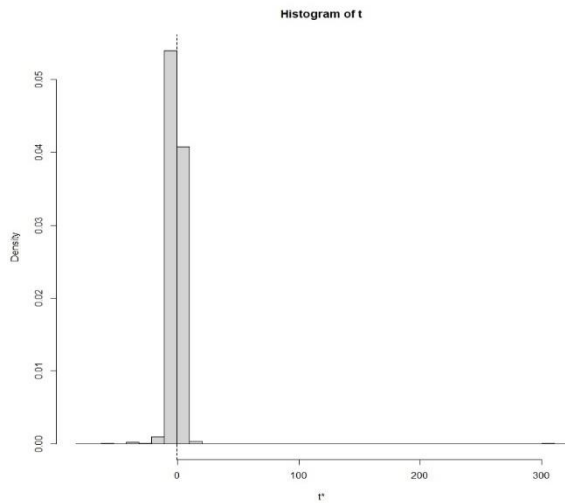
Distribution	Statistic	value	Method	LL	UL	bias	a	bse	ase	diff	p-value
Semi-circle	AISr	0.004	norm	-0.010	0.018	0.00002	0.00043	0.007	0.008	-0.0010	<0.001
	RISr	-0.507	BCa	-2.029	4.113	-0.0603	-0.00014	4.897	1.031	3.8669	<0.001
	PCS	0.006	norm	-0.015	0.026	0.00003	0.00044	0.011	0.012	-0.0015	<0.001
Normal	AISr	0.003	norm	-0.023	0.029	-0.0004	0.00163	0.013	0.015	-0.0018	<0.001
	RISr	-0.255	BCa	-3.759	12.332	-0.7699	-0.00212	38.998	1.374	37.6246	<0.001
	PCS	0.002	norm	-0.018	0.023	-0.0003	0.00162	0.011	0.012	-0.0015	<0.001
Logistic	AIS	-0.004	norm	-0.055	0.045	0.0011	0.00007	0.026	0.026	-0.0003	0.613
	RIS	0.169	BCa	-5.167	21.436	-0.1044	-0.00004	13.294	1.190	12.1045	<0.001
	PCS	-0.002	norm	-0.025	0.021	0.0005	0.00009	0.012	0.012	-0.0002	0.461

<sup>1</sup> Note. Statistic: *AISr* = Reversed Absolute Index of Skewness, *RISr* = Reversed Relative Index of Skewness, and *PCS* = Percentile Coefficient of Skewness. Method: norm = normal from library(boot) and BCa = bias-corrected and accelerated bootstrap from library 'bcjack'. *LL* = low limit and *UL* = upper limit of the 95% bootstrap confidence interval, *bias* = difference between the bootstrap estimate (mean in the bootstrap sampling distribution of the statistic) and the original sample estimate, *bse* = bootstrap standard error, *ase* = asymptotic standard error, *diff* = difference between errors: *bse* - *ase*, *p*-value = one-tailed *p*-value for chi-square test of a single variance (alternative hypothesis: bootstrap variance is lower (left tail) or higher (right tail) than the asymptotic variance expected from the normal approximation). Probability values less than 0.05, indicating the statistical test's significance, are highlighted in bold. Source: prepared by the author.

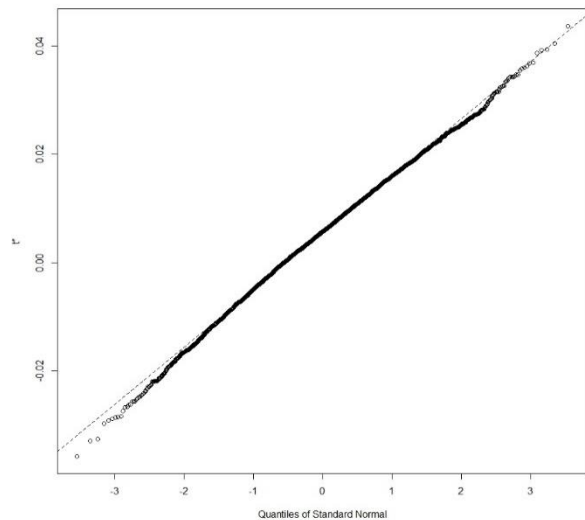
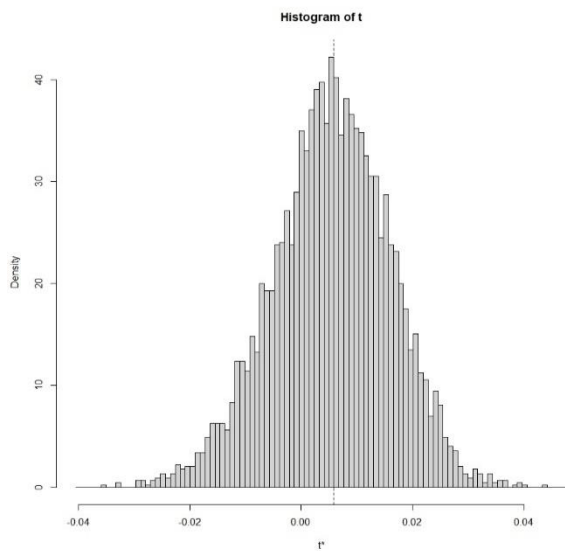


**Figure 4.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution of *AISr* from the 10,000-data sample drawn from a semicircle distribution. Source: prepared by the author.

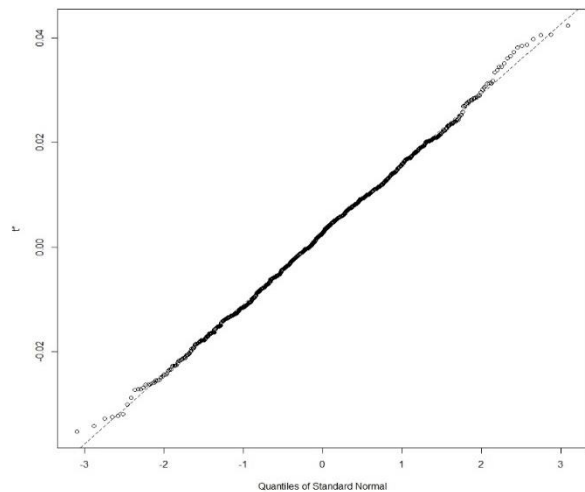
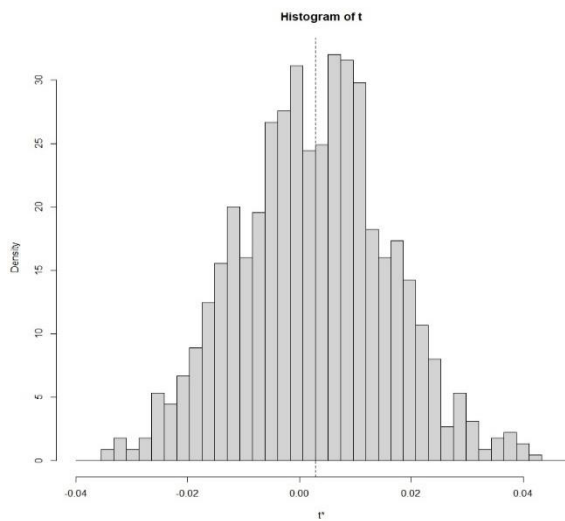




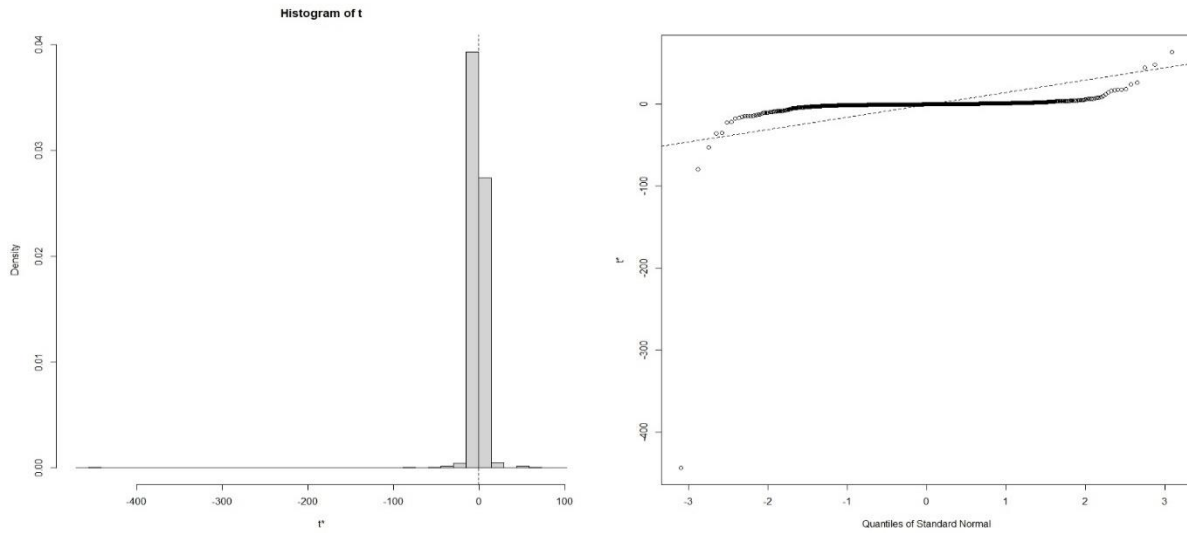
**Figure 5.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution of  $RISr$  from the 10,000-data sample drawn from a semicircle distribution. Source: prepared by the author.



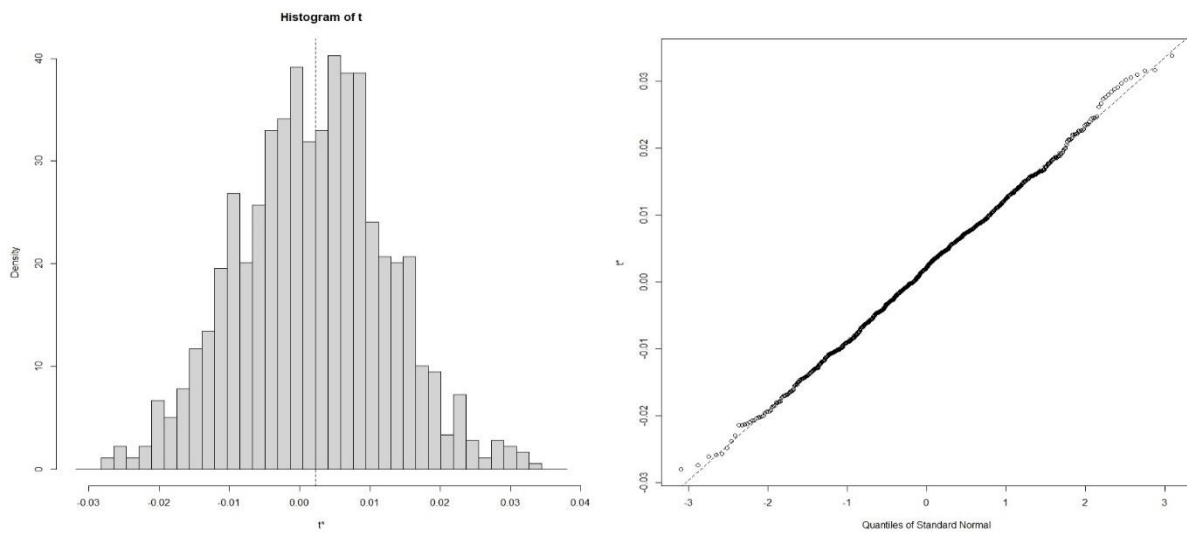
**Figure 6.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution of  $PCS$  from the 10,000-data sample drawn from a semicircle distribution. Source: prepared by the author.



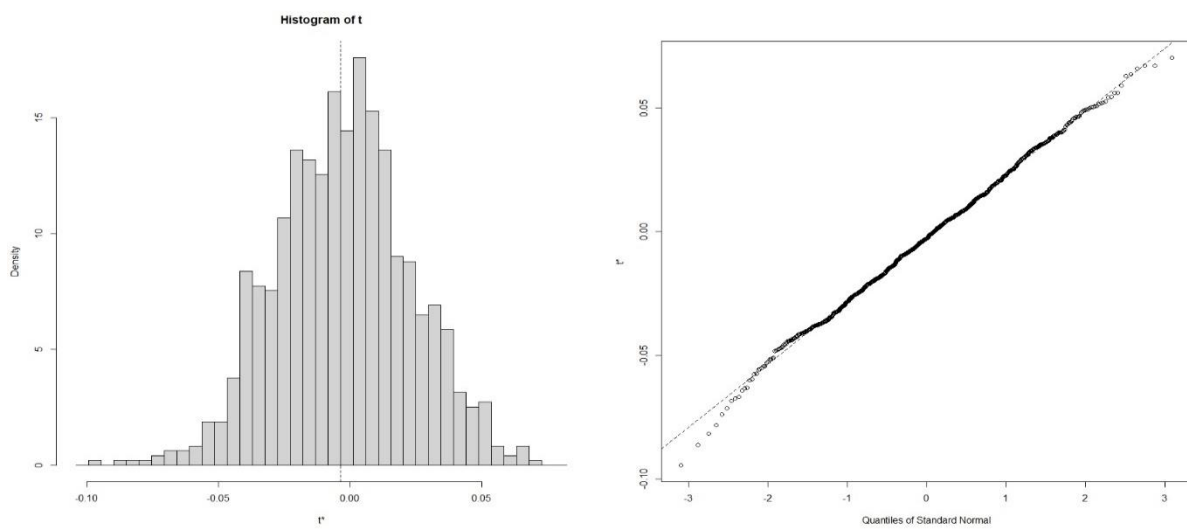
**Figure 7.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution of  $AISr$  from the 10,000-data sample drawn from a normal distribution. Source: prepared by the author.



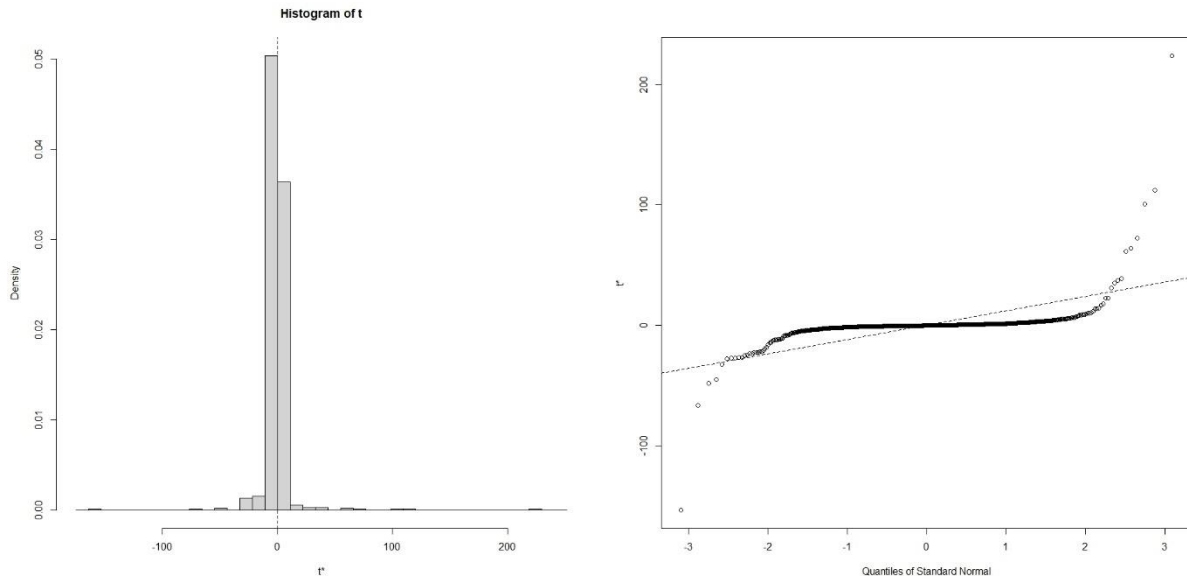
**Figure 8.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution of  $RISr$  from the 10,000-data sample drawn from a normal distribution. Source: prepared by the author.



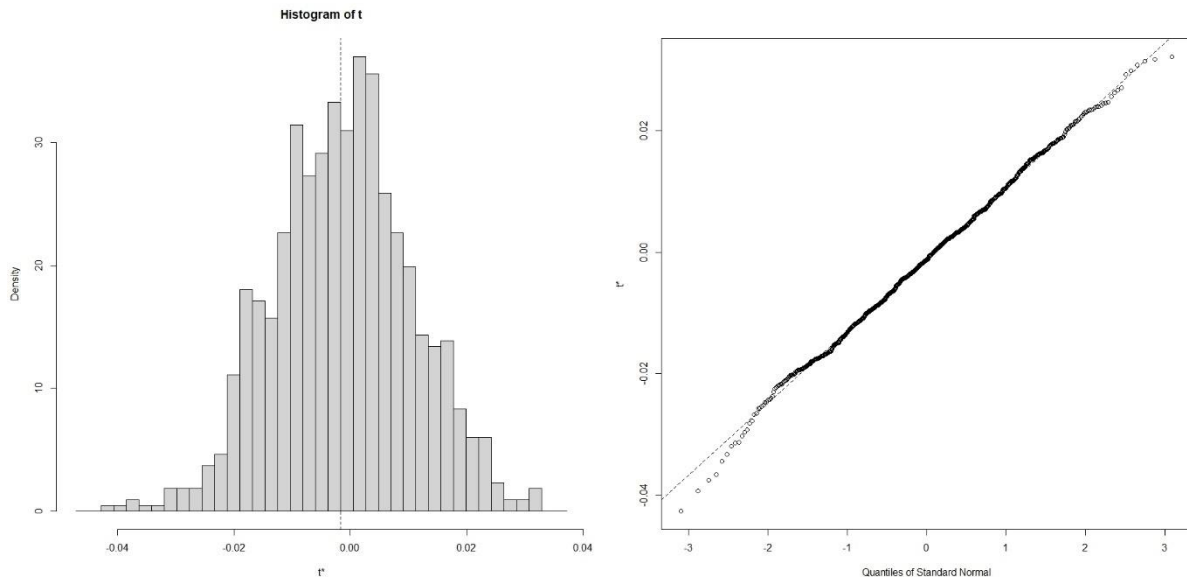
**Figure 9.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution of  $PCS$  from the 10,000-data sample drawn from a normal distribution. Source: prepared by the author.



**Figure 10.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution of  $AISr$  from the 10,000-data sample drawn from a logistic distribution. Source: prepared by the author.



**Figure 11.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution of  $RISr$  from the 10,000-data sample drawn from a logistic distribution. Source: prepared by the author.



**Figure 12.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution of  $PCS$  from the 10,000-data sample drawn from a logistic distribution. Source: prepared by the author.

## 5.2. Script for point and interval estimation of $AISr$ and $PCS$

### 5.2.1. Content and structure of the script

The randomness of the sample is a fundamental assumption for statistical inference [48], so the script starts by testing this assumption using the Wald-Wolfowitz runs test. Running the test in the R program requires downloading the 'randtests' library [49].

For this and the other inferential tests in the script, the significance level is set at 0.05, which is the most conventional value [50]. However, it can be modified considering the sample size [51]. It can be raised to 0.1 with small samples (20 to 29) or lowered to 0.01 with large samples (1000 or more). It can also be changed based on a theoretical approach [52] or to control for family-wise error rate [53].

The a posteriori statistical power ( $\Phi$ ) is calculated by bootstrapping. From the original sample, 1000 samples are generated, and for each of these 1000 samples, the exact probability of the two-tailed test ( $p\text{-value} = \text{'exact'}$ ) is calculated. However, one can switch to the asymptotic probability in the case of a large sample, where  $n_0$  (the number of data points smaller than the median or criterion) and  $n_1$  (the

number of data points larger than the median) are each greater than or equal to 20 (p-value = 'normal'). The number of probabilities less than the significance level, divided by the number of simulations, yields the statistical power or the probability of rejecting the null hypothesis when it is false [26]. See Equation 13, where  $I$  is the indicator function,  $B$  represents the number of simulations, and  $\alpha$  denotes the significance level.

$$\phi = \frac{\sum_{i=1}^B I(p_{two-sided\_exat} < \alpha)}{B} \quad (13)$$

The script continues by testing for normality. In case of a good fit, the use of the asymptotic error and confidence interval for  $AISr$  and  $PCS$  can be chosen. This distributional assumption is tested through four tests with different inferential foundations [54], and additionally, the statistical power is computed. Executing these tests in R requires loading the 'moments' [55] and 'nortest' [56] libraries.

The normality check starts with the D'Agostino-Berlanger-D'Agostino  $K^2$  test [44] based on the sum of squares of the standardized values of skewness [24] and kurtosis [40], whose statistic converges in distribution to a chi-square distribution with two degrees of freedom. With this test, a posteriori statistical power is calculated with the complementary cumulative distribution function of a non-central chi-square distribution.

The second normality test is the Lilliefors test [57], which assesses the maximum linear distance between the empirical and theoretical cumulative distribution functions. The third test is the Anderson-Darling test [41], which measures the standardized quadratic distance between these functions. The fourth and final test is the Shapiro-Francia test [42-43], based on the shared variance between empirical and theoretical quantiles. With these three tests, a posteriori statistical power is obtained through bootstrapping. They were chosen for their recommended power and adequacy [58-61].

The assessment of normality through inferential tests is complemented by graphical representation using a histogram with overlaid density and normal curves. The uniform amplitude and number of bins are determined using the Freedman-Diaconis optimization rule (breaks = "fd"), which imposes no distributional assumption. However, this can be switched to the Scott's rule (breaks = "scott") if the data fit well to a normal distribution [62]. The density for the overlaid curve is computed using the Epanechnikov's kernel function [63], known for its parabolic shape that minimizes integrated mean square error most effectively [64]. The bandwidth is computed using the Sheather-Jones procedure [65], which minimizes asymptotic integrated mean square error and is considered one of the most suitable methods for this purpose [66-67].

It should be noted that when running the script using the R program downloaded on your computer, the plot can be saved as a high-definition JPEG file. Simply expand the R graphics device window (ACTIVE), where in its toolbar under Files, you will find the 'Save As' function that offers this option [68].

After these checks, the script proceeds to the point and interval estimation of the additive complement of  $AIS$  and  $PCS$ . The additive complement of the  $IAS$  was chosen for two reasons: firstly, to maintain interpretative consistency with the classic index established by Pearson [3-4] — positive values indicate right-tailed skewness, and negative values indicate left-tailed skewness. Secondly, because  $PCS$  represents the standardized form of the additive complement of  $AIS$  once it is divided by the semi-percentile range. Quantiles were calculated using R's rule 8 [18-19].  $RISr$  is omitted from the script due to its excessive bootstrap standard error. However, its calculation with R applied to a sample of 60 data points is shown at the end of this Results section (sub-section 5.3).

The confidence level is set at 95%, corresponding to the 5% significance level chosen for the inferential tests. For the bootstrap confidence intervals, 1000 random samples were created by sampling with replacement from the original sample, and the statistics  $AISr$  and  $PCS$  were calculated for each of these 1000 samples [26]. Thus, bootstrap sampling distributions were generated for these two statistics. The symmetry of the bootstrap sampling distribution was assessed using the D'Agostino test [24], kurtosis was tested using the Anscombe-Glynn test [40], and normality was evaluated using the Shapiro-Francia test [42-43]. Additionally, bootstrap bias, bootstrap standard error ( $bse$ ), jackknife acceleration ( $a$ ), and the density histogram and Q-Q normal plot of the bootstrap sampling distributions were obtained [23].

Bootstrap confidence intervals are calculated using three methods: normal (norm), percentile (perc), and bias-corrected and accelerated percentile (BCa). The choice of method depends on the characteristics of the bootstrap sampling distribution. The normal method is preferred when the distribution is approximately normal. If there is no normality, bias ( $|b| \geq 0.05$ ) and asymmetry ( $|a| \geq 0.025$  and significance D'Agostino's skewness test [24]), then the BCa method is the best choice [25]. Alternatively, when there is no normality, but acceleration and bias are minimal, the percentile method remains suitable. To implement these methods, downloading the 'boot' library is necessary [69].

Finally, the bootstrap probability of retaining the null hypothesis of symmetry, assessed using *AISr* and *PCS*, was calculated. This is obtained by counting how many values are less than or equal (left tail) and how many values are greater than or equal (right tail) to the parameter (null hypothesis of symmetry) among the 1000 data from the statistic's bootstrap sampling distribution of *AISr* or *PCS*. The smallest probability is chosen and multiplied by 2. If it is less than or equal to the significance level ( $\alpha$ ), the null hypothesis of symmetry holds in a two-tailed test, otherwise it is rejected [70].

A random sample of at least 20 data points [71] and preferably at least 30 data points [72] from a quantitative variable  $X$  is recommended for using this script. The script is divided into three independent parts to make it lighter to execute. What can be changed in the script to match the user's data is marked in blue, such as the sample or vector of scores.

### 5.2.2. The script with its three parts

```
# First part of the script: randomness and normality of the sample or vector of scores x
x <- c(0.532, -0.248, 1.342, -0.096, -0.022, 0.105, -4.243, -0.084, -2.281, 2.821, -0.002, 2.623, -0.559, -1.586,
0.177, -0.074, 0.421, 0.101, 0.683, -0.005, -0.087, -0.499, -0.754, 0.043, -0.664, -0.445, -2.677, -1.197, 0.537,
0.239, 0.006, -0.070, -0.506, -0.234, 0.079, 5.489, -0.310, 0.357, 0.641, 0.388, 0.233, 1.431, 1.417, 0.919, 0.247,
0.227, 0.392, 3.590, 0.470, -0.132, 0.105, 2.120, -0.228, 0.203, -1.910, 0.338, 0.211, -0.322, -0.080, 0.963)

cat("Testing the randomness of the sample x using the Wald-Wolfowitz runs test", "\n")
library(randtests)
runse <- runs.test(x, alternative = "two.sided", threshold = median(x), pvalue = 'exact')
runsa <- runs.test(x, alternative = "two.sided", threshold = median(x), pvalue = 'normal')
alpha <- 0.05
ww_power <- function(x, alpha, B = 1000) {n <- length(x)
p_values <- numeric(B)
for (i in 1:B) {bootstrap_sample <- sample(x, replace = TRUE)
result <- runs.test(x, alternative = "two.sided", threshold = median(x), pvalue = 'exact')
p_values[i] <- result$p.value}
power <- mean(p_values < alpha)
return(power)}
set.seed(123)
power <- ww_power(x, alpha)
cat("Wald-Wolfowitz runs test. Criterion: median", "\n")
cat("Number of runs: r =", runse$runs, "\n")
cat("n_0 = #(x_i < mdn(x)) =", runse$parameter["n1"], "y", "n_1 = #(x_i > mdn(x)) =",
runse$parameter["n2"], "\n")
cat("n = n_0 + n_1 =", runse$parameter["n"], "\n")
cat("Two-tailed exact probability value: p =", round(runse$p.value, 3), "\n")
cat("Mean: M(R|n_0, n_1) =", runse$mu, "y", " Standard deviation: DE(R|n_0, n_1) =",
round(sqrt(runse$var), 3), "\n")
cat("Standardized number of runs: z_r =", round(runse$statistic, 3), "\n")
cat("Two-tailed asymptotic probability value: p =", round(runsa$p.value, 3), "\n")
cat("Statistical power for the Wald-Wolfowitz runs test using bootstrap simulation:  $\phi$  =", power, "\n")
cat("Testing for normality using four tests with different rationales:", "\n")
library(moments)
```

```

agostino <- agostino.test(x, alternative = "two.sided")
ag <- anscombe.test(x, alternative = "two.sided")
K2 <- (agostino$statistic["z"]^2 + ag$statistic["z"]^2)
p <- pchisq(K2, df = 2, lower.tail = FALSE)
power_K2 <- 1 - pchisq(qchisq(alpha, df = 2, lower.tail = FALSE), df = 2, ncp = K2, lower.tail = TRUE,
log.p = FALSE)
cat("Test for normality using the D'Agostino-Berlanger-D'Agostino test.", "\n")
cat("Test statistic: k^2(x) =", round(K2, 3), "\n")
cat("Asymptotic probability value in a two-tailed test: p_value =", round(p, 3), "\n")
if (p < alpha) {cat(sprintf("The null hypothesis of normality is rejected at a significance level of %.2f using
K squared test.", alpha), "\n")}
} else {cat(sprintf("The null hypothesis of normality is maintained at a significance level of %.2f using K
squared test.", alpha), "\n")}
cat("The right-tailed statistical power for the alternative hypothesis of non-normality for the K squared
test:  $\phi$  =", round(power_K2, 4), "\n")
library(nortest)
lillie.test(x)
result_lillie <- lillie.test(x)
if (result_lillie$p.value < alpha) {cat(sprintf("The null hypothesis of normality is rejected at a significance
level of %.2f using Lilliefors test.", alpha), "\n")}
} else {cat(sprintf("The null hypothesis of normality is maintained at a significance level of %.2f using
Lilliefors test.", alpha), "\n")}
lillie_power <- function(x, alpha, B = 1000) {n <- length(x)
p_values <- numeric(B)
for (i in 1:B) {bootstrap_sample <- sample(x, replace = TRUE)
result <- lillie.test(bootstrap_sample)
p_values[i] <- result$p.value}
power <- mean(p_values < alpha)
return(power)}
set.seed(123)
power <- lillie_power(x, alpha)
cat("Statistical power for the Lilliefors normality test using bootstrap simulation:  $\phi$  =", power, "\n")

ad.test(x)
result_ad <- ad.test(x)
if (result_ad$p.value < alpha) {cat(sprintf("The null hypothesis of normality is rejected at a significance
level of %.2f using Anderson-Darling test.", alpha), "\n")}
} else {cat(sprintf("The null hypothesis of normality is maintained at a significance level of %.2f using
Anderson-Darling test.", alpha), "\n")}
ad_power <- function(x, alpha, B = 1000) {n <- length(x)
p_values <- numeric(B)
for (i in 1:B) {bootstrap_sample <- sample(x, replace = TRUE)
result <- ad.test(bootstrap_sample)
p_values[i] <- result$p.value}
power <- mean(p_values < alpha)
return(power)}
set.seed(123)
power <- ad_power(x, alpha)
cat("Statistical power for the Anderson-Darling normality test using bootstrap simulation:  $\phi$  =", power,
"\n")

```

```

sf.test(x)
result_sf <- sf.test(x)
if (result_sf$p.value < alpha) {cat(sprintf("The null hypothesis of normality is rejected at a significance
level of %.2f using Shapiro-Francia test.", alpha), "\n")}
} else {cat(sprintf("The null hypothesis of normality is maintained at a significance level of %.2f using
Shapiro-Francia test.", alpha), "\n")}
sf_power <- function(x, alpha, B = 1000) {n <- length(x)
p_values <- numeric(B)
for (i in 1:B) {bootstrap_sample <- sample(x, replace = TRUE)
result <- sf.test(bootstrap_sample)
p_values[i] <- result$p.value}
power <- mean(p_values < alpha)
return(power)}
set.seed(123)
power <- sf_power(x, alpha)
cat("Statistical power for the Shapiro-Francia normality test using bootstrap simulation:  $\phi$  =", power,
"\n")

# Histogram with overlaid density and normal curves
# breaks = "scott" in case of normality or breaks = "sturges" in the case of symmetry
hist(x, breaks = "fd", col = "darkolivegreen2", border = "black", freq = FALSE, main = "Histogram with
overlaid normal and density curves", xlab = "X values", ylab = "Density", ylim = c(0, 0.8), cex.main = 1.5,
cex.lab = 1.5, cex.axis = 1.2)
lines(density(x, kernel = "epanechnikov", bw = "sj"), col = "darkblue", lwd = 4)
x_values <- seq(mean(x) - 4 * sd(x), mean(x) + 4 * sd(x), length = 1000)
y_values <- dnorm(x_values, mean = mean(x), sd = sd(x))
lines(x_values, y_values, col = "red", lwd = 4)

# Second part of the script: AISr
x <- c(0.532, -0.248, 1.342, -0.096, -0.022, 0.105, -4.243, -0.084, -2.281, 2.821, -0.002, 2.623, -0.559, -1.586,
0.177, -0.074, 0.421, 0.101, 0.683, -0.005, -0.087, -0.499, -0.754, 0.043, -0.664, -0.445, -2.677, -1.197, 0.537,
0.239, 0.006, -0.070, -0.506, -0.234, 0.079, 5.489, -0.310, 0.357, 0.641, 0.388, 0.233, 1.431, 1.417, 0.919, 0.247,
0.227, 0.392, 3.590, 0.470, -0.132, 0.105, 2.120, -0.228, 0.203, -1.910, 0.338, 0.211, -0.322, -0.080, 0.963)

library(boot)
library(moments)
library(nortest)
alpha <- 0.05

cat("Point estimate of the additive opposite of Kelley's absolute skewness index", "\n")
n <- length(x)
q0.1 <- quantile(x, 0.10, type = 8)
q0.5 <- quantile(x, 0.50, type = 8)
q0.9 <- quantile(x, 0.90, type = 8)
AISr = (q0.1 + q0.9) / 2 - q0.5
cat("Sample size: n =", n, "\n")
cat("Quantile of order 0.1 by the rule 8: q_0.1(x) =", round(q0.1, 3), "\n")
cat("Quantile of order 0.5 by the rule 8: q_0.5(x) =", round(q0.5, 3), "\n")
cat("Quantile of order 0.1 by the rule 8: q_0.9(x) =", round(q0.9, 3), "\n")
cat("Additive opposite of Kelley's absolute asymmetry index: AISr = ", round(AISr, 3), "\n")

```

```

cat("Asymptotic confidence interval of the additive opposite of Kelley's absolute skewness index", "\n")
ase <- 0.599143 * (quantile(x, 0.9, type = 8) - quantile(x, 0.10, type = 8)) / sqrt(length(x))
LL = AISr - qnorm(1-alpha/2) * ase
UL = AISr + qnorm(1-alpha/2) * ase
cat("Asymptotic standard error of AISr: ase(AISr) =", round(ase, 3), "\n")
cat("Asymptotic confidence interval", (1 - alpha) * 100, "% for AISr: 95% CI (" , round(LL, 3), ",", round(UL,
3),")\n")

# Acceleration (jackknife)
AISr_jack <- numeric(n)
for (i in 1:n) {x_jack <- x[-i]}
AISr_jack[i] <- quantile((quantile(x_jack, 0.10, type = 8) + quantile(x_jack, 0.90, type = 8))/2 - x_jack, 0.50,
type = 8)}
acel <- sum((mean(AISr_jack) - AISr_jack)^3) / (6 * sum((mean(AISr_jack) - AISr_jack)^2)^(3/2))
cat("Aceleración (jackknife): a =", round(acel, 6), "\n")
set.seed(123)
b <- boot(data = x, function(x, i) {(quantile(x[i], 0.10, type = 8) + quantile(x[i], 0.90, type = 8))/2 -
quantile(x[i], 0.50, type = 8)}, R=1000)
boot_est <- mean(b$t)
cat("Bootstrap estimation: m(b$t) =", round(boot_est, 3), "\n")
agostino_boot <- agostino.test(b$t, alternative = "two.sided")
ag_boot <- anscombe.test(b$t, alternative = "two.sided")
print(agostino_boot)
print(ag_boot)
cat("D'Agostino-Berlanger-D'Agostino normality test", "\n")
K2 <- (agostino_boot$statistic["z"]^2 + ag_boot$statistic["z"]^2)
p <- pchisq(K2, df = 2, lower.tail = FALSE)
cat("K^2 =", round(K2, 3), "\n")
cat("p_value =", round(p, 3), "\n")
sf.test(b$t)
b
plot(b)
boot.ci(b, conf = 0.95, type = c("bca", "perc", "norm"))
p_boot_left <- mean(b$t < 0)
p_boot_right <- mean(b$t > 0)
p_boot <- min(p_boot_left, p_boot_right)
cat("One-tailed bootstrap probability value for the null hypothesis of symmetry: p =", round(p_boot, 3),
"\n")
if (p_boot < alpha) {cat(sprintf("The null hypothesis of normality is rejected at a significance level of %.2f
using bootstrap probability.", alpha), "\n")}
} else {cat(sprintf("The null hypothesis of normality is maintained at a significance level of %.2f using
bootstrap probability.", alpha), "\n")}

# Third part of the script: PCS

x <- c(0.532, -0.248, 1.342, -0.096, -0.022, 0.105, -4.243, -0.084, -2.281, 2.821, -0.002, 2.623, -0.559, -1.586,
0.177, -0.074, 0.421, 0.101, 0.683, -0.005, -0.087, -0.499, -0.754, 0.043, -0.664, -0.445, -2.677, -1.197, 0.537,
0.239, 0.006, -0.070, -0.506, -0.234, 0.079, 5.489, -0.310, 0.357, 0.641, 0.388, 0.233, 1.431, 1.417, 0.919, 0.247,
0.227, 0.392, 3.590, 0.470, -0.132, 0.105, 2.120, -0.228, 0.203, -1.910, 0.338, 0.211, -0.322, -0.080, 0.963)

library(bcaboot)

```



```

library(boot)
library(moments)
library(nortest)

cat("Point estimate of the percentile coefficient of skewness", "\n")
PCS <- (quantile(x, 0.10, type = 8) + quantile(x, 0.90, type = 8) - 2* quantile(x, 0.50, type = 8)) / (quantile(x,
0.90, type = 8) - quantile(x, 0.10, type = 8))
cat("Percentage coefficient of skewness: PCS(x) = ", PCS, "\n")

cat("Asymptotic confidence interval of percentile coefficient of skewness", "\n")
ase <- 1.198286 / sqrt(length(x))
alpha <- 0.05
LL <- PCS - qnorm(1 - alpha / 2) * ase
UL <- PCS + qnorm(1 - alpha / 2) * ase
cat("Asymptotic standard error of PCS: ase(PCS) =", round(ase, 3), "\n")
cat("Asymptotic confidence interval", (1 - alpha) * 100, "% for PCS: 95% CI (", round(LL, 3), ",", round(UL,
3), ")\n")

cat("Bootstrap confidence interval of the percentile coefficient of skewness", "\n")
# Aceleración (jackknife)
n <- length(x)
PCS_jack <- numeric(n)
for (i in 1:n) {x_jack <- x[-i]
PCS_jack[i] <- (quantile(x_jack, 0.10, type = 8) + quantile(x_jack, 0.90, type = 8) - 2 * quantile(x_jack, 0.50,
type = 8)) / (quantile(x_jack, 0.90, type = 8) - quantile(x_jack, 0.10, type = 8))}
acel <- sum((mean(PCS_jack) - PCS_jack)^3) / (6 * sum((mean(PCS_jack) - PCS_jack)^2)^(3/2))
cat("Acceleration (jackknife): a =", round(acel, 9), "\n")
library(boot)
set.seed(123)
b<-boot(data=x, function(x, i) {(quantile(x[i], 0.10, type = 8) + quantile(x[i], 0.90, type = 8) - 2*quantile(x[i],
0.50, type = 8)) / (quantile(x[i], 0.90, type = 8) - quantile(x[i], 0.10, type = 8))}, R=1000)
boot_est <- mean(b$t)
cat("Bootstrap estimation: m(b$t) =", round(boot_est, 3), "\n")
agostino_boot <- agostino.test(b$t, alternative = "two.sided")
ag_boot <- anscombe.test(b$t, alternative = "two.sided")
print(agostino_boot)
print(ag_boot)
cat("D'Agostino-Berlanger-D'Agostino normality test", "\n")
K2 <- (agostino_boot$statistic["z"]^2 + ag_boot$statistic["z"]^2)
p <- pchisq(K2, df = 2, lower.tail = FALSE)
cat("K^2 =", round(K2, 3), "\n")
cat("p_value =", round(p, 3), "\n")
sf.test(b$t)
b
plot(b)
boot.ci(b, conf = 0.95, type=c("bca", "perc", "norm"))
p_boot_left <- mean(b$t < 0)
p_boot_right <- mean(b$t > 0)
p_boot <- min(p_boot_left, p_boot_right)
cat("One-tailed bootstrap probability value for the null hypothesis of symmetry: p =", round(p_boot, 3),
"\n")

```

```

if (p_boot < alpha) {cat(sprintf("The null hypothesis of normality is rejected at a significance level of %.2f
using bootstrap probability.", alpha), "\n")}
} else {cat(sprintf("The null hypothesis of normality is maintained at a significance level of %.2f using
bootstrap probability.", alpha), "\n")}

```

### 5.2.3. The output of the script applied to the vector of scores x

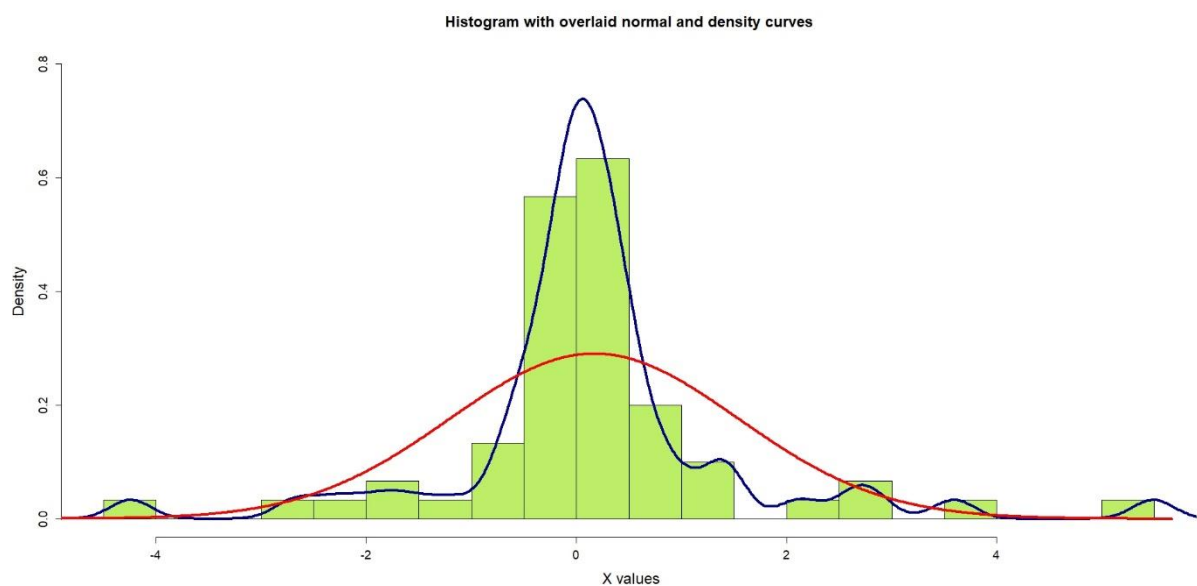
The 60 sample data points, shown in blue in the script, were simulated from a Cauchy distribution with parameters:  $x_0 = 0.05$  (location parameter corresponding to the median) and  $\gamma = 0.38$  (scale parameter corresponding to the semi-interquartile range) [73]. To give the data a concrete context within the field of social sciences, they are considered to correspond to stock price return data in volatile markets. For this type of data, the Cauchy distribution probability model has been found to exhibit good predictive ability [46-47]. The data were rounded to three decimal places. Next, the script for random sample generation is presented, which is stable due to having a fixed seed [74].

```

set.seed(123)
x <- rcauchy(60, location = 0.05, scale = 0.38)
print(round(x, 3))

```

Considering a significance level of 5%, the sample of variable X, composed of 60 data, is random by the Wald-Wolfowitz runs test ( $r = 27$ ,  $n_0 = 30$ ,  $n_1 = 30$ ,  $z(r) = -1.042$ , two-tailed exact probability value:  $p = 0.298$ ,  $\phi = 0$ ). It deviates from normality by the D'Agostino-Berlanger-D'Agostino ( $k^2(x) = 17.457$ ,  $p < 0.001$ ,  $\phi = 0.9705$ ), Lilliefors ( $d = 0.194$ ,  $p < 0.001$ ,  $\phi = 1$ ), Anderson-Darling ( $AD = 3.4071$ ,  $p < 0.001$ ,  $\phi = 1$ ), and Shapiro-Francia ( $w = 0.838$ ,  $p < 0.001$ ,  $\phi = 1$ ) tests. In the histogram with the overlaid density and normal curves (Figure 13), a leptokurtic profile is seen, rather symmetrical and with a peak in the center of the distribution. It is thinned in the shoulder zone, but thickened in the tail zone.



**Figure 13.** Histogram of X values with overlaid density and normal curves. Source: prepared by the author from the script in sub-section 5.2.2, executed with the R program.

The reversed absolute index of skewness ( $AISr$ ) reaches a value of 0.106, close to 0 (Table 4). When the asymptotic confidence interval is calculated at the 95% level, 0 is within the interval (95% CI (-0.267, 0.479);  $ase = 0.190$ ), indicating expected symmetry. However, since the sample of X values does not conform to a normal distribution and is assumed to follow a Cauchy distribution, which lacks finite moments, such an interval is inappropriate and should not be reported. Instead, bootstrap estimation should be used, with a small standard error ( $bse = 0.430$ ), making it reliable.

The bootstrap confidence intervals for the three methods include zero, with the most efficient, having the smallest width, obtained through the percentile method. Its width measures 1.606, compared to

1.664 for the BCa method and 1.685 for the normal method (Table 5). At a 5% significance level, the bootstrap sampling distribution ( $b\hat{t}$ ) exhibits symmetry according to the D'Agostino test and mesokurtosis according to the Anscombe-Glynn test, while conforming to normality according to the D'Agostino-Berlanger-D'Agostino test (Table 6). The histogram displays a bell-shaped profile, and the normal quantile-quantile plot shows good alignment of the data at 45 degrees, except for slight deviations at the extreme points (Figure 14). The Shapiro-Francia test confirms normality at a 1% significance level (Table 6), and the skewness and kurtosis values (Table 4) exceed 0.05. Therefore, the BCa confidence interval is deemed the most appropriate (Table 5). Consistent with the bootstrap confidence interval, the null hypothesis of normality is upheld at a 5% significance level using bootstrap probability (Table 5).

**Table 4.** Point estimation, bootstrap estimation, bootstrap bias, bootstrap standard error, and jackknife acceleration.

Statistic	point_est	boot_est	boot_bias	boot_se	jack_a
AISr	0.106	0.158	0.052	0.430	-0.053
PCS	0.086	0.115	0.029	0.295	-0.055
RISr	1.174	4.470	3.296	71.923	-0.038

<sup>1</sup> Note. Statistic: *AISr* = Reversed Absolute Index of Skewness, *CPS* = Percentile Coefficient of Skewness, and *RISr* = Reversed Relative Index of Skewness, *point\_est* = statistic value or point estimation in the original sample ( $t_0$ ), *boot\_est* =  $\text{mean}(b\hat{t})$  = bootstrap estimation or mean of the statistic in the 1000 bootstrap samples, *boot\_bias* = *boot\_est* -  $t_0$  = bootstrap = bootstrap bias estimation or difference between the bootstrap estimate and point estimation in the original sample, *boot\_se* =  $\text{sd}(b\hat{t})$  = bootstrap standard error or standard deviation of the statistic in the 1000 bootstrap samples, *jack\_a* = jackknife acceleration or measure of the rate of change in the standard deviation ( $\sigma[t_0]$ ) of the estimate ( $t_0 = f(x)$ ) as the data changes. Source: prepared by the author from the script in sub-section 5.2.2, executed with the R program.

**Table 5.** Bootstrap confidence interval for the three statistics and bootstrap probability for null hypothesis of symmetry.

Statistic of skewness	Bootstrap confidence interval at 95%			Bootstrap p_value
	norm	perc	BCa	
AISr	(-0.789, 0.896)	(-0.629, 0.977)	(-0.742, 0.922)	0.356
PCS	(-0.521, 0.634)	(-0.468, 0.612)	(-0.531, 0.556)	0.356
RISr	(-143.09, 138.84)	(-115.42, 158.42)	(-219.02, 51.65)	0.336

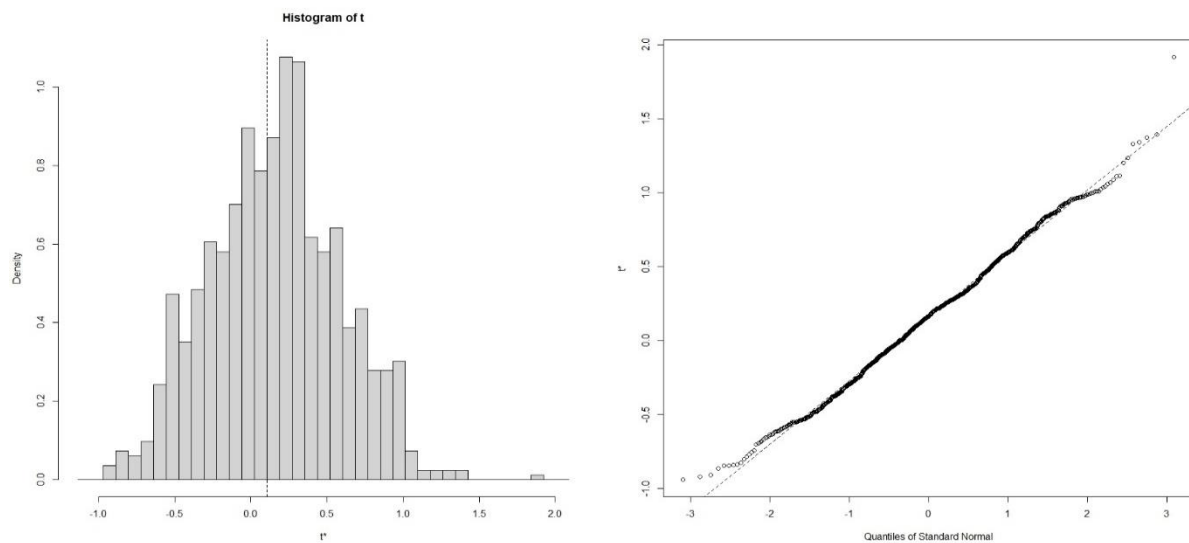
<sup>1</sup> Note. Statistic: *AISr* = Reversed Absolute Index of Skewness, *CPS* = Percentile Coefficient of Skewness, and *RISr* = Reversed Relative Index of Skewness. Confidence intervals using the normal (norm), percentile (perc), and bias-corrected and accelerated percentile (BCa) methods. Calculations based on 1000 extractions with replacement from the original sample. Probability values less than 0.05, indicating the statistical test's significance, are highlighted in bold. Source: prepared by the author from the script in sub-section 5.2.2, executed with R.

**Table 6.** Skewness and normality in the bootstrap sampling distribution of the three statistics of skewness.

Statistic of skewness	D'Agostino: skewness test			Anscombe-Glynn: kurtosis test			DBD test		Shapiro-Francia Royston	
	$\sqrt{b_1}$	z	p	$b_2$	z	p	$k^2$	p	w	p
AISr	0.108	1.400	0.162	2.846	-0.976	0.329	2.912	0.233	0.996	<b>0.010</b>
PCS	-0.233	-2.989	<b>0.003</b>	2.240	-8.221	<b>&lt;0.001</b>	76.522	<b>&lt;0.001</b>	0.982	<b>&lt;0.001</b>
RISr	1.282	13.135	<b>&lt;0.001</b>	44.534	18.365	<b>&lt;0.001</b>	509.83	<b>&lt;0.001</b>	0.441	<b>&lt;0.001</b>

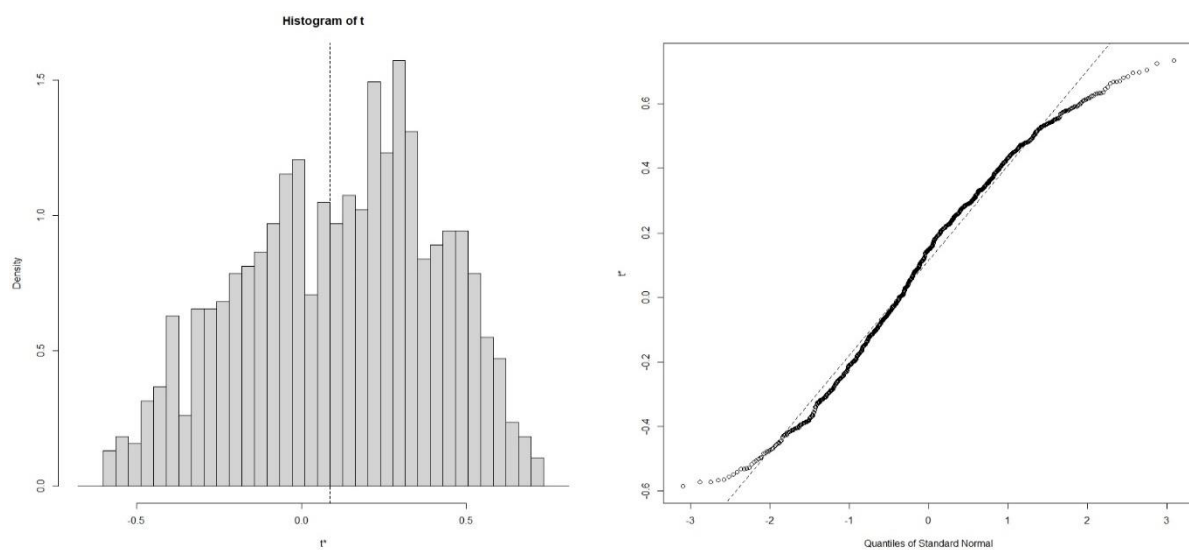
<sup>1</sup> Note. Statistic: *AISr* = Reversed Absolute Index of Skewness, *CPS* = Percentile Coefficient of Skewness, and *RISr* = Reversed Relative Index of Skewness, D'Agostino skewness test:  $\sqrt{b_1}$  = measure of skewness based on the third standardized central moment, z = test statistic or standardized value of  $\sqrt{b_1}$ , p = probability value in a two-tailed test. Anscombe-Glynn kurtosis test:  $b_2$  = measure of kurtosis based on the fourth standardized central moment, z = test statistic or standardized value of  $b_2$ , p = probability value in a two-tailed test, D'Agostino-Berlanger-D'Agostino DBD normality test:  $k^2$  = test statistic, p = probability value, Shapiro-Francia normality test using Royston's procedure: w = test statistic based on the square of the correlation between theoretical and empirical quantiles and p = probability value. Probability values less than 0.05, indicating the statistical test's significance,

are highlighted in bold. Source: prepared by the author from the script in sub-section 5.2.2, executed with the R program.



**Figure 14.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution for *AISr* derived from the 60-data sample of variable *X*. Source: prepared by the author from the script in sub-section 5.2.2, executed with the R program.

The percentile coefficient of skewness reached a value of 0.086, close to 0 (Table 4). The asymptotic confidence interval includes 0 (95% *CI*: (-0.217, 0.389), *ase* = 0.155). However, given the non-normal distribution assumed to follow a Cauchy distribution, this interval is not appropriate. All three bootstrap confidence intervals include 0 and are reliable, with small standard errors (*bse* = 0.295) and bias ( $0.029 < 0.05$ ). The interval with the smallest width is obtained by the percentile method ( $w = 1.080$ ), followed closely by the BCa method ( $w = 1.087$ ), and finally the normal method ( $w = 1.155$ ) (Table 5). The bootstrap sampling distribution of the statistic shows negative skewness and platykurtosis (Table 6), resulting in a histogram with thickened shoulders and shortened tails, and an elongated S-shape in the normal quantile-quantile plot (Figure 15). Given the bootstrap sampling distribution's deviation from normality due to negative skewness and high kurtosis ( $|a| = 0.055 > 0.025$ ), the BCa method is considered the most appropriate (Table 5). Consistent with the bootstrap confidence interval, the null hypothesis of normality is upheld at a significance level of 0.05 using bootstrap probability (Table 5).



**Figure 15.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution for *PCS* derived from the 60-data sample of variable *X*. Source: prepared by the author from the script in sub-section 5.2.2, executed with the R program.

### 5.3. Calculation of *RISr* in the sample of 60 data points using R

In the script, the option to compute the bootstrap confidence interval using the BCa method with the 'bcjack' library [72] was included due to potential memory limitations or timeouts when using the 'boot' library. The acceleration value is derived from computations with the 'bcjack' library, but the display of the resulting bootstrap confidence interval using the BCa method is disabled: `# print(bcjack_results)`. To activate it, uncomment `boot.ci(b, conf = 0.95, type = c("bca", "perc", "norm"))` by adding the symbol `#` in front and remove it from `print(bcjack_results)`. This package requires significantly fewer computational resources compared to the 'boot' library for computing the bootstrap confidence interval using the BCa method.

```
x <- c(0.532, -0.248, 1.342, -0.096, -0.022, 0.105, -4.243, -0.084, -2.281, 2.821, -0.002, 2.623, -0.559, -1.586,
0.177, -0.074, 0.421, 0.101, 0.683, -0.005, -0.087, -0.499, -0.754, 0.043, -0.664, -0.445, -2.677, -1.197, 0.537,
0.239, 0.006, -0.070, -0.506, -0.234, 0.079, 5.489, -0.310, 0.357, 0.641, 0.388, 0.233, 1.431, 1.417, 0.919, 0.247,
0.227, 0.392, 3.590, 0.470, -0.132, 0.105, 2.120, -0.228, 0.203, -1.910, 0.338, 0.211, -0.322, -0.080, 0.963)
```

```
library(bcaboot)
library(boot)
library(moments)
library(nortest)
```

```
cat("Point estimate of the additive opposite of Kelley's relative index of skewness", "\n")
RISr <- (quantile(x, 0.10, type = 8) + quantile(x, 0.90, type = 8)) / (2 * quantile(x, 0.50, type = 8)) - 1
cat("Opposite additive Kelley's relative index of skewness: RISr(x) =", round(RISr, 3), "\n")
```

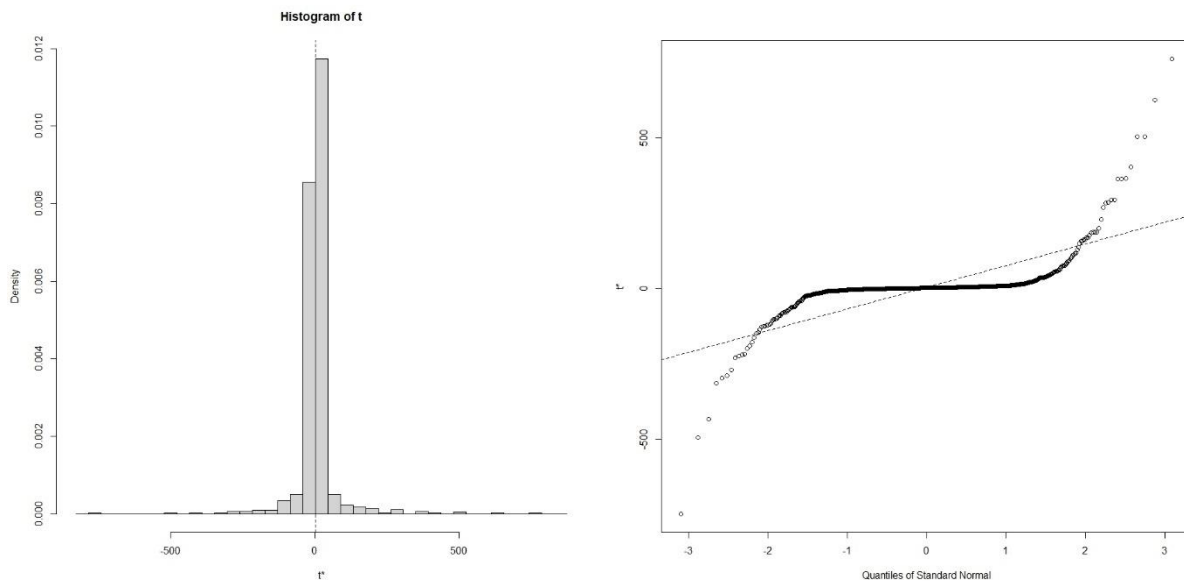
```
cat("Bootstrap confidence interval of the additive opposite of Kelley's relative index of skewness", "\n")
# Acceleration (jackknife)
set.seed(123)
bcjack_results <- bcjack(x, B = 1000, function(x, i) {(quantile(x[i], 0.10, type = 8) + quantile(x[i], 0.90,
type = 8)) / (2 * quantile(x[i], 0.50, type = 8)) - 1}, alpha = 0.025, verbose = TRUE)
accel <- bcjack_results$stats[1, "a"]
cat("Acceleration (jackknife): a =", round(accel, 6), "\n")
set.seed(123)
b <- boot(data=x, function(x, i) {(quantile(x[i], 0.10, type = 8)+quantile(x[i], 0.90, type = 8))/(2*
quantile(x[i], 0.50, type = 8)) - 1}, R=1000)
boot_est <- mean(b$t)
cat("Bootstrap estimation: m(b$t) =", round(boot_est, 3), "\n")
agostino_boot <- agostino.test(b$t, alternative = "two.sided")
ag_boot <- anscombe.test(b$t, alternative = "two.sided")
print(agostino_boot)
print(ag_boot)
cat("D'Agostino-Berlanger-D'Agostino normality test", "\n")
K2 <- (agostino_boot$statistic["z"]^2 + ag_boot$statistic["z"]^2)
p <- pchisq(K2, df = 2, lower.tail = FALSE)
cat("K^2 =", round(K2, 3), "\n")
cat("p_value =", round(p, 3), "\n")
sf.test(b$t)
b
plot(b)
boot.ci(b, conf = 0.95, type = c("bca", "perc", "norm"))
# print(bcjack_results) # Remove the '#' symbol if the 'boot' library cannot calculate the BCa confidence
interval
```

```

p_boot_izq <- mean(b$t < 0)
p_boot_der <- mean(b$t > 0)
p_boot <- min(p_boot_izq, p_boot_der)
cat("One-tailed bootstrap probability value for the null hypothesis of symmetry: p =", round(p_boot, 3),
"\n")
alpha <- 0.05
if (p_boot < alpha) {cat(sprintf("The null hypothesis of normality is rejected at a significance level of %.2f
using Bootstrap probability.", alpha), "\n")
} else {cat(sprintf("The null hypothesis of normality is maintained at a significance level of %.2f using
bootstrap probability.", alpha), "\n")}

```

Kelley's reversed relative index of skewness (*RISr*) reaches a value of 3.296 (Table 4). The bootstrap sampling distribution of the *RISr* statistic (*b\$t*) exhibits positive skewness and strong leptokurtosis, indicating departure from normality (Table 6). The histogram displays a pointed profile with thinned shoulders and elongated tails. The normal quantile-quantile plot flattens in the central area and deviates on both sides, forming a concave curvature to the left and a convex curvature to the right, describing a broad-shouldered shape (Figure 16). Moreover, it shows considerable bias and an acceleration greater than 0.025 (Table 4), making the BCa method the most appropriate for the bootstrap confidence interval, which includes 0 at a 95% level (Table 5). The interval width is notably large ( $w = 270.67$ ), similar to intervals obtained with the other two methods: 273.84 with the percentile method and 281.93 with the normal method (Table 5), due to a very large bootstrap standard error ( $bse = 71.923$ ) (Table 4), rendering it unreliable. Consistent with the bootstrap confidence interval, the null hypothesis of normality is upheld at a 5% significance level using bootstrap probability (Table 5), although this statistic should ultimately be disregarded due to its lack of reliability.



**Figure 16.** Histogram and normal quantile-quantile plot of the bootstrap sampling distribution for *RISr* derived from the 60-data sample of variable X. Source: prepared by the author from the script in this sub-section 5.3, executed with the R program.

## 6. Discussion

From the design of this study, source samples of 10,000 data points were generated using inverse transform sampling [37] from three continuous, symmetric distributions with different kurtosis levels (platykurtic, mesokurtic, and leptokurtic). The bootstrap sampling distribution of the absolute index and percentile coefficient of skewness converges to a normal distribution, whereas that of the relative index of skewness deviates due to pronounced leptokurtosis. Therefore, when developing the R script to obtain point and interval estimates for these indices, it includes errors and asymptotic intervals for the absolute index and percentile coefficient. The script allows for checking normality and considering

---

the sample distribution of variable  $X$  to choose between asymptotic or bootstrap alternatives. In cases of normality or near-normality (symmetry, mesokurtosis, and a bell-shaped profile) with a large sample size, the asymptotic confidence interval is highly suitable. It's worth noting that asymptotic and bootstrap standard errors for these two statistics are very similar across the three simulated distributions, all of which are symmetric and have finite moments.

Furthermore, the script allows computation of the bootstrap confidence interval for these two measures of skewness using three methods: normal, percentile, and BCa [25, 69]. To determine the most suitable method, the script checks the symmetry, kurtosis, and normality of the bootstrap sampling distribution, along with acceleration, skewness, and error statistics. The normal method is preferred if the distribution is normal [69]. If the distribution is non-normal but exhibits symmetry, low skewness, and low acceleration ( $< 0.05$  and  $< 0.025$ , respectively), the percentile method is suitable, especially if it provides a narrower interval than the BCa method. For non-normal distributions with significant skewness and kurtosis, along with substantial bias and acceleration, the BCa method is recommended [72]. If the error is very large—greater than a quarter of the variable's range or the sample size—the interval is questionable and should be approached with caution, or better discarded together with point estimate [23, 25-26].

In the example presented, originating from a distribution without finite moments [46-47], the asymptotic approximation cannot be applied. Moreover, due to the excessive bootstrap standard error, the  $RISr$  should be disregarded. Given the non-small bias and acceleration, and the less-than-ideal fit to normality of the bootstrap sampling distribution, the BCa method is the optimal choice for obtaining confidence intervals [72]. This method provides intervals with amplitudes very similar to those from the percentile method, which would otherwise be somewhat inadequate [25]. It can be concluded that both indices are reliable, given their small standard errors, and that the sample stems from a symmetric distribution, as evidenced by bootstrap confidence intervals encompassing 0 and bootstrap probability values exceeding the significance level of 0.05 ( $AISr = 0.052$ , 95% BCa bootstrap CI (-0.742, 0.922), two-tailed bootstrap  $p$ -value = 0.356;  $PCS = 0.029$ , 95% BCa bootstrap CI (-0.531, 0.556), two-tailed bootstrap  $p$ -value = 0.356). This significance level is appropriate for a sample size of 60, allowing for valid inference due to its random nature. Specifically, this conclusion holds true because the sample originates from a symmetric distribution, such as the Cauchy distribution [46].

Limitations of the study include the exclusion of asymmetric distributions with finite moments, such as gamma, loglogistic, or lognormal distributions [16]. Additionally, we did not vary the size of the source sample to assess convergence to normality (e.g., increments from 30 to 900 by 100, and from 1000 to 10000 by 1000). Instead, the study focused on three symmetric distributions with varying kurtosis and very large source samples, chosen to effectively represent the source distribution [17].

The question arises whether the script is applicable to ordinal variables. If the variable has a large number of ordered categories and assumes normality, a common practice in psychology and other social sciences, it could be feasible [75]. These skewness measures rely on quantiles, which are calculated robustly using R's method 8. However, using a rounding method (such as the 1 to 3 rules in R) for quantile calculation warrants further study [14, 20].

## 7. Conclusions

The sampling distribution of Kelley's absolute index of skewness converges to a normal distribution. Its additive complement, divided by the semi-percentile rank which corresponds to the coefficient of percentile skewness, serves as its appropriate standardization. The sampling distribution of this coefficient also converges to a normal distribution. It is inappropriate to standardize the absolute index using the median. The R script developed can be practically and didactically useful for point and interval estimation of both the absolute index and its standardized form.

**Funding:** This research did not receive any external funding.

**Acknowledgments:** The author expresses gratitude the reviewers and editor for their helpful comments.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

- [1] T. L. Kelley, "A new measure of dispersion," *Quar. Pub. Amer. Statist. Assoc.*, vol. 17, no. 134, pp. 743-749, June 1921. <https://doi.org/10.1080/15225445.1921.10503833>
- [2] T. L. Kelley, *Statistical Method*, The Macmillan Company, New York, 1923, pp. 75-77. <https://doi.org/10.1080/15225445.1921.10503833>
- [3] K. Pearson, "Contributions to the mathematical theory of evolution. I. On the dissection of asymmetrical frequency curves". *Phil. Trans. Roy. Soc. London A*, vol. 185, pp. 71-110, January 1894. <https://doi.org/10.1098/rsta.1894.0003>
- [4] K. Pearson, "Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material", *Phil. Trans. Roy. Soc. London A*, vol. 186, pp. 343-414, 1895. <https://doi.org/10.1098/rsta.1895.0010>
- [5] A. L. Bowley, *Elements of Statistics*, P. S. King and Son, London, 1901.
- [6] U. G. Yule, *An introduction to the Theory of Statistics*. Charles Griffin and Company Limited, London, 1912. <https://doi.org/10.1037/13786-000>
- [7] R. A. Fisher, "The moments of the distribution for normal samples of measures of departure from normality". *Proc Roy. Soc. London A*, vol. 130, no. 812, pp. 16-28, December 1930. <https://doi.org/10.1098/rspa.1930.0185>
- [8] D. Stout, "A question of statistical inference: E. G. Boring, T. L. Kelley, and the probable error". *Am. J. Psychol.*, vol. 102, no. 4, pp. 549-562, April 1989. <https://doi.org/10.2307/1423307>
- [9] R. Ihaka, R: past and future history, 2022. <https://cran.r-project.org/doc/html/interface98-paper/paper.html>
- [10] D. R. Bickel, "Robust estimators of the mode and skewness of continuous data," *Comput. Stat. Data Anal.*, vol. 39, no. 2, pp. 153-163, April 2002. [https://doi.org/10.1016/S0167-9473\(01\)00057-3](https://doi.org/10.1016/S0167-9473(01)00057-3)
- [11] G. Altinay, A simple class of measures of skewness. Munich Personal RePEc Archive, Paper No. 72353, pp. 1-13, September 2016. <https://mpira.ub.uni-muenchen.de/72353/>
- [12] A. Singh, L. Gewali, and J. Khatiwada, "New measures of skewness of a probability distribution," *Open J. Stat.*, vol. 9, no. 3, pp. 601-621, October 2019. <http://dx.doi.org/10.4236/ojs.2019.95039>
- [13] A. Eberl, and B. Klar, "Asymptotic distributions and performance of empirical skewness measures," *Comput. Stat. Data Anal.*, vol. 146, article 106939, June 2020. <https://doi.org/10.1016/j.csda.2020.106939>
- [14] R Development Core Team, *Quantile [stats]*. R Documentation. Sample Quantiles, 2024. <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/quantile.html>
- [15] K. Stapor, "Descriptive and inferential statistics," in *Introduction to Probabilistic and Statistical Methods with Examples in R*, vol 176, Intelligent Systems Reference Library, Cham, Switzerland: Springer, 2020, pp. 63-131. [https://doi.org/10.1007/978-3-030-45799-0\\_2](https://doi.org/10.1007/978-3-030-45799-0_2)
- [16] S. C. Gupta, and V. K. Kapoor, *Fundamentals of Mathematical Statistics (12 th ed.)*, New Delhi: Sultan Chand & Sons, 2020.
- [17] L. M. Chihara, and T. C. Hesterberg, *Mathematical Statistics with Resampling and R*. New York: John Wiley & Sons, 2022.
- [18] A. Linden, "CENTILE2: Stata module to enhance centile command and provide additional definitions for computing sample quantiles," *Statistical Software Components S459262*, Boston College Department of Economics, 2023.
- [19] D. I. Sukhoplyuev, and A. N. Nazarov, *Methods of descriptive statistics in telemetry tasks*, in *Proceedings of the 2024 Systems of Signals Generating and Processing in the Field of on-Board Communications*, Moscow, Russian Federation, (article 10496798), New Orleans, LA: Institute of Electrical and Electronics Engineers (IEEE), 2024. <https://doi.org/10.1109/IEEECONF60226.2024.10496798>
- [20] R. J. Hyndman, and Y. Fan, "Sample quantiles in statistical packages," *Am. Stat.*, vol. 50, no. 4, pp. 361-365, November 1996. <https://doi.org/10.2307/2684934>
- [21] J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, Reading, MA, 1977.
- [22] A. Roques, and A. Zhao, "Association rules discovery of deviant events in multivariate time series: an analysis and implementation of the sax-arm algorithm", *Image Processing On Line*, vol. 12, pp. 604-624, December 2022. <https://doi.org/10.5201/ipol.2022.437>
- [23] P. M. Dixon, "The bootstrap and the jackknife: describing the precision of ecological indices," in S. Scheiner, Ed., *Design and Analysis of Ecological Experiments*, London: Chapman and Hall/CRC, 2020, pp. 290-318. <https://doi.org/10.1201/9781003059813>
- [24] R. B. D'Agostino, "Transformation to normality of the null distribution of  $g_1$ ," *Biometrika*, vol. 57, no. 3, pp. 679-681, December 1970. <https://doi.org/10.1093/biomet/57.3.679>
- [25] B. Efron, and B. Narasimhan, "The automatic construction of bootstrap confidence intervals," *J. Comput. Graph. Stat.*, vol. 29, no. 3, pp. 608-619, March 2020. <https://doi.org/10.1080/10618600.2020.1714633>
- [26] G. Rousselet, C. R. Pernet, and R. R. Wilcox, "An introduction to the bootstrap: a versatile method to make inferences by using data-driven simulations," *Meta-Psychology*, 7, article 2019.2058, December 2023 <https://doi.org/10.15626/MP.2019.2058>
- [27] G. Sánchez-Barajas, and A. Gómez-Navarro, *Estadística General Aplicada con Excel [General Statistics Applied with Excel]*. Universidad Autónoma Nacional de México, Ciudad de México, 2017.
- [28] D. V. Hinkley, "On power transformations to symmetry," *Biometrika*, vol. 62, no. 1, pp. 101-111, April 1975. <https://doi.org/10.2307/2334491>



- [29] R. Chattamvelli, and R. Shanmugam, "Skewness," in *Descriptive Statistics for Scientists and Engineers. Applications in R*, Cham, Switzerland: Springer, 2023, pp. 91-110. [https://doi.org/10.1007/978-3-031-32330-0\\_4](https://doi.org/10.1007/978-3-031-32330-0_4)
- [30] S. Luo and D. Villar, The skewness of the price change distribution: A new touchstone for sticky price models. *Journal of Money, Credit and Banking*, vol. 53, no. 1, pp. 41-72, September 2021. <https://doi.org/10.1111/jmcb.12700>
- [31] P. L. Mamidi, N. D. Arigela, K.N.V.R. Lakshmi, and A. Srilakshmi, "Skewness corrected control charts: a new probability model," *Obstetrics and Gynaecology Forum*, vol. 34, no. 3s, pp. 775-779, May 2024. Retrieved from <https://obstetricsandgynaecologyforum.com/index.php/ogf/article/view/352>
- [32] F. Najafi, S. Naderpour, M. Moradinazar, M. Khoramdad, A. Vahedian-Azimi, T. Jamialahmadi, and A. Sahebkar, "Percentiles for anthropometric measures in 11-18 years-old students of 73 developing countries," *Diabetes Metab. Syndr.*, vol. 14, no. 6, pp. 1957-1962, Nov-Dec 2020. <https://doi.org/10.1016/j.dsx.2020.10.002>
- [33] M. Iseringhausen, I. Petrella, and K. Theodoridis, "Aggregate skewness and the business cycle," *Rev. Econ. Stat.*, 1-37, October 2023. [https://doi.org/10.1162/rest\\_a\\_01390](https://doi.org/10.1162/rest_a_01390)
- [34] K. Harmenberg, The labor-market origins of cyclical skewness, Technical Report, University of Copenhagen, February 2021. Available at [https://www.karlharmenberg.com/papers/skewness\\_harmenberg.pdf](https://www.karlharmenberg.com/papers/skewness_harmenberg.pdf)
- [35] R. A. Groeneveld, and G. Meeden, "Measuring skewness and kurtosis," *J. Roy. Stat. Soc. Ser. D Statistician*, vol. 33, no. 4, pp. 391-399, December 1984. <https://doi.org/10.2307/2987742>
- [36] S. Salgado, F. Guvenen, and N. Bloom, "Skewed business cycles," National Bureau of Economic Research, Working Paper 26565, December 2019. <https://doi.org/10.3386/w26565>
- [37] S. Chowdhury, *Monte Carlo Methods Utilizing Mathematica®: Applications in Inverse Transform and Acceptance-Rejection Sampling*, Berlin: Springer Nature, 2023. <https://doi.org/10.1007/978-3-031-23294-7>
- [38] R. Kimberk, "Beta distribution of long memory sequences," arXiv:2404.05736, March 2024. <https://doi.org/10.48550/arXiv.2404.05736>
- [39] F. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1-21, February 1969. <https://doi.org/10.1080/00401706.1969.10490657>
- [40] F. J. Anscombe, and W. J. Glynn, "Distribution of kurtosis statistics b2 for normal samples," *Biometrika*, vol. 70, no. 1, pp. 227-234, April 1983. <https://doi.org/10.1093/BIOMET/70.1.227>
- [41] T. W. Anderson, and D. A. Darling, "Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes," *Ann. Math. Stat.*, vol. 23, no. 2, pp. 193-212, June 1952. <http://dx.doi.org/10.1214/aoms/1177729437>
- [42] S. S. Shapiro, and R. S. Francia, "An approximate analysis of variance test for normality," *J. Am. Stat. Assoc.*, vol. 67, no. 337, pp. 215-216, April 1972. <https://doi.org/10.1080/01621459.1972.10481232>
- [43] J. P. Royston, "A toolkit for testing for non-normality in complete and censored samples," *J. Roy. Stat. Soc., Ser. D-Statistician*, vol. 42, no. 1, pp. 37-43, March 1993. <https://doi.org/10.2307/2348109>
- [44] R. B. D'Agostino, A. Berlangier, and R. B. Jr. D'Agostino, "A suggestion for using powerful and informative tests of normality," *Am. Stat.*, vol. 44, no. 4, pp. 316-321, November 1990. <https://doi.org/10.2307/2684359>
- [45] P. Kvam, B. Vidakovic, and S. J. Kim, *Nonparametric Statistics with Applications to Science and Engineering with R*, 2nd ed., Hoboken, NJ: John Wiley & Sons, 2022. <https://doi.org/10.1002/9781119268178>
- [46] A. Blöchliger, Gauss versus Cauchy: a comparative study on risk, in T. Hüttche, Ed., *Finance in Crises. Contributions to Finance and Accounting*, Cham, Switzerland: Springer, 2023, pp. 177-198. [https://doi.org/10.1007/978-3-031-48071-3\\_12](https://doi.org/10.1007/978-3-031-48071-3_12)
- [47] S. Nadarajah, and T. Hitchen, "Estimation of models for stock returns," *Computational Economics*, March 2024. <https://doi.org/10.1007/s10614-024-10580-x>
- [48] B. Coker, C. Rudin, and G. King, "A theory of statistical inference for ensuring the robustness of scientific results," *Management Science*, vol. 67, no. 10, pp. 6174-6197, October 2021. <https://doi.org/10.1287/mnsc.2020.3818>
- [49] F. Caeiro, and A. Mateus, *Randtests: Testing Randomness in R*, 2024. <https://cran.r-project.org/web/packages/randtests/index.html>
- [50] S. Kwak, "Are only p-values less than 0.05 significant? A p-value greater than 0.05 is also significant!," *J. Lipid. Atheroscler.*, vol. 12, no. 2, pp. 89-95, May 2023. <https://doi.org/10.12997/jla.2023.12.2.89>
- [51] G. Di Leo, and F. Sardanelli, "Statistical significance: p value, 0.05 threshold, and applications to radiomics – reasons for a conservative approach," *Eur. Radiol. Exp.*, vol. 4, article 18, March 2020. <https://doi.org/10.1186/s41747-020-0145-y>
- [52] J. H. Kim, and I. Choi, "Choosing the Level of Significance: A Decision-theoretic Approach," *Abacus*, vol. 57, no. 1, pp. 27-71, March 2021. <https://doi.org/10.1111/abac.12172>
- [53] K. J. Nicholson, M. Sherman, S. N. Divi, D. R. Bowles, and A. R. Vaccaro, "The Role of Family-wise Error Rate in Determining Statistical Significance," *Clin. Spine Surg.*, vol. 35, no. 5, pp. 222-223, June 2022. <https://doi.org/10.1097/BSD.0000000000001287>
- [54] C. Avram, and M. Mărușteri, "Normality assessment, few paradigms and use cases," *Rev. Romana Med. Lab.*, vol. 30, no. 3, pp. 251-260, July 2022. <https://doi.org/10.2478/rrlm-2022-0030>
- [55] L. Komsta, and F. Novomestky, *Moments: moments, cumulants, skewness, kurtosis and related tests*, 2022. <https://cran.r-project.org/web/packages/moments/index.html>
- [56] J. Gross, and U. Ligges, *Package 'nortest'*, 2022. <https://cran.r-project.org/web/packages/nortest/nortest.pdf>

- [57] H. W. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *J. Am. Stat. Assoc.*, vol. 62, no. 318, pp. 399-402, August 1967. <https://doi.org/10.2307/2283970>
- [58] S. Demir, "Comparison of normality tests in terms of sample sizes under different skewness and Kurtosis coefficients," *Int. J. Assess. Tool. Educ.*, vol. 9, no. 2, pp. 397-409, May 2022. <https://doi.org/10.21449/ijate.1101295>
- [59] N. Khatun, "Applications of Normality Test in Statistical Analysis," *Open J. Stat.*, vol. 11, no. 1, pp. 113-122, February 2021. <https://doi.org/10.4236/ojs.2021.111006>
- [60] I. Mala, V. Sladek, and D. Bilkova, "Power comparisons of normality tests based on l-moments and classical tests," *Math. Stat.*, vol. 9, no. 6, pp. 994-1003, November 2021. <https://doi.org/10.13189/ms.2021.090615>
- [61] D. K. Wijekularathna, A. B. Manage, and S. M. Scariano, "Power analysis of several normality tests: A Monte Carlo simulation study," *Commun. Stat-Simul C Journal*, vol. 51, no. 3, pp. 757-773, September 2020. <https://doi.org/10.1080/03610918.2019.1658780>
- [62] R. Sahann, T. Müller, and J. Schmidt, "Histogram binning revisited with a focus on human perception," in *Proceeding of the 2021 IEEE Visualization Conference (VIS)*, New Orleans, LA: Institute of Electrical and Electronics Engineers (IEEE), 2021, pp. 66-70. <https://doi.org/10.1109/VIS49827.2021.9623301>
- [63] V. A. Epanechnikov, "Nonparametric estimation of a multidimensional probability density," *Theory of Probability and its Applications*, vol. 14, no. 1, pp. 156-161, January 1969. <https://doi.org/10.1137/1114019>
- [64] N. Fadillah, P. A. Dariah, A. Anggraeni, N. Cahyani, and L. Handayani, "Comparison of Gaussian and Epanechnikov Kernels," *Tadulako Social Science and Humaniora Journal*, vol. 3, no. 1, pp. 13-22, September 2022. <https://doi.org/10.22487/sochum.v3i1.15745>
- [65] S. J. Sheather, and M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation," *J. Roy. Stat. Soc., Ser. B Stat. Methodol.*, vol. 53, no. 3, pp. 683-690, January 1991. <https://doi.org/10.1111/j.2517-6161.1991.tb01857.x>
- [66] W. N. Venables, and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed., Cham, Switzerland: Springer, 2002. <https://doi.org/10.1007/978-0-387-21706-2>
- [67] E. C. Ogwu, and H. I. A. Ojarikre, "Comparative study of the rule of thumb, unbiased cross validation and the Sheather Jones-direct plug-in approaches of kernel density estimation using real life data," *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 11, no. 3, pp. 1800-1809, March 2023.
- [68] R Development Core Team. *The R Manuals*, 2024. <https://cran.r-project.org/manuals.html>
- [69] A. Canty, B. Ripley, and A. R. Brazzale, *Package 'boot'*, 2024. <https://cran.r-project.org/web/packages/boot/boot.pdf>
- [70] G. Cavaliere, and I. Georgiev, "Inference under random limit bootstrap measures," *Econometrica*, vol. 88, no. 6, pp. 2547-2574, November 2020. <https://doi.org/10.3982/ECTA16557>
- [71] D. Lakens, "Sample size justification," *Collabra: Psychology*, vol. 8, no. 1, article 33267, January 2022. <https://doi.org/10.1525/collabra.33267>
- [72] B. Efron, and B. Narasimhan, *Package 'bcaboot'*. Bias corrected bootstrap confidence intervals, 2022. <https://cran.r-project.org/web/packages/bcaboot/bcaboot.pdf>
- [73] A. Pekgör, "A novel goodness-of-fit test for Cauchy distribution," *J. Math.*, vol. 2023, no. 1, article 9200213, March 2023. <https://doi.org/10.1155/2023/9200213>
- [74] B. Lawson, L. Leemis, and V. Kudlay, *Set.seed: Seeding random variate generators*, 2023. <https://www.rdocumentation.org/packages/simEd/versions/2.0.1/topics/set.seed>
- [75] T. L. Kelley, *Fundamentals of Statistics*, Cambridge, MA: Harvard University Press, 1947.

Research Article

Received: date:08.08.2024  
Accepted: date:16.12.2024  
Published: date:31.12.2024

# A Robust Initial Basic Subset Selection Method for Outlier Detection Algorithms in Linear Regression

Mehmet Hakan Satman\*

<sup>1</sup>Istanbul University, Department of Econometrics, İstanbul Türkiye; mhsatman@istanbul.edu.tr  
Orcid:0000-0002-9402-1982<sup>1</sup>

\*Correspondence: mhsatman@istanbul.edu.tr

**Abstract:** The main motivation of this study is to develop an efficient algorithm for diagnosing and detecting outliers in linear regression up to a reasonable level of contamination. The algorithm initially obtains a robust version of the hat matrix at the linear algebra level. The basic subset obtained in the first stage is improved through concentration steps as defined in the fast-LTS (Least Trimmed Squares) regression algorithm. The method can be plugged into another algorithm as a basic subset selection state. The algorithm is effective against outliers in both X and Y directions by a rate of 25%. The complexity of the algorithm increases linearly with the number of observations and parameters. The algorithm is quite fast as it does not require iterative calculations. The success of the algorithm against a specific contamination level is demonstrated through simulations.

**Keywords:** Linear regression, Outlier Detection, Robust Regression

## 1. Introduction

Suppose the linear regression model is

$$y = X\beta + \varepsilon$$

where  $y$  is the  $n$ -vector of the response variable,  $X$  is the  $n \times p$  design matrix,  $\varepsilon$  is the  $n$ -vector of the stochastic error term with zero mean and constant variance,  $n$  is the number of observations, and  $p$  is the number of regression parameters. The ordinary least squares estimate of  $\beta$ , noted as  $\hat{\beta}$ ,

$$\hat{\beta} = (X'X)^{-1}X'y$$

is the most efficient estimator among the linear and unbiased estimators, that is,  $\hat{\beta}$  minimizes

$$MSE(\hat{\beta}) = Var(\hat{\beta}) + Bias^2(\hat{\beta}). \quad (1)$$

When data contains unusual observations Eq. 1 may increase drastically depending on the direction and level of the contamination. The LAD (Least Absolute Deviations) estimator is resistant up to 50% of contamination on the vertical outliers since it minimizes the conditional median of the response variable, that is, the LAD estimator  $\hat{\beta}_{LAD}$  minimizes the objective function

$$\min \sum |y - X\hat{\beta}_{LAD}|$$

which in turn can be written as a linear programming problem

$$\min \sum_{i=1}^n e_i^- + e_i^+$$

Subject to:

$$X_{(i)}\beta_{LAD} + e_i^- - e_i^+ = y_i \quad (2)$$

$$e_i^-, e_i^+ \geq 0, i = 1, 2, \dots, n$$

$$\beta_{LAD} \in \mathbb{R}^p$$

where  $X_{(i)}$  is the  $i$ th row of  $X$ ,  $e_i^- \geq 0$  if  $i$ -th residual is located under the regression object, otherwise  $e_i^- = 0$ ,  $e_i^+ \geq 0$  if  $i$ -th residual is located above the regression object, otherwise  $e_i^+ = 0$ . If  $e_i^- = e_i^+ = 0$  then the regression equation exactly fits the  $i$ -th observation. Eq. 2 has an exact solution to the robust method LAD, however, LAD is not resistant to  $X$ -space outliers, namely, bad leverage points [1].

LTS (Least Trimmed Squares) estimator  $\hat{\beta}_{LTS}$  is resistant to both vertical and  $X$ -space outliers up to 50% contamination and it is defined as

$$\min \sum_{i=1}^h |y_i - X_{(i)}\hat{\beta}_{LTS}|^2$$

where  $|y_i - X_{(i)}\hat{\beta}_{LTS}|^2$  is the  $i$ -th ordered squared residual,  $h$  is at least half of the number of observations  $n$ . The Fast-LTS algorithm requires selecting many basic subsets and then iterating and enlarging the basic subset until some convergence criteria is met [2]. Since the algorithm is based on comprehensive iterations of calculations, more computation time is required to obtain parameter estimates. When the data is large, metaheuristics [18] can be used to minimize the objective function, however, the required computation time is still a problem just because the optimization problem has not a closed form as in many robust regression estimators.

In this paper, a robust version of vector and matrix algebra is defined to obtain a more robust version of the well-known hat matrix. This hat matrix is then used to construct a small basic subset which is supposed to be free of outliers. This basic subset is then fed into a robust regression estimator to obtain robust regression parameter estimates. The proposed method is based on a single-pass algorithm and fast. It's shown that the proposed method is resistant up to 25% level of contamination.

In Section 1 the problem is introduced. In Section 2, the basics of the robust hat matrix construction is introduced. In Section 3, the use of the robust hat matrix based initial basic subset is introduced in a robust regression estimator, LTS. In Section 4, a suite of simulation studies is performed to assess the performance of the algorithm for varying dimensions and levels of contamination. Finally, in Section 5, the results are discussed and we conclude.

## 2. Preliminaries

Suppose that the  $p$ -vectors  $x$  and  $y$  are defined as

$$x = [x_1, x_2, \dots, x_p]^T$$

and

$$y = [y_1, y_2, \dots, y_p]^T$$

The classical dot product sums up the products of corresponding elements of vectors  $x$  and  $y$ , that is,

$$x'y = \sum_{i=1}^p x_i y_i$$

The dot product  $x'y$  can also be written as

$$x'y = f(x_1 y_1, x_2 y_2, \dots, x_p y_p) \times p$$

where

$$f(x, y) = \sum_{i=1}^p (x_i y_i) / p.$$

In other terms, the sum is  $p$  times the sample mean. Now suppose that  $f^*$  is a robust location estimator. Then the robust dot product of  $x$  and  $y$  can be written as

$$x' \otimes y = f^*(x_1y_1, x_2y_2, \dots, x_py_p) \times p. \quad (3)$$

Since  $f^*(.)$  is a robust location estimator then  $f^*(.) \times p$  is a robust estimate of sum. Using  $x' \otimes y$  instead of  $x'y$  results in a more robust version of the vector product.

### 2.1. Robust Matrix Multiplication

Suppose that  $X$  and  $y$  are  $m \times n$  and  $n \times p$  matrices, respectively. Then the  $i$ -th row and the  $j$ -th column of the robust matrix multiplication of  $C = A \otimes B$  is defined as

$$c_{ij} = X_{(i)} \otimes Y_{(j)} \quad (4)$$

where  $X_{(i)}$  is the  $i$ -th row vector of  $X$  and  $y_{(j)}$  is the  $j$ -th column vector of  $y$ ,  $\otimes$  is the multiplication operator as defined in Equation 3, and  $f^*$  is a univariate robust location estimator.

### 2.2. Robust Hat Matrix

The original hat matrix of linear regression is based on the design matrix and it is defined as

$$H = X(X'X)^{-1}X'$$

and the diagonal elements of  $H$  are investigated for bad leverage points [3]. When the design matrix contains unusual observations, because of the summation operator, the corresponding diagonal elements of  $H$  tend to have smaller values and cause the masking effect. Replacing the matrix multiplication operator with its robust counterpart yields

$$H_R = X(X' \otimes X)^{-1}X'$$

where  $\otimes$  is the robust multiplication operator. Since the part  $(X' \otimes X)$  is supposed to be outlier free for some contamination level, it is expected that the unusual observations will have relatively larger values on the corresponding diagonal elements of  $H_R$ .

### 2.3. Trimean as a Robust Location Estimator

Trimean estimator is a robust estimator of the location parameter which is defined as

$$Trimean(x) = \frac{Q_{25}(x) + 2Q_{50}(x) + Q_{75}(x)}{4}$$

where  $Q_{25}(x)$  is the first quartile of  $x$ ,  $Q_{50}(x)$  is the sample median of  $x$ , and  $Q_{75}(x)$  is the third quartile of  $x$ , respectively [4]. The Trimean is less robust than the sample median, however, it's more efficient. The break-down point of Trimean is 25%, that is, it stays resistant to outliers when the fraction of contamination is up to 25% of data. Moreover, Trimean is a suitable estimator due to its efficiency resembling the arithmetic mean and its robustness akin to the median.

In Table 1, the use of the robust version of the Hat matrix ( $H_R$ ) is represented using a set of simulated data. The robust location estimator is selected as  $f^* = Trimean$ . The data consist of two variables, namely  $x_1$  and  $x_2$ , with the sample size of  $n = 10$ . The last two observations have unusual values. The 9th and 10th observations are manipulated by distorting the values corresponding to the first variable.  $Diag(H)$  and  $Diag(H_R)$  represent the diagonal elements of the hat matrix and robust hat matrix, respectively.

**Table 1.** Illustrative example of a set of simulated data.

Obs.	Cons.	$x_1$	$x_2$	$Diag(H)$	$Diag(H_R)$
1	1.0	0.493	0.520	0.319	0.649
2	1.0	0.332	0.534	0.307	0.648
3	1.0	0.895	0.797	0.121	0.236
4	1.0	0.910	0.545	0.258	1.023
5	1.0	0.779	0.967	0.345	0.871

6	1.0	0.201	0.765	0.164	0.867
7	1.0	0.622	0.760	0.122	0.130
8	1.0	0.311	0.960	<u>0.374</u>	1.769
9	1.0	2.900	0.756	0.283	<u>14.42</u>
10	1.0	4.300	0.764	<u>0.702</u>	<u>38.24</u>

In Table 1, it is shown that the two largest diagonal elements incorrectly belong to the observations 8 and 10, whereas, observations 9 and 10 have relatively large values on the diagonal elements of the Trimean based robust hat matrix. As it is expected, larger and smaller values are omitted in the summation part of the vector multiplication stage.

### 3. The Main Algorithm

Constructing a basic subset and then iterating and enlarging this subset until some convergence criteria is met is not a new concept in detecting outliers in linear regression [5, 6, 7]. [5] constructs a subset of  $p+1$  observations and then enlarges these subsets using the DFFITS [8] statistics. Similarly [6] and [7] enlarge the basic subset using multivariate statistics calculated on the design matrix in order to detect bad-leverage points. [9] and [10] construct and iterate an initial estimate and then update the weights of an iteratively weighted least squares estimate. [11] introduces a two-stage method for detecting outliers in linear regression. In the first stage of the method, a subset of outlier-free observations is created using a robust covariance matrix estimation inspired by the Comediance [12] statistic. Several iterations are performed for constructing initial subsets as the essential steps of the LMS [13], LTS [2], and LTA [14] estimators. However, the initially selected subsets may contain outliers, and in the later stages of the outlier detection algorithm (or the robust estimator), this could lead to clean observations being reported as outliers and outliers being reported as clean observations. The developed algorithm has been designed to find a better initial subset in a shorter time and to reduce the computational overhead in subsequent iterations of the detection algorithms

The steps of the devised initial basic subset construction algorithm is given as follows:

#### Main Algorithm:

**Step 0.** Construct the design matrix  $X$  with dimensions  $n$  and  $p$ .

**Step 1.** Compute the robust hat matrix using the formula  $X(X' \otimes X)^{-1}X'$  where the operator  $\otimes$  is defined as in Eq. 3 and Eq. 4, respectively.

**Step 2.** Let  $|h_{ii}|$  is the ordered diagonal elements of the robust hat matrix. Record the indices of first  $p + 1$  elements of  $|h_{ii}|$ .

**Step 3.** Perform C-Steps of the LTS algorithm for the initial basic subset obtained in the previous step. Report  $\hat{\beta}_{LTS}$  and the other necessary output of the final regression.

The main algorithm is basically based on constructing an initial basic subset of length  $p + 1$  which is supposed to be free of outliers. The final step is based on the concentration steps (C-Steps) of the LTS algorithm but only a single initial basic subset is fed for obtaining a better larger subset with size  $h$ . The algorithm differs from the LTS algorithm as it constructs a single basic subset instead of selecting many random basic subsets.

Iterating C-Steps is the inner part of the outlier detection algorithm. C-Steps, as defined in [2], start with estimating the linear regression coefficients using a  $p + 1$  subset of observations. The smallest  $h$  ordered absolute residuals obtained in a previous step are then used to estimate the regression parameters.  $h$  is a number that covers at least half of the observations. Subsequent iterations are performed until the difference of the last two objective function values are less than a small value. The Fast-LTS algorithm iterates these steps for many randomly drawn  $p + 1$  subsets. The devised algorithm is fast as it uses a single initial basic subset which is obtained using the robust hat matrix.

Table 2 represents regression coefficients estimated using the Hawkins, Bradu, Kass's (HBK) Artificial Data [15] by several estimators. HBK data consists of three independent variables and has 75 observations. The model is supposed to be  $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \varepsilon$ . LMS, LTS, BCH [7], and the devised algorithm  $R_{HAT}$  have similar results as  $\widehat{\beta}_0 < 0$ ,  $\widehat{\beta}_1 > 0$ ,  $\widehat{\beta}_2 > 0$ , and  $\widehat{\beta}_3 < 0$ . OLS and LAD end up with a different set of signs of estimations. This set of results is a clean indicator of the dataset

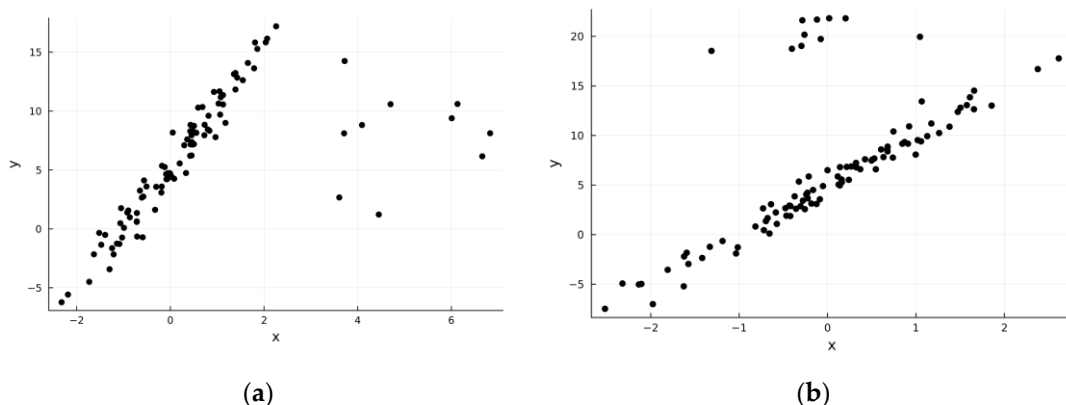
at hand has possible X-space outliers. The last column of Table 2 represents the relative time consumed by the estimators. The fastest estimator is OLS and its relative time is set to 1x. It can be said that the exact LAD estimation is 12.21 times slower than the OLS. The devised estimator is the fastest one among others as it is 1.52 times slower than the OLS estimator. Summarizing the results, it can be said that the devised algorithm is fast enough and it is capable of handling contaminated data with similar results obtained with the high breakdown estimators.

**Table 2.** Estimated regression coefficients using the Hawkins, Bradu, Kass's Artificial Data by several estimators.

Estimator	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	Time
OLS	-0.387550	0.239185	-0.334548	0.383341	1
LAD	-0.881474	0.091312	0.154760	0.214647	12.21
LMS	-0.677950	0.266808	0.112131	-0.144488	91.47
LTS	-0.637726	0.253676	0.108266	-0.139344	62.13
BCH	-0.535321	0.227364	0.049063	-0.097577	10.16
$R_{HAT}$	-0.578735	0.261031	0.045958	-0.103124	1.52

#### 4. The Simulation Study

A set of simulation studies is performed for measuring the performance of the devised algorithm. In each single simulation, a random regression data is created using a specific data generating process. The regression model has always an intercept term.  $x_i$  variables are drawn from a standard Normal distribution for  $i = 1, 2, \dots, p$ . The error term also follows a Normal distribution with zero mean and unit variance. Regression parameters are set to 5 for all  $\beta_i$ . Generated data is contaminated either in X or y directions. Contamination with a given level is always guaranteed by adding random noise to largest values on the corresponding set of variables. Simulations are repeated for null contamination level, that is, results contain the case of absence of outliers. LinRegOutliers [19] package of the Julia programming language [20] is used for the calculations. Figure 1 represents random regression data with contamination on either X-space (a) or vertical (b) direction. Note that the represented data has only a single independent variable, however, simulations are always performed for  $p > 2$ . In Figure 1 (a) 10 observations are located in a manner that is considered inappropriate as they don't follow the same structure with the remaining data points by considering the X-space. In Figure 1 (b), 10 observations are vertically incompatible with the regression structure and they have values larger than at least the maximum of the clean data points in vertical direction.



**Figure 1.** Simulation data with  $n = 100$ ,  $p = 2$ , and contamination level of 10% in either X or y directions, respectively. (a) represents the case of the contamination on X, (b) represents the case of the contamination on the response variable, y.

Figures 2 - 7 represent the simulation results. The horizontal axis shows the level of contamination, that is, when the contamination is 0%, the data is free of outliers, otherwise, data is contaminated by the ratio of the given level. In the vertical axis, average of the parameter estimates is given, that is, an unbiased

and robust estimator should have values around a constant line of  $y = 5$ . Moreover, if the fluctuation around the constant is low, then it can be concluded that the estimator has a smaller variance. The ideal graphics follow the  $\hat{\beta} = 5$  line with minimum variance.

Figures 2 and Figure 3 represent the simulation results for  $n = 50$  and  $p = 3$  when the direction of contamination is  $X$  and  $y$ , respectively. It is shown that the parameter estimates are very close to 5 which is the true regression parameters given in the data generating process. When the level of contamination increases, parameter estimates tend to have larger variances and a negative bias emerges in case of presence of  $X$  outliers. Bias of the estimates remain near to zero in the presence of  $y$ -outliers. In both Figure 2 and Figure 3, it is shown that the parameter estimates drastically disrupt when the level of contamination is larger than 25%.

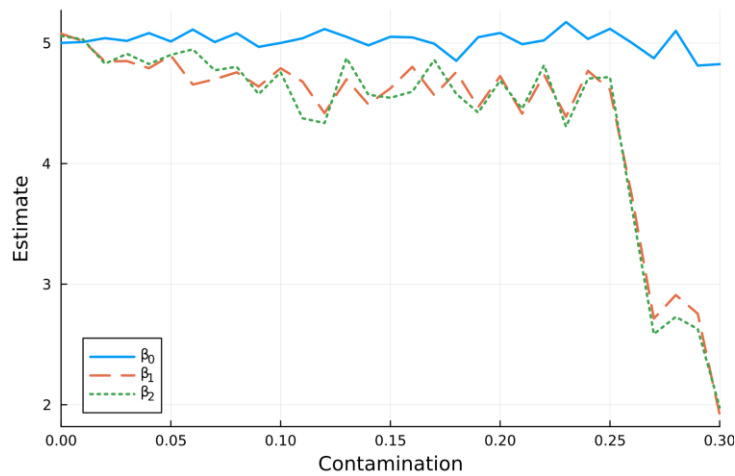


Figure 2. Parameter estimations for  $n = 50$ ,  $p = 3$ , and contamination in  $X$ -direction

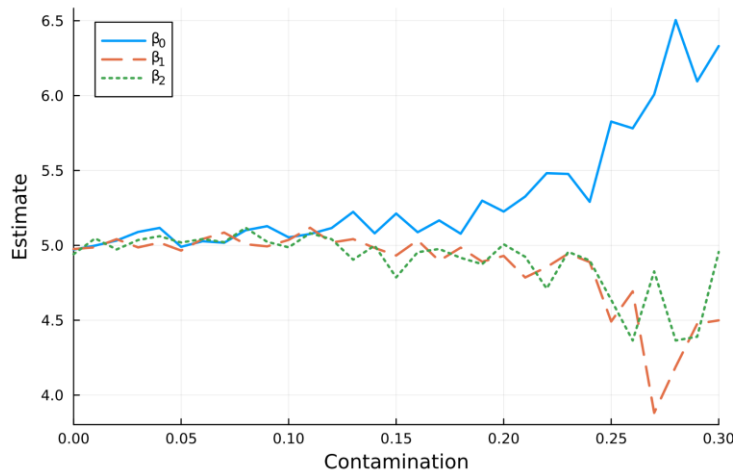


Figure 3. Parameter estimations for  $n = 50$ ,  $p = 3$ , and contamination in  $y$ -direction

Figures 4 and Figure 5 represent the simulation results for  $n = 100$  and  $p = 5$ . The results are similar when they are compared to the previous ones for  $n = 50$  and  $p = 3$ , however, in presence of  $X$ -outliers, the bias and variance seem to be more stable when the level of contamination is  $c \leq 0.25$ . The intercept parameter estimate remains unbiased with a small increase in the variance. When the data is contaminated in the vertical direction, the results obtained by the configurations  $n = 50$  and  $n = 100$  are similar.



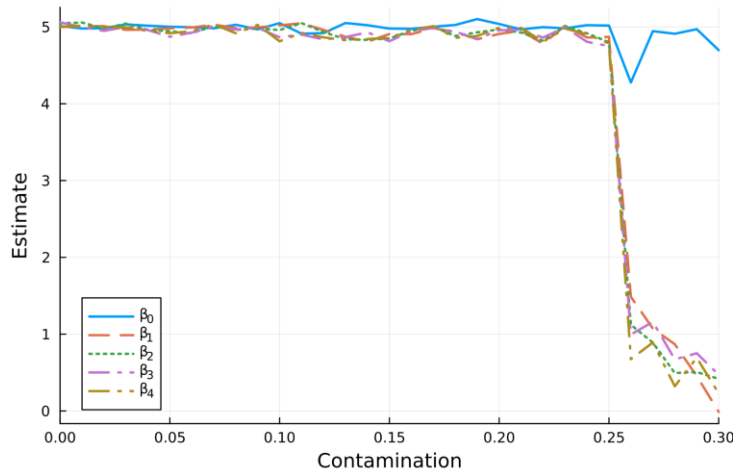


Figure 4. Parameter estimations for  $n = 100$ ,  $p = 5$ , and contamination in X-direction

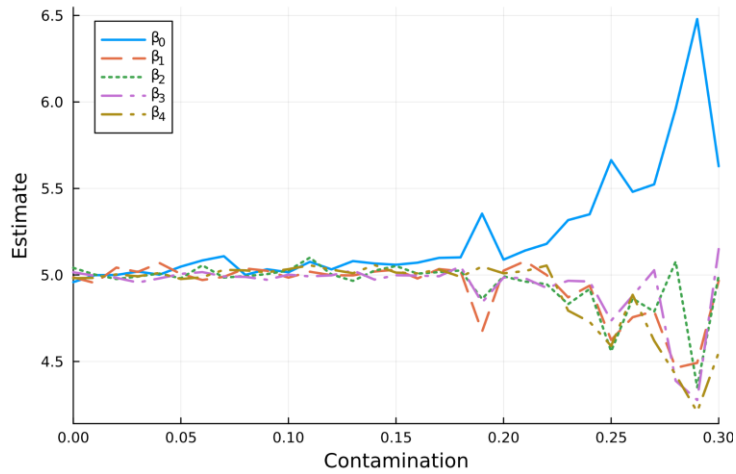


Figure 5. Parameter estimations for  $n = 100$ ,  $p = 5$ , and contamination in y-direction

Figure 6 and Figure 7 represent the simulation results for  $n = 500$  and  $p = 10$ . It is shown that the parameter estimates still remain stable for the contamination level is  $c \leq 0.25$ . As expected, when the contamination level is larger than  $c = 0.25$ , the estimates drastically disrupt. Differently, in the presence of y-outliers as it is shown in Figure 7, all of the parameter estimates take a value between a narrower range, e.g.  $4.96 \leq \hat{\beta}_i \leq 5.06$ , and the range still remains stable when the contamination level increases.

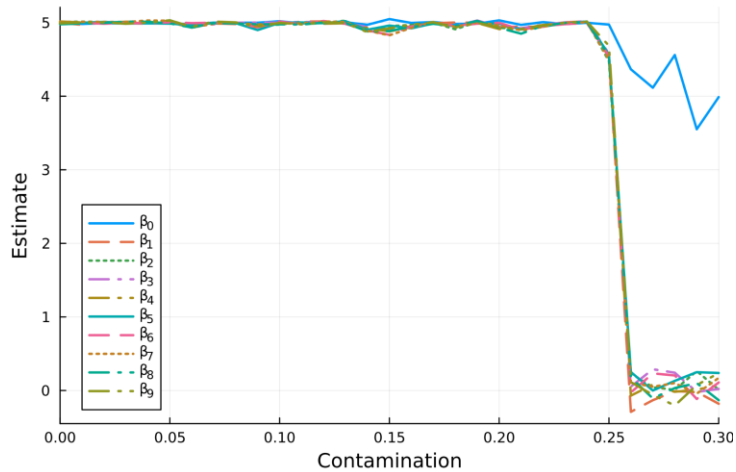
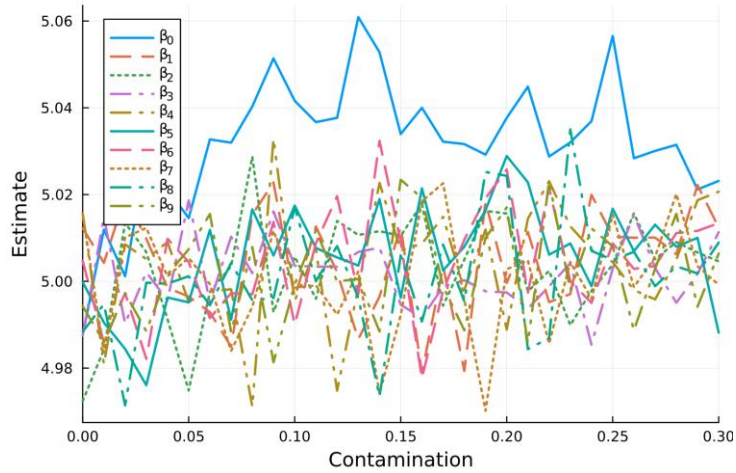
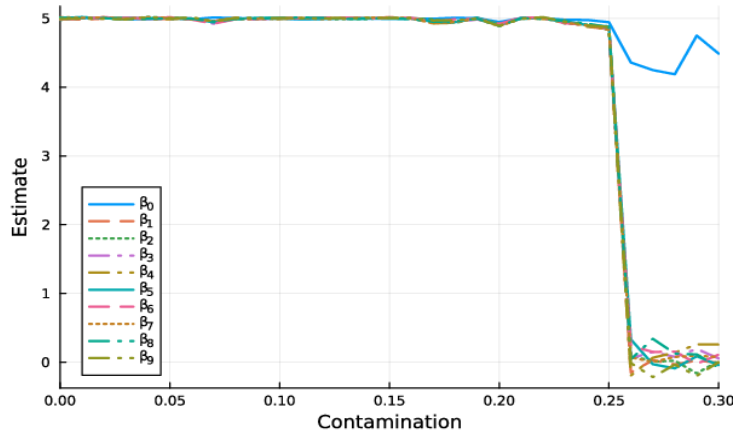


Figure 6. Parameter estimations for  $n = 500$ ,  $p = 10$ , and contamination in X-direction

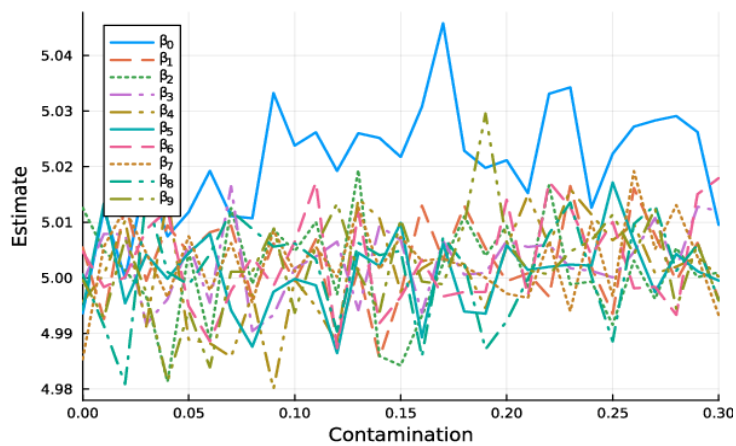


**Figure 7.** Parameter estimations for  $n = 500$ ,  $p = 10$ , and contamination in y-direction

To investigate the performance of the algorithm, we generate simulated datasets with  $n = 1000$  and  $p = 50$ . This configuration can be considered large for many research applications. Figure 8 and Figure 9 represent the simulation results for  $n = 1000$  and  $p = 10$ . Simulation results have shown that even as the number of observations and parameters increase, distortions in the parameters begin at the point where the contamination level reaches 25%. Similar to previous simulation designs, in the presence of y-outliers, as shown in Figure 9, all parameter estimates fall within a narrower range, e.g., 4.96 to 5.06. This range remains stable even as the contamination level increases.



**Figure 8.** Parameter estimations for  $n = 1000$ ,  $p = 10$ , and contamination in X-direction



**Figure 9.** Parameter estimations for  $n = 1000$ ,  $p = 10$ , and contamination in y-direction

## 5. Discussions and Conclusions

The robust statistics literature has a vast set of outlier detection and robust regression estimators including LMS, LTS, LTA, LAD, M-Estimators [16], S-Estimators [17], etc. Some of the estimators serve as high-breakdown estimators for datasets with a contamination level up to 50%. However, these methods and estimators are computationally heavy and they require more time to obtain a result.

The robust hat matrix based initial basic subset selection method is a single-pass and linear algebraic method that provides a clean input which is free of bad leverage points ( $X$ -space outliers) when the data is contaminated up to 25%. That property makes the devised method superior in datasets with a low contamination level. Note that the initial basic subset construction should be seen as an inner step of a detection algorithm and theoretical properties of the robust hat matrix are not investigated in this paper. Hence, it is not a standalone estimator but a subset selection tool only.

A suite of simulation study is performed to measure the performance of the devised method. Simulation data is generated for different levels of number of observations, parameters and several levels of contamination in both directions. The simulation results show that the method is applicable in all of the cases. The theoretical and known drawback is the maximum level of contamination. The simulation results also show that the bias and variance, hence MSE, of estimators reduce (towards zero) when the number of observations and parameters increases.

The robust hat matrix estimation is based on the Trimean estimator. The maximum breakdown is determined by this estimator. Using the sample median instead of the Trimean causes ending up with a singular matrix which prevents obtaining a matrix inverse. Selection of another robust location measure can be investigated in future works.

This paper is primarily motivated by the development of a clean starting set of points for later use in the Fast-LTS algorithm. However, this stage can be easily integrated into any other robust regression method and/or outlier detection methods. Future studies could investigate how the initial subset selection stage can be applied to other outlier detection methods. This would provide insights into whether the proposed approach can enhance the performance and accuracy of alternative algorithms. Additionally, its integration into robust regression methods could reveal potential improvements in handling datasets with challenging characteristics.

**Funding:** This research did not receive any external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Acknowledgments:** The author thanks the referees for their valuable suggestions which made this paper much improved.

## References

- [1] X. Gao and Y. Feng, "Penalized weighted least absolute deviation regression," *Statistics and its interface*, vol. 11, no. 1, pp. 79–89, 2018.
- [2] P. J. Rousseeuw and K. Van Driessen, "Computing LTS regression for large data sets," *Data mining and knowledge discovery*, vol. 12, pp. 29–45, 2006.
- [3] D. C. Hoaglin and R. E. Welsch, "The hat matrix in regression and anova," *The American Statistician*, vol. 32, no. 1, pp. 17–22, 1978.
- [4] J. W. Tukey et al., *Exploratory data analysis*. Reading, MA, 1977, vol. 2.
- [5] A. S. Hadi and J. S. Simonoff, "Procedures for the identification of multiple outliers in linear models," *Journal of the American statistical association*, vol. 88, no. 424, pp. 1264–1272, 1993.
- [6] N. Billor, A. S. Hadi, and P. F. Velleman, "Bacon: blocked adaptive computationally efficient outlier nominators," *Computational statistics & data analysis*, vol. 34, no. 3, pp. 279–298, 2000.
- [7] N. Billor, S. Chatterjee, and A. S. Hadi, "A re-weighted least squares method for robust regression estimation," *American journal of mathematical and management sciences*, vol. 26, no. 3-4, pp. 229–252, 2006.
- [8] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, 2005.
- [9] S. Barratt, G. Angeris, and S. Boyd, "Minimizing a sum of clipped convex functions," *Optimization Letters*, vol. 14, pp. 2443–2459, 2020.
- [10] S. Chatterjee and M. Mächler, "Robust regression: A weighted least squares approach," *Communications in Statistics-Theory and Methods*, vol. 26, no. 6, pp. 1381–1394, 1997.
- [11] M. H. Satman, "A new algorithm for detecting outliers in linear regression," *International Journal of statistics and Probability*, vol. 2, no. 3, p. 101, 2013.
- [12] L. Huo, T.-H. Kim, and Y. Kim, "Robust estimation of covariance and its application to portfolio optimization," *Finance Research Letters*, vol. 9, no. 3, pp. 121–134, 2012.

- 
- [13] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American statistical association*, vol. 79, no. 388, pp. 871–880, 1984.
- [14] D. M. Hawkins and D. Olive, "Applications and algorithms for least trimmed sum of absolute deviations regression," *Computational Statistics & Data Analysis*, vol. 32, no. 2, pp. 119–134, 1999.
- [15] D. M. Hawkins, D. Bradu, and G. V. Kass, "Location of several outliers in multiple-regression data using elemental sets," *Technometrics*, vol. 26, no. 3, pp. 197–208, 1984.
- [16] D. De Menezes, D. M. Prata, A. R. Secchi, and J. C. Pinto, "A review on robust m-estimators for regression analysis," *Computers & Chemical Engineering*, vol. 147, p. 107254, 2021.
- [17] P. Rousseeuw and V. Yohai, "Robust regression by means of s-estimators," in *Robust and Non-linear Time Series Analysis: Proceedings of a Workshop Organized by the Sonderforschungsbereich 123 "Stochastische Mathematische Modelle"*, Heidelberg 1983. Springer, pp. 256–272, 1984.
- [18] M.H. Satman. "A genetic algorithm based modification on the LTS algorithm for large data sets." *Communications in Statistics-Simulation and Computation* 41.5, pp. 644-652, 2012.
- [19] M.H. Satman, S. Adiga, G. Angeris, and E. Akadal. "LinRegOutliers: A Julia package for detecting outliers in linear regression." *Journal of Open Source Software* 6, no. 57: 2892, 2021.
- [20] J. Bezanson, S. Karpinski, V.B. Shah, V. B., and A. Edelman. *Julia: A fast dynamic language for technical computing*. arXiv preprint arXiv:1209.5145, 2012

# Residual Modelling as a New Approach for Variable Selection

Asli Nurefşan Kocak<sup>1</sup>, Muhammet Furkan Daşdelen<sup>2</sup>, Mehmet Kocak<sup>3\*</sup>

<sup>1</sup>Istanbul Medipol University, International School of Medicine, Istanbul, Turkey asli.kocak@std.medipol.edu.tr

<sup>2</sup>Istanbul Medipol University, International School of Medicine, Istanbul, Turkey muhammed.dasdelen@std.medipol.edu.tr

<sup>3</sup>Istanbul Medipol University, International School of Medicine, Istanbul, Turkey mehmetkocak@medipol.edu.tr

Orcid:0009-0000-5367-7443<sup>1</sup> Orcid:0000-0003-2251<sup>2</sup>-2093 Orcid:0000-0002-3386-1734<sup>3</sup>

\*Correspondence: mehmetkocak@medipol.edu.tr

**Abstract:** Variable selection in statistical model building still has challenges to overcome as the depth and breadth of the research data is expanding. To help reduce this challenge, we introduce a new approach in variable selection, called residual modeling, which can be applicable regardless of the number of predictors. We compare the statistical power and type-1 error retention of the forward, backward, and stepwise variable selection approaches with the proposed modeling strategy controlling for known predictors. In Residual Modeling, each predictor enters the model as a single predictor, whose resulting residuals become the dependent variable for the next predictor, and so on. We compare these models under different scenarios with varying sample sizes and various combinations of significant and insignificant predictors. When there exist known predictors from the literature, in identifying new significant predictors controlling for these known predictors, Residual Modelling shows higher statistical power especially as the number of predictors increases compared to the other variable selection methods used. It also has reduced bias in parameter estimation and reduced standard errors. The Type-1 error was retained at its nominal level for Residual Modelling while forward, backward, and stepwise variable selection approaches had slightly reduced Type-1 Error rates. When dealing with multiple predictors in the presence of known significant predictors, Residual Modelling offers a practical solution without causing loss of statistical power or increased Type-1 Error Rate.

**Keywords:** variable selection, dimension reduction, forward selection, backward selection, stepwise selection, residual modelling

## 1. Introduction

In statistical modelling, we often deal with multiple predictors (i.e., independent variables) for a given response variable (i.e., dependent variable). The immediate, and typically the easiest, approach would be to fit a model with all the predictors included in the model and assess the variables that show significant association with the response variable. Then, the model is 'cleaned' by removing the insignificant variables. The primary goal here is to choose the right set of variables for inclusion in a statistical model at the end. There are several approaches to variable selection, each with its own methodology and criteria for selection.

**Forward Selection:** This approach begins with an intercept-only model (i.e., no predictor in the model) and adds variables one at a time. At each step, the variable that provides the most significant improvement to the model is added, until no significant improvement can be made<sup>1</sup>.

**Backward Elimination:** In contrast to forward selection, backward elimination starts with all candidate variables included in the model. Variables are removed one at a time if they are not statistically significant, with the least significant variables removed first [1,2].

**Stepwise Selection (also known as Mixed Selection):** A combination of forward selection and backward elimination, stepwise selection involves adding variables as in forward selection but also includes a step where variables can be removed if they no longer provide a significant contribution to the model in the presence of newly entered predictors. That is, a variable can go in and out of the model as the model building progresses [2].

In the above selection approaches, instead of using the p-values, approaches like the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) can also be used to compare models. These criteria help in selecting a model that balances good fit with simplicity, penalizing models with excessive variables [3].

Each of these methods has its own set of advantages and disadvantages and can be chosen based on the specific requirements of the statistical analysis being performed, and each may affect the stability, unbiasedness, and validity of the final model [2,4]. All these approaches suffer from overfitting (especially with Backward selection approach), existing multicollinearity among predictors, inflated Type-1 Error due to multiple testing.

In this work, we introduce a new variable selection approach, called Residual Modelling. The term “Residual Modelling” has previously been used in the literature mainly within the context of forecasting [5,6] and voice and video editing [7,8]. We use the term specifically within the context of regression model building and variable selection.

## 2. Materials and Methods

This study has been carried out according to the Helsinki declaration with the Istanbul Medipol University Ethics Committee review and approval on September 30, 2019 (Ethics Committee Application No: 10840098-604.01.01-E.53819).

For a response variable  $Y$  with  $n$ -observations, the intercept-only regression model can be expressed mathematically as follows:

$$Y_i = \mu + \varepsilon_i, i = 1, 2, 3, \dots, n$$

If there are 5 predictors (i.e., independent variables) in addition to the known predictors (KP), for example,  $X_1, X_2, X_3, X_4,$  and  $X_5$ , the model that contains all predictors is called Full Model, and mathematically expressed as follows:  $Y_i = \beta_0 + \beta_{KP} * X_{KPi} + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \beta_3 * X_{3i} + \beta_4 * X_{4i} + \beta_5 * X_{5i} + \varepsilon_i$

where  $i = 1, 2, 3, \dots, n$ , and  $\beta_{KP}$  vector of parameters represents the effects of the known predictors  $X_{KP}$ . In the above model,  $\beta_0$  is the intercept of the model, representing the average value of the response variable  $Y$  when all the effects of the predictors are equal to zero. The parameter  $\beta_1$  expresses the predicted change in  $Y$  for each unit of change in  $X_1$  affects  $Y$  keeping all other predictors fixed,  $\beta_2$  shows expresses the predicted change in  $Y$  for each unit of change in  $X_2$  keeping all other predictors fixed, and so on. The error term  $\varepsilon$  represents random error, which is assumed to be independently and identically distributed Gaussian variates with zero mean and variance  $\sigma^2$ .

Setting entry and stay criteria around p-values, forward, backward, and stepwise variable selection strategies can be employed. Statistical packages also allow the user to set certain variables (e.g., known predictors-KP) to be kept in the model regardless of their significance. This is a critical functionality as we would be more interested in testing whether or not a given candidate predictor is significantly associated with the outcome variable controlling for the known factors and predictors.

### 2.1. A New Variable Selection Approach: Residual Modelling

In residual modelling, each predictor enters the model as a single predictor; then, the potential effect of this predictor is removed from the response variable by using the residuals of this model as the response variable for the next potential predictors. This process continues until all variables are tested. Table-1 shows an illustration of the Residual Modeling approach with 5 predictors.

**Table 1:** Residual Modelling Framework with five predictors ( $\beta_{NV}$  represents the effects of known predictors  $X_{NV}$ )

Step 1: $X_1$ is the only predictor in the model	$Y = \beta_{01} + \beta_{KP} * X_{KP} + \beta_1 * X_1 + \varepsilon_1$
Step 2: The residuals ( $\varepsilon_1$ ) from Step 1 are used as the response variable to assess the effect of $X_2$	$\varepsilon_1 = \beta_{02} + \beta_2 * X_2 + \varepsilon_2$
Step 3: The residuals ( $\varepsilon_2$ ) from Step 2 are used as the response variable to assess the effect of $X_3$	$\varepsilon_2 = \beta_{03} + \beta_3 * X_3 + \varepsilon_3$
Step 4: The residuals ( $\varepsilon_3$ ) from Step 3 are used as the response variable to assess the effect of $X_4$	$\varepsilon_3 = \beta_{04} + \beta_4 * X_4 + \varepsilon_4$
Step 5: The residuals ( $\varepsilon_4$ ) from Step 4 are used as the response variable to assess the effect of $X_5$	$\varepsilon_4 = \beta_{05} + \beta_5 * X_5 + \varepsilon_5$

The order of the variables entering the model selection can be done in a random fashion as well although it should not make a difference under the independence assumption. With Residual Modelling, we hope to answer the following question: Does  $X_2$  have significant association with  $Y$  after removing the association of  $X_1$  with  $Y$ . Then, we move to the next variable to assess its association with  $Y$  after removing the effects of  $X_1$  and  $X_2$ .

Similar to the competitive variable selection approaches, known predictors enter the Residual Modelling strategy in Step-1; that is, the effects of the known predictors (KPs) are removed from the response variable in Step-1 and the residuals from this model becomes the response variable for the next candidate predictor to be tested.

We compared the performance of these four approaches in terms of the statistical power, Type-1 error retention, estimation bias and standard errors through simulations.

**Simulation Design:**

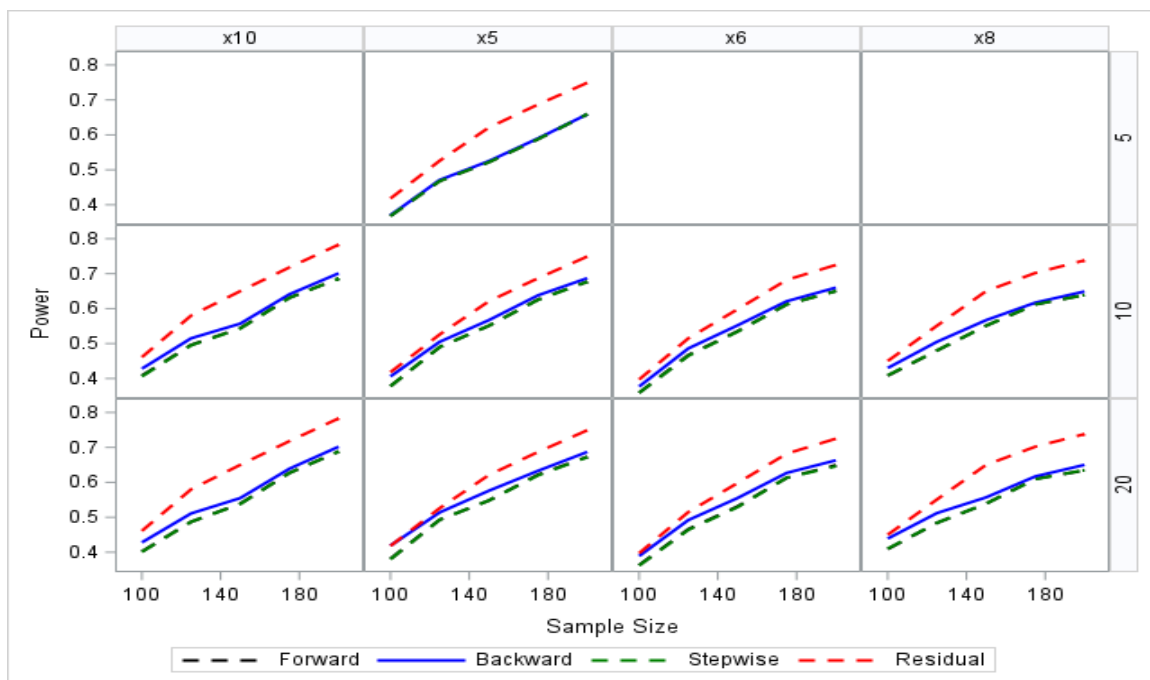
- Number of predictors: 5, 10, 20 ( $X_1 - X_{20}$ ) of whom two are binary factors ( $X_6, X_8$ )
- Response Variable-1:  
 $Y_1 = 120 + 0.2X_1 + 0.2X_5 + 0.4X_6 + 0.4X_8 + 0.2X_{10} + \epsilon$
- Response Variable-2:  $Y_2 = 120 + 0.2X_1 + 0.2X_2 + 0.2X_3 + 0.2X_4 + 0.2X_5 + 0.4X_6 + 0.2X_7 + 0.4X_8 + 0.2X_9 + 0.2X_{10} + \epsilon$
- Assumed Known Predictors:  $X_1$  only,  $X_{10}$  only,  $X_1$  and  $X_5$ ,  $X_1$  and  $X_5$  and  $X_{10}$
- Sample size: 100, 125, 150, 175, 200 (Considering at least 10 records per predictor)
- Number of simulation repeats: 1000

The simulations were summarized as the number of significant runs for each of the predictors, where the percentage of the runs with significant results for insignificant variables will represent the empirical Type-1 Error Rate and the percentage of the runs with significant results for significant variables will represent the empirical Statistical Power. Parameter estimates and their standard errors were also summarized as the averages across the simulation runs.

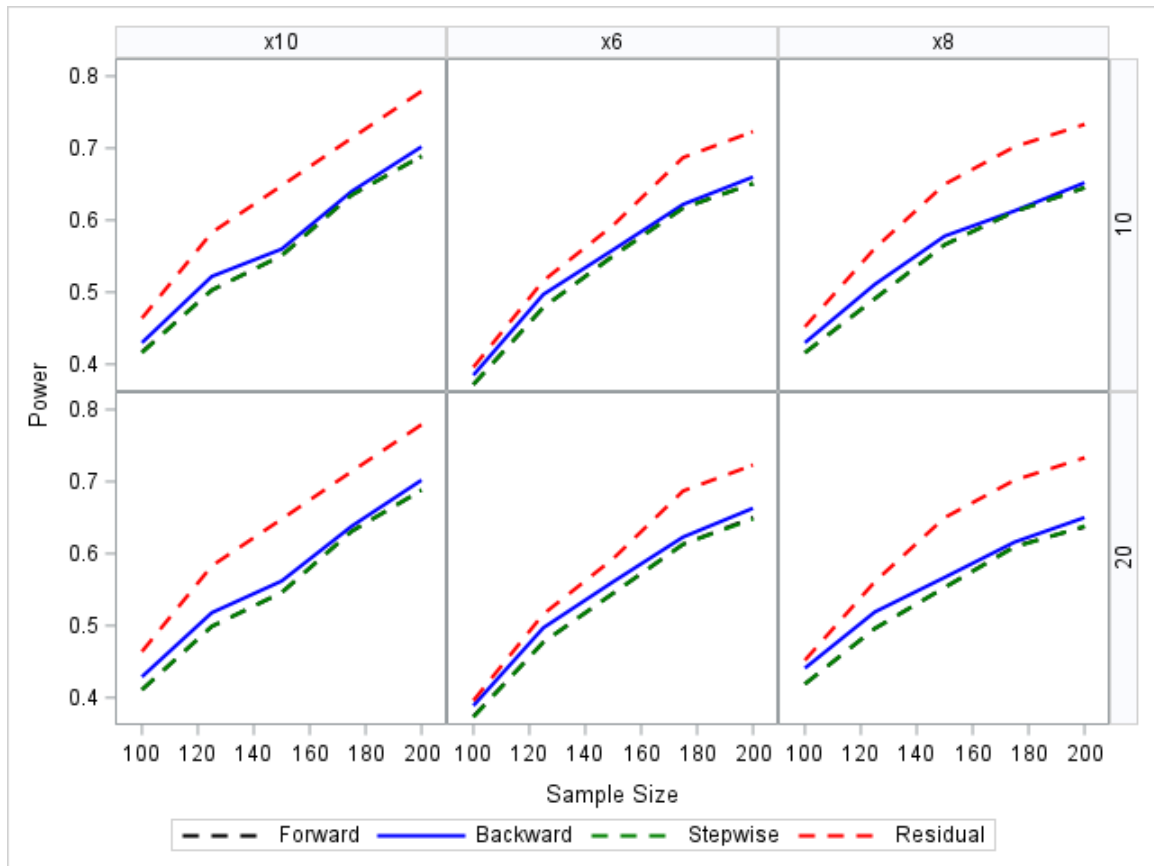
All analyses were conducted in a parallel manner in SAS® Version 9.4 (SAS Institute, Cary, North Carolina, USA).

**3. Results**

Empirical statistical power advantage of Residual Modelling was apparent especially as the sample size increases (Figure 1-2, Supplementary Figures A1-2). The performance of the Backward Elimination (BE) approach is the second best overall.

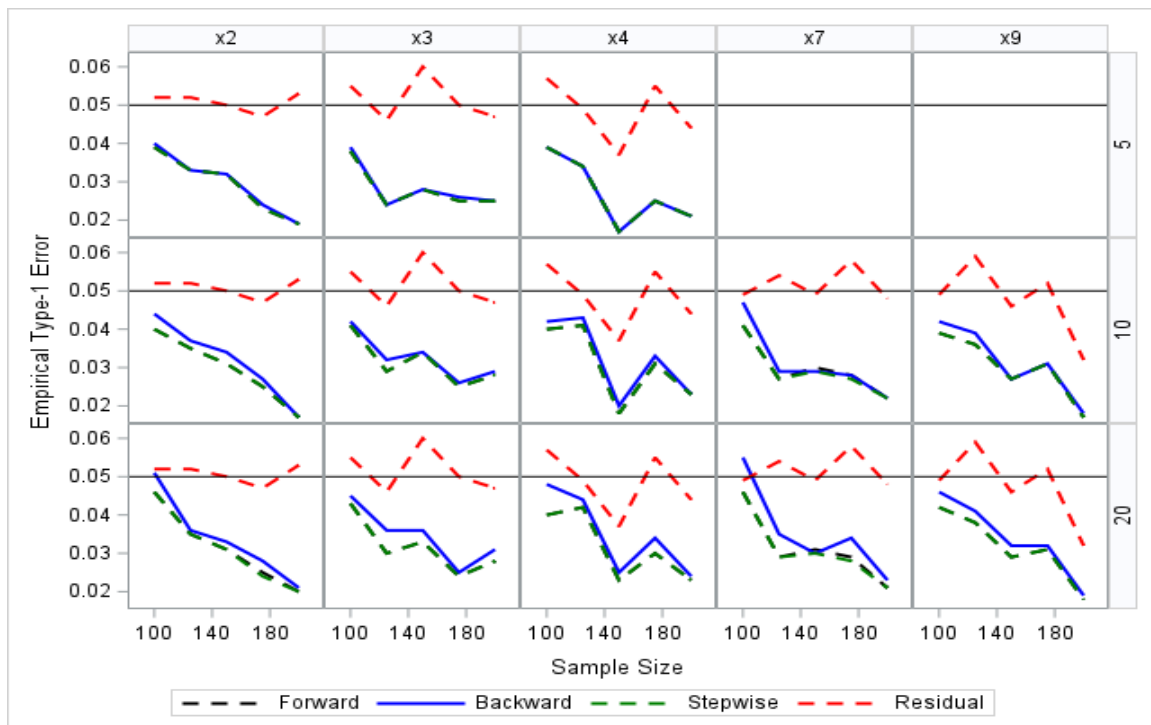


**Figure 1:** Empirical Power when  $X_1$  is the known significant predictor



**Figure 2:** Empirical Power when  $X_1$  and  $X_5$  are considered as the known significant predictors

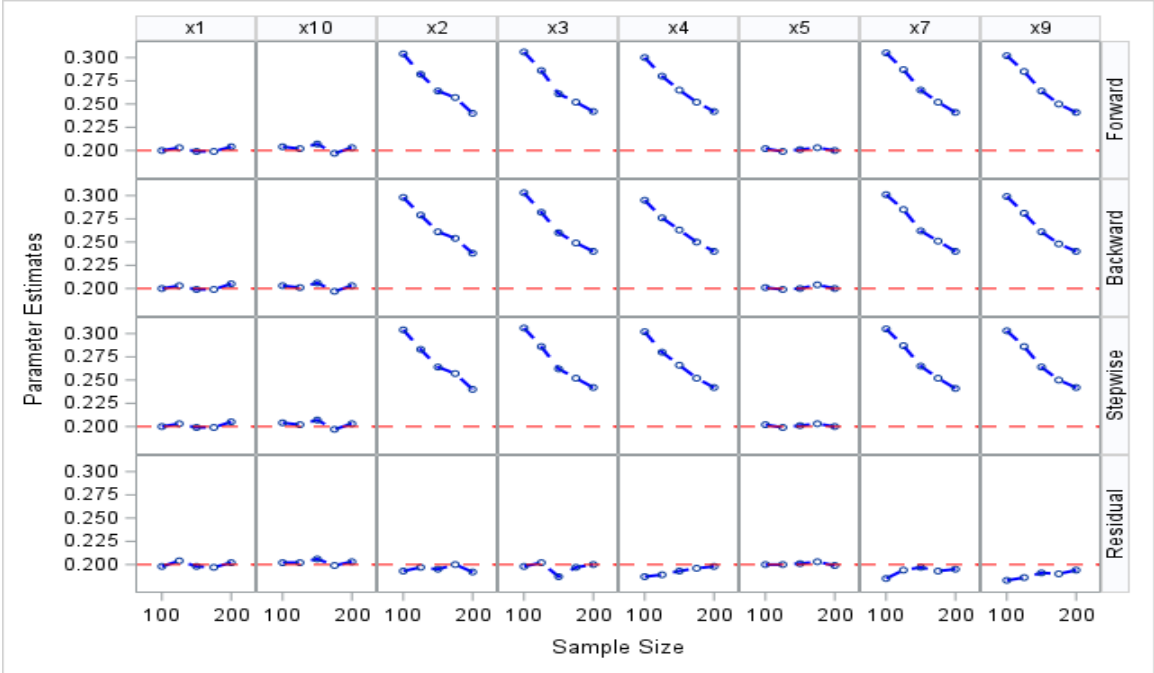
In Figure 3, we observe that the Type-1 Error Rate is retained by Residual Modelling at around the nominal level of 0.05. The other models underestimate the Type-1 Error rate and it seems to be a function of sample size as well. We observe an overall decline in Type-1 Error as the sample size increases.



**Figure 3:** Empirical Type-1 Error rate for representative insignificant variables

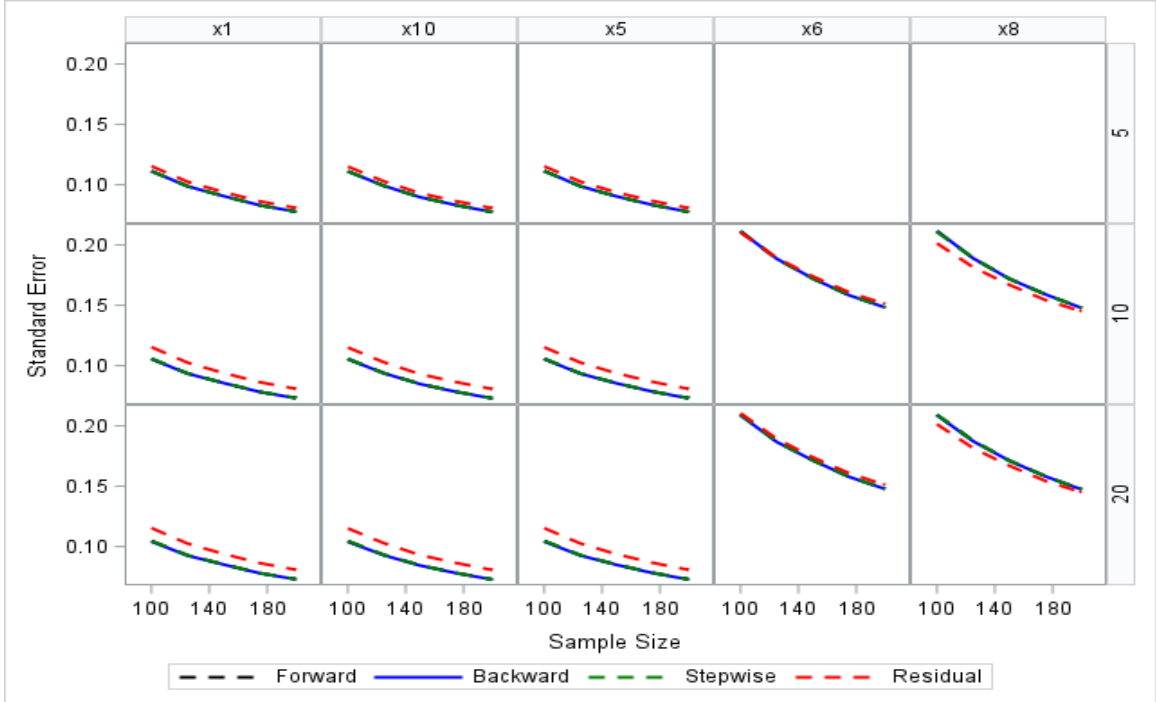


Residual Modelling resulted in stable parameter estimation both for the known and new significant predictors; although the parameter estimation of the known predictors by the other three variable selection approaches was stable for the known predictors, they overestimated the parameters for the new significant predictors, which improved as the sample size increases as shown in Figure-4. This situation was the same for the two binary predictors (Supplementary Figure A3).



**Figure 4:** Parameter Estimates for the continuous significant predictors by Variable Selection methods

Standard error was elevated for Residual Modelling especially with higher number of predictors in the model compared the other three approaches as shown in Figure-5. Standard error improves with the increases sample size as anticipated.



**Figure 5:** Percentage of simulation repeats showing significance rate of redundant predictors in both models.

The same set of simulations were run under two additional scenarios: Scenario-1: Known predictors were applied in all models and the residuals of these models were used as the response variables for

---

the forward, backward, and stepwise variable selection approaches. Scenario-2: Candidate variables were shuffled to assess the impact of random order of predictors on the performance. Both scenarios resulted in very similar performances in all models, indicating the robustness of Residual Modelling under random assessment of candidate predictors and the robustness of the other three variable selection approaches under the complete removal of the effects of the known predictors.

### 3.1. Real-World Application

The lung and throat cancer (LTC) deaths per 100,000 population were obtained for 81 provinces of Turkiye for Year 2019. Ever smoking, ever alcohol consumption, and ever exposure to second-hand smoking at home were obtained as the known factors for the LTC deaths. We wish to investigate if the yearly variation (using the coefficient of variation-CV) of humidity, air pressure, wind speed, SO<sub>2</sub> and Particulate Matter-10 are associated with the LTC deaths controlling for the known effects of smoking and alcohol. As a result, all for approaches identified the CV of humidity and windspeed as significant new predictors of the LTC deaths, suggesting that provinces with higher CV of humidity and windspeed reported relatively lower LTC deaths (Supplementary Figure A4).

## 4. Discussion

We have shown through simulations that the Residual Modelling performs better as the number of predictors increase while Full Modelling seems slightly more statistically powerful with small number of predictors.

In this investigation, we aimed to address a challenge in variable selection and model interpretation. Rather than focusing on statement like “Predictor  $X_1$  is significantly associated with the response variable  $Y$  after controlling the effect of  $X_2$ ”, we feel that a statement like “Predictor  $X_1$  is significantly associated with the response variable  $Y$  after removing the effect of  $X_2$ ” is better and more interpretable. For example, when studying lung cancer mortality, smoking is a known factor. When both smoking and other candidate factors are entered into the model, smoking may lose its significance just due to some intrinsic association of smoking with other factors and the research may conclude “Smoking was not found to be significantly associated with lung cancer mortality in the presence of such and such predictors.” Residual Modelling allows us to remove the effects of all known factors before we investigate the effects of other potential factors with the following question in mind: “Do these additional factors have any significant effect on  $Y$  beyond the known effects of smoking, etc.”

Another advantage of Residual Modelling is that it does not limit itself with the number of predictors at hand. For example, Residual Modelling approach can be applied to gene expression studies where the number of predictors far exceeds the number of patients or samples at hand, utilizing the full power of the entire sample size for each potential predictors, while Full Modelling requires that the sample size be much higher than the number of predictors. Here, the effects of the known factors are removed from the response variable and then the effect of each gene can be investigated on the residuals. Familywise error control measures can be taken on their p-values as usual.

In model building, the subject level expertise supported by a comprehensive literature search is indispensable to identify the known predictors of the response variable at hand. In Residual Modelling, this becomes much more critical as the effects of these known predictors are removed from the response variable literally before investigating the candidate predictors. This requires extra attention especially when the predictors have inherent correlations among them. Naturally, when we deal with a group of highly correlated predictors, when one is considered as a known factor and its effect on the response variable is removed, the other candidate predictors are highly likely to be found as insignificant. This is exactly where the subject knowledge is needed.

The user of Residual Modelling is free to combine the residual modelling approach with other variable selection approaches. For example, once the known effects are removed from the response variable, a stepwise variable selection can be carried out on the residual of this model. As the order of the candidate variables may be important especially when these predictors have a certain level of correlation among them, the user can carry out sensitivity analyses by shuffling the candidate variable list. Under the multicollinearity assumption, this would not matter; however, in real-life studies, we see at least mild level correlations among potential predictors more often than not.

Residual Modelling approach is also prone to similar weaknesses as in the other variable selection approaches such as overfitting (especially with Backward selection approach), existing multicollinearity among predictors, inflated Type-1 Error due to multiple testing. Therefore, as a future research, we plan to study who the regularisation techniques such as Lasso, Ridge, and Elastic Net are can be used as penalization and shrinkage approaches to improve the model performance of Residual Modelling [9-13].

Residual Modelling is proposed for continuous response variables specifically where we can obtain residuals that are continuous variables in nature as well. Therefore, in its current state, Residual Modelling is not applicable to other types of responses such as binary, count, time-to-event, etc.

## 5. Conclusions

We have shown that Residual Modelling has desirable power and type-1 error rate properties in model building when there are known predictors (KPs) we need to account for in testing the significance of a new set of candidate predictors.

**Author Contributions:** Conceptualization, ANK and MK.; methodology, MK.; software, MFD and MK.; validation, MFD and MK.; formal analysis, MK.; investigation, MK.; resources, MK.; data curation, MK.; writing—original draft preparation, ANK, MFD, and MK.; writing—review and editing, ANK, MFD, and MK.; visualization, MK.; supervision, MK.; project administration, MK.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** We have no additional acknowledgements beyond the author contributions to be declared.

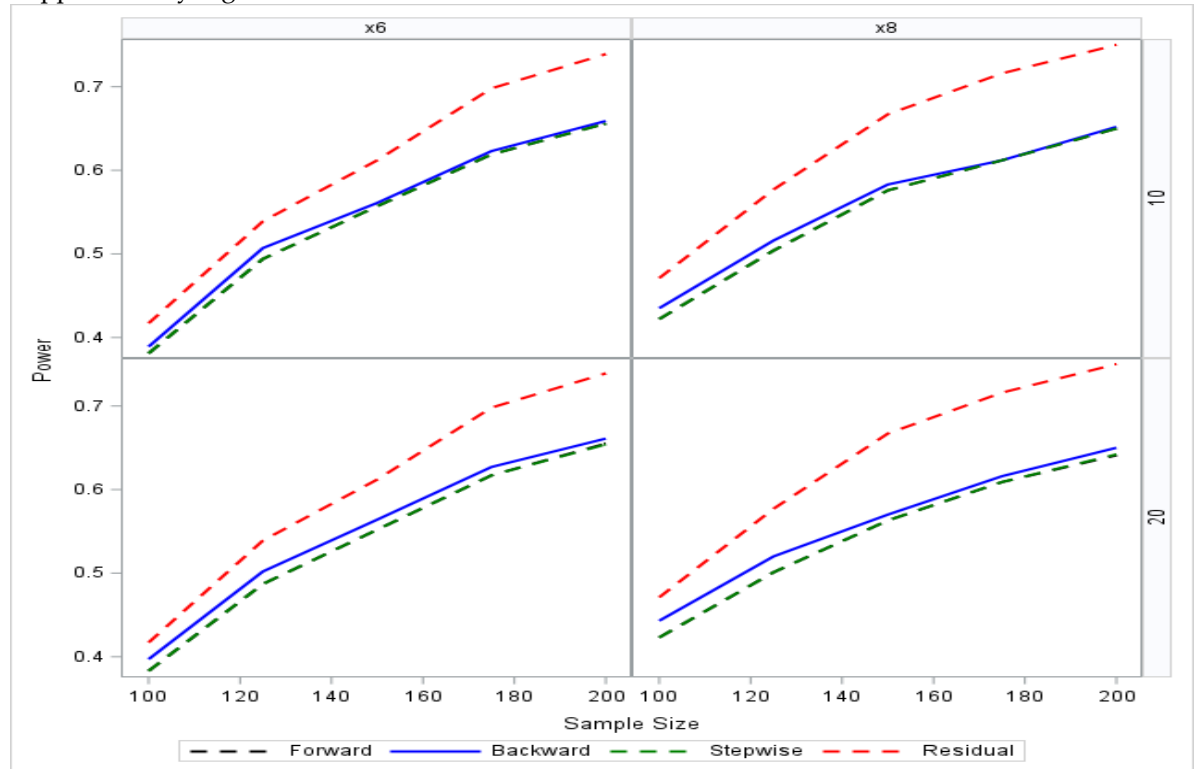
**Conflicts of Interest:** The authors declare no conflict of interest.

## References

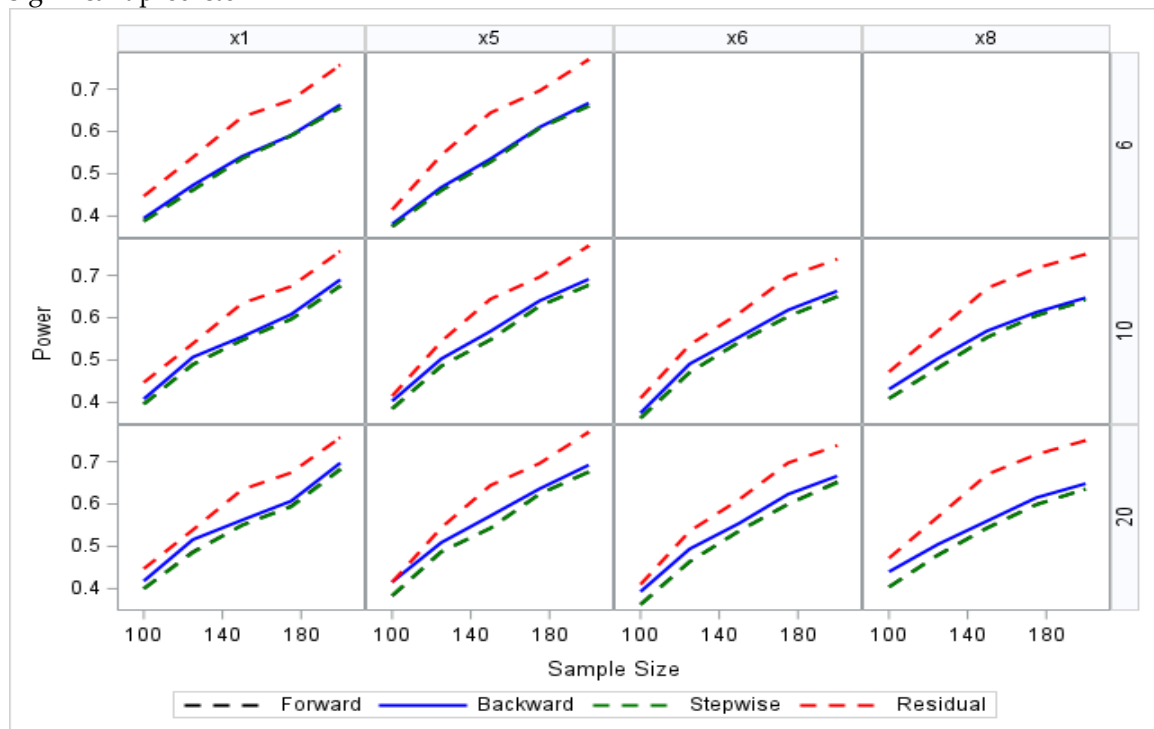
- [1] W. Sauerbrei, A. Perperoglou, M. Schmid, M. Abrahamowicz, H. Becher, H. Binder, D. Dunkler, F. E. Harrell, P. Royston, and G. Heinze, "State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues," *Diagnostic and Prognostic Research*, vol. 4, pp. 1–8, Dec. 2020. doi: 10.1186/s41512-020-00074-3.
- [2] M. Z. Chowdhury and T. C. Turin, "Variable selection strategies and its importance in clinical prediction modelling," *Family Medicine and Community Health*, vol. 8, no. 1, 2020. doi: 10.1136/fmch-2019-000262.
- [3] G. Claeskens, "Statistical model choice," *Annual Review of Statistics and Its Application*, vol. 3, pp. 233–256, Jun. 2016. doi: 10.1146/annurev-statistics-041715-033659.
- [4] G. Heinze, C. Wallisch, and D. Dunkler, "Variable selection—a review and recommendations for the practicing statistician," *Biometrical Journal*, vol. 60, no. 3, pp. 431–449, May 2018. doi: 10.1002/bimj.201700067.
- [5] Y. Wang, Q. Chen, N. Zhang, and Y. Wang, "Conditional residual modeling for probabilistic load forecasting," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7327–7330, Aug. 2018. doi: 10.1109/TPWRS.2018.2819668.
- [6] P. S. de Mattos Neto, G. D. Cavalcanti, D. S. O. Santos Júnior, and E. G. Silva, "Hybrid systems using residual modeling for sea surface temperature forecasting," *Scientific Reports*, vol. 12, no. 1, p. 487, Jan. 2022. doi: 10.1038/s41598-021-04342-8.
- [7] M. Goodwin, "Residual modeling in music analysis-synthesis," in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, May 1996, vol. 2, pp. 1005–1008. IEEE. doi: 10.1109/ICASSP.1996.543310.
- [8] X. Weng, Y. Li, L. Chi, and Y. Mu, "High-capacity convolutional video steganography with temporal residual modeling," in *Proceedings of the 2019 on International Conference on Multimedia Retrieval*, Jun. 2019, pp. 87–95. doi: 10.1145/3323873.3325047.
- [9] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. doi: 10.1111/j.2517-6161.1996.tb02080.x.
- [10] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970. doi: 10.1080/00401706.1970.10488634.
- [11] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005. doi: 10.1111/j.1467-9868.2005.00503.x.
- [12] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010. doi: 10.18637/jss.v033.i01.
- [13] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. New York: Springer, 2013, ch. 6. doi: 10.1007/978-1-4614-7138-7.

## Appendix A

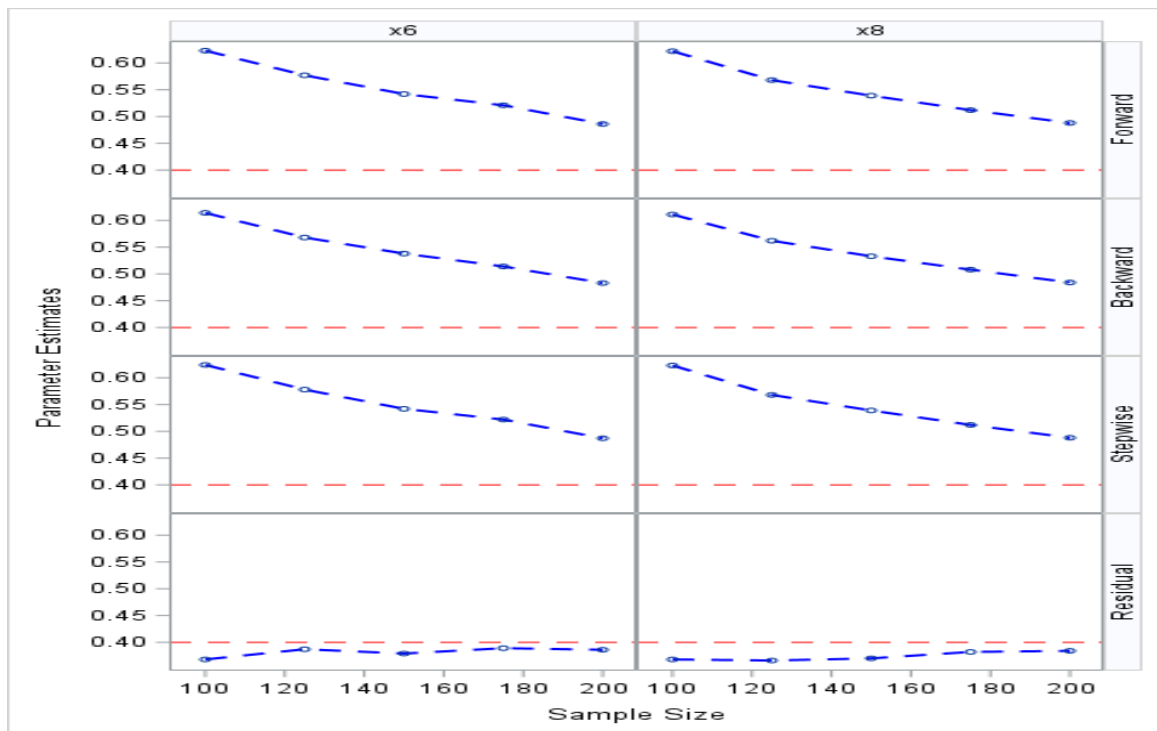
Supplementary Figures



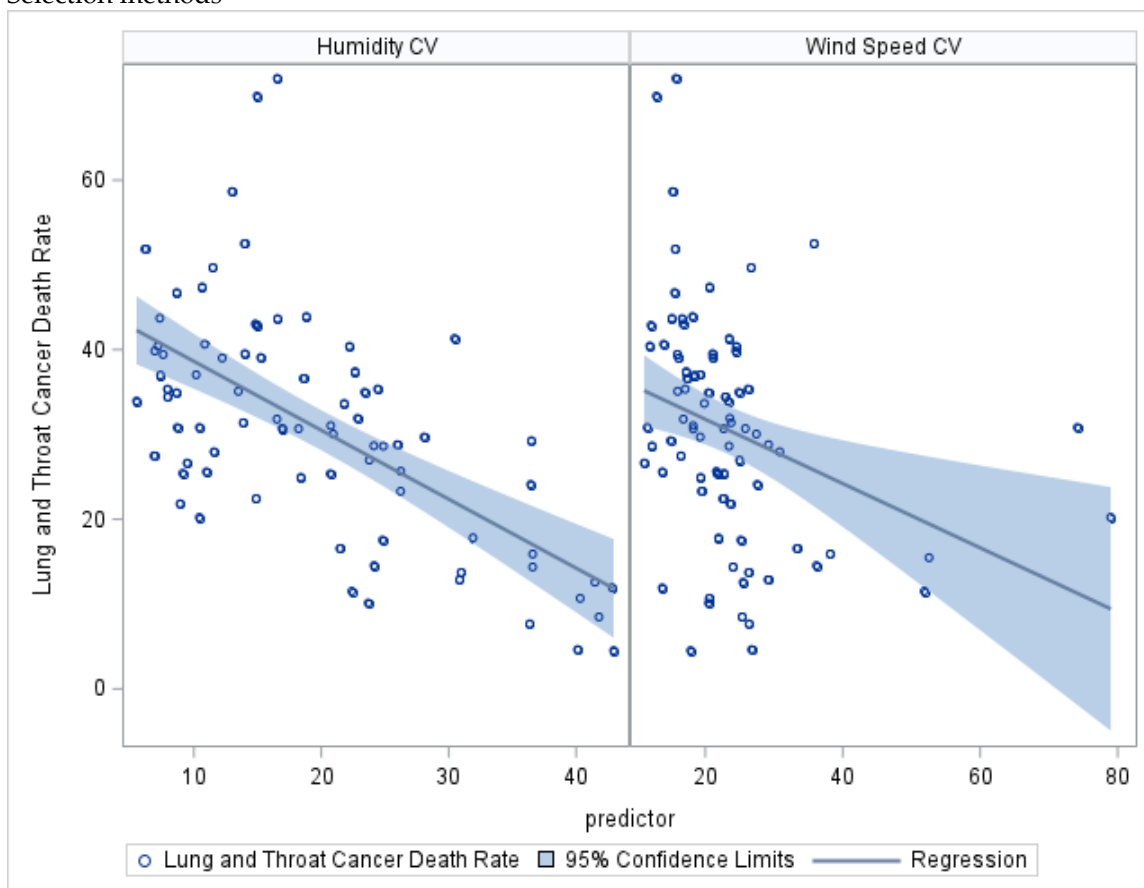
**Supplementary Figure A1:** Empirical Power when X1, X5, and X10 are considered to be the known significant predictor



**Supplementary Figure A2:** Empirical Power when X10 is considered to be the known significant predictor



**Supplementary Figure A3:** Parameter Estimates for the binary significant predictors by Variable Selection methods



**Supplementary Figure A4:** Association of Humidity and Wind Speed Coefficient of Variation (CV) with Lung and Throat Cancer Deaths per 100,000 population

Research Article

Received: date:02.09.2024  
Accepted: date:24.10.2024  
Published: date:31.06.2024

# Examination of Statistical Methods Used in Group Comparisons in Visual Analogue Scale (VAS) Data in Terms of Type I Error Rate and Power of Test: Monte Carlo simulation study for VAS data

Özgür Koşkan<sup>1</sup>, Malik Ergin<sup>2\*</sup>

<sup>1</sup>Isparta University of Applied Sciences, Faculty of Agriculture, Department of Animal Science, Isparta, Türkiye;  
ozgurkoskan@isparta.edu.tr

<sup>2</sup>Isparta University of Applied Sciences, Faculty of Agriculture, Department of Animal Science, Isparta, Türkiye;  
malikergin@isparta.edu.tr

Orcid: 0000-0002-5089-6250<sup>1</sup> Orcid: 0000-0003-1810-6754<sup>2</sup>

\*Correspondence: malikergin@isparta.edu.tr

**Abstract:** In this study, how the application changes with the size of the scale, type I error rate and power of test values were examined in different sample sizes. The material of this study is the random numbers generated according to different sample sizes and different group means of standard deviation out of populations that hold multinomial distribution. In the study, permutation test, F-test and Kruskal-Wallis (KW) test were examined using combinations with different sample sizes ( $n = 5, 10, 20, 30, 50$ ) and different p values (probability of occurrence of events which are 0.25, 0.50, 0.75) first interms of Type I error rate and then power of the test using different group averages ( $\Delta = 1.0, 1.5, 2.0$ ). As a result of the simulations, it is seen that with small sample sizes, Kruskal-Wallis test was unable to maintain Type I error rate at 0.05 level. In the evaluation of such Likert-type data, it can be stated that, permutation test, one of the distribution free tests, is more practical than other tests in terms of maintaining the Type I error rate at 0.05 level and high power of test values. The permutation test is satisfactory in terms of both the type I error rate and the power of the test. Also, permutation test is a distribution-free test. Therefore, can be used without prerequisites. In almost every combination ( $\pi, \Delta, k, n$ ), permutation test had similar or superior type 1 error and power values than the F and KW tests. It was observed that compared to the 5 visual analogue scale, when the data is measured in 10 and 20 visual analogue scale, the power of test values decreased. In other words, if more than 5 visual analogue scale are made on the visual analogue scale data, the results would negatively be affected.

**Keywords:** quantitized data; likert-type data; permutation test; monte carlo simulation; power of test values

## 1. Introduction

Visual analogue scale (VAS) data are the data sets obtained from scoring verbal predictions. The most widely used visual analogue scales are 5 and 10-point scales however larger scales are also often used too. The points can vary from no pain, little bit pain, medium pain, much pain, and very much pain. For evaluation, analysis is conducted by attributing numerical values to these expressions.

The technique's apparent simplicity and adaptability to a wide range of research settings have made it an appealing measurement option. VAS is preferred by researchers as it is simple, quick, easily understood by untrained staff and subjects, and allows the use of numerical values suitable for statistical analysis. Patient satisfaction is one of the most important evaluations and it will be even more important in the future. The scale needs to be practical, objective, and applicable for the measurement. In the literature, the distribution of the obtained data is reported to show a discrete uniform distribution [1-3].

---

The discrete uniform distribution of data, such as VAS values, usually has the assumption that each value will be selected from the sample with equal probability in a given draw. The probability mass function of the discrete uniform distribution is given by  $P(X=x)=1/(n+1), x=0,1,\dots,n$ , which means that  $X$  can take any integer value between 0 and  $n$  with equal probability. The mean and variance of the distribution are  $n/2$  and  $(n(n+1))/12$ , respectively. To generate a random number from the discrete uniform distribution, a random number ( $R$ ) can be drawn from the  $U(0,1)$  distribution,  $S=(n+1)R$  can be calculated, and the integer part of  $S$  can be drawn from the discrete uniform distribution [4]. In the literature, the power of ANOVA and PermANOVA tests are generally tested with symmetrical distributions such as normal distribution,  $t$ -distribution, and curved distributions such as Beta, Gamma and Chi-square, and both type I error and power of test values are tested. As it is known, all of these distributions are continuous distributions.

In fact, in the studies it was assumed that the distribution of VAS values approached normal [5-8]. However, likert type variables such as VAS values are discrete variables and show a discrete distribution. Likert types such as having 5 or 10 scale values are data sets with discrete uniform distribution. This study, which compares the methods used in group comparisons such as PermANOVA, Kruskal-Wallis, on the discrete uniform distribution obtained in different ways by changing the parameters, will shed light on the literature. Because, in practice, these tests are still widely used on data obtained in likert type, such as VAS values.

There are many studies showing that PermANOVA and Kruskal-Wallis tests are distribution-independent tests and are also used in ordinal data [9-12]. It is thought that choosing the distribution as discrete uniform and examining these tests in terms of both type I error and power of the test will provide an innovative approach to the literature.

Based on these principles, VAS data show a discrete uniform distribution. As it is known,  $F$ , Kruskal-Wallis (KW), and permutation tests are commonly used in group comparisons. If the data meets the prerequisites of parametric tests, the  $F$  test is used, if not, the KW test is used. The permutation test is a resampling method and does not require any prerequisites. With the development of cheap and fast computers, the permutation tests now tend to be used more often as fast computers are needed to compute these powerful tests whose calculations are easy. One of the advantages of the resampling method is that there is not much need for too much knowledge in mathematical and statistical formulas. Enough knowledge to ensure an understanding of the concepts and methods is sufficient for the application of the resampling approach [13]. While the expected results cannot be obtained using traditional methods, they can be obtained using the resampling method [14]. In addition, due to the limitations of traditional parametric and non-parametric significance tests, permutation testing is considered an important alternative [15].

The purpose of this study was to examine how type I error rate and power of test values change in different sample sizes when statistical methods comparing groups are used for visual analogue scale data.

## 2. Materials and Methods

The material of this study consisted of random numbers in three groups, generated using Microsoft Power Station Developer Studio and IMSL Library in terms of different sample sizes and different group means of standard deviation out of populations that hold discrete uniform distribution. Permutation,  $F$ , and Kruskal-Wallis (KW) tests were examined in terms of type I error rate and power of the test by using combinations with different sample sizes ( $n = 5, 10, 20, 30, 50$ ),  $\pi$  values (probability of occurrence of events which were 0.25, 0.50, 0.75) and group means ( $\Delta = 1.0, 1.5, 2.0$ ). In the study, 100.000 simulations were performed for each sample size ( $n$ ) and scale ( $k$ ). Among these, the distribution of random numbers was showed below with an example for  $n=50$  and  $k=5, 10, \text{ and } 20$ . These distributions represent scenarios that could be observed. The three distributions given in the figures are presented as representative in Figure 2. Each sample was drawn completely randomly from the population 100.000 times. Consequently, due to this random sampling, the distribution of VAS values will have varying frequencies in each sample. Similarly, by adjusting the distribution's parameter 'p' up and down, the

mean and standard deviation of VAS values decrease and increase. This allows for the creation of situations that are closer to reality.

**Permutation Test:** The permutation test is a resampling method that does not require any prerequisites. The permutation test is also called the distribution-free test, which does not require the normality assumption for the distribution of data. The permutation test can be an alternative to the F test, especially in studies with small sample sizes [16]. In the study, after the F test table value calculated for 3 groups was found, the F value was calculated again by resampling which means grouping the data of these three groups again randomly. When the F value was equal to or greater than the first calculated F value, it was counted. The total counted F value was divided to resampling iterated 10000 times to find the percentage ratio of the F value. When the percentage ratio of the F value was either equal to or smaller than 0.05, the  $H_0$  hypothesis is rejected by the permutation test. **F test (ANOVA):** As it is known, the variance analysis technique is the most widely used statistical method to examine whether the difference between the means of two or more independent groups happens coincidentally or not. Let formulize  $(i=1, \dots, n_i)$  and  $(j=1, \dots, k)$  and  $\sum nk=N$  to show the  $X_{ij}$  as  $i^{th}$  observation in the  $J^{th}$  treatment group. It is assumed that the  $X_{ij}$  here is distributed independently and normally with  $\mu_j$  and  $\sigma^2$  parameters. In this case,  $\mu_j$  and  $\sigma^2$  are the best linear prediction estimators. The data obtained at the end of the study can be identified with the  $Y_{ij}=\mu+\alpha_i+e_{ij}$  model. In this mathematical model;  $\mu$ ; is the overall mean, that is the population mean,  $\alpha_i$ ; is the effect of  $i^{th}$  treatment  $e_{ij}$ ; is the error term. It is assumed that all  $e_{ij}$  are independent of each other with normal distribution, and  $\sigma_j^2$  is pooled variance. For this purpose, to examine whether there is a difference between the aforementioned group means or not, the following control and alternative hypothesis need to be tested as:

$$H_0: \mu_1=\mu_2=\dots=\mu_k$$

$H_1$ : "the difference between the means of at least two groups is significant" and calculated as:

$$\text{Total Sum of Square} = \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X})^2 \quad (1)$$

$$\text{Between Sum of Square} = \sum_{i=1}^k n_i * (\bar{X}_i - \bar{X})^2 \quad (2)$$

$$\text{Within Sum of Square} = \sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X})^2 \quad (3)$$

$F = \frac{\sum_{i=1}^k n_i * (\bar{X}_i - \bar{X})^2}{\sum_{i=1}^k \sum_{j=1}^n (X_{ij} - \bar{X})^2}$  that calculated value (k-1) and (N-k) is tested with F table value with degrees of freedom to accept or reject the  $H_0$  hypothesis.

**Kruskal-Wallis Test**

It is well known that Kruskal-Wallis test is the most widely used statistical technique to examine whether the difference between the median of two or more independent groups stems from coincidence or not. The statistics applied for the Kruskal-Wallis test is H-type statistics. To apply this kind of statistics, the units of the experiment are supposed to be randomized through the population, observations are independent of each other, and the group rank means be compared using the rank values of these data. The following control and alternative hypothesis need to be tested.

$$H_0: M_1=M_2=\dots=M_k$$

$H_1$ : "the difference between the median of at least two groups is significant".

$$H = \frac{12}{N(N+1)} * \sum_{j=1}^k \left( \frac{(\sum R_j)^2}{n_j} \right) - 3(N + 1) \quad (4)$$

The calculated H statistics (k-1) value is checked using the table value of chi-square with degrees of freedom. Later, it is determined whether the difference between the rank means of the groups, in other words, medians is significant or not [9].

Power of test and type I error rate



Two types of error occur in the controlling hypothesis when the control hypothesis is decided by being tested with the alternative hypothesis. At the end of the controls, when the correct control hypothesis is rejected, this error is called the type I error. The second type error is the one that occurs in case the alternative hypothesis is actually true and when the control hypothesis is accepted. We obtain the power of test by subtracting this error from 1 [20]. At the end of 100 000 simulations performed in the study, type I error and power of test values were found.

### 3. Results

Type I error rate values for permutation, F, and KW tests are provided in Table 1. When Table 1 is examined, as it is observed in all scales (k) when the sample size (n) was 5, KW test was unable to keep the type I error rate at 0.05 level as determined at the beginning and had lower values. This case also continued relatively less when (n) was 10. However, when the sample size was larger than 20, the type I error rate determined at the beginning was maintained. The changes in the distribution of the probability of event occurrence ( $\pi$ ) did not affect the type I error. In small sample sizes, while the F test maintained the type I error at 0.05 level, KW test did not maintain the type I error at 0.05 level and had values lower than 0.05. When the probability of event occurrence ( $\pi$ ) was 0.25, the scale (k) was 5 and the sample size (n) was 5, the type I error rate for F, KW, and permutation test were 0.051, 0.042, and 0.050, respectively. This shows that the type I error rate at the beginning could not be maintained as 0.05 in KW test. A similar situation was observed when the sample size was 10; where type I error rate values were 0.054, 0.043, and 0.051 for F, KW, and permutation test respectively. Other cases also gave similar results for other  $\pi$  values. For sample size 20 and over, the type I error rate was maintained at 0.05 for F, KW, and permutation test.

**Table 1.** Type I error rate values for F, KW, and permutation tests.

$\pi$	k	n	F	KW	Permutation	
0.25	5	5	0.051	0.042	0.050	
		10	0.052	0.046	0.049	
		20	0.051	0.049	0.051	
		30	0.050	0.046	0.052	
		50	0.052	0.050	0.050	
	10	5	0.054	0.043	0.051	
		10	0.052	0.046	0.050	
		20	0.051	0.048	0.050	
		30	0.051	0.049	0.049	
		50	0.050	0.050	0.051	
	20	5	0.056	0.043	0.052	
		10	0.053	0.047	0.051	
		20	0.051	0.048	0.050	
		30	0.051	0.049	0.050	
		50	0.051	0.049	0.050	
	0.50	5	5	0.051	0.041	0.052
			10	0.053	0.047	0.049
			20	0.050	0.048	0.051
			30	0.051	0.049	0.050
			50	0.053	0.050	0.050
10		5	0.054	0.041	0.052	
		10	0.051	0.046	0.051	
		20	0.050	0.048	0.049	
		30	0.049	0.049	0.051	
		50	0.050	0.049	0.050	
20		5	0.055	0.042	0.054	
		10	0.052	0.047	0.052	
		20	0.051	0.049	0.050	
		30	0.051	0.049	0.051	
		50	0.051	0.049	0.050	

		5	0.050	0.041	0.049
		10	0.051	0.046	0.050
	5	20	0.052	0.048	0.051
		30	0.050	0.049	0.051
		50	0.053	0.049	0.051
0.75		5	0.055	0.042	0.052
		10	0.051	0.045	0.052
	10	20	0.052	0.049	0.051
		30	0.051	0.050	0.049
		50	0.051	0.050	0.050
		5	0.055	0.042	0.053
	20	10	0.052	0.046	0.051
		20	0.050	0.047	0.049
		30	0.051	0.049	0.050
		50	0.050	0.049	0.050

Power of test values for F, KW, and permutation tests when  $\Delta=1$ ,  $\Delta=1.5$ , and  $\Delta=2$  are provided in Table 2, Table 3, and Table 4, respectively. When Table 2, Table 3, and Table 4 are examined it is observed that the power of test was affected by changes in  $\pi$  value. When one standard deviation difference ( $\Delta$ ) was created among group means (Table 2), and the power of test is examined, it is observed that resampling methods such as the permutation test and F test had higher values compared to the Kruskal Wallis test especially when the sample size was larger than 20. When looking at Tables 2, 3 and 4, in general, it can be said that compared to normally distributed data, VAS values showing discrete uniform distribution had lower power of test values in all three tests. In addition, it is shown that when  $\pi$  value increased the difference between tests decreased (Table 3). However, it is also seen that the permutation test, known as an independent test, had greater power of test values than KW test in all cases (Table 2, 3, 4). When the difference among the group means was ( $\Delta$ ) 1.5 standard deviation, permutation, and F tests continued to be relatively more powerful in small sample sizes. However, this apparent situation decreased as the sample size increased in the groups.

When the difference among the group means was ( $\Delta$ ) 2 standard deviation, only in the cases where the sample size was 5 and 10, the permutation and F tests had higher power of the test values than KW test (Table 4). When the sample size increased, the difference between tests in terms of power of test decreased and when the sample size was 30, the power of test values of tests become similar. But it is observed that in all cases when evaluating VAS data, permutation, and F test had higher power of test values than KW test (Table 2, 3, 4). As can be seen in Tables 2, 3, and 4, where the power of test values are given, the power of all three tests relatively decreased as the  $\pi$  value increased. When the overall results are examined, it can be inferred that KW test had lower power of test values in the analysis of sensory data, in this study VAS data. However, especially in small sample sizes, the permutation and F tests had higher power of test values. It is seen that whatever the distribution is, as the difference among the group means increases, the power of all three tests increases. As it is known, the larger the sample size gets, the higher the power values become.

Consequently, we can say that in visual analogue scale type data analysis, the resampling method, which is also known as distribution independent tests, is a suitable test. It should be favored instead of KW test, especially in visual analogue scale data such as VAS. After an overall examination of this study, it is concluded that in almost every case, the permutation method, which is one of the resampling methods can be favored for visual analogue scale type data in terms of both type I error rate and power of test values.

**Table 2.** Empirical Test Power for Samples from Multinomial Distributions ( $\Delta=1$ )

$\Delta$	$p$	$k$	$n$	F	KW	Permutation
			5	0.136	0.110	0.129
			10	0.542	0.433	0.524
		5	20	0.971	0.946	0.968
			30	1.000	1.000	1.000

1.0	0.25	10	50	1.000	1.000	1.000
			5	0.070	0.057	0.066
			10	0.151	0.141	0.148
			20	0.600	0.520	0.061
			30	0.960	0.920	0.957
		50	0.999	0.999	0.999	
		20	5	0.059	0.045	0.053
			10	0.076	0.069	0.074
			20	0.171	0.158	0.174
			30	0.363	0.329	0.375
	50		0.824	0.716	0.827	
	0.50	5	5	0.148	0.121	0.151
			10	0.697	0.645	0.669
			20	0.994	0.983	0.994
			30	1.000	1.000	1.000
			50	1.000	1.000	1.000
		10	5	0.085	0.068	0.076
			10	0.249	0.217	0.236
			20	0.500	0.450	0.485
			30	0.930	0.882	0.923
50			1.000	1.000	1.000	
0.75	20	5	0.063	0.049	0.056	
		10	0.088	0.082	0.088	
		20	0.216	0.189	0.220	
		30	0.166	0.330	0.485	
		50	0.881	0.834	0.881	
	5	5	0.151	0.120	0.141	
		10	0.434	0.423	0.430	
		20	0.994	0.983	0.993	
		30	0.990	0.990	0.999	
		50	1.000	1.000	1.000	
0.25	10	5	0.077	0.059	0.067	
		10	0.173	0.160	0.179	
		20	0.550	0.490	0.540	
		30	0.880	0.830	0.879	
		50	0.999	0.999	0.999	
	20	5	0.060	0.047	0.054	
		10	0.078	0.070	0.074	
		20	0.181	0.163	0.181	
		30	0.343	0.311	0.340	
		50	0.824	0.765	0.823	

**Table 3.** Test Power for F, KW, and Permutation Tests with Standard Deviation Difference ( $\Delta = 1.5$ )

$\Delta$	$\pi$	$k$	$n$	F	KW	Permutation
1.0	0.25	5	5	0.485	0.409	0.480
			10	0.870	0.775	0.867
			20	1.000	1.000	1.000
			30	1.000	1.000	1.000
			50	1.000	1.000	1.000
		10	5	0.123	0.099	0.107
			10	0.278	0.254	0.275
			20	0.892	0.803	0.892
			30	0.999	0.997	0.999
			50	1.000	1.000	1.000
		20	5	0.063	0.049	0.061

1.5	0.50	20	10	0.105	0.095	0.102
			20	0.331	0.296	0.319
			30	0.631	0.585	0.627
			50	0.995	0.981	0.995
	0.50	5	5	0.248	0.212	0.220
			10	0.698	0.648	0.676
			20	1.000	1.000	1.000
			30	1.000	1.000	1.000
	0.50	10	5	0.106	0.083	0.099
			10	0.373	0.300	0.366
			20	0.898	0.852	0.900
			30	0.999	0.999	0.999
	0.50	20	5	0.068	0.052	0.063
			10	0.127	0.145	0.125
			20	0.423	0.366	0.425
			30	0.840	0.738	0.833
	0.75	5	5	0.151	0.121	0.160
			10	0.781	0.757	0.777
			20	1.000	1.000	1.000
			30	1.000	1.000	1.000
0.75	10	5	0.077	0.059	0.078	
		10	0.325	0.268	0.322	
		20	0.919	0.869	0.918	
		30	0.999	0.994	0.999	
0.75	20	5	0.065	0.050	0.061	
		10	0.105	0.093	0.105	
		20	0.351	0.298	0.351	
		30	0.652	0.606	0.651	
			50	0.985	0.966	0.985

**Table 4.** Test Power for F, KW, and Permutation Tests with Standard Deviation Difference ( $\Delta = 2$ )

$\Delta$	$\pi$	$k$	$n$	F	KW	Permutation
0.25	0.50	5	5	0.488	0.410	0.483
			10	0.959	0.873	0.954
			20	1.000	1.000	1.000
			30	1.000	1.000	1.000
			50	1.000	1.000	1.000
		10	5	0.128	0.103	0.123
			10	0.464	0.399	0.460
			20	0.985	0.954	0.983
			30	1.000	1.000	1.000
			50	1.000	1.000	1.000
	20	5	0.073	0.058	0.065	
		10	0.155	0.134	0.151	
		20	0.535	0.484	0.531	
		30	0.894	0.845	0.889	
		50	0.999	0.999	1.000	
	5	5	0.443	0.354	0.418	
		10	0.998	0.999	0.997	
		20	1.000	1.000	1.000	

2.0	0.50	10	30	1.000	1.000	1.000
			50	1.000	1.000	1.000
			5	0.179	0.145	0.167
			10	0.672	0.595	0.664
			20	0.992	0.980	0.992
		30	0.999	0.999	1.000	
		50	1.000	1.000	1.000	
		20	5	0.082	0.064	0.080
			10	0.198	0.179	0.194
			20	0.699	0.601	0.696
	30		0.980	0.937	0.980	
	50		1.000	1.000	1.000	
	0.75	5	5	0.520	0.424	0.499
			10	0.970	0.940	0.971
			20	1.000	1.000	1.000
			30	1.000	1.000	1.000
			50	1.000	1.000	1.000
		10	5	0.150	0.117	0.138
			10	0.527	0.403	0.520
			20	0.994	0.983	0.993
30			1.000	0.999	1.000	
50			1.000	1.000	1.000	
20	5	0.076	0.059	0.071		
	10	0.157	0.134	0.159		
	20	0.572	0.489	0.566		
	30	0.883	0.854	0.885		
	50	1.000	0.999	1.000		

#### 4. Discussion

In the analysis of sensory data, for every case, permutation and F tests maintained the type I error rate at 0.05 level [21,22]. The study by Routledge showed similar results in terms type I error rate [23]. The study showed that KW test was not able to maintain the type I error rate at 0.05 level. It can be said that in the analysis of sensory data, each of the three tests had low power of test values. It is considered that this situation stems from the abrupt distribution of the data. Hence it is seen that in the studies in which other kinds of continuous distribution other than normal distributions are applied, the power values take higher values in similar simulations. Still, when the differences between the group mean in terms of standard deviation increase, realized the power of test values increase.

Similar results to this can also be observed in the studies by Başpınar and Gürbüz, Koşkan and Gürbüz, Mendeş and Tekindal, and Weber [18,22,24,25]. It is also stated in many research that the resampling method should be favored in cases where the sample size is small and the distribution is not normal [18,26-30].

As is seen in Table 1, the type I error rate of the permutation and F tests holds its determined value. However, in small sample sizes, the KW test cannot maintain the determined type I error value.

Keskin and Mendeş compared variance analysis and some approach tests (Marascuilo, James's second degree and Alexander - Govern Tests) in terms of power values that occur empirically in the samplings from populations with exponential distribution (1.00) [31]. In their study, the number of groups was 4 and 5, and the number of observations both equal and different in each group was between 3 and 100. The differences among the group means were 0.5, 1.0, 1.5, and 2 standard deviations ( $\Delta$ ). The power values of the tests considered for each combination were obtained out of 100000 simulation tests. Their results showed that the power of tests changed regarding the number of observations, whether the number of observations within the groups was equal or not, and the difference among the group means. In a study reported by Vidoni which used 10 000 simulations, where the probability of occurrence was  $\pi=0.3, 0.5, \text{ and } 0.7$ , the sample size was 5 and 10, it was concluded that different models should be used

for discrete data when the sample size was 10 [32]. This agrees with our results. In order to have more reliable results in the evaluation of sensory data, it is essential to acquire the data with the help of an expert before the evaluation step and to work with more experts as much as possible. The problem remains that it is hard to find enough expert ratings. If this problem is eliminated, acquiring especially small-scale figures such as 5-point likert scales shows more reliable results. Another result is that for evaluating sensory data, resampling tests are alternatives for KW and F tests. Heller and Venkatraman, Ludbrook and Dudley, Good, Balasubramani et al., Pesarin and Salmaso, Koşkan and Gürbüz, Figueiredo reported that resampling approaches that are also named as distribution-free tests can be favored especially in small sample sizes instead of F test [18,26-30].

## 5. Conclusions

As a result of this study, it can be inferred that when comparing groups in visual analogue scale data, permutation test should be used. The permutation test is satisfactory in terms of both the type I error rate and the power of the test. Also, permutation test is a distribution-free test. Therefore, can be used without prerequisites.

In almost every combination ( $\pi$ ,  $\Delta$ ,  $k$ ,  $n$ ), permutation test had similar or superior type 1 error and power values than the F and KW tests. It was observed that compared to the 5 visual analogue scale, when the data is measured in 10 and 20 visual analogue scale, the power of test values decreased. In other words, if more than 5 visual analogue scale are made on the visual analogue scale data, the results would negatively be affected.

**Author Contributions:** Conceptualization, Ö.K.; methodology, Ö.K.; software, Ö.K.; validation, Ö.K., M.E.; formal analysis, Ö.K.; investigation, Ö.K., M.E.; resources, Ö.K.; data curation, Ö.K.; writing—original draft preparation, Ö.K.; writing—review and editing, M.E., Ö.K.; visualization, Ö.K.; supervision, Ö.K.; project administration, Ö.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- [1] D. L. Clason and T. J. Dormody, "Analyzing data measured by individual likert-type items," *J. Agric. Educ.*, vol. 35 no.4, pp. 31-35, 1994.
- [2] K. A. A. Small, "Discrete choice model for ordered alternatives," *Econometrica*, vol. 55 no. 2, pp. 409-424, 1987.
- [3] M. A. Garcia-Perez, "Are the steps on likert scales equidistant? Responses on visual analog scales allow estimating their distances," *Educ. Psychol. Meas.*, 1-32, 2023.
- [4] S. Sinharay, Discrete Probability Distributions, McGraw Group Pty, 2010.
- [5] O. Kavuncu, İstatistik Teorisi ve Teorik Dağılımlar, T.C. Ziraat Bank Publishing, 1995.
- [6] F. X. Lesage, S. Berjot, and F. Deschamps, "Clinical stress assessment using a visual analogue scale," *Occupational medicine*, vol. 62 no. 8, pp. 600-605, 2012.
- [7] M. Kanda, M. Matsuhashi, N. Sawamoto, T. Oga, T. Mima, T. Nagamine and H. Shibasaki, "Cortical potentials related to assessment of pain intensity with visual analogue scale (VAS)," *Clinical Neurophysiology*, vol. 113 no. 7, pp. 1013-1024, 2002.
- [8] A. Reich, M. Heisig, N. Q. Phan, K. Taneda, K. Takamori, S. Takeuchi and J. C. Szepietowski, "Visual analogue scale: evaluation of the instrument for the assessment of pruritus," *Acta Dermato Venereologica*, vol. 92 no. 5, pp. 497, 2012.
- [9] F. Dexter and D. H. Chestnut, "Analysis of statistical tests to compare visual analog scale measurements among groups," *Anesthesiology*, vol. 82, pp. 896-902, 1995.
- [10] M. Kliger, S. Stahl, M. Haddad, E. Suzan, R. Adler and E. Eisenberg, "Measuring the intensity of chronic pain: are the visual analogue scale and the verbal rating scale interchangeable?" *Pain Practice*, vol. 15, no. 6, pp. 538-547, 2015.
- [11] M. J. Hjerstad, P. M. Fayers, D. F. Haugen, A. Caraceni, G. W. Hanks, J. H. Loge, R. Fainsinger, N. Aass and S. Kaasa, "Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review," *Journal of Pain and Symptom Management*, vol. 41 no. 6, pp. 1073-1093, 2011.
- [12] H. Van Laerhoven, H. J. Van der Zaag-Loonen and B. H. Derkx, "A comparison of likert scale and visual analogue scales as response options in children's questionnaires," *Acta paediatrica*, vol. 93 no. 6, pp. 830-835, 2004.
- [13] J. Bielewicz, B. Daniluk, and P. Kamieniak, "VAS and NRS, same or different? Are visual analog scale values and numerical rating scale equally viable tools for assessing patients after microdiscectomy?" *Pain Res Manag*, (vol,page), 2022.
- [14] A. Butler, P. Rothery and D. Roy, "Minitab macros for resampling methods," *Teach. Stat.*, vol. 25, no. 1, pp. 22-25, 2003.
- [15] E. Stark and M. Abeles, "Applying resampling methods to neurophysiological data," *J Neurosci Methods*, vol. 145, pp. 133-14, 2005. doi: 10.1016/j.jneumeth.2004.12.005
- [16] D. S. Collingridge, "A primer on quantitized data analysis and permutation testing," *J Mix Methods Res*, vol. 7 no. 1, pp. 81-97, 2012.

- [17] A. Figueiredo, "Resampling methods in ANOVA for data from the von Mises-Fisher distribution," *Commun. Stat. Simul. Comput*, pp. 1-15, 2021.
- [18] Ö. Koşkan and F. Gürbüz, "Comparison of F test and resampling approach for type I error rate and test power by simulation method," *Journal of Agricultural Sciences*, vol. 15, no. 1, pp. 105-111, 2009.
- [19] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, 1987.
- [20] T. Kesici and Z. Kocabaş, *Biyoistatistik*, Ankara University Press, 1998.
- [21] D. J. K. Mewhort, "A comparison of the randomization test with the F test when error is skewed," *Behav. Res. Methods*, vol. 37 no.3, pp. 426-435, 2005.
- [22] M. Weber, *Robustness and Power of the t, Permutation T and Wilcoxon Tests*(Doctoral thesis, University of Wayne State University, Detroit, Michigan), 2006. Retrieved from <https://www.proquest.com/docview/304968464?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses>
- [23] R. D. Routledge, "P-values from permutation and F tests," *Comput Stat Data Anal*, vol. 24, pp. 379-386, 1997.
- [24] E. Başpınar, and F. Gürbüz, "The power of the test in the samples of various sample sizes were taken from the binary combinations of the normal, beta, gamma, and weibull distributions," *Journal of Agricultural Sciences*, vol. 6, no. 1, pp. 116-127, 2000.
- [25] M. Mendeş and B. Tekindal, "Determining the appropriate sample size in testing the difference between the means in normal and non-normal populations," *The Journal of the Industrial Arts Education Faculty of Gazi University*, vol. 111, pp. 25-38, 2002.
- [26] G. Heller and E. S. Venkatraman, "Resampling procedures to compare survival distributions in the presence of right-censored data," *Biometrics*, vol. 52, no. 4, pp. 1204-1213, 1996. doi: 10.2307/2532836.
- [27] J. Ludbrook and H. Dudley, "Why permutation tests are superior to t and F tests in biomedical research," *Am. Stat.*, vol. 52, no. 2, pp. 127-132, 1998. doi: 10.1080/00031305.1998.10480551.
- [28] P. Good, *Permutation Tests: Practical Guide to Resampling Methods for Testing Hypotheses*, Springer Verlag, New York, USA, 2000. doi: 10.1007/978-1-4757-2346-5.
- [29] G. K. Balasubramani, S. R. Wisniewski, H. Zhang, and H. F. Eng, "Development of an efficient SAS macro to perform permutation test for two independent samples," *Comput. Methods Programs Biomed.*, vol. 79, pp. 179-187, 2005. doi: 10.1016/j.cmpb.2005.03.010.
- [30] F. Pesarin and L. Salmaso, "Permutation test for univariate and multivariate ordered categorical data," *Austrian J. Stat.*, vol. 35, no. 2, pp. 315-324, 2006. doi: 10.17713/ajs.v35i2&3.378.
- [31] S. Keskin and M. Mendeş, "Comparison of the power of the test of one-way ANOVA method and some approximation tests for the samples drawn from the exponential distributed population," *J. Agric. Sci.*, vol. 8, no. 4, pp. 293-299, 2002.
- [32] P. Vidoni, "Improved prediction limits for continuous and discrete observations in generalised linear models," *Biometrika*, vol. 88, no. 3, pp. 881-887, 2001.

Research Article

Received: date:22.10.2024

Accepted: date:24.12.2024

Published: date:31.12.2024

# Makine Öğrenmesi Algoritmalarıyla Konut Fiyatlarının Tahmini: Giresun Örneği

Mehmet Akif KARA<sup>1\*</sup>

<sup>1</sup>Giresun University, Department of Business Administration, Giresun Türkiye; akifkara28@gmail.com

Orcid: 0000-0003-4308-9933<sup>1</sup>

\*Correspondence: akifkara28@gmail.com

**Özet:** İnşaat sektörü son 20 yılın en önemli sermaye birikim alanları arasında yer almaktadır. Konut satışları da bu alanın önemli bir parçasıdır. Konut fiyatlarının seyrinin tahmin edilmesi hem üretici hem de tüketiciler açısından önemlidir. Makine öğrenmesi algoritmaları tahmin çalışmalarında sıklıkla kullanılmakta ve iyi performans göstermektedir. Bu çalışmada Giresun örneği üzerinde 365 veri kullanılarak konut değerlendirme tahmini MATLAB 2023b yazılım programı kullanılarak çözümlenmektedir. Çalışmada RMSE, MAE,  $R^2$ , MSE metrik değerleri kullanılarak makine öğrenmesi algoritmaları tahmin başarılarına göre incelenmektedir. Çalışmada konut değerlemek için kullanılan "brüt m<sup>2</sup>, net m<sup>2</sup>, bina yaşı, bulunduğu kat, toplam kat, site içinde olup olmama, asansör, ısıtma, açık ve kapalı otopark," değişkenleri kullanılmıştır. Veriler Türkiye'de yaygın olarak kullanılan ve en çok ziyaret edilen bir e ticaret web sitesinden elde edilmiştir. Çalışma sonucunda Gaussian Process Regression algoritmasının daha iyi bir performans gösterdiği ortaya koyulmuştur. Bu algoritmanın bu veri seti üzerinde  $R^2$  değeri %89 olarak çok yüksek bulunmuştur.

**Keywords:** makine öğrenmesi, konut değerlendirme, tahmin.

## Prediction of Housing Prices Using Machine Learning Algorithms: The Case of Giresun

**Abstract:** The construction sector has been one of the most significant areas for capital accumulation over the past 20 years. Housing sales are also an important part of this sector. Predicting the trend of housing prices is crucial for both producers and consumers. Machine learning algorithms are frequently used in predictive studies and have shown good performance. This study focuses on housing valuation predictions using 365 data points from Giresun, analyzed with MATLAB 2023b software. The study examines the predictive success of machine learning algorithms using metrics such as RMSE, MAE,  $R^2$ , and MSE. The variables used for housing valuation include "gross m<sup>2</sup>, net m<sup>2</sup>, building age, floor location, total floors, whether it's within a site, presence of an elevator, heating, open and closed parking." The data was obtained from a popular e-commerce website widely used in Turkey. The results indicate that the Gaussian Process Regression algorithm performs better, achieving a high  $R^2$  value of 89% on this dataset.

**Keywords:** machine learning, residential property valuation, forecasting.

### 1. Giriş

Yapay Zeka'nın gelişimi ile birlikte geliştirilen birçok algoritma tahmin çalışmalarında da kullanılmaktadır. Makine Öğrenmesi algoritmaları bu bağlamda YZ (AI) yöntemi olarak tercih edilmektedir [1]. Makine öğrenmesi, giriş verileri ile çıkış verileri arasında doğrusal bir formül ile ifade edilemeyen problemleri



çözebilmek için geliştirilmiş olan güncel esnek hesaplama sistemlerine verilen genel bir adlandırmadır [2]. Makine öğrenmesi 3 alt sınıflandırma içermektedir. Bunlar; denetimli, denetimsiz ve yarım denetimli algoritmalarıdır. Denetimli öğrenme modelleri genel olarak sınıflandırma ve regresyon problemlerinde yoğun olarak kullanılırken, denetimsiz öğrenme modelleri ise kümeleme alanında daha çok uygulanmaktadır [3]. Makine öğrenmesi, bilgisayarların verilerden öğrenmesini ve belirli görevleri gerçekleştirmesini sağlamaktadır. Geleneksel programlama yöntemlerinin aksine, bilgisayara belirli kural ya da algoritma yerine, verilerle beslenerek kendi kendine öğrenme yeteneği kazandırılmaktadır.

Bu çalışmanın motivasyonu Türkiye başta olmak üzere gelişmekte olan ülkeler için önemli bir sermaye birikim alanı olan inşaat sektörünün dinamosu olan konut piyasalarının durumunu makine öğrenmesi algoritmalarıyla ortaya koymaktır. Klasik yöntemler genel itibariyle piyasa trendlerini ve tarihsel verileri göz önünde bulundurmaktadır. Makine öğrenmesi algoritmaları ise daha derinlemesine ve veri odaklı bir yaklaşım sunmaktadır. Bu bağlamda büyük veri setlerinden ilişkileri öğrenerek daha doğru tahmin yapma yeteneğine sahip olmaktadır. Ayrıca fiyat tahminlerinin doğruluğunu artırmak için farklı faktörleri de analiz edebilmektedir. Konut değerlendirme bir yanıyla son yılların en önemli sermaye birikim alanı olan inşaat sektörü bir yanıyla da barınma hakkı kavramsallaştırması bağlamında bireyleri doğrudan ilgilendiren bir konudur [4]. Barınma hakkı bağlamında tartışıldığında konutların fiyatlarının ve bu fiyatlara göre sağlıklı, insani ve sosyal olarak yaşanabilir durumda olmaları gerekmektedir [5]. Sermaye birikim açısından özellikle neoliberal politikaların kullanıldığı günümüzde Türkiye’de büyüme politikasının temel araçlarından birisi inşaat piyasasıdır [6]. Bu açıdan birbiriyle bağlantılı çok sayıda sektörü ilgilendirmektedir. İşletmeler, yöneticiler, karar alıcılar ve politika yapıcılar arasında da ilgi odağı haline gelmektedir. Dolayısı ile konut fiyatlarının takip edilmesi önemlilik arz etmektedir.

Çalışmanın amacı makine öğrenmesi algoritmalarıyla konut fiyat tahmini gerçekleştirmektedir. Algoritmaların belirli metrikler dâhilinde bu problem için nasıl performans gösterecekleri de ikincil amaçtır. Çalışmanın örneklemini Giresun ili Merkez ilçesi oluşturmaktadır. Giresun ili gibi yerel piyasalar çeşitli ekonomik, sosyal ve çevresel faktörlerin etkisi altında şekillenmektedir. Dolayısı ile Giresun ilindeki konut fiyatlarını etkileyen çeşitli unsurları belirlemek ve bu unsurları makine öğrenmesi algoritmaları kullanarak fiyat tahmininde kullanmak amaçlanmaktadır. Çalışma verilerinin Giresun ili Merkez ilçede 3+1 daireler ile kısıtlanması ise çalışmanın kısıtlılığını oluşturmaktadır. Çalışmada öncelikli olarak makine öğrenmesi algoritmaları ele alınmış, ardından literatür incelemesi ile alandaki boşluk ortaya koyulmuş ve uygulama gerçekleştirilmiştir. Bu sayede yalnızca yerel konut piyasasına ilişkin bilgiler elde edilmekle kalınmayacak aynı zamanda makine öğrenme algoritmalarının pratik uygulaması da ortaya konulacaktır. Çalışma sonucunda elde edilmesi muhtemel bulguların yerel yönetim ve yatırımcılar başta olmak üzere ilgili paydaşlar açısından stratejik kararlar alınırken yol göstereceği düşünülmektedir.

## 2. Makine Öğrenmesi

Günümüzde büyük verilerin giderek artış göstermesi ile birlikte bu verilerden yeni çıkarımlar elde edebilmek için makine öğrenmesi algoritmaları kullanılmaktadır. Bu alanda kullanılan algoritmalar, büyük veri kümelerindeki ilişki ve kalıpları tespit edebilmek için eğitilmektedir. Böylece en iyi karar ve tahminleri yapabilmektedir. Makine öğrenmesi uygulamaları, sahip oldukları veri miktarı ile orantılı olarak gelişmekte ve daha fazla veri ile daha iyi sonuçlar üretmeleri sağlanmaktadır. Alana ilişkin literatür incelendiğinde; makine öğrenmesinin üç ana kategoride ele alındığı görülmektedir [7],[8]. Bunlar; denetimli, denetimsiz ve yarı denetimli öğrenme olarak adlandırılmaktadır.

*Denetimli öğrenme*, makine öğreniminde en yaygın ve popüler yöntemlerden biridir. Bu yöntemde, her eğitim verisi bir çıktı etiketi üretmektedir. Sistem, eğitim süreci aracılığıyla sorunları öğrenir ve belirli bir hedef değişkenin tahmin edilmesi amacıyla etiketli veriler kullanarak model oluşturur. Bu süreç, girdiler (özellikler) ile çıktılar (hedef değişken) arasındaki ilişkileri anlamak için çeşitli algoritmalar kullanır. Denetimli öğrenme, genellikle sınıflandırma ve regresyon problemleri için tercih edilir [9].

*Denetimsiz makine öğrenmesi* ise, etiketlenmemiş veriler üzerinde içgörü ve ilişkilerin keşfedilmesine odaklanır. Bu yöntemde eğitim sırasında etiketlenmemiş veriler kullanılır ve sistemin, bu verilerden kalıpları keşfetmesi beklenir. Sonuçlar belirsizdir ve sistemden istenen çıktılar önceden bilinmemektedir. Dolayısıyla, bu model eğitim sırasında doğru cevaplarla yönlendirilmemiştir [10].

*Yarı denetimli makine öğrenmesi* ise, denetimli ve denetimsiz öğrenmenin bir kombinasyonunu temsil eder. Bu yöntemde bazı veriler etiketlenirken, bazıları etiketlenmez. Amaç, etiketlenmemiş küçük bir veri kümesinin etiketli veri ile bir arada kullanılmasıyla öğrenmede belirgin bir iyileşme sağlamaktır. Bu yaklaşım, her iki yöntemin avantajlarını birleştirerek daha etkili sonuçlar elde etmeyi hedefler [3].

Bu çalışmada denetimli makine öğrenmesi algoritmaları kullanılmıştır. Bu algoritmalar;

*Destek Vektör Makinaları (SVM)*: SVM genellikle sınıflandırma problemlerinin çözümünde kullanılan gözetimli öğrenme yöntemlerinden birisidir. Bu yöntemde, bir düzlem üzerine yerleştirilmiş noktaları ayırmak için bir doğru çizer. Bu doğrunun, iki sınıfının noktaları için de maksimum uzaklıkta olmasını amaçlar. Karmaşık ama küçük ve orta ölçekteki veri setleri için uygundur. SVM, el yazısı tanıma, çevre bilimleri, genelleştirilmiş tahmine dayalı kontrol uygulamaları, görüntü sınıflandırma vb. konularda uygulanmaktadır. SVM, genellikle sınıflandırma görevleri için kullanılsa da, regresyon problemlerinde de uygulanabilir. Destek Vektör Regresyonu (SVR), verileri en iyi ayıran hiper düzlemi bulmayı amaçlar ve bu süreçte hata marjını minimize etmeye odaklanır [10].

*Rastgele Orman (Random Forest)*: Rastgele ormanlar veya rastgele karar ormanları, eğitim zamanında çok sayıda karar ağacı oluşturularak ve sınıfların modu (sınıflandırma) veya ortalama/ortalama tahmini olan sınıfın çıktısını alarak çalışan sınıflandırma, regresyon ve diğer görevler için bir topluluk öğrenme yöntemidir. Rastgele ormanlar genellikle karar ağaçlarından daha iyi performans gösterir, ancak doğrulukları gradyan destekli ağaçlardan daha düşüktür [11].

*Doğrusal Regresyon (Linear Regression)*: Doğrusal regresyon, titizlikle çalışılan ve pratik uygulamalarda yaygın olarak kullanılan ilk tip regresyon analizidir. Bunun nedeni, bilinmeyen parametrelerine doğrusal olarak bağlı modellerin, parametreleriyle doğrusal olmayan ilişkili modellere göre daha kolay uydurulması ve elde edilen tahmin edicilerin istatistiksel özelliklerinin belirlenmesinin daha kolay olmasıdır [12].

*Karar Ağaçları (Decision Trees)*: Karar ağaçları, regresyon problemlerinde sürekli hedef değişkenleri tahmin etmek için kullanılır. Veriyi ağaç yapısına bölerek, her düğümde belirli kriterlere dayalı olarak kararlar alır. Bu yaklaşım, veri setindeki karmaşıklıkları yakalamada özellikle etkilidir [13].

*Gaussian Process Regression*: Bu istatistiksel model, verilerin sürekli bir dağılımını modellemek için kullanılır. Tahminlerindeki belirsizlikleri dikkate alarak sonuçları bir dağılım şeklinde sunar. Tahminler için güven aralıkları belirlemede özellikle etkilidir [14].

*Ensemble Bagged Tree*: Bu yöntem, birden fazla karar ağacını bir araya getirerek tahmin performansını artırmaya dayanır. Ağaçlar, verinin farklı alt kümeleri üzerinde eğitilir; bu da genel model performansını artırır ve aşırı uyum riskini azaltır [15].

Bu algoritmaların her biri, veri setinin özelliklerine ve ele alınan probleme bağlı olarak farklı avantajlar sunar. Uygun algoritmanın seçimi, genellikle bu faktörlere bağlı olarak değişir ve en iyi sonuçları veren algoritmayı belirlemek için deneme-yanılma süreci ile karşılaştırma gerektirir.

### 3. Literatür İncelemesi

Literatürde konut fiyatlarının tahmini üzerine özellikle ekonometrik yöntemlerle yapılmış çalışmalar fazlalık göstermektedir. Buna karşı makine öğrenmesi algoritmalarıyla yapılan çalışmalar sınırlı sayıdadır. Literatür incelemesinde Web of Science ve TRDizin veri tabanlarına bakılmıştır. Hem bu veri tabanlarında etki değeri yüksek dergilerin yayımlanması ve ilgili çalışmaların atıf sayılarının yüksek olması tercih sebebidir.

Burhan, çalışmasında bağımlı değişken olarak konut fiyatını, bağımsız değişkenler olarak net m<sup>2</sup>, bina yaşı, bulunduğu kat, site içinde olup olmaması gibi faktörleri almıştır. Çalışmada Naive Bayes yöntemi dışındaki k-En yakın komşu, karar ağaçları ve rastgele orman algoritmalarının doğruluk oranlarının %60 üzerinde olduğu ortaya koyulmuştur. En başarılı tahmin yönteminin ise Rastgele Orman algoritması olduğu görülmüştür [16]. Oral vd. çalışmalarında 11 kriter kullanarak konut fiyatı tahmini için makine öğrenmesi yöntemlerini kullanmıştır. Çalışmada model değerlendirmesi için RMSE, MSE ve  $R^2$  değerleri kullanılmıştır. Çalışma sonucunda en iyi performans değerine sahip yöntem Bagged Tree Ensemble olmuştur [17]. Doğan vd. çalışmalarında Ankara Keçiören ilçesi örnekleme dahilinde 11 kriter içeren 20 farklı yapay sinir ağları modelini kullanarak fiyat tahmininde bulunmuştur. Çalışma sonucunda YSA yaklaşımının fiyat tahmininde başarılı sonuç verdiği görülmektedir [18]. Kayakuş vd. çalışmalarında makine öğrenmesi algoritmalarından Rastgele Orman, Yapay Sinir Ağları ve DVM yöntemlerini kullanmıştır. 9 değişken ile Ocak 2013-Kasım 2020 arası aylık verileri içeren modeller çerçevesinde konut metrekare fiyatlarını tahmin etmiştir. Üç yöntemde başarılar

sonuçlar vermiştir [19]. Özdemir vd. Sakarya ilinde yer alan 4 ilçe için DÖ, Polinomsal Regresyon ve Rastgele Orman algoritmalarını kullandığı çalışmalarında 7 değişken içeren modeller çerçevesinde konut fiyat tahmininde bulunmuştur. Çalışma sonucunda en başarılı tahmini DÖ yöntemi ile elde etmişlerdir [20]. Garcia vd. Bu araştırma, konut fiyatlarını tahmin etmek için en iyi makine öğrenmesi algoritmalarını belirlemeyi ve COVID-19 pandemisinin bir İspanyol şehrindeki konut fiyatları üzerindeki etkisini nicel olarak ölçmeyi amaçlamaktadır. Metodoloji, veri hazırlama, özellik mühendisliği, hiperparametre eğitimi ve optimizasyonu, model değerlendirme ve seçimi ile nihayetinde model yorumlama aşamalarını içermektedir. Boosting tabanlı topluluk öğrenme algoritmaları (Gradient Boosting Regressor, Extreme Gradient Boosting ve Light Gradient Boosting Machine) ile bagging tabanlı algoritmalar (rastgele orman ve extra-trees regressor) kullanılarak bir doğrusal regresyon modeli ile karşılaştırılmaktadır. Araştırma, COVID-19 pandemisi ilan edilmeden önce ve sonra Alicante (İspanya) emlak pazarına ait coğrafi referanslı mikro verilerle birlikte, kadastro, sosyo-demografik ve ekonomik göstergeler ve uydu görüntüleri gibi diğer tamamlayıcı kaynaklardan elde edilen bilgilerle bir vaka çalışması geliştirmektedir. Sonuçlar, makine öğrenmesi algoritmalarının, emlak pazarı verileri gibi karmaşık verilerin doğrusal olmayan özelliklerine daha iyi uyum sağladıkları için geleneksel doğrusal modellerden daha iyi performans gösterdiğini ortaya koymaktadır. Bagging tabanlı algoritmalar (rastgele orman ve extra-trees regressor) aşırı öğrenme sorunları gösterirken, boosting tabanlı algoritmalar daha iyi performans ve daha düşük aşırı öğrenme göstermektedir [21]. Soltani vd. çalışmalarında çeşitli değişkenlerin konut fiyatı varyasyonları üzerindeki etkilerini keşfetmek için makine öğrenmesi algoritmalarını kullanmıştır. Avustralya'nın Metropolitan Adelaide şehrine ait 32 yıllık konut fiyatı veri seti ile çalışmada 428000 satış işlem kaydı ve 38 açıklayıcı değişken kullanmıştır. Gradient-Boosting ve Rastgele Orman gibi topluluk makine öğrenmesi tekniklerinin daha iyi sonuç verdiğini ortaya koymuştur. Modellerin tahmin başarılarını artırmak için mekansal-zamansal gecikme değişkeni eklenmiş ve başarılı sonuç vermiştir. Çalışma sonucunda bu tür değişkenlerin Makine öğrenmesi uygulamalarında yararlı bir moderatör olacağı vurgulanmıştır [22]. Park ve Bae çalışmalarında ABD'de bulunan Virginia Fairfax şehrindeki 5359 apartman dairesinin konu verilerini analiz etmiştir. Yazarlar, makine öğrenmesi algoritmalarına dayalı bir konut tahmin modeli geliştirmektedir. RIPPER algoritmasının konut fiyat tahmininde diğer modellere göre daha başarılı performans gösterdiğini ortaya koymuştur [23]. Truong vd. Çalışmalarında konut fiyat tahmini için farklı modelleri incelemektedir. Rastgele Orman, XGBoost ve Light GBM gibi üç farklı makine öğrenmesi algoritması ile Hibrit Regresyon ve Yığın Genelizasyon Regresyonu (Stacked Generalization Regression) gibi iki tekniği karşılaştırmıştır. Çalışma sonucunda yazarlar, kullanılan yöntemlerden arzu edilen sonuçlar elde ettiğini vurgulamıştır. Ancak her modelin kendi avantaj ve dezavantajlarını ortaya koymuşlardır. Özellikle zaman karmaşıklığı vurgusu yapılmaktadır [24]. Akay vd. Çalışmalarında ARIMA modeli ile Rassal Orman ve Hibrit Rassal Orman yöntemlerini kullanarak Türkiye Konut Fiyat Endeksi serisi için öngörü performansları karşılaştırılmıştır. Çalışma bulgularına göre hibrit modelin daha iyi performans gösterdiği görülmektedir [25].

#### 4. Materyal ve Yöntem

Çalışmada kullanılan veri seti, Türkiye'de en fazla ziyaret edilen bir satış web sitesi/uygulamasından elde edilmiştir. Veri toplama süreci 13.10.2024-16.10.2024 tarihlerini kapsamaktadır. Bu tarih aralığında web sitesinde bulunan ilanlar değerlendirmeye dâhil edilmiştir. Çalışmada örneklem olarak Giresun Merkez ilçesi seçilmiştir. Giresun'un Doğu Karadeniz Bölgesi'nde en fazla göç veren illerden birisi olması, konut fiyatlarının diğer bölge illerine göre yüksek seviyede olması tercih nedenidir. Aynı zamanda çalışmada 3+1 daireler tercih edilmiş ve sınırlandırma yapılmıştır. Bu bağlamda Giresun İli Merkez ilçesinde 440 ilana ulaşılmıştır. Daha sonra veriler ön işlem sürecinden geçirilmiştir. Farklı ilanlarda farklı emlak şirketleri tarafından paylaşıldığı tespit edilmiş ve veri setinde sadece bir tanesi kullanılmıştır. Analizde kullanılan veri sayısı 365 olarak tamamlanmıştır. Çalışmada kullanılan ve bağımsız değişken olarak ele alınan kriterler literatürde yapılan benzer çalışmalardan elde edilmiştir. Bu çalışmalardan farklı olarak Giresun ili özelinde yaşanan "araç parkı sorunu" bağlamında açık otopark, kapalı otopark gibi değişkenler de kriter olarak alınmış ve toplam kriter sayısı 11 olarak belirlenmiştir. Bağımlı değişken olarak konut fiyatları, bağımsız değişkenler olarak ise brüt m<sup>2</sup>, net m<sup>2</sup>, bina yaşı, bulunduğu kat, toplam kat sayısı, site içinde olup olmama, ısıtma, asansör, açık otopark, kapalı otopark olup olmaması olarak belirlenmiştir. Çalışmada MATLAB2023b yazılım programında yer alan Regression Learner Toolbox aracı kullanılmıştır. Öncelikle; veriler ön işleminden geçirilmiştir. Kategorik veriler

Tablo 2’de gösterildiği biçimde kodlanmış ve programa uygun hale getirilmiştir. Regression Learner Toolbox kendi içerisinde normalizasyon işlemini yaptığı için ayrıca normalizasyona ihtiyaç duyulmamıştır.

#### 4.1. Model Performansını Değerlendirme

Makine öğrenmesi çalışmalarında model performansları çeşitli metriklerle bakılarak değerlendirilmektedir. Bu metrikler arasında en sık kullanılanları, açıklamaları ve matematiksel eşitlikleri Tablo 1’de gösterilmektedir.

**Tablo 1.** Çalışmada kullanılan metrikler ve açıklamaları

Metrik	Açıklama	Eşitlik
$R^2$	Modeldeki bağımsız değişkenlere göre bağımlı değişkenin varyasyon oranını ölçer. Regresyon modelinin çok fazla bağımsız değişkeni varsa test verilerinde istenilen başarıyı göstermez. Bu durumlarda Düzeltilmiş $R^2$ kullanılır.	$1 - \frac{\sum_{t=1}^{ntest} (x_t - \hat{x}_t)^2}{\sum_{t=1}^{ntest} (x_t)^2}$
MSE	Gerçek değerler ile tahmin edilen değerler arasındaki hataların karesinin ortalaması ile tanımlanır	$\frac{1}{ntest} \sum_{t=1}^{ntest} (x_t - \hat{x}_t)^2$
RMSE	MSE’nin kareköküdür.	$\sqrt{\frac{1}{ntest} \sum_{t=1}^{ntest} (x_t - \hat{x}_t)^2}$
MAE	Gerçek değerler ile tahmin edilen değerler arasındaki hataların mutlak değerlerinin ortalaması ile tanımlanır.	$\sum_{i=1}^{ntest} \frac{ x_t - \hat{x}_t }{ntest}$

$R^2$  için 1’e en yakın değeri alması beklenirken MSE, RMSE ve MAE için en düşük değeri veren tahmin başarılı sayılacaktır.

#### 5. Bulgular

Çalışmada kullanılan veri setinin ilk 10 değeri Tablo 2’de gösterilmektedir. Tablo 2’de yer alan değerler MATLAB programına uygun hale getirilmiştir. Değişkenler arasında yer alan “bulunduğu kat”, “bina yaşı”, “site içinde bulunup bulunmama”, “ısıtma”, “asansör”, “açık otopark” ve “kapalı otopark” değişkenlerine sayısal karşılık atanmıştır ve analize uygun hale getirilmiştir.

**Tablo 2.** Veri seti (ilk 10 değer)

Brüt m2	Net m2	Bulunduğu Kat	Toplam Kat Sayısı	Bina Yaşı	Site İçinde	Isıtma	Asansör	Açık Otopark	Fiyat	Kapalı Otopark
170	160	bahçe katı	7	2	evet	dgaz	var	var	5950000	yok
165	150	3	7	0	hayır	dgaz	var	var	3350000	yok
140	135	5	9	1	hayır	dgaz	var	var	3850000	yok
140	110	1	10	0	hayır	dgaz	var	var	2900000	var
158	150	2	10	0	evet	dgaz	var	var	3500000	yok
145	128	3	3	1	evet	dgaz	var	var	3975000	yok
165	156	7	7	5-10 arası	evet	merkezi	var	var	3500000	yok
130	125	2	5	5-10 arası	evet	dgaz	var	var	2550000	yok
160	130	7	10	5-10 arası	evet	merkezi	var	var	4000000	var
135	125	9	10	5-10 arası	hayır	dgaz	var	var	4000000	yok

İlk olarak veri seti MATLAB yazılım programında analize uygun hale getirilmiştir. Örneğin; “site içinde olma” kriterine 1 atanırken “site içinde olmama kriterine” 0 değerleri atanmıştır. Daha sonra program içerisinde bulunan Regression Learner Toolbox ile analiz gerçekleştirilmiştir.

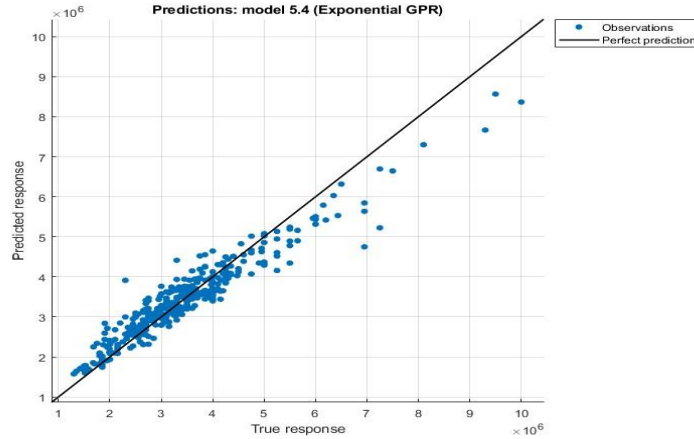
Regression Learner Toolbox ile yapılan analiz sonucunda algoritmaların RMSE,  $R^2$ , MSE ve MAE sonuçları

**Tablo 3.** Algoritmaların sonuçları

Methods	RMSE	$R^2$	MSE	MAE
Linear Regression	8.57	0.54	1.73	6.08
Regression Tree	6.68	0.72	4.46	4.81
SVM	7.52	0.65	5.66	3.27
Ensemble Bagged Tree (RO)	7.39	0.66	5.47	5.06
<b>Gaussian Process R.</b>	<b>4.17</b>	<b>0.89</b>	<b>1.73</b>	<b>2.98</b>
Neural Network	7.45	0.65	5.55	5.21

Tablo 3’de gösterildiği üzere en iyi performansı Gaussian Process Regression algoritması göstermiştir. Bu algoritmanın RMSE değeri diğer yöntemlere göre daha düşük olduğu için en iyi performansı gösterdiği ortaya koyulmaktadır. Bu yöntemin yanı sıra Regression Tree yöntemi de en iyi ikinci performansı göstermiştir.  $R^2$  değerleri incelendiğinde ise Linear Regression algoritması hariç tüm algoritmalarının modelin %60 ından fazlasını açıkladığı görülmektedir.  $R^2$  değeri modelin açıklama gücünü ifade etmektedir. Bu bağlamda GPR yönteminin modeli çok iyi açıkladığı görülmektedir.

Grafik 1’de GPR yönteminin tahmin grafiği gösterilmektedir.



**Grafik 1.** GPR ile elde edilen tahmin grafiği

Grafik 1 incelendiğinde Giresun ilinde konut fiyatlarının belirli bir nokta aralığında yoğunlaştığı görülmektedir.

## 6. Sonuç ve Tartışma

Bu çalışmada Giresun ili Merkez ilçesinde bir satış sitesinde ilana koyulan 3+1 daireler için konut fiyat tahmininde bulunulmuştur. Makine öğrenmesi algoritmalarının tahmin performansları karşılaştırılmıştır. Çalışmada MATLAB yazılım programı içerisinde bulunan Regression Learner Toolbox aracı kullanılmıştır. Yapılan analiz sonucunda en iyi performansı gösteren yöntem Gaussian Process Regression algoritması olmuştur. Yöntemin MAE, MSE ve RMSE değerleri sırasıyla 0.29, 0.17, 0.41,  $R^2$  değeri ise 0.89 olarak bulunmuştur. Bu yöntemin yanı sıra Neural Network, Support Vector Machine, Regression Trees gibi yöntemlerin de modeli açıklama dereceleri yüksek olarak bulunmuştur. Tüm bu sonuçlar makine öğrenmesi algoritmalarının konut değerlendirme çalışmalarında kullanılabileceğini ve iyi performans gösterdiklerini ortaya koymaktadır.

Konut fiyatlarının tahmin edilmesi birçok açıdan önemlidir. Benzer çalışmalar yapmayı planlayan muhtemel araştırmacıların;

- Farklı veri setleri ve yöntemler kullanarak tahmin performanslarını karşılaştırılması önerilmektedir.
- Aynı veri seti üzerinden çalışacak araştırmacıların ise kriter sayılarını (değişken) artırması ya da azaltması ile tahmin performansı gerçekleştirmeleri önerilebilir.

**Yazar Katkıları:** Çalışma tek yazarlıdır.

**Finansman:** "Bu araştırma dışarıdan fon almadı"

**Çıkar çatışmaları:** "Yazar çıkar çatışması beyan etmemektedir"

## Kaynaklar

- [1] E.Egrioglu, U.Yolcu ve E.Baş, Yapay Sinir Ağları, Ankara: Nobel, 2019.
- [2] M.Atalay ve E.Çelik, "Büyük veri analizinde yapay zeka ve makine öğrenmesi uygulamaları", Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Dergisi, C.9, s.22, ss.255-172, 2017.
- [3] M.Mohri, A.Rostamzadeh ve A.Talwalkar, Foundations of Machine Learning, USA: The MIT Press, 2012.

- [4] B.Caliskan, H.A.Özkan Yazar ve A.Keskin, "Understanding sociodemographic characteristics shaping the choice of large homes: lessons from Turkish metropolises", *International Journal of Housing Markets and Analysis*, <https://doi.org/10.1108/IJHMA-07-2024-0090>, 2024.
- [5] R. Rolnik, "Place, inhabitation and citizenship: the right to housing and the right to the city in the comtemporary urban world", *International Journal of Housing Policy*, c.14, s.13, ss.293-300, 2014.
- [6] S. Purkis, "İstanbul'da inşaat odaklı birikimin durdurulamayan yükselişi: konut fazlasına karşın artan konut açığı", *Mülkiye Dergisi*, c. 40, sy. 4, ss. 91-112, 2016.
- [7] O.A. Taiwo, "Types of machine learning algorithms", *New Advances in Machine Learning*, Y.Zhang (Ed.), InTech, University of Portsmouth United Kingdom, pp.3-31, 2010.
- [8] F.Y. Osisanwo et al., "Supervised machine learning algorithms: classification and comparison", *International Journal of Computer Trends and Technology (IJCTT)*, v.48, n.3, pp.128-138, 2017.
- [9] B.Mahesh, "Machine learning algorithms-a reiew", *International Journal of Science and Research*, v.9,n.1, pp.381-386, 2020.
- [10] D.A. Pisner and D.M. Schneyer, "Support vector machine", in *Machine Learning*, Academic Press, pp.101-121, 2020.
- [11] Ö. Akar and O. Güngör, "Rastgele Orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması", *hkmojd*, no. 106, pp. 139-146, December 2012, doi: 10.9733/jgg.241212.1t.
- [12] E.Namlı, R.Ünlü and E.Gül, "Fiyat tahminlemede makine öğrenmesi teknikleri ve doğrusal regresyon yöntemlerinin kıyaslanması: Türkiye'de satılan ikinci el araç fiyatlarının tahminlenmesine yönelik bir vaka çalışması", *KONJES*, vol.7, n.4, pp.806-821, 2019.
- [13] W.Y. Loh, "Classification and regression trees", *Wiley Interdisciplinary Reviews: Data mining and knowledge discovery*, v.1,n.1, pp.14-23, 2011.
- [14] M.Seeger, "Gaussian processes for machine learning", *International Journal of Neural Systems*, v.14,n.02, pp.69-106, 2004.
- [15] T.G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision tress: Bagging, boosting, and randomization", *Machine Learning*, 40, pp.139-157, 2000.
- [16] B.H. Arda, "Konut fiyatları tahmininde makine öğrenmesi sınıflandırma algoritmalarının kullanılması: Kütahya kent merkezi örneği", *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, s.76, ss.221-237, 2023.
- [17] M.Oral, E.Okatan, İ.Kırbaş, "Makine öğrenme yöntemleri kullanarak konut fiyat tahmini üzerine bir çalışma: Madrid örneği", 3 rd *International Young Researchers Student Congress*, pp. 263-272, 2021.
- [18] O.Doğan et al., "Konut değerlendirme tahmininde yapay sinir ağları ve çok regresyon analizi yöntemlerinin kıyaslanması: Yenimahalle/Ankara örneği", *Türkiye Arazi Yönetimi Dergisi*, v.6, n. 1, pp.21-31, 2024.
- [19] M.Kayakuş, M.Terzioğlu, F.Yetiş, "Forecasting housing prices in Turkey by machine learning methods", *Aestim*, 80, pp.33-44, 2022.
- [20] M.Özdemir, K.Yıldız ve B.Büyüktanır, "Housing price estimation with deep learning: A case study of sakarya Turkey", *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, c.9,s.1,ss.138-151, 2022.
- [21] R.T. Mora-Garcia, M.F. Cespedes-Lopez, V.R. Perez-Sanchez, "Housing price prediction using machine learning algorithms in COVID-19 times", *Land*, v.11, n.11, pp.1-32, 2022.
- [22] A. Soltani, M.Heydari, F.Aghaei, C.J.Pettit, "Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms", *Cities*, n.131, 2022.
- [23] B. Park and J.K. Bae, "Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data". *Expert systems with applications*, v.42, n. 6, pp.2928-2934, 2015.
- [24] Q.Truong, M.Nguyen, H.Dang, B.Mei, "Housing price prediction via improved machine learning techniques". *Procedia Computer Science*, 174, pp. 433-442, 2020.
- [25] E.Ç. Akay et al., "Türkiye konut fiyat endeksi öngörüsü: ARIMA, rassal orman ve arima-rassal orman". *PressAcademia Procedia*, c.10, s.1, ss.7-11, 2019.

Research Article

Received: date:22.09.2024  
Accepted: date:24.11.2024  
Published: date:31.12.2024

# Mühendislik Tasarım Becerileri Algı Testinin Geliştirilmesi (MüTasBAT)

Suat Türkoğuz<sup>1</sup>, Kübranur Sarı<sup>2\*</sup>

<sup>1</sup>Dokuz Eylül Üniversitesi, Buca Eğitim Fakültesi, İzmir Türkiye; suat.turkoguz@gmail.com  
<sup>2</sup>Dokuz Eylül Üniversitesi, Eğitim Bilimleri Enstitüsü, İzmir Türkiye 2; kubranursarii@gmail.com  
Orcid: 0000-0002-7850-2305<sup>1</sup> Orcid: 0000-0003-0372-033X<sup>2</sup>

\*Correspondence: kubranursarii@gmail.com

**Öz:** Bu çalışmada, ortaokul öğrencilerinin mühendislik tasarım süreçlerine yönelik becerilerini değerlendirebilmek amacıyla "Mühendislik Tasarım Becerileri Algı Testi" (MüTasBAT) geliştirilmiştir. Araştırma, tarama modeli çerçevesinde gerçekleştirilmiş ve İzmir'deki Bilim ve Sanat Eğitim Merkezlerinde (Bilsem) eğitim gören 7. sınıf öğrencileri üzerinde yürütülmüştür. Ölçeğin geliştirilme süreci, kapsamlı bir literatür taraması ve uzman görüşleri ile desteklenmiştir. İlk olarak, mühendislik tasarım becerilerine ilişkin araştırmalar incelenmiş ve teorik modeller değerlendirilmiştir. Ardından, Moazzen, Miller, Wild, Jackson ve Hadwin (2014) üniversite öğrencilerine yönelik "Engineering Design Survey" ölçeğinin tasarım-beceri maddeleri, ortaokul 7. sınıf öğrencilerine uygun hale getirilerek "Mühendislik Tasarım Becerileri Algı Testi (MüTasBAT)" olarak yeniden geliştirilmiştir. Geliştirilen test için alan uzmanlarının görüşlerine başvurularak testin kapsam geçerliliği sağlanmıştır. Yapı geçerliliğini test etmek amacıyla faktör analizi uygulanmış ve bu analiz sonucunda testin dört faktörlü bir yapıya sahip olduğu tespit edilmiştir. Bu faktörler sırasıyla "Kuluçka Tasarım Geliştirme Becerileri", "Yenilikçi Tasarım Geliştirme Becerileri", "Tasarım Değerlendirme Becerileri" ve "Tasarımı Sunma ve Raporlama Becerileri" olarak adlandırılmıştır. Ölçeğin güvenilirlik analizi, Cronbach alfa güvenilirlik katsayısı kullanılarak yapılmış ve testin genelinde yüksek bir iç tutarlılık elde edilmiştir. Bu da testin güvenilir bir ölçme aracı olduğunu göstermektedir. Bu çalışmanın mühendislik tasarım geliştirmelerinin değerlendirilmesine yönelik literatüre önemli bir katkı sağlaması beklenilmektedir.

**Anahtar kelimeler:** beceri, mühendislik tasarım, test geliştirme

## Development of Engineering Design Skills Perception Test (EnDeSPT)

**Abstract:** This study aimed to develop the "Engineering Design Skills Perception Test" (EnDeSPT) that evaluating by the engineering design process skills of middle school students. The study was carried out within the framework of the survey research and was conducted on 7th grade students attending at Science and Art Education Centers in Izmir. The development process of the test was supported by a comprehensive literature review and expert opinions. First, studies related to engineering design skills were examined, and theoretical models were evaluated. Subsequently, the design-skill items of the 'Engineering Design Survey' scale developed by Moazzen, Miller, Wild, Jackson, and Hadwin (2014) for university students were adapted for 7th grade secondary school students and re-developed as the 'Engineering Design Skills Perception Test' (EnDeSPT). To ensure the content validity of the developed test, opinions from field experts were consulted. To test the structural validity, factor analysis was conducted, and as a result of this analysis, it was determined that the test has a four-factor structure. These factors were named "Incubation Design Development Skills", "Innovative Design Development Skills", "Design Evaluation Skills" and "Design Presentation and Reporting Skills" respectively. The reliability analysis of the test was conducted using the Cronbach's alpha reliability coefficient, and high internal consistency was achieved throughout the test. This indicates that the test is a reliable measurement tool. It is expected that this study will make a significant contribution to the literature on the evaluation of engineering design developments.

**Keywords:** skill, engineering design, test development

## 1. Giriş

21. yüzyılda bilginin hızla ilerlemesi ile birlikte küreselleşen dünyaya uyum sağlamak giderek daha önemli hale gelmektedir. Bu nedenle ülkelerin, bilgi ve teknoloji üretme kapasitesine sahip, toplumsal olayları analiz edebilen ve günümüzün değişen koşullarına uyum sağlayabilen bireylere olan ihtiyacı gittikçe artmaktadır. Özellikle toplumda ve iş dünyasında etkili olabilmek için üretken bireyler yetiştirmek hedeflenmektedir. Dolayısıyla birçok ülke son yıllarda özellikle fen ve matematik eğitiminde yenilikçi yaklaşımlar ve yöntemler geliştirmeye yönelmektedir [1]. Kelley ve Knowles (2016), yenilikçi yaklaşımların mühendislik tasarımı eğitimine entegrasyonunun önemini vurgularken, fen ve matematik eğitimindeki bu ilerlemelerin mühendislik ile bütünleşmesinin öğrenci başarısını artırdığını belirtmektedir [2]. Bu noktada yenilikçi yaklaşımların mühendislik tasarımı eğitiminin de önemli bir parçası haline geldiği söylenebilir. Böylece yenilikçi yaklaşımlar yoluyla mühendislik tasarımı eğitiminin kalitesi artırılarak, öğrencilerin modern teknolojilere ve uygulamalara uyum sağlamaları desteklenebilmektedir.

Son yıllarda eğitim sistemlerinde yapılan yeniliklerle mühendislik tasarım becerilerinin önemi giderek artmaktadır. Özellikle STEM (Science, Technology, Engineering, and Mathematics) eğitimine yapılan vurguyla birlikte, bu beceriler, öğrencilere yalnızca teorik bilgi vermekle kalmayıp aynı zamanda onların problem çözme, yaratıcı düşünme gibi yeteneklerini de geliştirmeyi hedeflemektedir [3]. Mühendislik tasarım becerileri ile STEM eğitimi arasındaki ilişki, eğitimde bütünsel ve uygulamalı bir yaklaşımın temelini oluşturmaktadır. Temel eğitimde mühendislik eğitiminin ana bileşenleri olarak kabul edilen mühendislik tasarımı ve mühendislik düşüncesi, STEM disiplinleri arasında önemli bağlantılar oluşturarak yalnızca mühendislik ile sınırlı kalmamaktadır ([4, 5, 6]). Mühendislik tasarım süreci, STEM disiplinlerinin bütünleşmiş bir şekilde kullanıldığı bir çerçevedir. Bu süreçte, bilimsel bilgi ve teknoloji mühendislik yöntemleri aracılığıyla gerçek dünya problemlerini çözmek için bir araya getirilir. Nitekim, Next Generation Science Standards [NGSS], bilimin yanı sıra mühendislikten temel uygulamaları ve kavramları da içerir ve bilimin ve mühendislik eğitiminin birbiriyle ilişkili doğasını vurgulamaktadır [7]. Bilim, teknoloji, mühendislik ve matematiğin entegre edilmesiyle oluşturulan STEM eğitimi, 21. yüzyıl müfredatının merkezinde yer alan disiplinler arası öğrenme uygulamalarının bir örneği olarak karşımıza çıkmaktadır [8]. Mühendislik tasarım süreci, STEM disiplinleri arasında entegre bir köprü kurarak, öğrencilerin bilimsel ve matematiksel kavramları mühendislik ve teknoloji bağlamında nasıl kullanabileceklerini göstermektedir. Bu bağlamda, mühendislik tasarımı, STEM eğitiminin etkili ve kapsamlı bir şekilde gerçekleştirilmesini sağlayarak, öğrencilerin hem teorik hem de pratik bilgilerini birleştiren bir öğrenme deneyimi sunmaktadır. Mühendislik tasarım süreçleri STEM disiplinleri arasında önemli temel bağlantılar sağlamakta ve öğrencilerin birden fazla fikrin, yaklaşımın ve aracın birden fazla çözümü içeren karmaşık problemlere nasıl uygulanabileceğini anlamalarını sağlamaktadır [9]. Bu süreç, teorik bilgilerin pratik uygulamalarla pekişmesini sağlayarak, öğrencilerin daha derin ve anlamlı bir öğrenme deneyimi yaşamalarını desteklemektedir.

Eğitimde mühendislik tasarımının yeri, öğrencilerin teoriyi pratiğe dökme yeteneklerini geliştirmelerinde önemli bir rol oynamaktadır. Böylece öğrencilerin karmaşık problemleri daha düzenli ve etkili çözmelerini sağlamaktadır. Mühendislik tasarım süreci, problem tanımlama, çözüm geliştirme, prototip oluşturma, test etme ve iyileştirme aşamalarını içermektedir [10]. Bu aşamalar, öğrencilerin yaratıcı ve yenilikçi düşünme becerilerini geliştirmede önemli bir rol oynamakta ve aynı zamanda gerçek dünya problemlerine etkili çözümler üretme yeteneklerini artırmaktadır [11]. Bu bağlamda, mühendislik tasarım süreci, STEM eğitiminde uygulamalı bir araç olarak işlev görmektedir.

Eğitim süreçlerinde mühendislik tasarım becerilerinin öğrencilere kazandırılması, onların teorik bilgilerini pratiğe dönüştürmelerine ve yenilikçi projeler geliştirmelerine olanak tanımaktadır [12]. Bu süreç, öğrencilere fen derslerine gerçek dünya problemlerini çözmelerine yardımcı olacak süreçler sağlamayı amaçlamaktadır. Mühendislik tasarımına dayalı fen eğitimi, öğrencilerin mühendislik tasarım sürecinde karşılaştıkları problemleri çözebilmek için beyin fırtınası yapmalarını, eleştirel düşüncelerini ve iletişim kurmalarını gerektirerek 21. yüzyıl becerilerini geliştirmektedir [13]. Bu becerilerin öğrencilere kazandırılması, onların yalnızca mühendislik alanında değil, aynı zamanda farklı disiplinlerde de yaratıcı ve yenilikçi çözümler geliştirmelerine olanak tanımaktadır. Örneğin, öğrenciler mühendislik tasarım süreci aracılığıyla karmaşık problemlere çözüm bulma yeteneklerini geliştirmektedirler. Bu sayede hem teorik bilgilerini pratiğe dönüştürmede hem de gerçek dünya problemleri çözmeye yenilikçi projeler oluşturabilmektedirler [14]. Ayrıca, bu süreç öğrencilerin iş birliği yapma ve takım içinde etkili bir şekilde çalışma becerilerini



güçlendirmektedir. Çünkü mühendislik problemleri genellikle disiplinler arası bir yaklaşım gerektirmektedir [15]. Bu disiplinler arası yaklaşım, aynı zamanda problem çözme sürecinde esneklik ve yaratıcılık geliştirerek öğrencileri gelecekteki kariyerlerine hazırlayan önemli yetkinlikler sunabilmektedir. Mühendislik tasarım becerilerinin eğitim süreçlerine entegrasyonu, öğrencilere 21. yüzyıl iş gücü gereksinimlerine uygun yetkinlikler kazandırmakla birlikte onların hayat boyu öğrenme kapasitelerini de artırmaktadır. Dolayısıyla, mühendislik tasarım becerilerinin kapsamlı bir şekilde eğitim müfredatına dahil edilmesi, öğrencilerin sadece akademik başarılarını değil, aynı zamanda profesyonel ve kişisel gelişimlerini de desteklediği ve böylece onları dinamik ve hızla değişen bir iş gücü piyasasına daha iyi hazırlamış olduğu söylenebilir. Bu bağlamda, mühendislik tasarım eğitiminin öğrenciler için sunduğu geniş kapsamlı uygulama fırsatları ve problem çözme stratejileri iş gücü piyasasında rekabet avantajı sağlama ve onların sürekli değişen teknolojik ve sosyal koşullara uyum sağlama yeteneklerini artırması beklenmektedir.

Mühendislik tasarım eğitimi, öğrencilere yaratıcılıklarını kullanarak çeşitli problemlere çözüm bulma ve estetik açıdan kabul edilebilir çözümler geliştirme yetenekleri kazandırmayı amaçlayan bir eğitim alanıdır. Aynı zamanda mühendislik tasarım temelli öğretim bir tasarım süreci olmakla birlikte karşılaşılan problemleri tanımlamada, probleme dair alternatif çözümlerden en etkili çözümler üretmede iş birliğine yönelik bir çalışma süreci olarak tanımlanmaktadır [16, 17]. Dolayısıyla mühendislik tasarım eğitiminin amacının işlevsel ve estetik çözümler geliştirme becerisinin yanı sıra öğrencilere karşılaştıkları problemlerin çözümünde 21. yüzyıl becerilerinden biri olan inovasyon yetisinin kazandırılması olduğu söylenebilir. Kewalramani, Palaiologou ve Dardanou (2020) mühendislik tasarımının 21.yüzyıl becerilerini geliştirdiği, bir mühendis gibi düşünmeyi sağladığı, yaratıcı ve araştırma eğiliminin arttığı gerekçesiyle öğrenci eğitimine daha çok entegre edilmesinin faydalı olacağını belirtmiştir [18]. Bu nedenle, mühendislik tasarım becerileri, disiplinler arası yaklaşımlarla öğretim programlarına entegre edilmekte ve öğrencilerin geleceğin mühendisleri ve tasarımcıları olarak yetiştirilmeleri hedeflenmektedir.

Eğitim bilimleri alanındaki yeni araştırmalar ve pedagojik yaklaşımlar, daha etkili öğretim yöntemleri ve beceri gelişimi için önerilerde bulunmaktadır. Bu bilgiler ışığında programlar güncellenmektedir. Bu öğretim programları, öğrencilerin problem çözme, yaratıcı düşünme ve yenilikçi yaklaşımlar geliştirme becerilerini artırarak, onları geleceğin mühendisleri ve bilim insanları olmaya hazırlamayı amaçlamaktadır. Nitekim bu amaç doğrultusunda 21. yüzyıl fen eğitiminin odak noktasının mühendislik tasarım ve inovasyon becerilerinin geliştirilmesi olduğu görülmektedir. Türkiye’de 2006-2007 öğretim yılından itibaren fen öğretim programına ortaokul altı, yedi ve sekizinci sınıflarda zorunlu ders olarak ‘Teknoloji ve Tasarım’ dersi eklenmiştir. Daha sonraki yıllarda 2017 fen bilimleri dersi öğretim programında, mühendisliğe ‘Fen ve Mühendislik Uygulamaları’ başlığı altında yer verilmiştir [19]. Ayrıca 2018 yılında güncellenen program ile mühendislik ve tasarım sürecine ‘Fen, Mühendislik ve Girişimcilik Uygulamaları’ olarak ünitelerin tamamını kapsayacak şekilde yer verildiği görülmektedir [10]. Buradan yola çıkıldığında fen öğretim programında mühendislik tasarım temelli eğitimin yer alması 21. yüzyıl becerilerinin öğrencilere kazandırılma çalışmalarının tasarım odaklı olduğu somut olarak karşımıza çıkmaktadır. Bu durumda geleneksel fen eğitimine teknoloji ve mühendisliğin entegrasyonu ile disiplinler arası anlayışın benimsendiği ortaya çıkmaktadır. Bu entegrasyon, fen, teknoloji ve sanatın bir arada değerlendirildiği bir eğitim yaklaşımını teşvik ederken, mühendislik becerilerinin öğretim programlarında yer almasını sağlamaktadır. Mühendislik becerilerinin bu bağlamda eğitim müfredatına dâhil edilmesiyle birlikte, öğrencilerin bu alanlardaki becerilerini etkili bir şekilde değerlendirebileceğimiz ölçeklere duyulan ihtiyaç da artmıştır. Böylece, mühendislik ve teknoloji alanlarındaki yetkinliklerin sistematik bir şekilde ölçülmesi, öğrencilerin disiplinler arası becerilerini ve uygulama kapasitelerini daha iyi anlayabilmemize olanak tanımaktadır. Bu noktada alan yazın tarandığında 7.sınıflar için Fen, Mühendislik ve Girişimcilik Becerileri Değerlendirme Ölçeği [21], 4-5-6-7. ve 8. Sınıf Öğrencileri için Mühendislik Bilgi Düzeyi Ölçeği [22], 8.sınıflar için ısı transferi ve termal enerji konusuna ilişkin fen bilimlerinde mühendislik tasarımı ve kavramsal değişim [23] incelenmiş ve söz konusu ölçekler geliştirilmiştir. Konuya ilişkin çalışmalar incelendiğinde, mühendislik başlığı altında farklı beceri alanları için ölçek olmasına rağmen tasarım beceri alanına yönelik bir ölçek ile karşılaşılmamıştır. Alan yazında mühendislik tasarım becerilerinin kavramsal yapısının karmaşık ve belirsiz olduğu göz önüne alındığında, geliştirilen test sayesinde bu belirsizliğin nispeten ortadan kaldırılacağı söylenebilir. Bu nedenle bu çalışmanın alan yazına katkı sağlayacağı düşünülmektedir. Bu bağlamda çalışmanın amacı, Bilim ve Sanat Eğitim Merkezlerindeki

(BİLSEM) 7. sınıf öğrencileri üzerinde gerçekleştirilerek mühendislik tasarım becerilerini değerlendirmek ve bu becerilerin geliştirilmesine yönelik etkili bir ölçüm aracı sağlamaktır. Bu yaş grubunun daha önce mühendislik tasarımı konusunda bir deneyime sahip olması, çalışma sonuçlarının bu alandaki eğitim uygulamalarına ışık tutmasını ve öğretim stratejilerinin geliştirilmesini desteklemesini amaçlamaktadır.

## 1. Materyal ve Yöntem

### 2.1. Desen

Araştırmada betimsel araştırma modellerinden birisi olan tarama yöntemi kullanılmıştır. Tarama modeli, büyük örneklem kapsamında yürütülen ve bir grubun görüş, beceri gibi niteliklerini belirlemek için verilerin toplanmasını amaçlayan desendir [24, 25]. Benzer şekilde Creswell (2014) tarama modelini katılımcıların genelindeki eğilim ve algılarının nicel olarak betimlenmesini, bunun sonucunda birtakım çıkarımlar yapmayı sağlamak olarak ele almaktadır [26]. Tarama araştırmaları genellikle geniş bir kitleden bilgi toplanarak kullanılmaktadır. Bu desende araştırma konusu olduğu şekliyle hiçbir değişime neden olmadan tasvir edilir. Bu bağlamda araştırmada Bilim ve Sanat Eğitim Merkezlerinde (BİLSEM) öğrenim gören 7.sınıf öğrencilerinin mühendislik tasarım becerilerinden faydalanılarak test geliştirilmesi amaçlandığı için tarama deseni kullanılmıştır.

### 2.2. Evren ve Örneklem

Araştırmanın evreni, İzmir ili Konak ve Bornova ilçelerindeki Bilim ve Sanat Eğitim Merkezlerinde (BİLSEM) öğrenim gören 7. sınıf öğrencilerinden oluşmaktadır. Bu kurumlarda eğitim alan öğrenciler, yaratıcı problem çözme, inovatif düşünme ve disiplinler arası ilişkilendirme gibi becerileri kazanmak amacıyla yoğun bir eğitim programına dahil olmaktadır. Bu bağlamda, mühendislik tasarım becerileri ölçeğinin geliştirilmesinde, tasarım becerisi gelişimine odaklanmış bir eğitim programına dâhil olan öğrencilerden veri toplanması, ölçeğin geçerlilik ve güvenilirliği açısından büyük önem taşımaktadır. Tasarım etkinlikleri deneyimine sahip öğrencilere yönelik bu çalışma, 110 öğrenciyle gerçekleştirilmiş, 101 geçerli form değerlendirmeye alınmıştır. Bu öğrencilerin 56'sı kız, 45'i ise erkektir. Araştırmanın katılımcıları, seçkisiz olmayan örnekleme yöntemlerinden amaçsal örnekleme yöntemi ile belirlenmiştir. Amaçsal örnekleme, çalışma amacına göre araştırmacının bireysel yargısına bağlı örneklemin seçilmesidir [27]. Amaçlı bir şekilde seçilen örneklem, çalışmanın amacına bağlı olarak incelenen konu açısından zengin bilgiler verebilecek niteliktedir. Amaçlı örnekleme incelenen konu ya da olguya ilişkin önemli bilgi kaynaklarına ulaşmada araştırmacıya yardımcı olmakla birlikte araştırılan durumların derinlemesine incelenmesine imkân tanımaktadır [28]. Bu bağlamda araştırmanın katılımcılarını, İzmir ilinde yer alan Bilim ve Sanat Eğitim Merkezlerinde tasarım sürecine dahil olan 7. sınıf öğrencileri oluşturmaktadır. Bu araştırmada, ortaokul 7. sınıf öğrencilerinin mühendislik tasarım süreçlerine yönelik becerileri değerlendirildiği için çalışma ortaokul düzeyinde gerçekleştirilmiştir. Mühendislik tasarım becerileri hem soyut düşünme yetilerini hem de somut uygulama becerilerini kapsayan bir yapıya sahiptir. Ortaokul dönemi, öğrencilerin analitik düşünme, yaratıcılık ve problem çözme gibi üst düzey bilişsel becerilerini geliştirmeye başladıkları bir dönemdir. Bu nedenle, 7. sınıf öğrencilerinin seçilmesinin temel amacı, bu yaş grubunun mühendislik ve tasarım kavramlarına ilgi duymaya ve bu becerileri geliştirmeye yönelik yeterli bilişsel olgunluğa ulaşmış olmalarıdır. Dolayısıyla, mühendislik tasarım becerilerinin algılanması ve değerlendirilmesi için uygun bir yaş grubu olarak tercih edilmiştir.

### 2.3. Veri toplama araçları

#### 2.3.1. Mühendislik Tasarım Becerileri Algı Ölçeği (MüTasBAT)

Araştırmada veri toplama aracı, Moazzen, Miller, Wild, Jackson ve Hadwin (2014) [29] tarafından 'Engineering Design Survey' adıyla üniversite öğrencilerine yönelik geliştirilen ölçme aracının tasarım-beceri boyutu maddeleri dikkate alınarak Mühendislik Tasarım Becerileri Ölçeği (MüTasBAT) adında araştırmacı tarafından ortaokul 7. Sınıf öğrencileri için yeniden geliştirilmiştir (Ek-A). Bu testin geliştirilmesi için öncelikle literatür taraması yapılmış ve özellikle tasarım etkinlikleri yapan ortaokul 7. sınıf BİLSEM öğrencilerinin duygu ve düşünceleri görüşme yoluyla dikkate alınmıştır. Ölçeğin geliştirilmesi aşamasında sırasıyla, literatür taraması, madde havuzunun oluşturulması, uzman görüşlerinin alınması, pilot çalışma, testin geçerlik-güvenirlik çalışmaları gerçekleştirilmiştir.

Moazzen, Miller, Wild, Jackson ve Hadwin (2014) [29] tarafından 'Engineering Design Survey' adıyla üniversite öğrencilerine yönelik geliştirilen ölçeğin tasarım-beceri kapsamında yer alan ölçek maddeleri ortaokul

öğrencilerinin seviyesine uygun olacak şekilde beceri algısı biçiminde ifade edilmiştir. Böylelikle 62 maddelik madde havuzu oluşturulmuştur. Ölçeğin taslak formunda yer alan maddelerin ölçülmek istenilen amaca uygunluğu ile ilgili olarak eğitim bilimleri alanında uzman olan 5 kişinin görüşü alınmıştır. Uzman görüşleri için 'Uygundur', 'Uygun değildir', 'Düzeltilmeli' ve 'Önerim' şeklinde dört maddeli bir form oluşturulmuştur. Uzmanlardan gelen görüşler doğrultusunda bazı maddeler düzenlemeye gidilmiştir. Maddeler yazılırken ölçülmek istenen özelliğe ilişkin doğru cevap verildiğine ilişkin tutarlılık sağlamak amacıyla olumlu ve olumsuz maddeler oluşturulmasına ve maddelerin anlaşılabilir olmasına özen gösterilmiştir. Ölçekte cevaplanması istenilen maddeler için 'Hiç katılmıyorum (1)', 'Katılmıyorum (2)', 'Orta düzeyde katılıyorum (3)', 'Katılıyorum (4)' ve 'Tamamen katılıyorum (5)' ifadeleri şeklinde beşli likert tipinde hazırlanmıştır. Değerlendirme aşamasında ise olumsuz maddeler üzerinde bu puanlama ters şekilde yapılmıştır.

Ölçeğin geçerlilik ve güvenilirliğini değerlendirmek amacıyla, 62 maddelik taslak bir test, 12 kişilik bir ön görüşme ile elde edilen geri bildirimler doğrultusunda hazırlanmıştır. Ardından, tasarım deneyimlerine sahip 101 ortaokul 7. sınıf BİLSEM öğrencisine uygulanmıştır. Ön görüşmelerde, test maddelerinin anlaşılabilirliği ve uygulama sürecindeki etkinliği hakkında geri bildirim alınmıştır. Katılımcıların "prototip" ve "malîyet" gibi terimlerde anlamada zorluk yaşadığı gözlemlenmiştir. Elde edilen geri bildirimler doğrultusunda, maddelerde dil ve ifade netleştirilmiş ve tekrarlı cümleler düzeltilmiştir.

BİLSEM'den elde edilen veriler, testin yapı geçerliğini test etmek amacıyla Açıklayıcı Faktör Analizi (AFA) ve Doğrulayıcı Faktör Analizi (DFA) ile değerlendirilmiştir. Bu analizler, mühendislik tasarım beceri algılarını ölçmeye yönelik geliştirilen testin geçerli ve güvenilir olduğunu ortaya koymuştur. Analizler sonucunda, test 29 maddeye indirgenmiş ve dört alt boyut halinde yapılandırılmıştır: Kuluçka Tasarım Geliştirme Becerileri, Yenilikçi Tasarım Geliştirme Becerileri, Tasarım Değerlendirme Becerileri ve Tasarımı Sunma ve Raporlama Becerileri. Nihayetinde, testin 29 maddelik 4 faktörlü yapısının son hali belirlenmiştir.

#### 2.4. Verilerin analizi

Ölçeğin yapı geçerliğinin sağlanması için öncelikle SPSS programı kullanılarak Açıklayıcı Faktör Analizi (AFA) yapılmış, ardından AMOS programı kullanılarak Doğrulayıcı Faktör Analizi (DFA) gerçekleştirilmiştir. Bu bağlamda, öncelikle elde edilen verilerin faktör analizine uygun olup olmadığının belirlenmesi için Kaiser-Meyer-Olkin (KMO) ve Bartlett küresellik testi uygulanmıştır. Hazırlanan ve katılımcılara uygulanan testin güvenilirlik çalışması için tüm teste ve alt boyutlara ait Cronbach alfa katsayıları hesaplanmıştır.

## 2. Bulgular

Araştırmanın bu bölümünde açıklayıcı faktör analizi, doğrulayıcı faktör analizi ve güvenilirliğe yönelik bulgular sunulmuştur.

62 maddelik teste açıklayıcı faktör analizi yapılmıştır. Bu analiz, benzer konulara sahip maddeleri bir araya getirerek, maddeler arasındaki ilişkileri belirlemiş ve yorumlamayı kolaylaştırmıştır. Faktör analizi çalışmalarının ilk adımında, test uygulamasından elde edilen veriler Kaiser-Meyer-Olkin (KMO) ve Bartlett's testleri ile test edilmiştir. Yapılan testler sonucunda elde edilen değerler Tablo 1'de görüldüğü gibidir.

**Tablo 1.** Mühendislik tasarım becerileri algı testinin faktör ve madde analizi sonuçları

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Örnekleme Yeterliliği	0,805	
	Ki-Kare Değeri	1231,246
Bartlett'in Küresellik Testi	Serbestlik derecesi	406
	p	0,000

Ölçme aracına faktör analizi yapıldığında KMO (Kaiser-Meyer-Olkin) değerinin 0,805 ve Bartlett küresellik testinin anlamlı [ $\chi^2=1231,246$  ( $p<0,000$ )] olarak gözlenmiştir. Bu sonuca göre testteki maddelerden elde edilen verilerin faktör analizi için uygun olduğu ve değişkenlerin gruplandırılarak azaltılabileceği ortaya çıkmıştır. Ölçme aracına Maximum Likelihood tahminlemesi ve Equamax dik döndürme tekniği uygulanmış ve faktör yükleri 0,40 altında olan maddeler ihmal edilerek test yapısı incelenmiştir. Analiz sonucunda testin 29 madde içeren 4 alt faktörde toplanmış olduğu belirlenmiştir.

Tablo 2'de görüldüğü gibi 29 maddeden oluşan dört alt faktörlü nihaî bir test geliştirilmiştir.

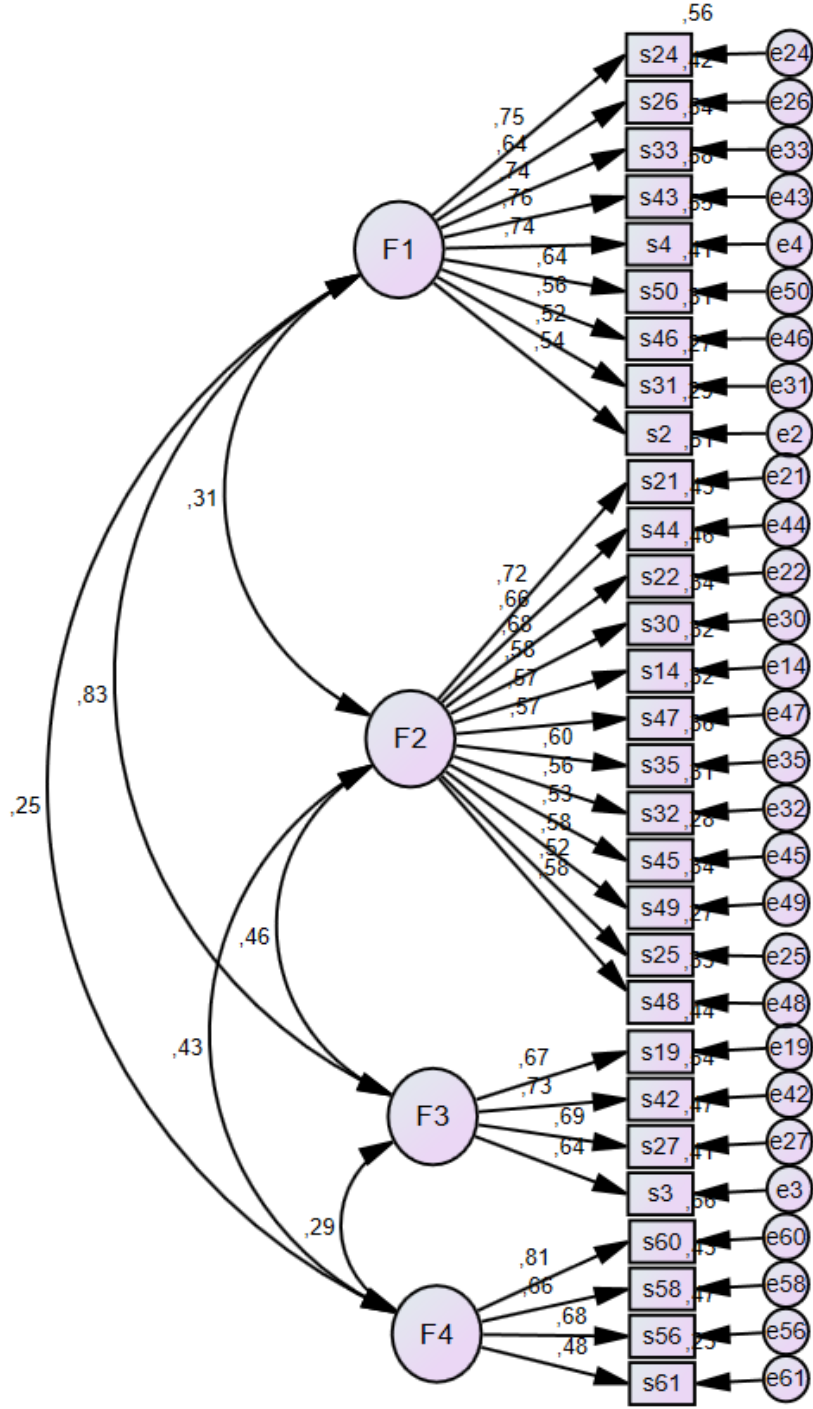
**Tablo 2.** Mühendislik tasarım becerileri algı testinin faktör ve madde analizi sonuçları

Madde No	1.Faktör yükü	2. Faktör yükü	3. Faktör yükü	4. Faktör yükü
s24	0,688			
s26	0,657			
s33	0,652			
s43	0,644		0,411	
s4	0,624			
s50	0,586			
s46	0,576			
s31	0,514			
s2	0,501			
s21		0,719		
s44		0,673		
s22		0,567	0,381	
s30		0,552		
s14		0,532		
s47		0,531		
s35		0,529		
s32		0,528		
s45		0,520		
s49		0,505		
s25		0,488		
s48		0,477		
s19			0,679	
s42	0,423		0,589	
s27	0,364		0,586	
s3			0,460	
s60				0,778
s58				0,712
s56				0,653
s61				0,431

Faktörlerde yer alan maddelerin faktör yüklerinin 0,40 değerinden yüksek olması, bu maddelerin güçlü bir faktör yapısı sağladığına yönelik bir kanıt olabilir. Birinci faktör içinde yer alan maddeler; öğrencilerin tasarım geliştirme sürecinde problem üzerinde bir süre ön çalışma yapma, fikir sunma ve problem çözümlerini bir anlamda kuluçkaya yatırdıkları süreç belirlenme çalışılmıştır. Dolayısıyla birinci faktör “Kuluçka tasarım geliştirme becerileri” şeklinde adlandırılmıştır. Birinci faktör 9 maddeden oluşmaktadır. İkinci faktörde yer alan maddeler incelendiğinde, öğrencilerin yaratıcı karar ve çözüm ile alternatif çözüm üretme, çözümlerini revize etme gibi ifadeler yer aldığından ikinci faktör “Yenilikçi tasarım geliştirme becerileri” biçiminde adlandırılmıştır. İkinci faktör 12 maddeden oluşmaktadır. Üçüncü faktördeki maddelerde ise değerlendirme kriterleri oluşturma ve değişkenleri ölçme gibi ifadeler yer almaktadır. Bu nedenle üçüncü faktör “Tasarım değerlendirme becerileri” şeklinde adlandırılmıştır. Üçüncü faktör 4 maddeden oluşmaktadır. Dördüncü faktör kapsamındaki maddeler incelendiğinde öğrencilerin tasarımlarını açıklama ve sunmalarını içeren ifadeler yer aldığından dördüncü faktör “Tasarımı sunma ve raporlama becerileri” olarak adlandırılmıştır. Dördüncü faktör 4 maddeden oluşmaktadır.

Çalışmada açımlayıcı faktör analizi sonucunda geliştirilen MüTasBAT'nün dört faktörlü yapısı için 101 öğrenciye uygulama yapılmıştır. Elde edilen veriler üzerinden doğrulayıcı faktör analizi yapılmıştır. Oluşan doğrulayıcı faktör analiz modeli Şekil 1'de gösterilmiştir.

Şekil 1'de tüm alt faktörler arasında korelasyon değeri verilmektedir.



Figür 1. Tüm alt faktörler arası korelasyon değeri

Yapılan doğrulayıcı faktör analizi sonucunda elde edilen bulgulara göre  $\chi^2/df = 1,182$ ,  $GFI = 0,797$  ve  $RMSEA = 0,043$  değerlerine ulaşılmıştır. Ayrıca, MüTasBAT’de alt faktörler arasında ve her bir faktörün kendi içindeki maddeler arasında yüksek korelasyonlar gözlemlenmiştir. Şekilde yer verilen değerlere göre alt boyutlar arasındaki ilişkinin 0,25 ile 0,83 arasında değiştiği görülmektedir. Bu durum, testin iç tutarlılığının yüksek olduğunu göstermektedir. Bu sonuçlar maximum likelihood yöntemiyle gerçekleştirilen faktör analizinden elde edilmiştir. Ayrıca Şekil 1’de maddelerin alt boyutlarla olan ilişkileri gösterilmiş olup, elde edilen sonuçlar istatistiksel olarak anlamlı bulunmuştur. Bu bulguların ardından, maddelerin alt boyutlarla olan ilişkilerini daha ayrıntılı incelemek amacıyla, parametrelerin regresyon tahminlerine ilişkin gösterimler Tablo 3’te sunulmuştur.

**Tablo 3.** Faktörlere karşılık gelen regresyon ağırlıkları

			Estimate (Tahmin)	S.E. (Standart Hata)	C.R. (Kritik Oran)	p
s24	<---	F1	1,000			
s26	<---	F1	,792	,126	6,303	***
s33	<---	F1	,980	,135	7,263	***
s43	<---	F1	,993	,132	7,498	***
s4	<---	F1	,886	,122	7,295	***
s50	<---	F1	,810	,130	6,221	***
s46	<---	F1	,684	,127	5,385	***
s31	<---	F1	,706	,141	4,992	***
s2	<---	F1	,570	,110	5,192	***
s21	<---	F2	1,000			
s44	<---	F2	,911	,149	6,133	***
s22	<---	F2	,996	,158	6,304	***
s30	<---	F2	,836	,154	5,446	***
s14	<---	F2	,778	,147	5,288	***
s47	<---	F2	,743	,141	5,283	***
s35	<---	F2	,899	,161	5,597	***
s32	<---	F2	,694	,133	5,221	***
s45	<---	F2	,738	,150	4,909	***
s49	<---	F2	,885	,163	5,427	***
s25	<---	F2	,748	,155	4,816	***
s48	<---	F2	,744	,139	5,371	***
s19	<---	F3	1,000			
s42	<---	F3	1,040	,172	6,061	***
s27	<---	F3	,923	,160	5,779	***
s3	<---	F3	,854	,157	5,444	***
s60	<---	F4	1,000			
s58	<---	F4	,805	,141	5,703	***
s56	<---	F4	,848	,144	5,871	***
s61	<---	F4	,578	,136	4,243	***

Faktörlere karşılık gelen maddelerin regresyon katsayı ağırlıkları Tablo 3'te verilmiştir. Elde edilen bilgilere göre ilk faktör için en anlamlı madde s43 sorusudur "*Sınıfta verilen mühendislik projesinin problem durumuna göre istenen tasarımı yapabilirim.*" ve istatistiksel anlamlılık tüm maddeler tarafından sağlanmıştır. İkinci faktörde en önemli maddenin s22 "*Sınıfta verilen mühendislik projesini gerçekleştirirken tasarımıyla ilgili sıra dışı olan yaratıcı fikirleri önemsemem.*" olduğu görülürken; üçüncü faktör için s42 "*Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen tasarımlarla ilgili meydana getirilen ürünü test edebilirim.*"; dördüncü faktör için s56 "*Mühendislik proje tasarımlarının sunum raporunu etkili bir dille anlatımda zorlanabilirim.*" sırasıyla ilgili faktörlerin en önemli maddeleridir.

MüTasBAT için faktör analizi yapıldıktan sonra maddelerin güvenilirlik testleri yapılmıştır. Test sonucu Cronbach Alpha değeri  $\alpha = 0.895$  olarak bulunmuştur. Ölçekteki her bir faktörün güvenilirlikleri hesaplanmıştır. Faktörlerin güvenilirlik katsayı sonuçlarından elde edilen bulgular Tablo 4'te yansıtılmıştır.

**Tablo 4.** Faktörlerin Cronbach's Alpha Değeri

Faktör numarası	Faktör Adı	Cronbach's Alpha
1	Kuluçka tasarım geliştirme becerileri	0,871
2	Yenilikçi tasarım geliştirme becerileri	0,867
3	Tasarım değerlendirme becerileri	0,776
4	Tasarımı sunma ve raporlama becerileri	0,746

Tablo 4'te görüldüğü üzere "*Kuluçka tasarım geliştirme becerileri*" faktörünün Cronbach Alpha güvenilirlik katsayısı  $\alpha = 0,871$ 'dir. "*Yenilikçi tasarım geliştirme becerileri*" faktörünün Cronbach Alpha güvenilirlik katsayısı  $\alpha = 0,867$ 'dir. "*Tasarım değerlendirme becerileri*" faktörünün Cronbach Alpha güvenilirlik katsayısı  $\alpha = 0,776$ 'dır. "*Tasarımı sunma ve raporlama becerileri*" faktörünün Cronbach Alpha güvenilirlik katsayısı  $\alpha = 0,746$ 'dır. Görüldüğü üzere her bir faktörün güvenilirliği oldukça yüksek değerleri içermektedir.

### 3. Sonuç ve Tartışma

MüTasBAT'ın faktör analizinin uygunluğu testlerinden sonra 62 maddelik bu teste faktör analizi yapılmıştır. Bu analiz sonucunda, test 4 ana faktör etrafında gruplanmış ve 29 maddeye indirilmiştir. Bu durum, testin daha öz ve geçerli hale gelmesini sağlayabilir. Faktör analizi sonuçları, çıkarılan maddelerin fazla tekrar içerdiği veya test maddelerinin bazıları için yetersiz anlam netliği ya da öğrencilerin cümleleri tam olarak anlamadığı anlamına gelebilir. Bu, testin belirlenen faktörlerin kapsamını doğru bir şekilde yansıttığını gösterir. Sonuçlar, testin geçerliliği hakkında olumlu bir işaret olabilir.

Bu çalışmada faktör analizi sonucunda belirlenen birinci faktör, öğrencilerin tasarım geliştirme sürecinde problem üzerinde ön çalışma yapma, fikir sunma ve problem çözümlerini kuluçka sürecine yatırma becerilerini kapsamaktadır. Bu becerilerin bir araya gelmesi, öğrencilerin yaratıcı düşünme ve yenilikçi çözümler üretme kapasitelerini artırdığı için eğitimde önemli bir yer tutmaktadır. Özellikle kuluçka aşaması, yaratıcı düşünme süreçlerinde kritik bir rol oynamaktadır. Wallas (1926) ve Ludwig (1989), yaratıcı süreçlerde kuluçkanın rolünü vurgularken, Yeşilyurt (2020) da kuluçka aşamasının yaratıcı düşünme sürecinde kritik bir rol oynadığını belirtmektedir [30, 31, 32]. Literatürde, kuluçka döneminin yaratıcı süreçlerdeki önemi vurgulanmıştır [33, 34, 35]. Dolayısıyla birinci faktör "Kuluçka tasarım geliştirme becerileri" şeklinde adlandırılmıştır. Kuluçka dönemi, tasarımcının bilinçli olarak bilgi toplama etkinlikleriyle düşüncelerini şekillendirdiği, ancak bu bilgilerin bilinç dışında yeniden düzenlendiği ve çözüm üretmeye katkı sağladığı bir süreçtir [36, 37]. Bu dönemde bireylerin bilinçdışı devreye giren düşünceler, yeni fikirlerin ve çözümlerin ortaya çıkmasına olanak tanımaktadır. Bu durumda MüTasBAT alt faktörlerinden biri olan kuluçka tasarım becerileri, yaratıcı düşüncenin önemli bir aşaması olan kuluçka dönemiyle yakından ilişkilidir. Li, Zhang, Zhang ve Luo (2022) kuluçka dönemine ilişkin araştırmaların, yalnızca yaratıcı süreçlerin gizemini açığa çıkarmakla kalmayıp, aynı zamanda bu süreçlerin uygulamalı alanlarda kullanımını teşvik ettiğini belirtmektedir [38]. Bu bağlamda, kuluçka tasarım becerilerinin mühendislik problemlerine yenilikçi ve etkili çözümler geliştirme yeteneğini desteklediği ve yaratıcı sürecin mühendislik uygulamalarında nasıl hayata geçirilebileceği ortaya konulmaktadır. Birinci faktörün "Kuluçka Tasarım Geliştirme Becerileri" olarak adlandırılması, öğrencilerin bu süreci ne kadar iyi yönettiklerini ve bu süreçteki çeşitli becerilerini detaylı bir şekilde ölçmeye olanak tanımaktadır. Bu faktörün 9 maddeden oluşması, bu becerinin karmaşıklığını ve çeşitliliğini yansıtmaktadır.

Bu çalışmada faktör analizi sonucunda belirlenen ikinci faktör, öğrencilerin yaratıcı kararlar alma, alternatif çözümler üretme ve çözümlerini revize etme becerilerini kapsamaktadır. Bu becerilerin, öğrencilerin yenilikçi düşünme ve problem çözme kapasitelerini artırdığı için, bu faktör "Yenilikçi tasarım geliştirme becerileri" olarak adlandırılmıştır. İkinci faktör 12 maddeden oluşmaktadır. Mühendislik tasarımı, mühendislerin belirli sınırlamalar ve gereksinimler çerçevesinde sorunları çözmeye ve bu sorunlara yönelik çözümler geliştirme sürecidir [39]. Bu süreç, mühendislerin karşılaştıkları problemlere yaratıcı ve yenilikçi çözümler bulmalarını gerektirir. Khaerudin, Hendrilia, Dewi, Erwin ve Al-Amin (2023) çalışmalarında, problem çözme becerilerinin tasarım alanındaki bireyler için vazgeçilmez olduğu ve bu becerilerin öğrencilerin yaratıcı düşünme yeteneklerini geliştirerek yenilikçi tasarım çözümleri üretmelerine olanak tanıdığı vurgulanmaktadır [40]. Benzer şekilde, Pusca ve Northwood (2018), tasarım düşüncesinin çok yönlü bir problem çözme süreci sunduğunu, bunun öğrencilerin karmaşık problemleri daha iyi kavramalarına yardımcı olduğunu ve problemleri yenilikçi yollarla çözmek için tasarım ilkelerinin gerekliliğini ortaya koyduğunu belirtmektedir [41]. Yenilikçi tasarım geliştirme becerileri, öğrencilerin problem çözme sürecinde aktif olmalarını, karmaşık durumlarla karşılaştıklarında alternatif çözümler üretebilmelerini ve bu çözümleri etkili şekilde uygulayabilmelerini sağlamaktadır. Bu beceriler, mühendislik ve tasarım disiplinlerinde, hızlı değişim ve yenilik gerektiren ortamlarda başarı için kritik bir rol oynamaktadır. Ayrıca, yenilikçi tasarım geliştirme becerileri, hızlı bilimsel ilerlemeler ve teknolojik keşiflerle desteklenerek mühendislerin yaratıcı ve analitik yeteneklerini ön plana çıkarır, bu da rekabetçi ortamlarda yeni ürünler geliştirilmesine olanak tanır [42]. Sonuç olarak, öğrencilerin problem çözme, yaratıcı kararlar alabilme ve alternatif çözümler üretebilme yetenekleri, yenilikçi tasarım sürecinin temel bileşenlerindedir ve tasarım sürecinde öngörülemeyen zorluklarla başa çıkmada kritik bir rol oynamaktadır.

Bu çalışmada faktör analizi sonucunda belirlenen üçüncü faktör, öğrencilerin değerlendirme kriterleri oluşturma ve değişkenleri ölçme becerilerini kapsamaktadır. Bu becerilerin, öğrencilerin tasarım süreçlerinde objektif ve sistematik değerlendirmeler yapabilmeleri açısından kritik öneme sahip olduğu görülmektedir. Bu

nedenle üçüncü faktör "Tasarım değerlendirme becerileri" şeklinde adlandırılmıştır. Üçüncü faktör 4 maddeden oluşmaktadır. Mühendislik tasarımı, belirli hedeflere ulaşan ve kısıtlamalara uyan ürünlerin sistematik bir şekilde üretilmesi ve bu süreçte yapılan değerlendirmelerin önemli olduğu bir süreçtir [43]. Öğrencilerin tasarım süreçlerinde objektif kriterler oluşturabilme ve bu kriterlere göre değerlendirme yapabilme yetenekleri, başarılı bir mühendislik tasarımı için temel bileşenlerdendir. Dym, Agogino, Eris, Frey ve Leifer (2005) çalışmalarında öğrencilerin tasarımlarını objektif bir şekilde değerlendirebilmeleri için gerekli kriterlerin belirlenmesi ve bu kriterler doğrultusunda tasarımın analiz edilmesi gerektiği vurgulamışlardır [14]. Benzer şekilde, Hynes (2009), tasarım sürecinde çözümü test etme ve değerlendirme adımlarının, taslak ürünlerin veya prototiplerin iyileştirilmesi için önemli bir rol oynadığını ifade etmiştir [44]. Mühendislik tasarım sürecinde tasarım değerlendirme becerisi, öğrencilerin tasarım süreçlerini daha bilinçli bir şekilde ele almalarına olanak tanır. Ayrıca, öğretmenlerin öğrencileri bu yönde yönlendirmesi, prototiplerin değerlendirilmesi ve yeniden tasarlanması süreçlerini destekleyerek öğrenme sürecini daha etkili hale getirebilir. Bu bağlamda, tasarım değerlendirme becerileri, öğrencilerin tasarım odaklı düşünme yaklaşımını kazanmaları açısından temel bir odak noktası olarak ele alınmalıdır.

Bu çalışmada faktör analizi sonucunda belirlenen dördüncü faktör, öğrencilerin tasarımlarını açıklama ve sunma yeteneklerini kapsamaktadır. Bu beceriler, öğrencilerin tasarımlarını etkili bir şekilde iletişim kurabilme ve sonuçları sistematik bir şekilde raporlama yeteneklerini ifade etmektedir. Bu nedenle, dördüncü faktör "Tasarımı Sunma ve Raporlama Becerileri" olarak adlandırılmıştır. Bu faktör, 4 maddeden oluşmaktadır. Bu bulgular, mühendislik eğitimcilerinin öğrencilerin tasarım sunma ve raporlama becerilerini geliştirmelerine yardımcı olacak yöntemler geliştirmeleri gerektiğini göstermektedir. Bazı araştırmalar, etkili iletişim ve raporlama becerilerinin mühendislik kariyerinde başarının önemli bir göstergesi olduğunu ortaya koymaktadır [45]. Bu nedenle, mühendislik süreçlerinde iletişim ve raporlama becerilerinin geliştirilmesi büyük bir önem taşımaktadır. Ürün geliştirme süreçlerinin STEM proje döngüsü ve mühendislik aşamalarına uygun olarak planlanması, yürütülmesi ve değerlendirilmesi gerekmektedir. Temel hedef, 21. yüzyıl becerilerinin gelişimini desteklemek olduğundan, mühendislik süreçlerinde raporlama ve sunma aşamaları kritik bir öneme sahiptir [46]. Genel anlamda bakıldığında MüTasBAT'nün tüm alt faktörleri, tasarım sürecinin anlaşılabilirliğini artırması ve tasarımların şeffaf bir şekilde iletilmesini sağlaması umulmaktadır. Bu dört faktör, mühendislik eğitiminde öğrencilerin hem teknik hem de iletişim becerilerini geliştirmelerine yönelik bütünsel bir yaklaşım sunması beklenmektedir.

MüTasBAT için faktör analizi yapıldıktan sonra maddelerin güvenilirlik testleri yapılmıştır. Test sonucu Cronbach Alpha değeri  $\alpha=0.895$  olarak bulunmuştur. MüTasBAT'nün tüm alt faktörler arasında, alt faktörlerin kendi aralarında yüksek korelasyona sahip olması ve testteki maddelerin buldukları faktör içinde yüksek korelasyonlara sahip olması testin iç tutarlılığının yüksek olduğunu göstermektedir. Ayrıca, içerik geçerliliği, uzman görüşleriyle desteklenmiş ve testin kapsamının yeterli olduğu doğrulanmıştır. Bu sonuçlar, testin hem güvenilir hem de geçerli olduğunu ve araştırma bağlamında güvenle kullanılabilirliğini ortaya koymaktadır. Bu özellikler, testin elde ettiği verilerin hem tutarlı hem de doğru olduğunu ve bu verilerin bilimsel analizlerde sağlam bir temel oluşturduğunu söylemek mümkündür.

#### 4. Öneriler

Çalışma, 7. sınıf öğrencileriyle sınırlı olduğundan, sonuçların genellenebilirliğini artırmak için testin araştırmalarda farklı ortaokul seviyelerinde (örneğin, 5., 6. ve 8. sınıf) ve çeşitli öğrenci gruplarında (örneğin, genel eğitim sınıfları, özel yetenek sınıfları) uygulanması önerilebilir.

Ölçeğin kullanımı ile mühendislik tasarım becerilerinin eğitimde daha etkin bir şekilde değerlendirilmesi sağlanabilir. Sonuçlar, öğrencilerin hangi tasarım aşamalarında daha fazla desteğe ihtiyaç duyduklarını belirlemede öğretmenlere yol gösterebilir. Bu, eğitim programlarının tasarımında ve mühendislik odaklı derslerin içeriğinin geliştirilmesinde kullanılabilir.

Çalışmada geliştirilen test (MüTasBAT), 7. sınıf öğrencilerinin mühendislik tasarım becerilerini geliştirmede hangi alanlarda daha fazla uygulamaya ve geri bildirim ihtiyacı duyduklarını belirlemek için kullanılabilir. Özellikle, "Yenilikçi Tasarım Geliştirme" ve "Tasarımı Sunma ve Raporlama" gibi faktörlerde belirlenen eksiklikler, öğrencilere sunulacak ek etkinliklerle giderilebilir.



MüTasBAT'ın, öğretmenlerin mühendislik tasarım süreçlerini nasıl öğrettiklerini değerlendirmek için bir araç olarak kullanılması önerilebilir. Bu bağlamda, öğretmenlerin eğitim programlarında ölçüğe dayalı beceri geliştirme etkinliklerine yer verilmesi sağlanabilir.

Ölçeğin, fen, matematik ve sanat gibi farklı disiplinlerle entegre edilen mühendislik tasarım süreçlerinde uygulanması önerilebilir. Bu, mühendislik tasarımının disiplinler arası eğitime katkısını anlamaya yardımcı olabilir.

Çalışmanın sonuçları, mühendislik tasarım becerilerini geliştiren eğitim müfredatının iyileştirilmesine yönelik öneriler sunabilir. Özellikle BİLSEM öğrencilerine yönelik mühendislik eğitimi programlarının daha hedefe yönelik bir şekilde geliştirilmesi sağlanabilir.

MüTasBAT'ın takım çalışmaları ve proje tabanlı öğrenme ortamlarında uygulanabilmesi için öğrencilerin takımdaki etkileşimlerini ölçmeye yönelik bir 'Mühendislik Takım Becerileri Ölçeği' geliştirilmesi önerilmektedir. Bu tür bir ölçek, takım çalışmalarında öğrencilerin iş birliği, iletişim ve problem çözme becerilerinin değerlendirilmesine olanak tanıyabilir. Bu sayede, öğrencilerin bireysel ve grup içi mühendislik tasarım becerilerindeki farklılıklar karşılaştırılabilir.

**Yazar Katkıları:** Kavramsallaştırma, S.T. ve K.S.; metodoloji, K.S.; yazılım, S.T. ve K.S.; doğrulama, S.T.; biçimsel analiz, S.T.; soruşturma, S.T.; kaynaklar, K.S.; veri iyileştirme, A.A.; yazı- orijinal taslak hazırlama, A.A., S.T. ve K.S.; yazma-gözden geçirme ve düzenleme, S.T. ve K.S.; görselleştirme, S.T. ve K.S. Tüm yazarlar makalenin yayınlanan versiyonunu okudu ve kabul etti.

**Finansman:** Bu araştırma dışarıdan fon almadı.

**Çıkar çatışmaları:** Yazarlar çıkar çatışması beyan etmemektedir.

**Bilgilendirme:** Bu makale, Dokuz Eylül Üniversitesi Eğitim Bilimleri Enstitüsü Matematik ve Fen Bilimleri Eğitimi Anabilim Dalı Fen Bilgisi Öğretmenliği Bilim Alanında 'Bilim ve Sanat Entegrasyonunda Leonardo Etkisi: Mühendislik Tasarım Becerileri ile Yaratıcılıklarının İncelenmesi' adlı doktora tezinin verilerinden üretilmiştir.

## Kaynaklar

- [1] Şahin, F. (2022). *7. sınıf fen bilimleri dersi hücre ve bölünmeler ünitesinde learningapps uygulaması kullanımının öğrencilerin başarı ve derse yönelik tutumlarına etkisi*. Fen Bilimleri Enstitüsü. (Yayımlanmamış Doktora tezi)
- [2] Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, 3(1), 1-11. <https://doi.org/10.1186/s40594-016-0046-z>
- [3] Schweingruber, H., Pearson, G., & Honey, M. (Eds.). (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. National Academies Press.
- [4] Bryan, L. A., Moore, T. J., Johnson, C. C., & Roehrig, G. H. (2015). Integrated STEM education. In C. C. Johnson, E. E. Peters-Burton, & T. J. Moore (Eds.), *STEM roadmap: A framework for integration* (pp. 23-37). London: Taylor & Francis
- [5] Lucas, B., Claxton, G., & Hanson, J. (2014). Thinking Like an Engineer: Implications for the education system. <http://www.raeng.org.uk/news/news-releases/2014/may/do-you-think-like-an-engineer>
- [6] Next Generation Science Standards [NGSS] (USA, 2014). <http://www.nextgenscience.org/> OECD (2013). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. OECD Publishing (<http://www.oecdilibrary.org/content/book/9789264190511-en>).
- [7] English, L. D. (2016). STEM education K-12: Perspectives on integration. *International Journal of STEM education*, 3, 1-8. DOI 10.1186/s40594-016-0036-1
- [8] Beers, S. (2011). 21st century skills: Preparing students for their future. [https://cosee.umaine.edu/files/coseeos/21st\\_century\\_skills.pdf](https://cosee.umaine.edu/files/coseeos/21st_century_skills.pdf)
- [9] Purzer, S., Hathaway Goldstein, M., Adams, R., Xie, C., & Nourian, S. (2015). An exploratory study of informed engineering design behaviors associated with scientific explanations. *International Journal of STEM Education*, 2(9), doi:10.1186/s40594-015-0019-7.
- [10] Wendell, K. B. (2008). The theoretical and empirical basis for design-based science instruction for children. *Unpublished Qualifying Paper, Tufts University*
- [11] Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering Design Processes: A Comparison of Students and Expert Practitioners. *Journal of Engineering Education*, 96(4), 359-379.
- [12] Yıldız Aslan, Ö. (2024). *Mühendislik tasarım temelli fen öğretiminin 5. sınıf öğrencilerinin problem çözme becerilerine, mühendislik ve mühendis algılarına etkisinin incelenmesi*. Necmettin Erbakan Üniversitesi Eğitim Bilimleri Enstitüsü. (Yayımlanmamış Yüksek Lisans tezi)
- [13] Başpınar, P., Çakıroğlu, J., & Karahan, E. (2024). The effect of engineering design-based science instruction on 6th-grade students' astronomy understandings. *Science Insights Education Frontiers*, 24(1):3835-3857. DOI: 10.15354/sief.24.or628
- [14] Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. *Journal of Engineering Education*, 94(1), 103-120.
- [15] National Research Council (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. The National Academies Press.

- [16] Ercan, S. & Şahin, F. (2015). Fen eğitiminde mühendislik uygulamalarının kullanımı: tasarım temelli fen eğitiminin öğrencilerin akademik başarıları üzerine etkisi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED)*, 9(1), 128-164.
- [17] Öznlü, Ö., & Çepni, S. (2023). Türkiye’de Mühendislik Tasarım Temelli Öğretim İle İlgili Fen Eğitimi Alanında Yapılan Çalışmaların Tematik Analizi. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*, (56), 890-910.
- [18] Kewalramani, S., Palaiologou, I., & Dardanou, M. (2020). Children’s engineering design thinking processes: the magic of the ROBOTS and the power of BLOCKS (electronics). *Eurasia Journal of Mathematics, Science and Technology Education*, 16(3), Article em1830. <https://doi.org/10.29333/ejmste/113247>
- [19] MEB. (2017). İlköğretim kurumları fen bilimleri dersi öğretim programı. Ankara: Talim ve Terbiye Kurulu Başkanlığı.
- [20] MEB. (2018). *Fen bilimleri dersi öğretim programı (İlkokul ve Ortaokul 3, 4, 5, 6, 7 ve 8. Sınıflar)*. Ankara: Millî Eğitim Bakanlığı Temel Eğitim Genel Müdürlüğü. Taslak Program
- [21] Özlüceci, M., & Kayacan, K. (2023). Fen, mühendislik ve girişimcilik becerileri değerlendirme ölçeği: geçerlilik ve güvenilirlik çalışması. *Millî eğitim dergisi*, 493-512.
- [22] Aydın, G., Saka, M., & Guzey, S. (2018). 4-5-6-7. ve 8. Sınıf Öğrencileri İçin Mühendislik Bilgi Düzeyi Ölçeği. *İlköğretim Online*, 17(2).
- [23] Schnittka, C., & Bell, R. (2011). Engineering design and conceptual change in science: Addressing thermal energy and heat transfer in eighth grade. *International Journal of Science Education*, 33(13), 1861-1887.
- [24] Büyükoztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş. & Demirel, F. (2012). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- [25] Fraenkel R.J. & Wallen E.N. (2006). *How to Design and Evaluate Research in Education*. McGraw-Hill, New York.
- [26] Creswell, J. W. (2014). *Nitel, nicel ve karma yöntem yaklaşımları araştırma deseni* (Çev. Ed. S. B. Demir). Ankara: Eğiten Kitap.
- [27] Yıldız, S. (2017). Sosyal bilimlerde örnekleme sorunu: nicel ve nitel paradigmalardan örnekleme kuramına bütüncül bir bakış. *Kesit Akademi Dergisi*, (11), 421-442.
- [28] Patton, M. Q. (2018). *Nitel araştırma ve değerlendirme yöntemleri*. (M. Bürün ve S.B. Demir, Çev. Ed.). Ankara: Pegem Akademi.
- [29] Moazzen, I., Miller, M., Wild, P., Jackson, L., & Hadwin, A. (2014). Engineering design survey. *Canadian Engineering Education Association (CEE14) Conference, Canmore*. Doi: <https://doi.org/10.24908/pceea.v0i0.5892>.
- [30] Wallas, G. (1926). *The Art Of Thought*. New York, NY: Harcourt Brace
- [31] Ludwig, A. M. (1989). *The Price of Greatness: Resolving the Creativity and Madness Controversy*. New York: Guilford Press.
- [32] Yeşilyurt, E. (2020). Yaratıcılık ve yaratıcı düşünme: Tüm boyut ve paydaşlarıyla kapsayıcı bir derleme çalışması. *OPUS International Journal of Society Researches*, 15(25), 3874-3915. DOI: 10.26466/opus.662721
- [33] Tsenn, J., Atilola, O., McAdams, D. A., & Linsey, J. S. (2014). The effects of time and incubation on design concept generation. *Design Studies*, 35(5), 500-526.
- [34] Yalçın, M. M. (2021). *Öğretimde yaratıcılık ölçeği’nin geçerlik-güvenirlik çalışması ve okul öncesi öğretmenlerinin 48-72 aylar arasındaki çocukların yaratıcılıklarını destekleme durumlarının farklı değişkenler açısından incelenmesi* (Master’s thesis, Necmettin Erbakan University (Turkey)).
- [35] Gilhooly, K. J., Georgiou, G., & Devery, U. (2013). Incubation and creativity: Do something different. *Thinking & Reasoning*, 19(2), 137-149. <https://doi.org/10.1080/13546783.2012.749812>
- [36] Smith, S. M., & Blankenship, S. E. (1991). Incubation and the persistence of fixation in problem solving. *The American journal of psychology*, 61-87. <https://doi.org/10.2307/1422851>
- [37] Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin*, 135(1), 94-120. <https://doi.org/10.1037/a0014212> PMID: 19210055
- [38] Li, Z., Zhang, Z., Zhang, Y., & Luo, J. (2022). 创造性思维的酝酿效应 [Yaratıcı düşüncenin kuluçka etkisi]. *Psikolojik Bilimdeki Gelişmeler*, 30(2), 291-307. <https://doi.org/10.3724/SP.J.1042.2022.00291>
- [39] Ayverdi, L. (2018). *Özel yetenekli öğrencilerin fen eğitiminde teknoloji, mühendislik ve matematiğin kullanımı: FeTeMM yaklaşımı*. Balıkesir Üniversitesi Fen Bilimleri Enstitüsü İlköğretim Anabilim Dalı (Yayımlanmamış Doktora tezi)
- [40] Khaerudin, K., Hendrilia, Y. ., Ayu Pawestri Kusuma Dewi, R. ., Erwin, E., & Amin, A.-. (2023). Innovative Approaches To Research Skill Development Frameworks For Evaluating Problem-Solving Skills In Design Projects In Indonesian Higher Education. *Indonesian Journal of Education (INJOE)*, 3(3), 582-598. Retrieved from <https://injoe.org/index.php/INJOE/article/view/90>
- [41] Pusca, D., & Northwood, D. O. (2018). Design thinking and its application to problem solving. *Global Journal of Engineering Education*, 20(1), 48-53. <http://www.wiete.com.au/journals/GJEE/Publish/vol20no1/06-Pusca-D.pdf>
- [42] Haik, Y., Şahin, T. M., & Sivaloganathan, S. (2011). *Engineering design process*. Stamford: Cengage Learning.
- [43] Jin, S. H., Song, K. I., Shin, D. H., & Shin, S. (2015). A performance-based evaluation rubric for assessing and enhancing engineering design skills in introductory engineering design courses. *International Journal of Engineering Education*, 31(4), 1007-1020.
- [44] Hynes, M. M. (2009). *Teaching middle-school engineering: An investigation of teachers' subject matter and pedagogical content knowledge*. (Unpublished Doctoral Dissertation, Tufts University)
- [45] Miller, R. L., & McFarlane, D. A. (2009). Improving engineering student presentation skills through an integrated curriculum approach. *Journal of Engineering Education*, 98(1), 83-92.
- [46] Coşkun, V. ve Özkaya, A. (2020). Öğretmen eğitiminde mühendislik odaklı disiplinlerarası iş birliğine dayalı STEM uygulaması ve ders izlencesi. *İhlara Eğitim Araştırmaları Dergisi*, 5(2), 327-361. <https://doi.org/10.47479/ihed.823328>

## Ek A

### Mühendislik Tasarım Süreci Becerileri Ölçeği

TEST MADDELERİ	1	2	3	4	5
1. Sınıfta verilen mühendislik projesini gerçekleştirirken etkili teknikler kullanarak özgün tasarım fikirleri üretebilirim.	1	2	3	4	5

2. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla beyin fırtınası yaparak tasarımlar üretebilirim.	1	2	3	4	5
3. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen tasarımlar arasından uygun olan birisini eldeki bilgi ve verilerle seçebilirim.	1	2	3	4	5
4. Sınıfta verilen mühendislik projesinin problem durumuna göre istenen tasarımı yapabilirim.	1	2	3	4	5
5. Sınıfta verilen mühendislik projesini gerçekleştirirken tasarımın amacına ulaşması için gerekli koşulları tanımlayabilirim.	1	2	3	4	5
6. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen tasarımlarla ilgili ön deneme sonuçlarına bağlı olarak tasarımlarda değişikliğe gidebilirim.	1	2	3	4	5
7. Sınıfta verilen mühendislik projesini gerçekleştirirken beklenmeyen durumlar ortaya çıktığında tasarımla ilgili yeni ölçülebilir değişkenler belirleyebilirim.	1	2	3	4	5
8. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen her tasarımın çevreye olası etkisini tahmin edebilirim.	1	2	3	4	5
9. Sınıfta verilen mühendislik projesinin problem durumundan yapılacak tasarımın amacını belirleyebilirim.	1	2	3	4	5
10. Sınıfta verilen mühendislik projesini gerçekleştirirken tasarımla ilgili karşılaşılan sorunlara alternatif çözümler üretmede zorlanabilirim.	1	2	3	4	5
11. Sınıfta verilen mühendislik projesini gerçekleştirirken beklenmeyen durumlar ortaya çıktığında tasarımdaki sorunu tekrardan revize etmede zorlanabilirim.	1	2	3	4	5
12. Sınıfta verilen mühendislik projesini gerçekleştirirken tasarımla ilgili sıra dışı olan yaratıcı fikirleri önemsemem.	1	2	3	4	5
13. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen her tasarımın topluma olası yansımalarını tahmin etmede zorlanabilirim.	1	2	3	4	5
14. Sınıfta verilen mühendislik projesini gerçekleştirirken tasarımla ilgili ihtiyaç duyulan bilgi kaynaklarına (kitap, dergi, kütüphane vs) nasıl ulaşılacağını bilemeyebilirim.	1	2	3	4	5
15. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen tasarımlarla ilgili yine de yeni bilgiler toplamaya devam etmede zorlanabilirim.	1	2	3	4	5
16. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen tasarımları detaylandırarak geliştiremeyebilirim.	1	2	3	4	5
17. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen tasarımlar arasından uygun olanlardan birisini ihtiyaç analizi yaparak seçmede zorlanabilirim.	1	2	3	4	5
18. Sınıfta verilen mühendislik projesini gerçekleştirirken beklenmeyen durumlar ortaya çıktığında tasarımın amacını yeniden revize edemeyebilirim.	1	2	3	4	5
19. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen tasarımlarla ilgili yine de yeni kavram ve fikirler üretmeye devam edemem.	1	2	3	4	5
20. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen tasarımlarla ilgili grubun değişen kararlarına uymak istemem.	1	2	3	4	5
21. Sınıfta verilen mühendislik projesini gerçekleştirirken kendimce bireysel olarak tasarımlar geliştiremeyebilirim.	1	2	3	4	5
22. Sınıfta verilen mühendislik projesini gerçekleştirirken tasarımla ilgili olarak mühendis, bilim insanı vs. gibi deneyimli kişilerden bilgi edinebilirim.	1	2	3	4	5
23. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen tasarımlarla ilgili meydana getirilen ürünü test edebilirim.	1	2	3	4	5
24. Sınıfta verilen mühendislik projesini gerçekleştirirken grupla geliştirilen tasarımlar için değerlendirme kriterleri oluşturabilirim.	1	2	3	4	5
25. Sınıfta verilen mühendislik projesinin problem durumundan yapılacak tasarımla ilgili hangi değişkenlerin ölçülmesi gerektiğini belirleyebilirim.	1	2	3	4	5
26. Mühendislik proje tasarımlarının raporunu dinleyicilerin seviyesine göre açıklamakta zorlanabilirim.	1	2	3	4	5
27. Mühendislik proje tasarımlarının raporunu genellikle çizim veya şekillerle anlatmayı tercih etmem.	1	2	3	4	5
28. Mühendislik proje tasarımlarının sunum raporunu belirli kısa bir zaman diliminde dinleyicilere aktaramayabilirim.	1	2	3	4	5
29. Mühendislik proje tasarımlarının sunum raporunu etkili bir dille anlatımda zorlanabilirim.	1	2	3	4	5

Research Article

Received: date:11.29.2024

Accepted: date:24.12.2024

Published: date:31.12.2024

# Child Suicide News: A Comprehensive Case Study from Turkish Media

Nazan Çakırer Çalbayram <sup>1\*</sup>, Gamze Balta<sup>2</sup>

<sup>1</sup>Canakkale Onsekiz Mart University, Department of Nursing, Çanakkale, Turkey; nazançalbayram@gmail.com

<sup>2</sup>Dörtçelik Children's Diseases Hospital, Bursa, Turkey; ggamzebalta@hotmail.com

Orcid: 0000-0003-2048-608X<sup>1</sup> Orcid: 0000-0002-6212-7491<sup>2</sup>

\*Correspondence: nazançalbayram@gmail.com

**Abstract:** This study aims to analyze news on child suicides and suicide attempts in Turkey's high-circulation national newspapers over a ten-year period. This qualitative study systematically content-analyzed news on suicidal behavior in three high-circulation national newspapers in Turkey from January 1, 2013, to December 31, 2022. Data were collected by searching for the keywords "child" and "suicide" and analyzed. The study found that 60.2% of suicidal children were male, and 39.8% were female, with a mean age of 15.09±2.16 years. The majority were high school (56.2%) and middle school (38.1%) students. Suicidal behavior peaked in 2022, with the highest incidences occurring in winter (29.9%) and spring (26.2%). Depression (30%) was the most frequently reported reason for suicidal behavior, followed by family problems (19.8%) and the Blue Whale Challenge (16.5%). Hanging (47.5%) was the most common method of suicide, followed by jumping from heights (23.8%). The increasing trend of suicidal behavior among children, particularly adolescents, highlights the need for targeted prevention strategies. The findings suggest that mental health support, family interventions, and societal awareness are crucial in addressing this issue. The influence of digital challenges and the need for gender-specific approaches also warrant further attention.

**Keywords:** child, news, suicide, content analysis

## 1. Introduction

Suicide, a significant public health issue, occurs when an individual intentionally harms themselves with the intent to end their life, resulting in death [1]. Durkheim defined suicide as "any death which results directly or indirectly from a positive or negative act of the victim themselves, which they know will produce this result" [2]. Conversely, when an individual harms themselves with the intent to end their life but does not result in death, it is referred to as a suicide attempt. It is important to note that not all suicide attempts result in injury [1].

According to the World Health Organization (WHO), over 700.000 individuals die by suicide annually. This equates to one person dying by suicide every 40 seconds, with suicide accounting for approximately 1.3% of all deaths globally [3]. In the United States, reports from the Centers for Disease Control and Prevention (CDC) indicate that suicide rates increased by 36% between 2000 and 2021, making it the second leading cause of death among children aged 10-14 years as of 2023 [4]. In Turkey, the crude suicide mortality rate stands at 4.88 per 100.000, with an observed increase in suicides among the 0-19 age group compared to previous years [5]. Studies have shown a corresponding rise in suicidal thoughts and attempts, corroborating these statistics [6, 7, 8].

Suicides occur when multiple factors converge over time, making it crucial to identify the temporal and causative factors that lead individuals to decide to commit suicide [9, 10, 11]. Literature indicates that childhood suicide attempts are associated with various factors, including mental disorders, family conflicts, sexual abuse, peer victimization, and online virtual games [12, 13, 14, 15, 16]. Suicides, often triggered by a negative and precipitating event, can be either planned or impulsive [17]. Children who

---

attempt suicide employ various methods, such as hanging, jumping from heights, and suffocation [18, 19].

The greatest risk factor for suicides, which can often be prevented with timely and appropriate interventions, is a previous suicide attempt [1, 20]. Previous studies have established that children and adolescents who attempt suicide often have a history of previous attempts [21, 22, 23]. For instance, Doğan et al. (2018) found that 8% of children who attempted suicide did not regret their attempt and continued to harbor a desire for death [24]. Moreover, a notable factor in children who have attempted suicide is the presence of psychiatric diagnoses [25]. Research has shown that adolescents who attempt suicide frequently have a diagnosis of depression [26, 27]. These findings underscore the necessity of improved monitoring of suicide and suicide attempts, which the WHO has identified as a public health priority [20].

To prevent suicide and suicide attempts, it is essential to identify their causes. Family members, teachers, and health professionals should remain vigilant and consider all potential factors. A comprehensive approach requires determining when, where, and why suicides and attempts occur, and identifying influencing factors.

This study aims to analyze print media reports on child suicides and attempts to inform better prevention strategies. By examining the portrayal and frequency of child suicides in the media, we seek to understand the media's role in public perception and its potential impact on preventive measures.

## **2. Materials and Methods**

### **2.1. Study Design and Scope**

This study covers the period from January 1, 2013, to December 31, 2022, and employs a qualitative research design. It involves a systematic content analysis of news reports on suicidal behavior from three national newspapers with high circulation in Turkey (names of the newspapers are withheld for ethical reasons). Content analysis is a versatile method used across various disciplines, providing researchers with a tool to conduct an in-depth analysis of texts and uncover significant features within the content.

### **2.2. Data Collection**

Newspapers were selected based on their daily circulation figures. The newspapers were accessed online, and searches were conducted using the keywords "child" and "suicide." The resulting news reports were analyzed and digitized according to predetermined variables. Duplicate reports were identified and removed to ensure the uniqueness of the data set. A total of 244 news reports, deemed relevant for the study's objectives, were selected for content analysis.

### **2.3. Variables Analyzed**

The analysis evaluated the following variables:

- Age of the child
- Gender
- Education status
- Place of residence
- Year of suicide
- Season of suicide
- Method of suicide
- Cause of suicide

### **2.4. Data Analysis**

The data were analyzed using SPSS (Statistical Package for the Social Sciences) version 23.0. Descriptive statistics were employed to provide a comprehensive overview of the current situation. Descriptive statistics aim to depict the existing conditions based on the collected data.

### **2.5. Limitations of the Study**

This study has several limitations. Accessing official data was challenging, and many suicide incidents remain unreported. Some incidents are not covered by the media due to perceived lack of

newsworthiness, which may affect the completeness of the data. Therefore, the findings may not represent a complete national picture. Additionally, the study only includes three high-circulation national newspapers, excluding other newspapers and communication mediums such as radio and television. Consequently, the results should be interpreted within these limitations.

### 3. Results

Among the children who committed suicide, 60.2% were male and 39.8% were female, with a mean age of 15.09±2.16 years (range: 8-18 years). The educational status of 43% of the children was reported in the news. Of these, 56.2% were high school students, and 38.1% were middle school students. Additionally, 77.5% of the children resided in city centers. Regionally, 19.3% lived in the Aegean, 15.2% in the Marmara, and 14.8% in the Mediterranean regions (Table 1).

**Table 1.** Socio-Demographic Characteristics of Children Who Committed Suicidal Behaviour (n=244)

Characteristics	Groups	
	X±Sd	min.- max.
Age	15.09±2.16	8-18
	<b>n</b>	<b>%</b>
12 years and older	230	94.3
Below 12 years of age	14	5.7
<b>Gender</b>		
Female	97	39.8
Male	147	60.2
<b>Education Status</b>		
Not specified	139	57
Specified	105	43
High School	59	56.2
Middle School*	40	38.1
Primary School*	5	4.8
University*	1	1
<b>Variables</b>	<b>n</b>	<b>%</b>
<b>Place of Residence</b>		
City centre	189	77.5
Other segments	55	22.5
<b>Geographical Regions</b>		
Aegean region	47	19.3
Marmara region	37	15.2
Mediterranean region	36	14.8
Black Sea region	34	13.9
Southeastern Anatolia	34	13.9
Central Anatolia	32	13.1
Eastern Anatolia	24	9.8

\*Percentage distribution of educational status in the news content (n=105)

The provinces with the highest frequency of suicidal behaviors were Istanbul, Adana, Şanlıurfa, and İzmir. Conversely, the least common provinces were Ankara, Eskişehir, and Balıkesir.

**Table 2.** Variables Related to Children's Suicidal Behaviour (n= 244)

Variables	n	%
<b>Years</b>		
2022	36	14.8
2021	33	13.5
2020	26	10.7
2019	23	9.4
2014	23	9.4
2018	22	9
2013	22	9
2016	21	8.6
2017	20	8.2
2015	18	7.4
<b>Seasons</b>		
Winter	73	29.9
Spring	64	26.2
Summer	54	22.1
Autumn	53	21.7
<b>Patterns of suicidal behaviour</b>		

Hanging	116	47.5
Firearm	58	23.8
Jumping from height	57	23.4
Other	11	4.5
Suffocation in Water	2	0.8
<b>Causes of suicidal behaviour</b>		
Not specified	153	62.7
Specified	91	37.3
<b>Reasons *</b>		
Depression	30	33
Other	20	22
Family problems	18	19.8
Blue Whale Challenge	15	16.5
Sexual abuse	8	8.8
<b>Consequences of suicidal behaviour</b>		
Death	227	93.0
Survival	17	7.0

\*: Percentage distribution of reasons for suicidal behaviour in the news content (n=91)

Suicidal behaviors were found to occur most frequently in 2022 (14.8%), followed by 2021 (13.5%) and 2020 (10.7%). Seasonally, 29.9% of the suicides occurred in winter and 26.2% in spring (Table 2). The months with the highest incidences of suicide were February (9.8%), November (9.8%), and December (9.8%), while August had the lowest incidence (4.5%). Regarding the methods of suicide, 47.5% of the children chose hanging, 23.8% chose jumping from heights, and 23.4% used firearms (Table 2).

It was found that 57.4% of the children performed suicidal behaviors at home, while 42.6% did so in out-of-home environments. Additionally, 14.3% of the children left a note, whereas 85.7% did not. It was observed that 93% of suicidal behaviors resulted in death, while 7% resulted in survival (Table 2). The reasons for 37.3% of the suicidal behaviors were mentioned in the news reports. Among these, 33% of the children attempted suicide due to depression, 20% due to other reasons (such as peer relations and forced marriage), and 19.8% due to family problems (Table 2).

**Table 3.** Distribution of Children's Methods and Reasons for Suicidal Behaviour by Gender (n=244)

Ways of performing suicidal behaviour	Gender				
	Female		Male		
	n	%	n	%	
Hanging	34	35.1	82	55.8	
Firearm	37	38.1	20	13.6	
Jumping from height	22	22.7	36	24.5	
Other	1	1	1	0.7	
Suffocation in Water	3	3.1	8	5.4	
<b>Causes of suicidal behaviour</b>					
Not specified	63	64.9	90	61.2	
Specified	Depression	9	9.3	21	14.3
	Other	9	9.3	9	6.1
	Family problems	2	2.1	13	8.8
	Blue Whale Challenge	6	6.2	2	1.4
	Sexual abuse	8	8.2	12	8.2
<b>Consequences of suicidal behaviour</b>					
Death	86	37.9	141	62.1	
Survival	11	64.7	6	35.3	

According to Table 3, which shows the distribution of the methods and reasons for suicidal behaviors by gender, 38.1% of the female children committed suicide by jumping, while 55.8% of the male children committed suicide by hanging. The study found that females frequently committed suicidal behavior due to depression and family problems, whereas males were more likely to commit suicidal behavior due to depression and the Blue Whale Challenge. It was also determined that 37.9% of females and 62.1% of males committed suicidal behavior resulting in death, indicating a significant difference between genders (U: 6612.0; p<0.05).

---

#### 4. Discussion

Suicides, which continue to increase over the years, have far-reaching physical, emotional, and economic consequences [20]. In our study, the mean age of children who exhibited suicidal behavior was  $15.09 \pm 2.16$  years (range: 8-18 years), with 94.3% being 12 years or older. This aligns with Uran (2019), who reported a mean age of  $15.29 \pm 1.78$  years (range: 7-18 years) among children and adolescents who attempted suicide, with the 15-year-old age group being the most frequent [25]. Similarly, Talay and Kanğın (2023) found that 96.4% of self-destructive poisonings involved children over 13 years of age [28]. Li et al. (2023) also identified the highest prevalence of suicidal thoughts among 12-year-old children [29]. These findings suggest that suicidal behavior in children increases with age and is more prevalent during adolescence.

Our study found a higher frequency of suicidal behavior among males (60.2%) compared to females (39.8%). Among those who exhibited suicidal behavior, 62.1% of males resulted in death, while 64.7% of females survived. No directly comparable studies were found in the literature. It is hypothesized that female children, often stigmatized as fragile and weak, may attempt suicide to draw attention to their problems, whereas male children, whose failures are less socially accepted, may resort to more lethal methods. Additionally, considering that child marriages significantly contribute to suicidal behavior among females, raising awareness about this issue across all societal segments is crucial.

The majority of children in our study were high school (56.2%) and middle school (38.1%) students. Other studies also found that high school students frequently attempt suicide [21, 25]. The high prevalence of suicidal behaviors in middle and high school ages, coinciding with adolescence, may be attributed to physiological and psychological changes and the effort for self-realization during this period. Adolescence is marked by significant stressors, including family problems, neglect and abuse, peer issues, academic challenges, and future anxiety. Adolescents who cannot find alternative solutions to these stressful situations may resort to suicidal behavior as a coping mechanism.

Our study observed a continuous increase in the rates of suicidal behavior over the years, reaching the highest level in 2022. This trend is consistent with other studies showing increased suicide attempts. Plemmons et al. (2018) reported an increase in suicidal thoughts and attempts among children over the last decade [8], and Burstein et al. (2019) noted a doubling of emergency room visits for similar reasons [7]. Seasonally, our study found that children most frequently engaged in suicidal behavior during winter (29.9%) and spring (26.2%). Similar seasonal patterns were reported by Kırıcı et al. (2023) and Elmas and Uluğ (2019) [21, 30]. Although no studies have shown a statistically significant difference between seasons and suicidal behaviors, it is hypothesized that increased exposure to school and peer-related problems during the winter may contribute to this pattern.

Before the suicide attempt, Dinleyici et al. (2018) found that 30.4% of children had a psychiatric diagnosis [31], while Görücü et al. (2020) reported a rate of 50.8% [32]. Depression was identified as the most common diagnosis in several studies [27, 33, 34]. Many children had previously attempted suicide one or more times [22, 23, 35]. In our study, while the previous medical histories of the children were not available from the news content, it was determined that depression was the reason for suicidal behavior in 30% of the cases. This suggests that observations and interventions for depressed children by families, teachers, and healthcare workers may be insufficient.

Family problems (19.8%), the Blue Whale Challenge (16.5%), sexual abuse (8.8%), and peer relations were significant factors in the suicidal behavior of children. Similarly, other studies have found that children frequently attempt suicide due to domestic problems [16, 33, 34]. Kavurma et al. (2023) reported significant negativity in family assessment scales, higher divorce rates among parents, and lower parental attachment scores in adolescents with self-harm behavior [36]. Destructive attachment patterns developed in infancy were also associated with suicidal and self-harming behaviors in childhood and adolescence [37]. Sexual abuse has been identified as a contributing factor to suicide in multiple studies [14, 38, 39]. Kılıç (2016) found higher rates of suicide attempts (32%) and suicidal ideation (30%) in sexually abused children compared to a control group.

The Blue Whale Challenge, identified as the second leading cause of suicidal behavior in our study, involves children and adolescents completing tasks with threatening content over 50 days. Talu (2019)



---

analyzed this phenomenon using Shneidman's "suicide cube" model, finding that the tasks overlapped with factors of pain, pressure, and restlessness [41]. The game induces psychological depression and culminates in the final task of suicide [13, 42]. The immersiveness of the Blue Whale Challenge may drive children with computer game addiction to attempt suicide.

In terms of methods, 47.5% of children in our study performed suicidal behavior by hanging, 23.8% by jumping from heights, and 23.4% by firearms. This is consistent with Ruch et al. (2021), who found that 78.4% of children used hanging [17], and Laido et al. (2017), who reported that 57.4% used hanging and 17% used jumping [18]. In contrast, other studies in Turkey have found drug overdose to be the most common method [21, 25, 43], likely due to the accessibility of drugs.

Durkheim (2013) posited that religion has a protective effect against suicides, with higher suicide rates observed among non-religious individuals [2]. Although the majority of the Turkish population is Muslim (90%), and suicide is explicitly forbidden in the Qur'an, our study found an increasing rate of suicide among children each year. This discrepancy suggests that the protective influence of religious beliefs may be waning or insufficient in preventing suicidal behavior among children.

## 5. Conclusions

This study highlights the increasing trend of suicidal behavior among children in Turkey over a decade, with significant findings regarding age, gender, educational status, and seasonal patterns. The majority of suicidal behaviors were observed among adolescents, particularly high school and middle school students, indicating that the transition period from childhood to adulthood is a critical phase requiring focused attention. The higher frequency of suicidal behavior among males, and the lethality of their methods compared to females, underscores the need for gender-specific prevention strategies.

The significant role of depression, family problems, and the Blue Whale Challenge as contributing factors suggests that targeted interventions addressing mental health, family dynamics, and digital safety are essential. This study emphasizes the importance of comprehensive observation and intervention by families, educators, and healthcare professionals for children showing signs of depression and other mental health issues.

Despite the high prevalence of religious beliefs in Turkey, the increasing rate of suicide suggests that additional cultural, social, and psychological factors need to be addressed. This discrepancy highlights the potential waning of the protective influence of religious beliefs or their insufficiency in preventing suicidal behavior among children.

Overall, this study underscores the urgent need for multifaceted suicide prevention strategies that include mental health support, family interventions, educational system involvement, and societal awareness. Developing and implementing comprehensive suicide prevention programs tailored to the specific needs of different age groups and genders is crucial. Furthermore, enhancing digital literacy and safety, particularly regarding online challenges like the Blue Whale Challenge, is imperative to protect vulnerable children.

Future research should focus on longitudinal studies to better understand the evolving patterns of suicidal behavior and the effectiveness of various intervention strategies. Collaboration between policymakers, mental health professionals, educators, and community leaders is essential to create a supportive environment that fosters the well-being and resilience of children and adolescents.

**Author Contributions:** Conceptualization, N.Ç.Ç.; methodology, N.Ç.Ç.; software, G.B.; validation, N.Ç.Ç., G.B.; formal analysis, G.B.; investigation, N.Ç.Ç., G.B.; resources, N.Ç.Ç.; data curation, G.B.; writing—original draft preparation, N.Ç.Ç.; writing—review and editing, G.B., N.Ç.Ç.; visualization, G.B.; supervision, N.Ç.Ç.; project administration, N.Ç.Ç. All authors have read and agreed to the published version of the manuscript.

**Funding:** No external funding.

**Conflicts of Interest:** The authors have no conflict of interest.

**Acknowledgments:** The authors would like to extend their gratitude to the national newspapers that contributed to this research by providing access to their archives, which were instrumental in the collection of data.

**Information Note:** It was previously presented at the 3rd International Mediterranean Pediatric Nursing Congress.

## References

- [1] National Institute of Mental Health. Suicide Prevention. Available From: [https://www.nimh.nih.gov/health/topics/suicide-prevention#part\\_9890](https://www.nimh.nih.gov/health/topics/suicide-prevention#part_9890). Date of Access: 25.07.2023.
- [2] Durkheim E, İlgelen Z. (Çev.) İntihar (1.Baskı). Pozitif Yayınları, İstanbul, 2013, s.3.
- [3] World Health Organization. Suicide Worldwide İn 2019, Global Health Estimates. 2019. Available from: <https://www.who.int/publications/i/item/9789240026643>.
- [4] Centers for Disease Control and Prevention. Facts About Suicide Preventing Suicide 2023. Available From: <https://www.cdc.gov/suicide/facts/index.html>.
- [5] Türkiye İstatistik Kurumu İntihar İstatistikleri. 2022. Available From: <https://data.tuik.gov.tr/Bulten/Index?p=Olum-ve-Olum-Nedeni-Istatistikleri-2022-49679>.
- [6] Gaylor E.M, Krause K. H, Welder L. E, Cooper A. C, Ashley C, Mack K. A, Crosby A. E, Trinh E, Ivey-Stephenson A. Z, Whittle L. Suicidal thoughts and behaviors among high school students, youth risk behavior survey, United States, 2021. *MMWR Suppl*, 2023; 72(Suppl-1): 45-54.
- [7] Burstein B, Agostino H, Greenfield B. Suicidal attempts and ideation among children and adolescents in US emergency departments, 2007-2015. *JAMA Pediatrics*, 2019; 173(6): 598-600.
- [8] Plemmons G, Hall M, Douppnik S, Gay J, Brown C, Browning W, Williams D. Hospitalization for suicide ideation or attempt: 2008-2015. *Pediatrics*, 2018; 141(6).
- [9] Ruch D. A, Heck K. M, Sheftall A. H, Fontanella C. A, Stevens J, Zhu M, ... Bridge J. A. Characteristics and precipitating circumstances of suicide among children aged 5 to 11 years in the United States, 2013-2017. *JAMA network open*, 2021; 4(7): e2115683-e2115683.
- [10] Erdoğan B, Celbiş O, Öner B.S. Psychological dynamics and risk factors of child and adolescent suicide. Öztürk (Ed.), *Aile psikopatolojisi*, 1. Baskı, Türkiye Klinikleri, Ankara, 2021, 127-132.
- [11] Klonsky E. D, May A. M, Saffer B. Y. Suicide, suicide attempts, and suicidal ideation. *Annual Review Of Clinical Psychology*, 2016; 12: 307-330.
- [12] Hıdır M, Kılınç F. E. Content analysis of studies on sexual abuse in childhood. *Turkish Journal of Health Research*, 2020; 1(1): 23-40.
- [13] Yılmaz M, Candan F. Oyun sanal intihar gerçek: "the blue whale challenge/mavi balina" oyunu üzerinden kurulan iletişimin neden olduğu intiharlar üzerine kuramsal bir değerlendirme. *Akdeniz Üniversitesi İletişim Fakültesi Dergisi*, 2018; (29. Özel Sayısı), 270-283.
- [14] Soylu N, Alpaslan A. H, Aslan M, Kaya İ. Suicide probability and related factors in sexually abused adolescents. *International Social Research Congress*, 2017; 1611.
- [15] Van Geel M, Vedder P, Tanilon J. Relationship between peer victimization, cyberbullying, and suicide in children and adolescents: a meta-analysis. *JAMA Pediatrics*, 2014; 168(5): 435-442.
- [16] Ünlü G, Aksoy Z, Ersan E. E. Evaluation of child and adolescents with attempted suicide. *Pamukkale Medical Journal*, 2014; 7(3): 176.
- [17] Ruch DA, Heck KM, Sheftall AH, Fontanella CA, Stevens J, Zhu M, . . . Bridg JA. Characteristics and precipitating circumstances of suicide among children aged 5 to 11 years in the United States, 2013-2017. *JAMA network open*, 2021; 4(7), e2115683-e2115683.
- [18] Laido Z, Voracek M, Till B, Pietschnig J, Eisenwort B, Dervic K, Niederkrotenthaler T. Epidemiology of suicide among children and adolescents in Austria, 2001-2014. *Wiener Klinische Wochenschrift*, 2017; 129: 121-128.
- [19] Skinner R, Mcfaull S. Suicide among children and adolescents in Canada: trends and sex differences, 1980-2008. *CMAJ*, 2012; 184(9): 1029-1034.
- [20] World Health Organization. *Suicide*. 2021. Available from: <https://www.who.int/news-room/fact-sheets/detail/suicide>.
- [21] Elmas B, Uluğ N. Evaluation of patients under 18 years of age who were admitted to the emergency department with suicidal attempt. *Journal of Biotechnology And Strategic Health Research*, 2019; 3(1): 38-43.
- [22] Akin E, Berkem M. Anger and Impulsivity in Adolescents Attempting Suicide. *Marmara Medical Journal*, 2012; 25(3): 148-152.
- [23] Hawton K, Bergen H, Kapur N, Cooper J, Steeg S, Ness J, Waters K. Repetition of self-harm and suicide following self-harm in children and adolescents: findings from the multicentre study of self-harm in England. *Journal of Child Psychology and Psychiatry*, 2012; 53(12): 1212-1219.
- [24] Doğan M, Öztürk S, Esen F, Demirci E, Öztürk M. A. Evaluation of child and adolescents who attempted suicide. *Bozok Tıp Dergisi*, 2018; 8(3): 30-34.
- [25] Uran P. Multidimensional assessment of the psychopathologies sociodemographic and suicide behavior characteristics of the children and adolescents referred for attempting suicide *Akdeniz Medical Journal*, 2019; 5(1): 94-103.
- [26] Fonseca-Pedrero E, Al-Halabí S, Pérez-Albéniz A, Debbané M. Risk and protective factors in adolescent suicidal behaviour: a network analysis. *International Journal of Environmental Research And Public Health*, 2022; 19(3): 1784.
- [27] Stewart J. G, Esposito E. C, Glenn C. R, Gilman S. E, Pridgen B, Gold J, Auerbach R. P. Adolescent self-injurers: comparing non-ideators, suicide ideators, and suicide attempters. *Journal of Psychiatric Research*, 2017; 84: 105-112.

- [28] Talay M. N, Kağın M. A single center experience in poisoning cases followed in the pediatric intensive care unit. *Harran Üniversitesi Tıp Fakültesi Dergisi*, 2023; 20(1): 45-51.
- [29] Li P, Wang Y, Kovess-Masfety V, Su X, Hou X, Li M, ... Xu G. Suicidal ideation and its relationship with mental health reported by different informants among children aged 6–12. *Journal of Affective Disorders*, 2023; 334: 271-277.
- [30] Kırıcı G. S, Öztürk D. U, Bölükbaşı H. S, Çakır H, Özer E. Çocuk acil servisine 2016-2021 yılları arasında başvuran intoksikasyon olgularının değerlendirilmesi. *Bulletin of Legal Medicine/Adli Tıp Bülteni*, 2023; 28(1).
- [31] Dinleyici M, Kırıl E, Yazar A. S, Şahin S. Evaluation of risk factors for suicide attempts among adolescent with drug intoxication. *Osmangazi Tıp Dergisi*, 2018; 41(2): 166-172.
- [32] Görücü R. A, Eraslan A. N, Göker Z. The effect of the seasons on impulsive suicide attempt: a descriptive study. *Turkish Journal of Pediatric Disease*, 2020; 15(6): 482-487.
- [33] Akman A. Ö. Evaluation of Suicide Attempts by Drug Overdose in the Adolescent Age Group. *Turkish Journal of Pediatric Disease*, 2019; 13(1): 7-12.
- [34] Özsoylu S, Kondolot M, Akyıldız B, Özmen S, Demirci E. İntihar girişimi nedeniyle yoğun bakıma yatırılan çocuk ve ergenlerin değerlendirilmesi. *Medical Journal of Bakirkoy*, 2019; 15(2).
- [35] Czyn E. K, King C. A. Longitudinal trajectories of suicidal ideation and subsequent suicide attempts among adolescent inpatients. *Journal of Clinical Child & Adolescent Psychology*, 2015; 44(1): 181-193.
- [36] Kavurma C, Bilaç Ö, Önder A, Doğan Y, Uzunoğlu G, Adanır A, Kandemir H. Evaluation of family functioning, attachment to parents and friends, decision-making styles and problem solving skills in adolescents with non suicidal self-injury behavior. *Turkish Journal of Clinics and Laboratory*, 2023; 14(2): 312-320.
- [37] Silva Filho O. C. D, Avanci J. Q, Pires T. D. O, de Vasconcellos Carvalhaes Oliveira R, Assis S. G. Attachment, suicidal behavior and self-harm in childhood and adolescence: a study of a cohort of Brazilian schoolchildren. *BMC pediatrics*, 2023; 23(1): 403.
- [38] Çetin K, Altın H. Sociodemographical evaluation of children with sexual abuse. *Necmettin Erbakan University Ereğli Faculty of Education Journal*, 2019; 1(2): 135-147.
- [39] İmren S. G, Ayaz A, Yusufoglu C, Arman A. R. Clinical features and risk factors related with suicide attempts in sexually abused children and adolescents. *Marmara Medical Journal*, 2013; 26(1): 11-16.
- [40] Kılıç, Yasemin. Cinsel istismar mağduru olan ve olmayan 11 ve 18 yaş arasındaki bir grup kız çocuğunun sosyo-demografik özellikler ve psikolojik belirtiler bakımından karşılaştırılması. 2016. Master's Thesis. Fatih Sultan Mehmet Vakıf Üniversitesi, Sosyal Bilimler Enstitüsü.
- [41] Talu E. Bir oyundan ötesi: "mavi balina" the blue whale adlı oyunun shneidman'ın "intihar küpü" modeli bağlamında analizi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 2019; 20(3): 1044-1078.
- [42] Çetin S. Online games and suicide: Blue whale example. *Milli kültür araştırmaları dergisi*, 2022; 6(1): 61-73.
- [43] Gizli Çoban Ö, Gül M, Önder A. Evaluation of children and adolescents admitted to emergency service with suicide attempt. *Journal of Pediatric Emergency and Intensive Care Medicine (Turkey)*, 2022; 9(3).

Review Article

Received: date:04.04.2024  
Accepted: date:24.11.2024  
Published: date:31.12.2024

# Yenilenebilir Enerji Kaynaklarının Hizmet Sektöründe Kullanımı: Literatür Çalışması

Afra Sudenaz Uslu<sup>1</sup>, Yavuz Özdemir<sup>2\*</sup>, Mustafa Yıldırım<sup>3</sup> Hakan Kavak<sup>4</sup>

<sup>1</sup>Industrial Engineering, Istanbul Health and Technology University, Istanbul Türkiye; afrasudenaz@gmail.com

<sup>2</sup>Industrial Engineering, Istanbul Health and Technology University, Istanbul Türkiye; yavuz.ozdemir@istun.edu.tr

<sup>3</sup>Industrial Engineering, Istanbul Health and Technology University, Istanbul Türkiye; mustafa.yildirim@istun.edu.tr

<sup>4</sup>Industrial Engineering, Istanbul Health and Technology University, Istanbul Türkiye; hakankavak34@gmail.com

Orcid: 0009-0006-7464-111X<sup>1</sup>, Orcid: 0000-0001-6821-9867<sup>2</sup>, Orcid: 0000-0001-5709-4421<sup>3</sup>, Orcid: 0009-0004-8753-5696<sup>4</sup>

\*Correspondence: yavuz.ozdemir@istun.edu.tr

**Öz:** Bu literatür makalesi, yenilenebilir enerji kaynaklarının hizmet sektöründeki etkilerini kapsamlı bir şekilde incelemektedir. Makale çalışması, enerji kavramının temel ilkelerinden başlayarak, yenilenebilir enerjinin perakende ticareti ve ulaştırma hizmetlerindeki uygulamalarına odaklanmaktadır. Ayrıca, Türkiye ve Dünya genelindeki perakende sektöründe yürütülen yenilenebilir enerji projeleri analiz edilmiştir. Güneş enerjisi, rüzgâr enerjisi, biyokütle ve biyoenerji projeleri ile jeotermal enerji kullanımı gibi farklı yenilenebilir enerji türlerinin hizmet sektöründeki uygulama ve etkileri üzerinde durulmuştur. Literatür taramasında yapılan çalışmaların yıllara göre konu kapsam dağılımlarına yer verilmiştir. Enerji kavramı üzerine yapılan çalışmaların çoğunda, yenilenebilir enerjiye genel olarak değinilmiştir. Jeotermal enerji, geçmiş yıllarda daha fazla incelenirken, son dönem çalışmaların farklı enerji kaynaklarına yoğunlaştığı görülmüştür.

**Anahtar kelimeler:** Yenilenebilir enerji, hizmet sektörü, perakende ticareti

## Use of Renewable Energy Resources in The Service Sector: Literature Review

**Abstract:** This literature article comprehensively examines the effects of renewable energy sources in the service sector. The article study focuses on the applications of renewable energy in retail trade and transportation services, starting from the basic principles of the concept of energy. In addition, renewable energy projects carried out in the retail sector in Turkey and around the world have been analyzed. The application and effects of different types of renewable energy such as solar energy, wind energy, biomass and bioenergy projects and geothermal energy use in the service sector have been emphasized. The scope of the studies conducted in the literature review is given according to years. In most of the studies conducted on the concept of energy, renewable energy has been generally addressed. While geothermal energy has been studied more in the past years, it has been observed that recent studies have focused on different energy sources.

**Keywords:** Renewable energy, service sector, retail trade

### 1. Giriş

Küresel nüfus artışıyla birlikte, enerji talebi ve insan ihtiyaçları hızla yükselmektedir. Ancak, bu artan enerji ihtiyacını karşılamak için sıklıkla kullanılan fosil kaynaklar, doğanın dengesini bozup ve çevreyi kirleterek ciddi sorunlara sebep olmaktadır. Fosil kaynakların hızla tükenmesi ve çevresel etkilerin artması, birçok ülkeyi yenilenebilir enerji kaynaklarına yönlendirmiştir. Ülkeler, enerji politikalarını belirlerken sadece sosyal faktörleri değil, aynı zamanda ekonomik, teknik ve çevresel

faktörleri de göz önünde bulundurmaktadır. Bu faktörler, enerji yatırımlarının temel kriterlerini oluşturmaktadır [1,2].

Enerji, iş yapma kapasitesi olarak tanımlanmaktadır ve enerji kaynakları, kullanım amaçlarına ve dönüştürülebilirlik özelliklerine göre sınıflandırılmaktadır. Birincil enerji kaynakları, doğrudan doğada bulunan ve dış müdahaleye gerek duymadan kullanılabilir olan kaynaklardır. Ham petrol, kömür, doğal gaz, rüzgar, jeotermal enerji ve güneş enerjisi bu kaynaklar arasında yer almaktadır. Bu birincil kaynaklar, uygun teknoloji ve süreçlerle ikincil enerji kaynaklarına dönüştürülerek farklı enerji ihtiyaçlarını karşılamak üzere kullanılmaktadır. Dönüşüm sürecinde, birincil enerji kaynakları, petrol ürünleri, elektrik, ısı veya biyoyakıtı dönüşmektedir. Yenilenebilir enerji kaynakları, genellikle doğrudan veya dolaylı olarak güneş enerjisinden elde edilir ve daha sonra elektrik veya ısı gibi ikincil enerji türlerine dönüştürülür [3, 4].

Perakende sektörü, hem geniş sosyal hedeflere katkı sağlama hem de işletme değerini artırma amacıyla gelir artışı, maliyet azaltma, marka geliştirme ve risk yönetimi gibi fırsatlara sahiptir. Perakendeciler, enerji tüketimi açısından özellikle mağazalar, dağıtım merkezleri ve ulaşım olmak üzere üç ana faaliyet kategorisinde etkin rol oynamaktadırlar. Bu alanlarda operasyonel verimliliği artırarak ve tüketici enerji verimliliğini iyileştirerek, yenilenebilir enerji kullanımını artırma ve yeni işletme değeri yaratma konusunda çeşitli fırsatlara sahiptirler [5,6].

Dünya genelinde, birincil enerji kaynaklarının yaklaşık beşte biri, ulaştırma sektörü tarafından tüketilmektedir [7]. Türkiye, gelişmekte olan bir ülke olarak, hızla artan nüfusu ve büyüyen ekonomisiyle birlikte araç sahipliği ve ulaşım ihtiyacında önemli bir artış yaşamaktadır. Özellikle şehir içi ulaşımda motorlu taşıtların kullanımı ve yapılan yolculuk sayıları gözle görülür biçimde artmaktadır. Ancak, Gayri Safi Milli Hasıla (GSMH) artış hızının ulaşım talebinin gerisinde kalması, ekonomik parametrelerin dengesizliğini ortaya koymaktadır [8].

Sektörel enerji talebinin etkin bir şekilde yönetilebilmesi için önerilen iki senaryo, yenilenebilir enerji kullanım oranını artırmak ve ulaşım talebini toplu taşımacılığa yönlendirilmesi şeklindedir. İlk senaryo, araç stoğundaki kullanım oranını yenilenebilir enerji kaynaklarına yönlendirmeyi amaçlamaktadır. Ancak, bu uygulama henüz ülkemizde yaygın değildir. İkinci senaryo ise, ulaşım talebinin enerji verimli sistemlere sahip toplu taşıma sistemlerine kaydırılmasını öngörmektedir. Her iki senaryo da zorlu ve maliyetli olabilir, ancak sürdürülebilir enerji politikalarının gerekliliği açısından değerlendirilmeleri önemlidir.

Önerilen stratejiler arasında, karayolu ulaşım talebinin %10'unun diğer ulaşım sistemlerine kaydırılması da bulunmaktadır. Yenilenebilir enerji kaynaklarının hizmet sektöründeki kullanımı ve uygulamaları, güneş enerjisi, rüzgar enerjisi, biyokütle ve jeotermal enerji üzerinden incelenmiştir. Türkiye'nin güneş enerjisi potansiyeli GEPA Albümü ile değerlendirilmiştir [9,10]. Rüzgar enerjisi kullanım potansiyeli Türkiye'de önemli ölçüde artmıştır ve rüzgar enerjisi potansiyel atlası teşvik amacıyla oluşturulmuştur [11].

Biyokütle ve biyoenerji projeleri, biyolojik kökenli kaynakların enerji üretimindeki önemini vurgulayarak, özellikle odunsu materyallerin günümüzdeki artan enerji talebinde etkili bir kaynak olduğunu belirtmiştir. Jeotermal enerji, yer kürede depolanan ısı enerjiden elde edilir ve bu enerji türünün bilimsel tanımı üzerinde durulmaktadır, jeotermal enerjinin sürdürülebilir enerji kaynakları arasında önemli bir potansiyele sahip olduğu vurgulanmıştır [12].

## **2. Literatür Taraması**

Özsümbül 1999 yılında yaptığı çalışmada; Türkiye'nin enerji potansiyeli, gelişimi ve gelecekteki enerji portresinin çıkarılması konusunda yüksek lisans tezi yapmış olup, ülkemiz açısından gerekli olan enerjinin en kolay şekilde nasıl sağlanacağını önemini belirterek nükleer santraller üzerinde durmuştur [13].

Mahmud 2000 yılında yaptığı çalışmada; Enerjinin üretimdeki rolünü yeniden incelenmiştir. Kısmi denge yaklaşımı kullanılmıştır. Enerji ve enerji dışı girdiler arasındaki farkın çok sınırlı olduğu ve bu nedenle enerji fiyatlarındaki artışlar üretim maliyetini doğrudan etkileyebildiği üzerine durulmuştur [14].

Hayli 2001 yılında yaptığı çalışmasında; Dünya nüfusunun ve sanayileşmenin gelişimine bağlı olarak artan enerji ihtiyacı yanında, fosil yakıtların tükenme riskleri ve artan çevre bilincinin oluşturduğu baskılarla, temiz ve yenilenebilir enerji kaynaklarına olan ilgi gün geçtikçe artmakta olduğunu çalışmasında anlatmıştır. Rüzgâr enerjisi, sadece gelişmiş ülkelerin değil, gelişmekte olan ülkelerin de sıcak baktıkları bir enerji kaynağı olduğunu vurgulamıştır. Geleceğe yönelik rüzgâr enerjisi santralleri projelerinin de sayıları hızla artmakta olduğunu, yapılan hesaplamalara göre 21.yüzyıl; tüm yenilenebilir enerji kaynaklarıyla birlikte rüzgâr enerjisinin tahmin edilenden daha fazla kullanılacağı ve yenilenebilir enerji kaynaklarının sıçrama yapacağı bir yüzyıl olacağı sonucuna varmıştır [15].

Chungpaibulpatana vd. 2003 yılında yaptıkları çalışmada; Üç ekonomik sektördeki enerji kullanım modellerinin kısa bir incelemesi; yani bu yazıda konut, sanayi ve ulaştırma sektörlerine yer vermiştir. Ulaştırma sektörü Tayland'da en büyük enerji tüketen sektör olup, bunu sırasıyla sanayi ve konut sektörleri izlemiştir. Hem ithal enerjinin hem de çevresel emisyonların azaltılması amacıyla enerji tasarrufu programları uygulanacağını anlatmıştır [16].

Etemoğlu vd. 2004 yılında yaptıkları çalışmalarında; Kendi öz varlığımız olan, dışa bağımlı olmayan, teknik ve ekonomik avantajlara sahip olan, çevre dostu jeotermal kaynakların maksimum ölçüde değerlendirilmesi enerji ekonomisi çalışmalarında önemli yer tuttuğunu, bu doğrultuda, tüm diğer ısı sistemlerinde olduğu gibi termodinamiğin ikinci kanununun jeotermal kaynaklar için uygulanması, sistem verimliliği, kullanılabilirliği ve sınıflandırılması için sağlam bir temel oluşturduğunu çalışmasında bahsetmiştir [12].

Kumbur vd. 2005 yılında yaptığı çalışmasında; Enerjinin insan ihtiyaçlarını karşılamada ve sürdürülebilir gelişmenin sağlanmasında oynadığı kilit rol vurgulanmıştır. Çalışmasında, enerji üretim ve tüketim süreçlerinin çevresel etkilerinin değerlendirildiği bir analiz yapılmıştır. Hidroelektrik santrallerin kurulma aşamasındaki hava kirliliği ve su kalitesinin düşürülmesi gibi sorunlara vurgu yapılmıştır. Yenilenebilir enerji kaynaklarına yönelme ve çevresel etkileri en az olan enerji kaynaklarını kullanma vurgulanmıştır. Güneş enerjisi, rüzgar enerjisi, jeotermal enerji ve biyokütle enerjisi gibi yerli ve çevre dostu enerji kaynaklarına yönelimin önemine vurgu yapılmıştır. Sonuç olarak ise, Türkiye'nin enerji politikalarında yenilenebilir enerji kaynaklarına daha fazla vurgu yapması, yerli üretimi teşvik etmesi, teknolojik gelişmeye odaklanması ve çevresel etkileri en aza indirecek stratejileri benimsemesi gerektiği belirtilmiştir. Bu çerçevede, yerli üretim imkanlarının organize edilmesi, ulusal enerji kullanım programlarının oluşturulması ve AR-GE çalışmalarının desteklenmesi önermiştir [17].

Kaymakçoğlu vd. 2005 yılında yaptıkları çalışmada; Jeotermal enerji ile ilgili yeni gelişmeler yakından takip edilmesinin gerekli olduğunu, yeni sahaların araştırmalarına destek olunması gerektiğini bunun yanında mevcut sahaların özellikleri ile buna bağlı kapasite ve kullanım olanakları tam olarak belirlenme çalışmalarının yapılması ve belirlenen alanlarda çalışmalar yapılması konusundan bahsetmiştir. Entegre tesislerin planlamasında da yüksek fayda olacağını anlatmıştır [18].

Ceylan 2005 yılında yaptıkları çalışmada; Enerji talebi denklemlerinin üç formu geliştirilmiştir. Genetik algoritma (GA) kavramına dayalı olarak geleceğe yönelik projeksiyonlar için ulaşım enerji talebi tahmin verimliliğinin artırılması anlatılmıştır. Burada önerilen tüm denklemler doğrusaldır ve doğrusal olmayandır bunlardan biri doğrusal, ikincisi üstel ve üçüncüsü ise ikinci derecedendir. Genetik Algoritma Taşıma Enerjisi Talep Tahmini (GATEDE) modelinin ikinci dereceden formu, gözlemlenen verilere daha iyi uyum sağlayan bir çözüm yolu sağladı ve yüksek oranlarda kullanılabilir olduğunu çalışmasında anlatmıştır [19].

Mathiesen vd. 2008 yılında yaptıkları çalışmada; Dünya çapındaki devletler ve şirketler, CO2 emisyonlarını azaltmayı ve yenilenebilir enerjiyi yaygınlaştırmayı hedeflediğini ama hedefe ulaşmanın temel unsuru, ulaşımında yenilenebilir enerjinin kullanılması olduğunu çalışmasında anlatmıştır. Taşımacılığı yenilenebilir enerjiye entegre etmeye yönelik tutarlı bir çaba gösterilmesi ve planlanması yapılmasını önerdiğini çalışmasında anlatmıştır [20].

Külekçi 2009 yılında yaptığı çalışmasında; Enerji kavramı ve enerji kaynaklarının sürdürülebilirliği geçmişten bugüne dünyanın en önemli konularından ve sorunlarından biri olduğunu anlatmıştır. Enerji kaynaklarının hızla tükenmesi, petrol, kömür, nükleer enerji gibi kendini yenileme durumu olmayan kaynakların bilinçsizce kullanılması, bu kaynakların çevreye ve atmosfere verdiği kirlilik gibi

etkenler insanları yenilenebilir enerji kaynaklarını kullanmaya yönlendirmeye yönelik çalışmasını yapmıştır. Çalışma kapsamında jeotermal enerjinin sürdürülebilirliği, jeotermal enerjinin çevreye etkisi gibi konular ülkemizdeki doğal kaynakların korunması ve çevre dostu kullanımlar açısından değerlendirilmiştir [21].

Al-Mofleh vd. 2009 yılında yaptıkları çalışmada; Enerji tüketiminin mevcut durumunun ve çeşitli enerji tasarrufunun kapsamlı bir resmi Malezya ortamı için uygun seçenekler araştırılmıştır. Tüketim modelini değerlendirmek için ayrıntılı bir araştırma yapılmış ve enerji verimliliğine yönelik mevcut teknikler anlatılmıştır. Ankete dayanarak, iyileştirmenin fizibilitesi Mevcut sistemler ve farklı sektörlerde yeni programların benimsenmesini anlatmıştır [22].

Abu-Taha 2011 yılında yaptığı çalışmasında; Enerji insanoğlu için bir ihtiyaç olduğunu ancak mevcut önümüzdeki dönemde enerji kaynaklarının sınırlı olacağı öngörüldüğünü eğer çevre için uyumlu kullanılmazsa belirli yıkıcı sonuçları olacağını dile getirmiştir. Yenilenebilir enerji çözüm olarak ortaya çıkıyor sürdürülebilir, çevre dostu ve uzun vadeli bir maliyet geleceğin etkili enerji kaynağı olacağını çalışmasında vurguluyor. Yenilenebilir enerji Alternatifler geleneksel kaynakların yerini alabilecek kapasitede olduğunu ve bunun kullanılması üzerine çalışmalarda bulunulması sonucuna varmıştır [23].

Karayılmazlar vd. 2011 yılında yaptığı çalışmasında; Enerji tüketimi ülkelerin gelişmişlik düzeylerinin bir göstergesi, bireylerin rahat yaşam sürmeleri için vazgeçilmez olduğunu anlatmıştır.

Gelişen teknoloji ve artan nüfusla birlikte enerji tüketimindeki artış enerjii tüm dünyada olduğu gibi ülkemizde de önemli bir problem olarak karşımıza çıkarmakta olduğunu anlatmıştır. Sanayileşme, nüfus artışı, kentleşme ve yaşam standartlarının yükselmesi gibi etkenler enerji tüketimini artırırken, enerji kaynaklarının hızla tükenmesine yol açmış olduğunu vurgulamıştır. Enerji açığını karşılamak için dünyada biyokütle çalışmalarına büyük hız verildiğini dile getirmiştir. Bu büyük potansiyelin yanı sıra biyokütlenin ekonomik ve çevresel açıdan olumlu özellikleri de göz önüne alındığında, biyoenerji konusuna ilgi giderek arttığını, biyokütle dünyada dördüncü en büyük enerji kaynağını oluşturması yönüyle önemli bir enerji kaynağı konumunda olduğunu ve sektörlerde kullanımının biz katkıları anlatmış ve bununla ilgili çalışmaların hızlandırılması gerektiği sonucuna varmıştır [24].

Erdal 2011 yılında yaptığı çalışmasında; Türkiye ve Dünyadaki enerji kaynaklarının durumunu ortaya koymak, mevcut ve gelecekteki olası sorunları irdelemek ve önerilen politikaları tartışmak; enerji arz güvenliğini etkileyen faktörleri belirlemek ve bunların Türkiye için geçerliliğini ampirik olarak test etmeyi amaçlamış ve önerilerde bulunmuştur [25].

Mahmutoğlu 2013 yılında yaptığı çalışmasında; Yenilenebilir enerji günümüzde ısıtma, ulaştırma, elektrik üretimi gibi birçok alanda kullanılmakta olduğunu bizlere anlatmıştır. Yenilenebilir enerji kaynaklarının yerli nitelikli oluşu, yakıt maliyetinin düşük olması, temiz ve çevreci özellikleri ile elektrik üretiminde son yıllarda öne çıktığını vurgulamıştır. Çalışmasında, yenilenebilir enerji kaynaklarının elektrik üretimi amaçlı kullanımı ağırlıklı analizlere yer vermektedir. Türkiye’de orta vade olarak nitelenen gelecek 10 yıllık dönemde yenilenebilir enerji kaynaklı elektrik üretiminin, artan elektrik enerjisi talebini ne derece karşılayabileceği konusunu çalışmasında irdelemiştir. Çalışmasında aynı zamanda Türkiye’nin enerjide dışa bağımlılığının azaltılmasına ilişkin alternatif çözüm önerileri sunmuştur. Enerji politikası uygulayıcıları, bir an önce enerji sektöründe mevcut dışa bağımlı yapının değiştirilmesi için gereken önlemleri alması gerektiği sonucuna varmıştır [26].

Ellabban vd. 2014 yılında yaptığı çalışmalarında; Yenilenebilir enerji kaynaklarının potansiyeli çok büyük olduğunu, çünkü dünyanın enerji talebini katlanarak aşmakta olduğunu bunun çözümünün ise bu tür kaynaklara sahip olacak geleceğin küresel enerji portföyünde önemli bir paya sahip olduğunu çalışmasında anlatmıştır. Yenilenebilir enerji kaynakları havuzunu geliştirdiklerine değinmiştir. Sonuç olarak ise, enerji kaynaklarının halihazırda kullanılmakta olduğunu, kullanımlarının iyileştirilmesine yönelik bilimsel gelişmeler, gelecekle ilgili umut vadettiğini anlatmıştır [27].

Vardar vd. 2014 yılında yaptıkları çalışmasında; “Renewable energy sources and Turkey” isimli akademik makalelerinde Türkiye’de yenilenebilir enerji kaynaklarının potansiyelleri üzerinde durmuşlardır [28].

Alrikabi 2014 yılında yaptığı çalışmada; Yenilenebilir kaynaklar uygun maliyetli olduğunu, aynı zamanda da kullanıcı dostu olduğunu çalışmada anlamıştır. Fosil yakıtları kolaylıkla yenilebilir enerji kaynağı olduğunu eğer teşvik edersek kirliliğe sebep olan etmenleri engelleyebileceğimizi çalışmada anlatmıştır. Ülke Ekonomisi arttırmaya katkı sağlayacağını anlatmıştır [29].

Mathiesen vd. 2014 yılında yaptıkları çalışmada; Metrikler arazi kullanımı, kaynak kullanılabilirliği, verimlilik, maliyetler ve genel enerji sistemi entegrasyonudur. Temelli bu ölçümlere göre ulaşımda mümkün olan her yerde elektriğin kullanılması gerektiği sonucuna varılmıştır. Elektriğin uygun olmadığı diğer alanlarda biyoenerji (bio) veya CO<sub>2</sub> (syn) bazlı yakıtlar kullanılabileceğini anlatmıştır. Buhar elektrolizinden elde edilen hidrojen ile güçlendirilen yakıtlar kullanılabileceğini dile getirmiştir [30].

Şenal vd. 2015 yılında yaptıkları çalışmada; Ülkemizde milli rüzgâr enerji santrali geliştirme çalışmaları hızlandırılarak büyük güç kapasiteli rüzgâr türbinleri ve türbin elemanlarının üretiminin bir an önce gerçekleştirilmesi gerektiğini anlatmıştır. Bu sebeple, yerli rüzgâr türbini teknolojisi alıcı ve satıcılarına KDV ve gelir vergisi indirimleri sağlanarak yerli üretim daha fazla teşvik edilmesinin önemli olduğunu anlatmış, böylece hem yerli rüzgâr türbin teknolojisinin gelişeceği hem de ülkemiz ekonomik büyümesine büyük katkı sağlanacağı asla göz ardı edilmemesi gerektiği sonucuna varmıştır [11].

Doğan vd. 2015 yılında yaptığı çalışmada; Yerli kaynaklardan enerji talebinin karşılanmasının önemine vurgu yapılmıştır. Aynı zamanda, enerjiyi verimli kullanmanın da büyük bir öneme sahip olduğu belirtilmiştir. Sanayi sektöründeki enerji yoğunluğunun azaltılması, özellikle demir çelik, çimento, cam, petrokimya ve petrol sektörlerindeki tasarrufun önemine vurgu yapılmıştır. Ulaşım sektöründe raylı sistemlerin ve toplu taşımanın teşvik edilmesinin, büyük ölçekli tesislerde kojenerasyon sistemlerinin kullanılmasının ve vatandaşların enerji tasarrufu konusunda bilinçlendirilmesinin önemine değinilmiştir. Sonuç olarak ise Türkiye'nin geleceğe yönelik planlamalar yaparken toplumun ihtiyaçlarını doğru analiz ederek ekonomik ve sosyal kalkınma açısından sürdürülebilir, bütünsel bir yaklaşım sergilemesi ve etkili bir enerji politikası oluşturması gerektiği belirtilmiştir [31].

Mohtasham 2015 yılında yaptığı çalışmada; Yenilenebilir enerjinin avantajlarını ve dezavantajlarını üzerine çalışma yapılmıştır. enerji kaynaklarının faydalarına dayanarak fosil yakıtlar yerine yenilenebilir enerjilerin kullanılmasının ülkemizin çevresel, sosyal ve ekonomik sorunlarının kontrolünde iyi bir çözüm olacağını anlatmıştır. Dünyanın kaynaklarını dağıtırken ve tüketirken önlem almamız gerektiğini anlatmıştır. Yenilenebilir enerji kullanımının ödüllendirici hale geldiği bir ortam yaratmamız gerektiğini, yani vergi teşviki gibi kullanımlarla insanları bunları kullanmaya teşvik etmememizin önemli olduğunu anlatmış ve ABD'deki örneğin sağladığı yararları bahsetmiş sonuç olarak ise, diğer ülkelerde de yaygınlaştırılmasının önemine değinmiştir [32].

Torunoğlu 2015 yılında yaptığı çalışmada; Günümüzde enerji tüketiminin hızla artmasına paralel olarak alışlagelen enerji kaynaklarının yakın bir gelecekte tükeneceği bilimsel bulgularla ispatlanmış bir gerçek olduğunu çalışmada anlatmıştır. Yenilenebilir enerjinin günümüzde ısıtma, ulaştırma, elektrik üretimi gibi birçok alanda kullanılmakta olduğunu anlatmıştır. Ama son yıllarda yenilenebilir enerji kaynaklarının yerli nitelikli oluşu, yakıt maliyetinin düşük olması, temiz ve çevreci özellikleri öne çıktığını çalışmada değinmiştir. Genel bir değerlendirmeye çalışmada yer vermiş, sonrasında Türkiye'de son dönemde ve gelecekte enerji alanında gerçekleşmesi muhtemel projeler üzerinde durmuştur. Sonuç olarak ise, doğadan aldığımız kadar, yine doğaya kazandırmak koşulu ile, tek yönlü yani dönüşümsüz dengeyi bozmasak, çevreden aldığımızı yararlandıktan sonra yine aynı özellikte geri verirse, doğal dengenin kurulmasına imkân ve zaman bırakırsak, doğa kendi kendini yenileyebilir ve eksilen bileşenini telafi edebileceğini anlatmıştır [33].

Kılıç 2015 yılında yaptığı çalışmada; Güneş enerjisine genel bir bakış, Türkiye'de güneş enerjisinin son durumu, güneş enerjisi potansiyelleri, üretimleri, devlet teşvikleri, kullanım sahaları ve güneş enerjisi teknolojileri incelemiştir. Böylece, yenilenebilir enerjilerdeki gelişmelere katkı sağlanması ve konuya yönelik bilinçlenmenin artırılması gerektiğini çalışmada amaçlanmıştır. Ülkemiz yenilenebilir enerji kaynakları bakımından Avrupa'da lider olabilecek kadar iyi bir konuma sahip



olduğunu çalışmasında vurgulamıştır. Yapılacak yeni kanuni düzenlemeler, çalışmalar ve yatırımlarla, bu yüksek potansiyellerdeki kendi yerel kaynaklarımızı, teknolojimizi ve teknik kadrolarımızı hayata geçirerek ve bilinçlenmeyi artırarak enerjide önemli adımlar atılabileceği sonucuna varmıştır [34].

Çelik 2015 yılında yaptığı çalışmasında; Fosil yakıtlar yerine kullanılacak çevre dostu, sınırsız ömre sahip ve düşük maliyetli yenilenebilir enerji kaynaklarının Türkiye açısından önemini ortaya koymaktadır. Birinci bölümde enerji kaynaklarının tanımlarına, özelliklerine, avantaj ve dezavantajlarına yer verilmiştir. İkinci bölümde yenilenebilir enerji kaynaklarının Dünya ve Türkiye açısından potansiyeli, kurulu gücü ve yatırım maliyetleri incelenmiştir. Son bölümde ise fosil yakıtlar ile yenilenebilir enerji kaynaklarının üretimi, tüketimi, ithalattaki payları ve yatırım maliyetleri karşılaştırılarak analiz edilmiştir. [35].

Mathiesen vd. 2015 yılında yaptıkları çalışmada; Yenilenebilir enerji kaynaklarının mevcut veya gelecekteki %100 yenilenebilir enerji kaynaklarına entegrasyonuna yönelik en düşük maliyetli çözümlerin belirlenmesi amacıyla, bir kişinin hangi adımları atması gerektiğini açıklamıştır. Akıllı Enerji Sistemleri yaklaşımının kullanımın önemini belirtmiştir. Makale de bunun neden ve nasıl yapılacağını anlatmıştır [36].

Şeker 2016 yılında çalışmasında; Günümüz dünyasının en önemli konularından ve sorunlarından biri, enerji ve enerji kaynaklarının hızlı bir şekilde tükenmeye başlaması, özellikle de fosil yakıt olarak adlandırılan kömür, petrol, doğalgaz gibi kendini yenileme özelliği olmayan kaynakların bilinçsizce ve hiç tükenmeyecekmiş gibi kullanılıyor olması hem bu kaynakların tükenmeye başlamasına hem de doğal çevrenin zarar görmesine ve ekolojik dengenin bozulmaya başlamasına neden olduğu. Bu çalışmada yenilenebilir enerji kavramı, yenilenebilir enerji kaynakları, yeşil pazarlama ve yenilenebilir enerji kaynaklarının pazarlanması değerlendirilmeye çalışılmışlardır [1].

Han vd. 2017 yılında yaptıkları çalışmada; Entegre yenilenebilir enerji kaynaklarına (RES) dayalı bir enerji sistemi tasarlamak ve işletmek için optimizasyona dayalı yeni bir yaklaşım önermeyi amaçlamıştır. Hedefe ulaşırken öncelikle farklı kaynakları (rüzgar, güneş, biyokütle ve karbondioksit), farklı enerji dönüşüm teknolojilerini ve farklı enerji taleplerini (elektrik, hidrojen ve yakıtlar) içeren RES tabanlı enerji üst yapısını oluşturmayı planlamıştır. Daha sonra ulaştırma sektörüne enerji tedarik sistemini tasarlamak ve analiz etmek için karma tamsayılı doğrusal programlamayı (MILP) kullanarak bir ağ optimizasyon modeli geliştirmeyi hedeflemiştir [37].

Heikki vd. 2017 yılında yaptıkları çalışmasında; Yenilenebilir enerji kapasitesindeki artışın elektrik hizmetlerinin mali performansını (FP) etkileyip etkilemeyeceği hakkındaki tartışma yapılmıştır. Panel üzerine yapılan bir veri çalışması yapılmıştır [38].

Ásványi vd. 2017 yılında yaptığı çalışmasında; Yenilenebilir enerjinin hizmet sektöründe ve özellikle turizm endüstrisinde giderek artan önemi, yeni bir araştırma konusu haline geldiği konularına değinmiştir. Yenilenebilir enerji turizmde pek çok açıdan, özellikle de enerji verimliliği, sürdürülebilirlik ve maliyetlerin azaltılması gibi konularda önemli bir rol oynadığını, ancak bugüne kadar konuyla ilgili literatürün sistematik olarak derlenmiş bir incelemesi ortaya çıkmadığını çalışmasında anlatmış ve dolayısıyla bu çalışma, yenilenebilir enerjinin yenilenebilir enerji kullanımına odaklanan araştırma yayınlarının içerik analizi yoluyla literatürdeki bu boşluğu doldurmaya çalışmıştır [39].

Karaca vd. 2017 yılında yaptıkları çalışmada; Yenilenebilir enerji yatırım hedefinin gerçekleştirilmesi için ülkenin sahip olduğu önemli bir potansiyel mevcut olduğunu, ancak bu potansiyelin hayata geçirilmesi her üyeden önce hükümetin yenilenebilir enerji kaynaklarının yaygınlaştırılması yönünde atacağı adımlar ile mümkün olunacağını çalışmasında anlatmıştır [40].

Koç vd. 2018 yılında yaptığı çalışmasında; Enerji kaynaklarının Dünya ve Türkiye için değerlendirilmesi incelemiştir. Bu enerji kaynaklarının küresel rezerv miktarları ve buna bağlı olarak kalan ömürleri, üretim ve tüketim değerler belirtilmiş olup, bu kaynaklardan elde edilen enerjinin üretim oranları ile tüketim oranları hakkında karşılaştırmalar yapılarak, kurulu güç miktarları hakkında bilgi vermiştir. Sonuç olarak ise, genel olarak ülkemiz yenilenebilir enerji açısından yüksek kaynak potansiyeline sahip olduğunu bununla birlikte Türkiye yenilenebilir enerji sistem kurulum ve yatırımlarında daha çok yabancı teknoloji ve kaynaklarına ihtiyaç duymakta olup kurulum süreci

maliyet bakımından oldukça fazla olduğuna bu yenilenebilir enerji potansiyeli değerlendirmek için devlet desteği ile yatırım yapılması sağlanarak altyapı çalışmalarının ve bu kaynakların kullanımının artırılması gerektiği sonucuna varmıştır [3].

John vd. 2018 yılında yaptıkları çalışmada; Birincil enerji tedarikinde yakıt kaynakları arasındaki karmaşık ilişki İklim değişikliğinin azaltılması ve enerji sistemlerinin geliştirilmesi vurgu yapılmıştır. Orta Doğu'da enerji gelişiminin geleceği konusunda büyük belirsizlik olduğunu anlatmıştır. Bu belirsizlik büyük ölçüde artan enerji talepleri ve bölgenin ekonomik durumuyla bağlantılı olduğunu çalışmasında dile getirmiştir [41].

Paramati vd. 2018 yılında yaptığı çalışmasında; Sanayi ve hizmet sektörlerindeki yenilenebilir enerji tüketiminin ve ekonomik faaliyetlerin dinamikleri üzerine çalışmıştır, enerji-büyüme bağlantısını ampirik olarak araştırmıştır.

Bu çalışmalar, bu iki değişken arasındaki bağlantıyı açıklamak için dört hipotez kümesini ortaya koymuştur. Hem yenilenebilir enerji tüketimi hem de ekonomik büyüme birbirini tetikliyorsa bu duruma değişkenler arasında çift yönlü (geri beslemeli) nedensellik dendiğini açıklamış. Bu durumda yenilenebilir enerji kullanımı ekonomik büyümede önemli bir rol oynamaktadır ve bunun tersi de geçerlidir.

Eğer nedensellik yenilenebilir enerji tüketiminden ekonomik büyümeye doğruysa, yenilenebilir enerji tüketimindeki herhangi bir azalma ekonomik çıktıyı engelleyeceğini anlatmıştır [42].

Eren vd. 2018 yılında yaptıkları çalışmada, enerji arzının güvenliği ve çevresel sürdürülebilirlik gibi önemli sorunları beraberinde getirdiğini ve fosil yakıtların yoğun kullanımı iklim değişikliği gibi büyük çevresel sorunlara neden olduğunu, aynı zamanda bu kaynakların jeopolitik gerilimlere de katkıda bulunduğunu anlatmıştır.

Dünyanın enerji ihtiyacını karşılamak için daha sürdürülebilir ve çeşitli enerji kaynaklarına yönelmenin önemine değinmiştir [43].

Alper 2018 yılında yaptığı çalışmasında; Günümüzde enerji talebinin hızla artmasına paralel olarak, geleneksel enerji kaynaklarının sınırlı olması, ülkeleri enerji üretim politikalarını değiştirmeye ve alternatif enerji kaynaklarını kullanmaya teşvik etmiştir.

Bu amaçla yapılan çalışmada, Türkiye'de yenilenebilir enerji kullanımı ile ekonomik büyüme arasındaki ilişki ve ilişkinin yönü, Bayer-Hanck eşbütünleşme testi ve Toda-Yamamoto nedensellik testi yardımıyla incelemiştir. Yapılan nedensellik test sonuçlarına göre, ekonomik büyümeden yenilenebilir enerji kullanımına doğru tek yönlü bir nedensellik vardır ve Türkiye'de koruma hipotezi geçerli olmuştur [44].

García-Olivares vd. 2018 yılında yaptıkları çalışmada; %100 yenilenebilir bir ekonominin, iklim değişikliği, enerji güvenliği, sürdürülebilirlik ve kirlilik gibi küresel zorluklara kalıcı bir çözüm sunabileceği tezini savunmuştur. Ancak, mevcut ulaşım sistemini dönüştürmenin, bu yenilenebilir geçişin en zor yönlerinden biri olduğunu öne sürmüştür.

Çalışma, fosil yakıtla dayalı ulaşım sistemlerine alternatif olarak önerilen veya kanıtlanmış teknolojileri ve sistemleri incelemekte ve bunların karbon sonrası döneme geçişteki umutlarını teknolojik ve enerji açısından değerlendirmektedir. Analiz, %100 yenilenebilir taşımacılığın mümkün olduğunu ancak kaynak tüketiminde belirsiz bir artışla tam anlamıyla uyumlu olmadığını belirtmektedir. Ayrıca, her ulaşım sektörünün bu geçişte karşılaşılabileceği başlıca malzeme ve enerji kısıtlamalarını ve engellerini detaylı bir şekilde ele almıştır [45].

Avcılar ve Yenilmez 2019 yılında yaptıkları çalışmada, Dünya genelinde yaşanan hava kirliliği, küresel ısınma ve enerji güvenliği gibi sorunlar, dünyanın enerji altyapısının, sıfır emisyon üreten %100 temiz ve yenilenebilir enerjiye hızlı bir şekilde dönüşmesini zorunlu hale getirdiğini çalışmasında değinmiştir.

Yenilenebilir enerji kullanımı dünya genelinde hızlı bir artış gösterdiğine ve yenilenebilir enerji tüketiminde gerçekleşen bu büyüme, sadece karbon emisyon azaltma zorunluluğunu yerine getiren işletmelere değil aynı zamanda büyük perakendeciler de dahil olmak üzere kurumsal organizasyonların da dahil olduğunu anlatmıştır.

Perakendeciler marka değerini ve satış gelirlerini artırmak, enerji maliyetlerini azaltmak ve artan enerji fiyatlarına yönelik riskleri en aza indirmek için yenilenebilir enerjiye yatırım yapılması gerektiğini anlatmıştır. Doğa ve çevre dostu olduğu için yenilenebilir enerji kaynakları ekolojik olarak nitelendirildiği ve yenilenebilir enerji mevcut kaynakların kullanımı ile tekrar tekrar üretilen ve sürdürülebilir olan bir enerji kaynağı olduğu sonucuna varılmıştır [46].

Kayısoğlu vd. 2019 yılında yaptıkları çalışmada; Yenilenebilir enerji kaynağı potansiyeli bulunan ülkemizde, bu potansiyelin verimli bir şekilde değerlendirilmesi ve ülkemizin enerji probleminin çözümüne katkıda bulunması gerektiğini çalışmada anlatmıştır. Yenilenebilir enerji sektörü uluslararası kural ve düzenlemeler çerçevesinde yeniden yapılandırılması gerektiğini anlatmıştır.

Büyük oranda yabancı teknolojiye ihtiyaç duyulan yenilenebilir enerji sektöründe, teknik ekipmanların yerli üretimi teşvik edilmesi gerektiğine ve bizzat devlet tarafından desteklenmesinin gerekli olduğuna çalışmada değinmiştir.

Yenilenebilir enerji yatırımlarına çevre etkileri de hesaba katılacak şekilde gerçekleştirilmesi gerektiğine değinmiştir. Yenilenebilir enerji konusunda daha fazla bilgilendirme yapılmalı ve bu konuda farkındalık artırılarak AR-GE çalışmalarına ağırlık verilmesi sonuçlarına varmıştır [47].

Aşkın vd. 2019 yılında yaptıkları çalışmada; Türkiye'nin sahip olduğu enerji portföyünü belirlemek ve enerji arzı güvenliği kapsamında karşı karşıya olduğu tehdit ve fırsatlar üzerine bir değerlendirme yapılmıştır. Türkiye'nin enerjide dışa bağımlı olması, enerji arz güvenliği bakımından, Türkiye'yi ekonomik krizlere açık hale getirdiğini çalışmada anlatmıştır.

Bu bağlamda Türkiye enerji portföyünde yerli enerji kaynaklarına yönelmeli ve enerji politikalarında yenilenebilir enerji kaynaklarına öncelik verilmesi gerektiği sonucuna varılmıştır [48].

Buonocore vd. 2019 yılında yaptıkları çalışmada; Sürdürülebilir Kalkınma Hedefleri ile anlamlı bir şekilde uyumlu hale getirilmelerine olanak tanınrsa, Ülke düzeyindeki elektrik üretimi ve kara taşımacılığı verilerini kullanarak, yenilenebilir enerji yatırımlarının artırılmasının ve kara taşımacılığındaki geliştirmelerin, karbon emisyonlarını azaltma potansiyelinin değerlendirilmesi için basit ve kullanıcı dostu bir dizi ölçüm geliştirilebileceğini çalışmada bahsetmiştir.

Hava kirleticilerin emisyonları ve bu emisyonların sağlık üzerindeki olası etkilerini değerlendirerek, çevresel sürdürülebilirliğin sağlanmasına yönelik adımlar atılması gerektiğini açıklamıştır [49].

Hasanuzzaman vd. 2020 yılında yaptıkları çalışmada; Farklı sektörlerin enerji talebinin iyi anlaşılması, bir ülkenin enerji planlaması ve politikası için önemli bir bileşen olduğunu anlatmıştır. Enerji talebi nüfus, kentleşme, sanayileşme, net sermaye geliri ve teknolojilerin gelişimi gibi farklı sosyoekonomik faktörlere bağlı olduğuna değinmiştir. Farklı sektörlerin enerji talebi hesaplama yöntemleri anlatılıp çalışmalardan bahsedilmiştir [50].

Aydoğdu 2021 yılında yaptığı çalışmada; Enerji, günümüzde modern sanayi ve hizmetler sektörünün hayati bir parçası haline geldiği, enerjiye erişim, enerjinin maliyeti ve çevresel etkileri, yenilenebilir enerjinin ön plana çıkmasına ve bunun ulusal ve uluslararası alanda desteklenmesine neden olması konusuna çalışmada anlatmıştır. Bu destekler, yenilenebilir enerjinin ve enerji verimliliğinin sağlanması, teknolojinin geliştirilmesi ve yaygınlaştırılması için önem arz ettiğine değinmiştir.

Yenilenebilir enerji ve enerji verimliliğine yönelik destekler, hemen hemen bütün ülkeler tarafından uygulanmakta olduğuna ve bu çerçevede, birçok program ve yöntem geliştirilmekte olduğuna ve bunların farklı etkileri söz konusu olduğunu, destekler kimi zaman yenilenebilir enerji üretiminde kullanılan makine ve ekipman üretimini teşvik için imalat sektörüne, kimi zaman doğrudan enerji üretimi için enerji sektörüne, kimi zaman enerji tüketiminde verimlilik sağlamak için sanayiye, kimi zaman da hem enerji üretimine hem de tüketimine yönelik olarak konutlara ve tarım sektörüne sağlanması gerektiğini anlatmıştır [4].

Güllü vd. 2021 yılında yaptıkları çalışmada; Dünya ve Türkiye genelinde enerji talebinin yükselmesi, rüzgar, güneş, hidrolik, jeotermal ve biyokütle gibi yenilenebilir enerji kaynaklarına olan ilgiyi hızla artırmaktadır. Türkiye'de yenilenebilir enerjiden elektrik üretiminin artmasının istihdam üzerindeki etkileri ve bu etkilerin yönü literatürde daha fazla yer bulmaya başlamıştır. Bu makalede, her bir alt sektör için istihdamın nicel etkileri, (JEDI) modeli kullanılarak incelenmiştir. Ayrıca, her alt sektör için

MW başına istihdam ölçümlenerek hangi sektörün istihdam açısından daha etkili olduğu karşılaştırılmıştır [51].

Binod vd. 2021 yılında yaptıkları çalışmada; Yenilenebilir enerjinin etkin entegrasyonunda akıllı şebekelerle artan enerji talebi yönetimini karşılanabileceğini çalışmasında anlatmıştır. Artık artan nüfusun enerji talebini akıllı teknolojiler kullanarak yönetilebileceğini çalışmasında dile getirmiştir [52].

Østergaard 2022 yılında yaptığı çalışmasında; Yenilenebilir enerji kaynaklarının uygulanması kapsamındaki araştırmaların durumunu ele almıştır.

Yenilenebilir teknolojilerin durumuna, yenilenebilir enerji kaynaklarının sürdürülebilir kalkınma hedeflerine ulaşmadaki rolüne, yenilenebilir enerji sistemlerinin sürdürülebilirliğine yönelik araştırmaların durumuna ve son olarak yenilenebilir enerji teknolojilerinin düşük karbonlu enerji sistemlerine entegrasyonuna odaklanmıştır. Yenilenebilir enerjiye dayalı enerji sistemlerine geçiş için önemli olan teknolojiler ve sistemler söz konusu olduğunda literatürdeki boşlukları doldurmaya yönelik önemli bir araştırma ilerlemesi ortaya koymuştur [53].

Gündüz vd. 2022 yılında yaptığı çalışmasında; Gelir gruplarına göre sınıflandırılmış ülkelerin, enerji durumu incelemiş, yenilenebilir enerji teknolojilerine olan yatkınlıkları, güçlü ve zayıf yönleri tespit etmişler. Türkiye üzerinden değerlendirilmelerde bulunmuşlar.

Üniversitelerdeki bilgi birikiminin, sektörde yer alan firmaların sermaye gücüyle birleşmesi, enerji üretimindeki verimliliği arttıracak ve önemini vurgulamışlardır. Doğal kaynakları enerjiye çevirmek, bugün de gelecekte de yurt dışı kaynaklardan temin edilen fosil kaynaklardan çok daha iyi olabileceği öngörülme bulgusunda bulunmuşlardır [2].

Maaouane vd. 2022 yılında yaptıkları çalışmada; Gelişmekte olan ülkelerde, ulusal düzeydeki kurumlar genellikle temel ulaşım enerji verimliliği göstergeleri ile sınırlı olduğunu çalışmasında anlatmıştır. Bu boşluğu doldurmak için 40 yapay sinir ağını temel alan bir ulaşım modeli geliştirip, enerji verimliliğine ilişkin veriler Sosyoekonomik değişkenleri kullanarak bu göstergeleri tahmin etmeye yönelik bir model geliştirilmesi üzerine çalışmasında anlatmıştır [54].

Aliabadi vd. 2023 yılında yaptıkları çalışmada; Ulaştırma sektörü, dünya çapında sera gazı emisyonlarına artan katkısı nedeniyle enerji geçişini hızlandırma konusundaki tartışmaların merkezinde yer alıyor; bu nedenle AB, Yenilenebilir Enerji Direktifi aracılığıyla yenilenebilir enerjinin ulaşımında kullanımına yönelik bağlayıcı hedefler belirlemiştir. Bu hedeflerin ekonomik etkisini analiz etmek için biyo yakıt ve elektriğe dayalı yakıt seçeneklerini, çeşitli ulaştırma sektörlerini ve gelecekteki politika gerekliliklerini dikkate alan bir optimizasyon modeli geliştirmeyi çalışmasında anlatmıştır [55].

Ahmet Kılmaz 2023 yılında yaptığı çalışmada, insanlık tarihi boyunca enerji kullanımı önemli olduğunu, teknolojinin ilerlemesiyle enerji talebi arttığını dile getirmiştir. Fosil yakıtların tükenmesi ve çevresel etkileri göz önüne alındığında, yenilenebilir enerji kullanımı önemini vurgulamıştır. Gelecekte akıllı şebekelere geçiş kaçınılmaz olduğundan teknolojinin entegrasyonu ile elektrik sistemleri daha verimli hale geleceğini, yenilenebilir enerji ve akıllı şebekelere geçiş, çevreci elektrik üretimini teşvik edeceğini anlatmıştır [56].

**Tablo 1.** Literatür Konu Dağılımları

Yazar Adı	Yıl	Konu										
		Yenilenebilir Enerji	Enerji Kavramı	Enerji Kullanım	Enerji Potansiyeli	Enerji Verimliliği	Ulaşım Sektörü	Sürdürülebilirlik	Güneş Enerjisi	Hidrolik Enerji	Rüzgâr Enerjisi	Jeotermal Enerji
Özsümbül, B. [13]	1999				x							
Mahmud, Syed F. [14]	2000											
Hayli, Selçuk [15]	2001										x	
Supachart Chungpaibulpatana vd.[16]	2003			x			x					

Etemoğlu, Akın vd. [12]	2004							x
Kumbur, Halil vd. [17]	2005							
Kaymakçioğlu, Fatih vd. [18]	2005							
Halim, Ceylan [19]	2005							x
Mathiesen, Brian [20]	2008						x	
Külekçi, Özlem Candan [21]	2009						x	
Al-Mofleh, Anwar vd. [22]	2009	x						x
Abu-Taha, Rimal [23]	2011						x	
Karayılmazlar, Selman vd. [24]	2011	x						
Erdal, Leman [25]	2011							
Mahmutoğlu, Murat [26]	2013	x						
Ellabban, Omar vd. [27]	2014	x					x	
Vardar Ali vd. [28]	2014						x	
Alrikabi, N. K. M. A.m [29]	2014	x						
Mathiesen, Brian vd. [30]	2014						x	x
Doğan, Hikmet vd. [31]	2015						x	x
Mohtasham, Javid [32]	2015	x						
Torunoğlu Gedik, Özge [33]	2015					x		x
Kılıç, Fatma Çanka [34]	2015	x						x
Çelik, Selvi Nur [35]	2015	x						
Mathiesen, Brian vd. [36]	2015	x						
Şenal, Mahmut Can vd. [11]	2015							x
Şeker A. [1]	2016	x						
Han, Seulki vd. [37]	2017	x						x
Heikki, Lehkonen vd. [38]	2017	x						
Ásványi, Katalin vd. [39]	2017	x					x	x
Karaca, Coşkun vd. [40]	2017	x						
John E. vd. [41]	2018	x	x					
Koç Ali, vd. [3]	2018							x
Paramati, Sudharshan Reddy, vd. [42].	2018	x						x
Eren Alper vd. [43]	2018							x
Alper, Fındık Özlem [44]	2018							x
García-Olivares, Antonio vd.[45]	2018	x						x
Avcılar, Mutlu Yüksel vd. [46]	2019	x						
Kayıoğlu, Birol vd. [47]	2019	x						x
Aşkın, Elif vd. [48]	2019	x						
Buonocore, Jonathan J vd. [49]	2019							x
Hasanuzzaman, M vd. [50]	2020					x	x	
Güllü, Mustafa vd. [51]	2021							x
Aydoğdu, Çağkan [4]	2021	x	x					
Binod Kumar Sahu vd. [52]	2021	x						
Gündüz, Necla Kübra vd. [2]	2022	x	x	x				
Østergaard, Poul Alberg, [53]	2022	x						
Maaouane, Mohamed vd. [54]	2022	x					x	x
Aliabadi, Danial Esmaeili vd. [55]	2023							x
Ahmet Kılmaz [56]	2023							x

Tablo 1’de görüldüğü üzere, literatür çalışmalarında çoğu enerji kavramı üzerine durulmuştur. Son zamanlarda yapılan literatür çalışmalarının arttığı gözlemlenmiştir. Jeotermal enerji konusu geçmiş

senelerde daha çok incelenmiş olup günümüze yakın olan tarihlerde yapılan çalışmalar farklı konular üzerine yoğunlaşmıştır.

### 3. Sonuç

Bu çalışmada, hizmet sektöründe yenilenebilir enerji kullanımının avantajları ele alınmış ve perakende ticaret ile ulaştırma hizmetlerindeki potansiyel vurgulanmıştır. Güneş enerjisi, rüzgâr enerjisi, biyokütle, biyoenerji ve jeotermal enerji gibi çeşitli kaynaklar detaylı şekilde incelenmiş, sektörel enerji talebinin belirlenmesi ve yönetimi konusunda önemli bulgular sunulmuştur. Çalışmanın bulguları, hizmet sektöründe yenilenebilir enerji uygulamalarının yaygınlaştırılmasına yönelik politika önerilerini ve geleceğe dair öngörülerini paylaşmıştır.

Literatür incelemelerinde, sürdürülebilirlik ve enerjinin üretimdeki rolü üzerine yapılan çalışmaların sınırlı olduğu gözlemlenmiştir. Enerji kavramı üzerine yapılan çalışmaların çoğunda, yenilenebilir enerjiye genel olarak değinilmiştir. Jeotermal enerji, geçmiş yıllarda daha fazla incelenirken, son dönem çalışmaların farklı enerji kaynaklarına yoğunlaştığı görülmektedir. Gelecek çalışmalar kapsamında, enerji yönetiminin sürdürülebilirlik bağlamında daha derinlemesine analiz edilmesi planlanmaktadır. Bu doğrultuda, hizmet sektöründeki yenilenebilir enerji kullanımının uzun vadeli etkileri ve bu alanlardaki politika geliştirme süreçlerinin detaylı bir şekilde incelenmesi yapılması planlanmaktadır.

**Yazar Katkıları:** Kavramsallaştırma Y.Ö, M.Y, H.K ve A.S.U. Metodoloji Y.Ö, M.Y. Biçimsel Analiz A.S.U ve H.K. Kaynaklar Y.Ö, M.Y, A.S.U. Yazı orijinal taslak hazırlama A.S.U ve H.K, gözden geçirme ve düzenleme Y.Ö, M.Y, A.S.U Tüm yazarlar makalenin yayınlanan versiyonunu okudu ve kabul etti.

**Finansman:** Bu araştırma dışarıdan fon almadı.

**Çıkar çatışmaları:** Yazarlar çıkar çatışması beyan etmemektedir.

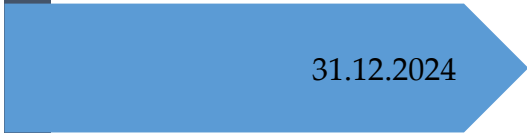

### Kaynaklar

- [1] Şeker, A., "Yenilenebilir enerji, Türkiye'de yenilenebilir enerji potansiyeli ve yeşil pazarlama ve yenilenebilir enerjinin pazarlanması," *Journal of International Social Research*, cilt 9, no. 46, 2016.
- [2] Gündüz, N. K., ve Bicil, İ., "Yenilenebilir enerji yatırımlarını belirleyen faktörler ve gelir gruplarına göre seçilmiş ülkelerin analizi," *Van Yüzcüncü Yıl Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, cilt 7, no. 13, ss. 1-28, 2022.
- [3] Koç, A., Koç, Y., ve Uğurlu, İ., "Dünyada ve Türkiye'deki enerji görünümünün genel değerlendirilmesi," *Mühendis ve Makina*, cilt 59, no. 692, ss. 86-114, 2018.
- [4] Aydoğdu, Ç., "Yenilenebilir enerji sektöründe ve enerji verimliliğinde kamusal destekler ve Türkiye'de yansımaları," *Akademik İzdüşüm Dergisi*, cilt 6, no. 1, ss. 52-74, 2021.
- [5] Accenture, "Sustainable Energy for All: *Opportunities for the Retail Industry*," 2012.
- [6] Accenture, "Uluslararası Arama Odaklı Geliştirme Çalıştayı," 2012.
- [7] DEK-TMK, "Dünya Enerji Konseyi, Türk Millî Komitesi, Enerji İstatistikleri," Ankara, 2006.
- [8] Haldenbilen, S., ve Ceylan, H., "The development of a policy for road tax in Turkey, using a genetic algorithm approach for demand estimation," *Transportation Research Part A: Policy and Practice*, cilt 39, ss. 861-877, 2005.
- [9] Enerji ve Tabii Kaynaklar Bakanlığı (ETKB), "Güneş," <http://www.enerji.gov.tr/tr-TR/Sayfalar/Gunes>, erişim tarihi: 06.06.2015.
- [10] Enerji ve Tabii Kaynaklar Bakanlığı (ETKB), 2007.
- [11] Şenel, M. C., ve Erdem, K. O. Ç., "Dünyada ve Türkiye'de rüzgar enerjisi durumu-Genel değerlendirme," *Mühendis ve Makina*, cilt 56, no. 663, ss. 46-56, 2015.
- [12] Etemoğlu, A. B., Can, M., ve Kılıç, M., "Ülkemiz jeotermal kaynaklarının ikinci kanun verim değerlerine bağlı sınıflandırılması," 2004.
- [13] Özsümbül, B., "Türkiye'nin enerji potansiyeli, gelişimi ve gelecekteki enerji portresinin çıkarılması," Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Enerji Sistemleri Mühendisliği Anabilim Dalı, Ankara, 1999.
- [14] Mahmud, S. F., "The energy demand in the manufacturing sector of Pakistan: some further results," *Energy Economics*, cilt 22, no. 6, ss. 641-648, 2000.
- [15] Hayli, S., "Rüzgâr enerjisinin önemi Dünya'da ve Türkiye'deki durumu," *Fırat Üniversitesi Sosyal Bilimler Dergisi*, cilt 11, no. 1, ss. 1-26, 2001.
- [16] Tanatvanit, S., Limmeechokchai, B., ve Chungpaibulpatana, S., "Sustainable energy development strategies: implications of energy demand management and renewable energy in Thailand," *Renewable and Sustainable Energy Reviews*, cilt 7, no. 5, ss. 367-395, 2003.
- [17] Kumbur, H., vd., "Türkiye'de geleneksel ve yenilenebilir enerji potansiyelinin ve yoğunluğunun karşılanması," *Yenilenebilir Enerji Kaynakları Sempozyumu ve Sergisi*, ss. 21, 2005.
- [18] Kaymakçioğlu, F., vd., "Jeotermal Enerjinin Değerlendirilmesi ve Elektrik Üretimi," *III. Yenilenebilir Enerji Kaynakları Sempozyumu*, Mersin: EMO Y, ss. 1-5, 2005.

- [19] Haldenbilen, S., ve Ceylan, H., "Genetic algorithm approach to estimate transport energy demand in Turkey," *Energy Policy*, cilt 33, no. 1, ss. 89-98, 2005.
- [20] Mathiesen, B. V., Lund, H., ve Nørgaard, P., "Integrated transport and renewable energy systems," *Utilities Policy*, cilt 16, no. 2, ss. 107-116, 2008.
- [21] Külekçi, Ö. C., "Yenilenebilir enerji kaynakları arasında jeotermal enerjinin yeri ve Türkiye açısından önemi," *Ankara Üniversitesi Çevre Bilimleri Dergisi*, cilt 1, no. 2, ss. 83-91, 2009.
- [22] Al-Mofleh, A., vd., "Analysis of sectoral energy conservation in Malaysia," *Energy*, cilt 34, no. 6, ss. 733-739, 2009.
- [23] Abu-Taha, R., "Multi-criteria applications in renewable energy analysis: A literature review," 2011 Proceedings of PICMET'11: Technology Management in the Energy Smart World (PICMET), ss. 1-8, 2011.
- [24] Karayılmazlar, S., vd., "Biyokütlelerin Türkiye'de enerji üretiminde değerlendirilmesi," *Bartın Orman Fakültesi Dergisi*, cilt 13, no. 19, ss. 63-75, 2011.
- [25] Erdal, L., "Enerji arz güvenliğini etkileyen faktörler ve yenilenebilir enerji kaynakları alternatifi," 2011.
- [26] M. Mahmutoğlu, "Türkiye elektrik sektöründe yenilenebilir enerjinin rolü," Yüksek Lisans Tezi, Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Ankara, 2013.
- [27] O. Ellabban, H. Abu-Rub, and F. Blaabjerg, "Renewable energy resources: Current status, future prospects and their enabling technology," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 748-764, 2014.
- [28] A. Vardar and O. Taskin, "Renewable energy sources and Turkey," *International Journal of Energy and Power Engineering*, vol. 3, pp. 245-249, 2014.
- [29] N. K. M. A. Alrikabi, "Renewable energy types," *Journal of Clean Energy Technologies*, vol. 2, no. 1, pp. 61-64, 2014.
- [30] B. V. Mathiesen *et al.*, "CEESA 100% Renewable Energy Transport Scenarios towards 2050: Technical Background Report Part 2," 2014.
- [31] H. Doğan and N. Yılankırkan, "Türkiye'nin enerji verimliliği potansiyeli ve projeksiyonu," *Gazi Üniversitesi Fen Bilimleri Dergisi Bölüm C: Tasarım ve Teknoloji*, vol. 3, no. 1, pp. 375-384, 2015.
- [32] J. Mohtasham, "Renewable energies," *Energy Procedia*, vol. 74, pp. 1289-1297, 2015.
- [33] Ö. Torunoğlu Gedik, "Türkiye'de yenilenebilir enerji kaynakları ve çevresel etkileri," Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, 2015.
- [34] F. Ç. Kılıç, "Güneş enerjisi, Türkiye'deki son durumu ve üretim teknolojileri," *Mühendis ve Makina*, vol. 56, no. 671, pp. 28-40, 2015.
- [35] S. N. Çelik, "Türkiye'nin enerjide dışa bağımlılığının azaltılmasında yenilenebilir enerji kaynaklarının önemi," Yüksek Lisans Tezi, Anadolu Üniversitesi, 2015.
- [36] B. V. Mathiesen *et al.*, "Smart Energy Systems for coherent 100% renewable energy and transport solutions," *Applied Energy*, vol. 145, pp. 139-154, 2015.
- [37] S. Han and J. Kim, "Optimization-based integration and analysis of a complex renewable energy system for the transportation sector," *Chemical Engineering Research and Design*, vol. 128, pp. 1-14, 2017.
- [38] S. Ruggiero and H. Lehtonen, "Renewable energy growth and the financial performance of electric utilities: A panel data study," *Journal of Cleaner Production*, vol. 142, pp. 3676-3688, 2017.
- [39] K. Ásványi *et al.*, "Literature review of renewable energy in the tourism industry," *Journal of Environmental Management & Tourism*, vol. 8, no. 2 (18), p. 476, 2017.
- [40] C. Karaca, A. Ulutaş, and M. Eşgünoğlu, "Türkiye'de optimal yenilenebilir enerji kaynağının COPRAS yöntemiyle tespiti ve yenilenebilir enerji yatırımlarının istihdam artırıcı etkisi," *Maliye Dergisi*, vol. 172, pp. 111-132, 2017.
- [41] N. Bayomi and J. E. Fernandez, "Trends of energy demand in the Middle East: A sectoral level analysis," *International Journal of Energy Research*, vol. 42, no. 2, pp. 731-753, 2018.
- [42] S. R. Paramati, N. Apergis, and M. Ummalla, "Dynamics of renewable energy consumption and economic activities across the agriculture, industry, and service sectors: evidence in the perspective of sustainable development," *Environmental Science and Pollution Research*, vol. 25, pp. 1375-1387, 2018.
- [43] E. A. Yılmaz and C. A. N. Hatice, "Türkiye'nin yenilenebilir enerji potansiyeli ve gelecek hedefleri," *Ordu Üniversitesi Sosyal Bilimler Enstitüsü Sosyal Bilimler Araştırmaları Dergisi*, vol. 8, no. 3, pp. 525-535, 2018.
- [44] F. Ö. Alper, "Yenilenebilir enerji ve ekonomik büyüme arasındaki ilişki: 1990-2017 Türkiye örneği," *Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 8, no. 2, pp. 223-242, 2018.
- [45] A. García-Olivares, J. Solé, and O. Osychenko, "Transportation in a 100% renewable energy system," *Energy Conversion and Management*, vol. 158, pp. 266-285, 2018.
- [46] Temiz Enerji Portalı, "Perakendecilerin kurumsal sürdürülebilirlik stratejileri kapsamında yenilenebilir enerji kaynak kullanımını düzeylerine ilişkin literatür incelemesi," 8. Ulusal Lojistik ve Tedarik Zinciri Kongresi, p. 36, 2019.
- [47] B. Kayışoğlu and B. Diken, "Türkiye'de yenilenebilir enerji kullanımının mevcut durumu ve sorunları," *Tarım Makinaları Bilimi Dergisi*, vol. 15, no. 2, pp. 61-65, 2019.
- [48] E. Aşkın and U. Aşkın, "Yeşil işler: Yenilenebilir enerji sektörünün istihdam yaratma potansiyeli," *International Congress of Energy Economy and Security*, 2019.
- [49] J. J. Buonocore *et al.*, "Metrics for the sustainable development goals: renewable energy and transportation," *Palgrave Communications*, vol. 5, no. 1, 2019.
- [50] M. Hasanuzzaman *et al.*, "Energy demand," *Energy for Sustainable Development*, pp. 41-87, Academic Press, 2020.
- [51] M. Güllü and Z. Kartal, "Türkiye'de yenilenebilir enerji kaynaklarının istihdam etkisi," *Sakarya İktisat Dergisi*, vol. 10, no. 1, pp. 36-65, 2021.
- [52] S. Panda, P. K. Rout, and B. K. Sahu, "Residential sector demand side management: A review," 2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON), IEEE, 2021.
- [53] P. A. Østergaard *et al.*, "Renewable energy for sustainable development," *Renewable Energy*, vol. 199, pp. 1145-1152, 2022.
- [54] M. Maaouane *et al.*, "Using neural network modelling for estimation and forecasting of transport sector energy demand in developing countries," *Energy Conversion and Management*, vol. 258, 115556, 2022.

- [55] D. E. Aliabadi *et al.*, "Future renewable energy targets in the EU: Impacts on the German transport," *Transportation Research Part D: Transport and Environment*, vol. 124, 103963, 2023.
- [56] A. Kızmaz, "Akıllı şebekelere yenilenebilir enerji kaynaklarının entegrasyonunun çevre açısından incelenmesi," 2023.





31.12.2024

Journal of Statistics & Applied Sciences, Issue – 10  
ISSN 2718-0999

