

İSTATİSTİK ARAŞTIRMA DERGİSİ

JOURNAL OF STATISTICAL
RESEARCH

JSRTR

ISSN 2791-7614

Cilt
Volume **12**

Sayı
Issue **01**

Yıl
Year **2022**

**İSTATİSTİK ARAŞTIRMA DERGİSİ
ULUSLARARASI HAKEMLİ DERGİ**

JOURNAL OF STATISTICAL RESEARCH
INTERNATIONAL PEER-REVIEWED JOURNAL

**ISSN: 2791-7614
CİLT/VOL. 12 • SAYI / ISSUE 1 • 2022**

**Türkiye İstatistik Kurumu Adına Sahibi / Owner on Behalf of Turkish Statistical Institute
Erhan ÇETİNKAYA**

**Baş Editör / Editor - in - Chief
Prof. Dr. Selahattin GÜRİŞ**

**Editör Yardımcısı / Assistant Editor
Prof. Dr. Nurcan METİN
Eğitim Daire Başkanı Orçun AYDIN**

**Sorumlu Müdür / Responsible Manager
Stratejik Planlama, İzleme ve Değerlendirme Grup Başkanı Gönül KALENDER**

**Yabancı Dil Editörü / Foreign Language Editor
TÜİK Uzmanı Nilgün DORSAN**

İletişim / Communication
Türkiye İstatistik Kurumu
Devlet Mah.
Necatibey Cad. No: 114
06420 Çankaya/Ankara/TÜRKİYE

Tel. / Phone: +90 312 454 73 50
Belgeç / Fax: +90 312 454 82 96
E-posta / E-mail: journal@tuik.gov.tr
İnternet Sayfası / Web Site:
<https://journal.tuik.gov.tr>

Türkiye İstatistik Kurumu yayını olan İstatistik Araştırma Dergisi, yılda iki kez yayımlanan uluslararası hakemli bir dergidir. Makalelerin bilimsel ve etik kurallara uygunluğu yazarların sorumluluğundadır; Türkiye İstatistik Kurumu sorumlu tutulamaz.

Journal of Statistical Research, the publication of the Turkish Statistical Institute, is an international peer-reviewed journal published twice a year. Compliance of the articles with scientific and ethical rules is the responsibility of the authors; Turkish Statistical Institute can not be held responsible.

Baş Editör'den

Değerli okuyucular,

İstatistik Araştırma Dergisi, Türkiye İstatistik Kurumu'nun 2002 yılında yayınlanmaya başladığı uluslararası akademik bir dergisidir.

Yayın politikası yenilenen ve kısa bir aradan sonra Temmuz 2021'de yeniden yayınlanmaya başlayan İstatistik Araştırma Dergisi'nde istatistik yanında ekonometri, yöneylem araştırması ve istatistiğin araç olarak kullanıldığı diğer bilim dallarında teorik çalışmalara yer verdiğimiz gibi, uygulamalı orijinal çalışmalara yer verilmektedir.

İstatistik Araştırma Dergisi'nin hedefi söz ettiğimiz konularda yayınlayacağı makaleler ile literatüre katkı sağlamaktır. Bu doğrultuda, derginin farklı endekslerde taranmasına yönelik çalışmalara da yer verilmektedir.

Bu sayıda makaleleri ile destek veren yazarlarımıza, dergimizin bilimsel çizgisine yön veren hakemlerimize, dergimizin Editörler Kuruluna üye olmayı kabul ederek bizi onurlandıran değerli bilim insanlarına, desteklerinden ötürü TÜİK Başkanı Sayın Erhan ÇETİNKAYA'ya ve emeği geçen herkese teşekkürlerimi sunarım.

Dergimizin Temmuz 2022 (Cilt 12, Sayı 1) sayısının bilim dünyasına katkı sağlamasını diler, bundan sonraki sayılarımıza da ilgi ve desteğinizin devamını diler, saygılarımı sunarım.

Prof. Dr. Selahattin GÜRİŞ
Baş Editör

Editörler / Editors

Prof. Dr.	Ebru ÇAĞLAYAN AKAY, Marmara Üniversitesi, İstanbul Fikri AKDENİZ, Çığ Üniversitesi, Mersin Cem CANEL, University of North Carolina Wilmington, Amerika Birleşik Devletleri Mehmet Ali CENGİZ, Ondokuz Mayıs Üniversitesi, Samsun Meral ÇETİN, Hacettepe Üniversitesi, Ankara Özlem İLK DAĞ, Orta Doğu Teknik Üniversitesi, Ankara Burak GÜRİŞ, İstanbul Üniversitesi, İstanbul Jamal HUSEİN, Angelo State University, Amerika Birleşik Devletleri Cem KADILAR, Hacettepe Üniversitesi, Ankara Safet KOZAREVIC, University of Tuzla, Bosnia and Herzegovina, Bosna-Hersek Sakkthivel Annamalai MANICKAM, Skyline University College, Birleşik Arap Emirlikleri Ünal Halit ÖZDEN, İstanbul Ticaret Üniversitesi, İstanbul Ömer ÖZTÜRK, The Ohio State University, Amerika Birleşik Devletleri Muthucattu Thomas PAUL, Papua New Guinea University of Technology, Papua Yeni Gine Mustafa SEVÜKTEKİN, Bursa Uludağ Üniversitesi, Bursa Ram SHANMUGAM, Texas State University, Amerika Birleşik Devletleri Kutluk Kağan SÜMER, İstanbul Üniversitesi, İstanbul Semra ERPOLAT TAŞABAT, Mimar Sinan Güzel Sanatlar Üniversitesi, İstanbul Mustafa TEKİN, İstanbul Üniversitesi, İstanbul Münevver TURANLI, İstanbul Ticaret Üniversitesi, İstanbul
Prof.	Emeritus Barry C. ARNOLD, University of California, Amerika Birleşik Devletleri Hanna DUDEK, Warsaw University of Life Sciences, Polonya Nikolai KOLEV, University of Sao Paulo, Brezilya Debasis KUNDU, Indian Institute of Technology, Hindistan W. Y. Wendy LOU, University of Toronto, Kanada
Doç. Dr.	İbrahim DEMİR, Türkiye İstatistik Kurumu Başkanlığı, Ankara
Dr.	Ojonugwa USMAN, Federal College of Education (Technical), Nijerya Subhash Kumar YADAV, Babasaheb Bhimrao Ambedkar University, Hindistan

Dergimize gönderilen makalelerin tümü intihal tespit aracıyla incelenmektedir.

All articles submitted to our journal are analyzed by plagiarism detection tools.

İÇİNDEKİLER / CONTENTS

ARAŞTIRMA MAKALELERİ / ORIGINAL RESEARCH ARTICLES

- 1-15 Öznur ÖZGÜR, Aycan HEPSAĞ
Doğrusal ve Doğrusal Olmayan Birim Kök Testlerinin Gecikme Uzunluğuna Olan Duyarlılığı /
Linear and Nonlinear Unit Root Tests Sensitivity to Lag Length
- 16-38 Ebubekir ERDOĞAN, Cavit YEŞİLYURT
Cinsiyete Dayalı Ücret Ayrımcılığının ISCO Meslek Gruplarına Göre Analizi / Gender Wage
Discrimination Analysis by ISCO Occupational Groups
- 39-55 Ersin ERCAN, Mehmet Mustafa YATARKALKMAZ, Onur Fatih BULUT
The Impact of Income Level on Household Greenhouse Gas Emissions: A Case Study for Turkey
/ Gelir Düzeyinin Hanehalkı Sera Gazı Emisyonlarına Etkisi: Türkiye için Bir Vaka Çalışması
- 56-67 Selin TEKİN, Birkan ERGÜÇ, Serdar ÖZ
Açık Veri ve Coğrafi Bilgi Sistemleri Kullanarak Türkiye için Sürdürülebilir Kalkınma
Göstergeleri Tahmini / Estimating Sustainable Development Indicators Using Open Data and
Geographic Information Systems for Turkey
- 68-81 Osman SERT, Alpaslan AKÇORAOĞLU
Tabaka Sınırlarının Belirlenmesinde, Kümeleme Analiz Yöntemleri ve Tabaka Sınırı Belirleme
Yöntemlerinin Karşılaştırılması / Comparison of Cluster Analysis Methods and Strata Boundary
Determination Methods for Determination of Strata Boundaries



Doğrusal ve Doğrusal Olmayan Birim Kök Testlerinin Gecikme Uzunluğuna Olan Duyarlılığı

Öznur ÖZGÜR

İstanbul Üniversitesi / Doktora Öğrencisi
oznurozgur8787@gmail.com
Orcid No: 0000-0003-3058-6017

Aycan HEPSAĞ

İstanbul Üniversitesi / Doç.Dr.
hepsag@istanbul.edu.tr
Orcid No: 0000-0001-6223-6914

Özet

Çalışmanın amacı Dickey-Fuller (1979), Kruse (2011), KSS (2003) ve Sollis (2009) tarafından önerilen birim kök testlerinin boyut ve güç özellikleri üzerindeki etkisini analiz etmektir. Birim kök testlerinin gecikme uzunlukları kullanılarak uygulanan çalışmada simülasyon analizinden yararlanılmaktadır. 1 ila 12'ye kadar gecikme uzunluğu verilerek 100 ile 250 gözlem için 0.9 ve 0.5 otokorelasyon katsayıları ile simülasyon çalışması yapılmaktadır. Her uygulanan analiz neticesinde testlerin gücünde azalış gözlemlenmektedir. Elde edilen ampirik bulgulara göre küçük (1.) ve büyük (12.) gecikme uzunluklarının Dickey-Fuller (1979) doğrusal birim kök testinde ve Kruse (2011), KSS (2003) ve Sollis (2009) doğrusal olmayan birim kök testlerinde ciddi güç kaybı ve ciddi boyut bozulmaları sonucuna varılmaktadır.

Anahtar Kelimeler: Boyut Bozulmaları, Doğrusal Birim Kök Testleri, Doğrusal Olmayan Birim Kök Testleri, Testin Gücü

Sorumlu Yazar / Corresponding Author: 1-Öznur ÖZGÜR, İstanbul Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü.

2-Aycan HEPSAĞ, İstanbul Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü.

Atıf / Citation: ÖZGÜR Ö., HEPSAĞ A. (2022). Doğrusal ve Doğrusal Olmayan Birim Kök Testlerinin Gecikme Uzunluğuna Olan Duyarlılığı. İstatistik Araştırma Dergisi, 12 (1), 1-15.

Linear and Nonlinear Unit Root Tests Sensitivity to Lag Length

Abstract

The purpose of the study is to analyze the effect of the unit root tests recommended by Dickey-Fuller (1979), Kruse (2011), KSS (2003) and Sollis (2009) on size and power characteristics. Simulation analysis is utilized in the study, which was applied using the delay lengths of unit root tests. Simulation study is conducted with 0.9 and 0.5 autocorrelation coefficients for 100 and 250 observations, with a delay length of 1 to 12. As a result of each applied analysis, decrease in the power of the tests was observed. According to the empiric findings, for Dickey-Fuller (1979) linear unit root test and for Kruse (2011), KSS (2003) and Sollis (2009) nonlinear unit root tests, there are serious power loss and size degradation at small(1st) and large (12st) delay lengths.

Keywords: *Dimension Distortions, Linear Time Series Unit Root Tests, Nonlinear Time Series Unit Root Tests, Power Loss*

1. Giriş

Doğrusal modellerde zaman içerisinde yapısal kırılmalar meydana gelebileceği gibi doğrusal olmayan modellerde de benzer etkilerle dönüm noktaları oluşabilmektedir. Bu dönüm noktalarında farklı rejimler ortaya çıkmaktadır. Bu durum farklı rejim modellerinin kullanılmasını ortaya çıkarmaktadır ve aynı zamanda modellerin kullanılabilmesi için rejim sayısının tespiti önemli bir rol oynamaktadır.

Doğrusal olmayan zaman serilerinde 1980'lerde büyük bir yönelim meydana gelerek ve birçok ilerleme göstererek çalışmalarda kendini göstermiştir. Yapılan bu çalışmalarda testin gücü test edilmeyip, uygulama sonucunda elde edilen bulgular değerlendirilmektedir. Ancak bu uygulamada hem doğrusal birim kök testi hem de doğrusal olmayan birim kök testi gücü simülasyon yardımı ile analiz edilmeye çalışılmaktadır.

Doğrusal olmayan zaman serileri varyansta doğrusal olmayan ve ortalamada doğrusal olmayan olmak üzere ikiye ayrılmaktadır. Bu çalışmada ortalamada doğrusal olmayan zaman serilerinden yararlanılmaktadır. Uygulamanın amacı simülasyon çalışması uygulanarak doğrusal birim kök testleri ve doğrusal olmayan birim kök testlerinin gücü incelenerek, elde edilen sonuçlara göre gecikme uzunluğundaki her bir artışın birim kök testlerinin gücünde azalışlar ve boyutta da ciddi bozulmalara neden olduğunu araştırmaktır.

2. Metodoloji

Doğrusal birim kök testlerinden Dickey ve Fuller (1979) birim kökü test etmek için $Y_t = \rho Y_{t-1} + \varepsilon_t$ AR(1) modelinde $\rho = 1$ hipotezini, $\Delta Y_t = \delta Y_{t-1} + \varepsilon_t$ modelinde ise $\delta = 0$ hipotezini tahmin etmektedir (Dickey ve Fuller, 1979). Ancak Y_{t-1} 'in elde edilmiş katsayısının sıfır olup olmadığını kavramak için bilindik t dağılımı uygulanmamaktadır. Çünkü $\delta = 0$ yani $(\rho = 1)$ önsavı ile Y_{t-1} 'in tahmin edilmiş katsayısının t değeri büyük örneklerde bile t dağılımı ile bağdaşmamaktadır. Bu yüzden Dickey-Fuller t değerinin τ (tau) istatistiğine uygun olduğunu ifade etmektedir, bu durum "Dickey-Fuller sınaması" olarak adlandırılmaktadır. Fakat $\delta = 0$ reddedilirse yani seri durağan ise $student - t$ sınaması kullanılmaktadır. Nedeni ise Dickey- Fuller testinin tek yanlı olarak ifade edilmesidir yani $\delta < 0$ ya da $\rho < 1$ şeklinde olmaktadır. Dickey- Fuller üç farklı şekilde tahmin edilmektedir.

Y_t rassal yürüyüş ise;

$$\Delta Y_t = \delta Y_{t-1} + \varepsilon_t$$

Y_t sabit terimli bir rassal yürüyüş ise;

$$\Delta Y_t = \beta_1 + \delta Y_{t-1} + \varepsilon_t$$

Y_t trendli bir rassal yürüyüş ise;

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \varepsilon_t$$

Burada t zaman ya da trend değişkeni olarak ifade edilmektedir. Üç farklı tahmin için hipotezler aşağıda verilmektedir (Gujarati ve Porter, 2012).

Sıfır hipotezi ; $H_0: \delta = 0$ (*seri birim köklüdür, seri durağan değildir*)

Alternatif hipotez ; $H_1: \delta < 0$ (*seri birim köklü değildir, seri durağandır*)

Doğrusal olmayan birim kök testlerinden Yumuşak Geçişli Otoregresif Model (STAR) Birim Kök Testi Kapetanios, Shin ve Snell (KSS) tarafından 2003 yılında literatüre kazandırılmıştır. Genişletilmiş Dickey-Fuller testinin doğrusal olmayan süreçlerde durağanlığı test etmede çok fazla güçlü olmadığı ve bu durumu ortadan kaldırmak için KSS (2003) yeni bir test geliştirmiştir. Bu test, doğrusal olmayan Üssel Yumuşak Geçişli Otoregresif (ESTAR) sürecinin birim köklü yani durağan olmayan sıfır hipotezini, alternatif hipotez ise durağanlığı ifade etmektedir. Yumuşak Geçişli Otoregresif (STAR) model aşağıda gösterilmektedir.

$$y_t = \beta y_{t-1} + \gamma y_{t-1} \Theta(\theta; y_{t-d}) + \varepsilon_t \quad t = 1, \dots, T \quad (1)$$

burada $\varepsilon_t \sim IID(0, \sigma^2)$ ve β ile γ bilinmeyen parametreleri belirtmektedir. STAR modelinden sonra burada benimsenen geçiş üssel formdadır yani;

$$\Theta(\theta; y_{t-d}) = 1 - \exp(-\theta y_{t-d}^2) \quad (2)$$

Burada $\theta \geq 0$ ve $d \geq 1$ gecikme parametresini ifade etmektedir. 1. ve 2. eşitlikler kullanarak Tek geçişli Üssel STAR (ESTAR) modeli oluşturulmaktadır.

$$y_t = \beta y_{t-1} + \gamma y_{t-1} [1 - \exp(-\theta y_{t-d}^2)] + \varepsilon_t$$
$$\Delta y_t = \phi y_{t-1} + \gamma y_{t-1} [1 - \exp(-\theta y_{t-d}^2)] + \varepsilon_t \quad (3)$$

Burada $\phi = \beta - 1$ olmaktadır. θ , geçiş parametresi ve iki rejim arasındaki geçişin hızını belirleyen parametre olarak anlamlandırılmaktadır (Hepsağ ve Akçalı, 2015, s.78).

ESTAR model (3)'te $\phi = 0$ ve $d = 1$ ise,

$$\Delta y_t = \gamma y_{t-1} [1 - \exp(-\theta y_{t-1}^2)] + \varepsilon_t \quad (4)$$

eşitliğini oluşturmaktadır. Burada θ parametresi, alternatif hipotezin pozitif değer alan belirli bir parametreye odaklanılmasını ifade etmektedir. Test için gerekli hipotezler aşağıda gösterilmektedir.

$$H_0: \theta = 0$$

$$H_1: \theta > 0$$

Burada sıfır hipotezinin doğrudan test edilebilmesi için γ 'nın sıfır hipotez altında tanımlanması gerekmektedir. Bu durumdan dolayı ESTAR modeline birinci dereceden Taylor açılımı uygulanarak aşağıda ifade edilen yardımcı regresyon oluşturulmaktadır (Kapetanios, Shin ve Snell, 2003).

$$\Delta y_t = \delta y_{t-1}^3 + error \quad (5)$$

$\delta = 0$ sıfır hipotezi alternatif hipoteze yani $\delta < 0$ 'a karşı test edilmektedir. t -istatistik değerinin KSS tarafından elde edilen kritik değerden küçük olması birim köklü sıfır hipotezinin reddedilemeyeceğini ifade etmektedir. Eğer t -istatistik değeri kritik değerden büyük ise birim köklü sıfır hipotezi reddedilmektedir. Yani alternatif hipotez ESTAR durağanlık elde edilmektedir.

Sollis (2009) çalışmasında geliştirilen, doğrusal olmayan ESTAR modelinin simetrik veya asimetrik özellik gösterdiği ve zaman serisinin doğrusal olmayan ESTAR durağan süreci olduğu ifade edildiyse ESTAR sürecinin simetrik veya asimetrik olduğu sonucuna varılmaktadır ki bu durum Sollis (2009) testinin avantajını ifade etmektedir. Geliştirilmiş ESTAR modeli için hem üssel bir fonksiyon hem de lojistik fonksiyon bir arada ifade edilmektedir. Model aşağıda gösterilmektedir.

$$\Delta y_t = G_t(\theta_1, y_{t-1}) \{S_t(\theta_2, y_{t-1}) \gamma_1 + (1 - S_t(\theta_2, y_{t-1})) \gamma_2\} y_{t-1} + \varepsilon_t$$
$$G_t(\theta_1, y_{t-1}) = (1 - \exp(-\theta_1 (y_{t-1}^2))) \quad \gamma_1 \geq 0$$
$$S_t(\theta_2, y_{t-1}) = [1 + \exp(-\theta_2 y_{t-1})]^{-1} \quad \gamma_2 \geq 0 \quad (6)$$

Burada $\varepsilon_t \sim IID(0, \sigma^2)$ olarak ele alınmaktadır. Modelde γ_1 geçiş parametresini ve iki rejim arasındaki geçiş hızını belirleyen parametre olarak anlandırılmaktadır. Sıfır hipotezi $\theta_1 = 0$ iken, alternatif hipotez ise $\theta_1 > 0$ şeklinde oluşturularak birim kök sınaması yapılmaktadır. Ancak $\theta = 0$ hipotezini doğrudan sınavabilmek için γ_1, γ_2 ve θ_2 parametrelerinin tanımlı olması gerekmektedir. Sollis (2009), bu sorunu ortadan kaldırabilmek için KSS birim kök testinde olduğu gibi birinci dereceden Taylor açılımı uygulayarak aşağıda ifade edilen yardımcı regresyonu oluşturmaktadır.

$$\Delta y_t = \phi_1 y_{t-1}^3 + \phi_2 y_{t-1}^4 + \varepsilon_t \quad (7)$$

Yukarıda ifade edilen yardımcı regresyon modeli en küçük kareler (EKK) yöntemi ile tahmin edilmektedir.

Birim köklü durağan olmayan sıfır hipotezi $\phi_1 = \phi_2 = 0$ biçiminde ifade edilirken, doğrusal olmayan simetrik ve asimetric ESTAR durağanlığı ifade eden alternatif hipotez ise;

$$\phi_1 \neq \phi_2 \neq 0$$

şeklinde gösterilmektedir.

Kruse (2011) çalışmasında doğrusal olmayan kısıtlama testi olan Wald testini düzenlemiştir. Bu düzenlemedeki amaç üssel geçiş fonksiyonunda sıfıra eşit olmayan bir c parametresi elde edebilmek ve doğrusal olmayan zaman serisi modelini ifade etmektedir.

$$\Delta y_t = \gamma y_{t-1} (1 - \exp\{-\theta(y_{t-1} - c)^2\}) + \varepsilon_t \quad (8)$$

$\theta = 0$ ve $G(y_{t-1}; \theta, c) = 1 - \exp(-\theta(y_{t-1} - c)^2)$ ise (8)'deki eşitliğe birinci Taylor açılımı uygulanarak, (9)'daki regresyon modeli elde edilmektedir.

$$\Delta y_t = \beta_1 y_{t-1}^3 + \beta_2 y_{t-1}^2 + \beta_3 y_{t-1} + u_t \quad (9)$$

Kruse (2011) burada $\beta_3 = 0$ ifadesini kullanarak KSS testine göre daha güçlü bir test meydana getirmektedir.

$$\Delta y_t = \beta_1 y_{t-1}^3 + \beta_2 y_{t-1}^2 + u_t$$

Burada, $\beta_1 = \gamma\theta$ ve $\beta_2 = -2c\gamma\theta$ 'yi ifade etmektedir. Yardımcı regresyona ait hipotezler;

$$H_0: \beta_1 = \beta_2 = 0$$

$$H_1: \beta_1 < 0, \beta_2 \neq 0$$

biçiminde gösterilmektedir. c 'nin gerçek değerler almasından dolayı H_1 alternatif hipotezinde β_2 çift taraflı olmaktadır. Bu testin sorunu bir parametrenin H_1 alternatif hipotezinin hem tek taraflı hem de çift taraflı olarak ifade edilmesidir. Kruse (2011) çalışmasında Abadir ve Distaso (2007) yöntemlerini kullanarak standart bir wald testi uygun olmadığından dolayı tek taraflı parametrenin çift taraflı olana göre ortogonalleştirilmektedir.

Kruse (2011) çalışmasında birim köklü sıfır hipotezini ile durağan ESTAR süreci alternatif hipotezini karşılaştırmak için τ istatistiğini ileri sürmektedir. Aynı zamanda τ istatistiği için kritik değerler oluşturmaktadır.

$$\tau = t_{\beta_1=0}^2 + 1(\beta_1 < 0)t_{\beta_1=0}^2 \quad (10)$$

3. Veri ve Ampirik Bulgular

Bu çalışmada öncelikle; Dickey Fuller(1979) doğrusal birim kök testinin ve KSS (2003), Sollis (2009), Kruse (2011) doğrusal olmayan birim kök testlerinin ampirik boyutunu farklı gözlemler için yani T=100, T=250 olarak %5 anlamlılık düzeyinde inceliyoruz. Gecikme uzunluğu (k) 1'den 12'ye kadar ele alınarak 10.000 tekrara dayalı simülasyon ile boyut deneyini winRATs programı aracılığı ile analiz ediyoruz. İncelenen tüm birim kök testleri için otokorelasyon katsayısı ρ , 0.5 ve 0.9 olarak ele alınmıştır. Veri oluşturma süreci aşağıdaki şekilde ele alınmıştır.

$$y_t = y_{t-1} + \varepsilon_t \quad \text{ile} \quad \varepsilon_t = \rho\varepsilon_{t-1} + u_t \quad (11)$$

burada u_t hata terimi normal dağılmaktadır. Bir AR(1) süreci olan $\rho = 0.5$ ve $\rho = 0.9$, ε_t hata terimini ifade etmekte olup, ε_t hata teriminin otokorelasyonlu olmasına izin verildiği için bu şekilde kullanılmıştır.

Dickey Fuller doğrusal birim kök testi Tablo 1, KSS doğrusal olmayan birim kök testi Tablo 3, Sollis doğrusal olmayan birim kök testi Tablo 5 ve Kruse doğrusal olmayan birim kök testi Tablo 10'daki değerlere göre gecikme uzunluğu arttıkça boyut bozulmalarının oluştuğu gözlemlenmiştir.

KSS (2003), Sollis (2009) ve Kruse (2011) doğrusal olmayan birim kök testlerinin güçlülük özelliğini belirlemek için sırasıyla aşağıda ifade edilen veri oluşturma süreçleri ele alınmıştır.

$$\Delta y_t = \gamma y_{t-1} [1 - \exp(-\theta y_{t-1}^2)] + \varepsilon_t \quad (12)$$

$$\Delta y_t = [1 - \exp(-\theta_1 (y_{t-1}^2))] \{ [1 + \exp(-\theta_2 y_{t-1})]^{-1} \gamma_1 + (1 - ([1 + \exp(-\theta_2 y_{t-1})]^{-1})) \gamma_2 \} y_{t-1} + \varepsilon_t \quad (13)$$

$$\Delta y_t = \gamma y_{t-1} (1 - \exp\{-\theta (y_{t-1} - c)^2\}) + \varepsilon_t \quad (14)$$

KSS doğrusal olmayan birim kök testinin güçlülük özelliğini sınavabilmek için $\rho = 0.5$ ve 0.9 , $\gamma = -1.5$; $\theta = 1.0, 0.1, 0.01$ için ifade edilen parametre aralığı belirlenmiştir.

Kruse doğrusal olmayan birim kök testinin güçlülük özelliğini sınavabilmek için c lokasyon parametresi alt ve üst sınır olarak sırasıyla -5 ve 5 olarak atanmıştır. θ parametresi ise; benzer şekilde alt ve üst sınır olarak 0.001 ve 0.01 değerleri ile ele alınmıştır.

Sollis doğrusal olmayan birim kök testinin güçlülük sınavası için ise; θ_1 parametresi 0.01 ve 1.0 değerleri, θ_2 parametresi 1.0 değeri, γ_1 parametresi $-0.1, -0.3$ ve -0.7 değerleri, γ_2 parametresi $-0.3, -0.7$ ve -1 değerleri atanarak belirlenmiştir.

Dickey Fuller doğrusal birim kök testi Tablo 2, KSS doğrusal olmayan birim kök testi Tablo 4, Sollis doğrusal olmayan birim kök testi Tablo 6, Tablo 7, Tablo 8, Tablo 9 ve Kruse doğrusal olmayan birim kök testi Tablo 11'de gecikme uzunluğu arttıkça testlerin gücünde önemli bir şekilde güç kaybı oluştuğu gözlemlenmiştir.

Tablo 1. Dickey-Fuller Birim Kök Testi (Sabit Terimli-Trendli)

k	$\rho = 0.5$ için		$\rho = 0.9$ için	
	T = 100	T = 250	T = 100	T = 250
1	0.04840	0.05560	0.04840	0.05560
2	0.04640	0.05190	0.04640	0.05190
3	0.04880	0.05280	0.04880	0.05280
4	0.04920	0.05070	0.04920	0.05070
5	0.04800	0.04950	0.04800	0.04950
6	0.04600	0.04850	0.04600	0.04850
7	0.04350	0.04960	0.04350	0.04960
8	0.04140	0.04720	0.04140	0.04720
9	0.04130	0.04850	0.04130	0.04850
10	0.03990	0.04680	0.03990	0.04680
11	0.04380	0.04590	0.04380	0.04590
12	0.04270	0.04610	0.04270	0.04610

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

Tablo 1'de (gözlem sayısı) T=100 ve T=250 için hem (otokorelasyon katsayısı) $\rho = 0.5$ hem de $\rho = 0.9$ 'da 1'den 12'ye kadar olan k (gecikme uzunluğu) artan her bir gecikmede testin boyutunda bozulmalara neden olduğu görülmektedir. Hem $\rho = 0.5$ hem de $\rho = 0.9$ 'daki değerlerin aynı değeri vermesi, otokorelasyon katsayısının sabit terimli-trendli Dickey-Fuller birim kök testine etkisinin olmadığı sonucuna varılmaktadır.

Tablo 2. Dickey-Fuller Birim Kök Testinin Güçlülük Özelliği (Sabit Terimli-Trendli)

<i>k</i>	$\rho = 0.5$ için		$\rho = 0.9$ için	
	<i>T</i> = 100	<i>T</i> = 250	<i>T</i> = 100	<i>T</i> = 250
1	0.04690	0.05030	0.04760	0.05110
2	0.04420	0.04670	0.04310	0.04880
3	0.04460	0.04660	0.04390	0.04920
4	0.04150	0.04280	0.04040	0.04810
5	0.04070	0.04220	0.03880	0.04760
6	0.03650	0.04060	0.03430	0.04200
7	0.03480	0.03970	0.03260	0.04350
8	0.03220	0.04040	0.02820	0.04210
9	0.03050	0.04100	0.02760	0.04310
10	0.02720	0.03960	0.02580	0.04450
11	0.02430	0.03900	0.02250	0.04200
12	0.02280	0.03830	0.02240	0.04240

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

Tablo 2’de (gözlem sayısı) *T*=100 ve *T*=250 için *k* (gecikme uzunluğu) artan her bir gecikmenin testin gücünü azalttığı tablodaki değerlerde açıkça görülmektedir. Otokorelasyon katsayısı hem $\rho = 0.5$ hem de $\rho = 0.9$ ’da hesaplanan her bir değer aynı sonucu vermesi, otokorelasyon katsayısının sabit terimli ve trendli Dickey-Fuller birim kök testinin güçlülük özelliğine etkisinin olmadığını ifade etmektedir.

Tablo 3. Trendden Arındırılmış KSS Birim Kök Testi

<i>k</i>	$\rho = 0.5$ için		$\rho = 0.9$ için	
	<i>T</i> = 100	<i>T</i> = 250	<i>T</i> = 100	<i>T</i> = 250
1	0.05000	0.05320	0.05000	0.05320
2	0.04620	0.05000	0.04620	0.05000
3	0.04850	0.04950	0.04850	0.04950
4	0.04940	0.04950	0.04940	0.04950
5	0.04630	0.05010	0.04630	0.05010
6	0.04300	0.04910	0.04300	0.04910
7	0.04300	0.04720	0.04300	0.04720
8	0.04070	0.04610	0.04070	0.04610
9	0.04250	0.04390	0.04250	0.04390
10	0.03850	0.04360	0.03850	0.04360
11	0.04040	0.04380	0.04040	0.04380
12	0.03620	0.04100	0.03620	0.04100

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

Tablo 3’te, *T*=100 ve 250 için gecikme uzunluğu (*k*) 1’den 12’ye arttıkça KSS Birim Kök testinin boyut bozulmalarına neden olduğunu görülmektedir. KSS Birim Kök Test sonuçları ile Kruse Birim Kök Test sonuçları karşılaştırıldığında gecikme sayısı arttıkça önemli derecede boyut bozulmalarının meydana geldiği gözlemlenmektedir.

Tablo 4. Trendden Arındırılmış KSS Birim Kök Testinin Güçlülük Özelliği

$\rho = 0.5$ için						
$\gamma = -1.5$						
	$\theta = 1.0$		$\theta = 0.1$		$\theta = 0.01$	
k	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$
1	0.97910	0.98000	0.99560	1.00000	0.04690	0.05030
2	0.82240	0.99780	0.98040	1.00000	0.04420	0.04670
3	0.62730	0.98350	0.95220	1.00000	0.04460	0.04660
4	0.45700	0.94890	0.91880	1.00000	0.04150	0.04280
5	0.35780	0.90410	0.87970	1.00000	0.04070	0.04220
6	0.27980	0.85430	0.84060	1.00000	0.03650	0.04060
7	0.23740	0.81430	0.79640	0.99980	0.03480	0.03970
8	0.19030	0.74630	0.75340	0.99950	0.03220	0.04040
9	0.16860	0.71050	0.71000	0.99920	0.03050	0.04100
10	0.14230	0.66860	0.67750	0.99850	0.02720	0.03960
11	0.11890	0.62580	0.65060	0.99870	0.02430	0.03900
12	0.10750	0.58550	0.60850	0.99750	0.02280	0.03830

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

$\rho = 0.9$ için						
$\gamma = -1.5$						
	$\theta = 1.0$		$\theta = 0.1$		$\theta = 0.01$	
k	$T = 100$	$T = 250$	$T = 100$	$T = 250$	$T = 100$	$T = 250$
1	0.94900	0.99990	0.99880	1.00000	0.04760	0.05110
2	0.63580	0.98610	0.99090	1.00000	0.04310	0.04880
3	0.39290	0.92070	0.97400	1.00000	0.04390	0.04920
4	0.24500	0.80310	0.95220	1.00000	0.04040	0.04810
5	0.16850	0.68000	0.92240	1.00000	0.03880	0.04760
6	0.12070	0.57180	0.88180	1.00000	0.03430	0.04200
7	0.09010	0.49100	0.84040	0.99990	0.03260	0.04350
8	0.07070	0.41850	0.79980	0.99990	0.02820	0.04210
9	0.05970	0.36300	0.77540	0.99980	0.02760	0.04310
10	0.04680	0.32160	0.73360	0.99930	0.02580	0.04450
11	0.04170	0.28880	0.70960	0.99890	0.02250	0.04200
12	0.03750	0.25530	0.66400	0.99870	0.02240	0.04240

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

KSS Birim kök testi güçlülük özelliği için $\rho = 0.5$ ile 0.9 , $\gamma = -1.5$ ve $\theta = 1.0, 0.1, 0.01$ için parametre aralığı oluşturulmaktadır. %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıkları hesaplanmaktadır. $T=100$ ve 250 gözlem sayısı ile $\rho = 0.5$ ve 0.9 otokorelasyon sayısı kullanılarak KSS testinin sonuçları Tablo 4'te gösterilmektedir. Örneğin; ilk tablo için $\rho = 0.5$, $\theta = 1.0$ ve $T=100$ iken KSS birim kök testinin gücü $k=1$ (1 gecikme) için %98 iken $k=12$ (12 gecikme) için %11'e kadar azalma olduğu ifade edilmektedir. İkinci tablo için $\rho = 0.9$, $\theta = 1.0$ ve $T=250$ ile KSS birim kök testinin gücü $k=1$ için %99 iken $k=12$ için %26'ya kadar azalış göstererek testin gücünde önemli bir derecede azalmaya neden olduğu net bir şekilde görülebilmektedir.

Tablo 5. Trendden Arındırılmış Sollis Birim Kök Testi

k	$\rho = 0.5$ için		$\rho = 0.9$ için	
	T = 100	T = 250	T = 100	T = 250
1	0.04910	0.05190	0.04910	0.05190
2	0.04560	0.05150	0.04560	0.05150
3	0.04930	0.05110	0.04930	0.05110
4	0.04810	0.05150	0.04810	0.05150
5	0.04850	0.04890	0.04850	0.04890
6	0.04480	0.05100	0.04480	0.05100
7	0.04270	0.05030	0.04270	0.05030
8	0.03990	0.05000	0.03990	0.05000
9	0.04180	0.05030	0.04180	0.05030
10	0.03980	0.04570	0.03980	0.04570
11	0.03970	0.04650	0.03970	0.04650
12	0.03730	0.04370	0.03730	0.04370

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

Trendden arındırılmış Sollis Birim Kök Testinde de hem gözlem sayıları hem de gecikme sayıları diğer testlerde olduğu gibi aynı olup gecikme uzunluğu arttıkça ciddi derecede boyut bozulmaları meydana gelmektedir. Dickey-Fuller Birim kök testinde olduğu gibi hem $\rho = 0.5$ hem de $\rho = 0.9$ 'daki değerlerin aynı değeri vermesi, otokorelasyon katsayısının trendden arındırılmış Sollis Birim Kök Testini etkilemediğini ifade etmektedir.

Tablo 6. Trendden Arındırılmış SOLLIS Birim Kök Testinin Güçlülük Özelliği

		$\rho = 0.5$											
		$\theta_1 = 0.01$						$\theta_2 = 1.0$					
k		1	2	3	4	5	6	7	8	9	10	11	12
$\gamma_1 = -0.1$	$T=100$	0.08430	0.07800	0.07210	0.06570	0.06310	0.06000	0.06160	0.05640	0.05340	0.04820	0.04640	0.04190
$\gamma_2 = -0.3$	$T=250$	0.19640	0.18300	0.17450	0.16270	0.16000	0.14890	0.14370	0.13530	0.13420	0.12680	0.12540	0.11740
$\gamma_1 = -0.1$	$T=100$	0.10770	0.09660	0.08940	0.08010	0.07810	0.07550	0.07320	0.06650	0.06540	0.05750	0.05540	0.05000
$\gamma_2 = -0.7$	$T=250$	0.30690	0.28260	0.26970	0.25270	0.23890	0.22790	0.21970	0.21260	0.20360	0.19570	0.18800	0.17620
$\gamma_1 = -0.1$	$T=100$	0.12380	0.10940	0.10120	0.09230	0.08860	0.08420	0.08250	0.07320	0.07180	0.06440	0.06180	0.05640
$\gamma_2 = -1$	$T=250$	0.37200	0.33860	0.32570	0.30360	0.29270	0.27440	0.26690	0.25420	0.24450	0.23960	0.22580	0.21620
$\gamma_1 = -0.3$	$T=100$	0.10140	0.09260	0.08380	0.07800	0.07460	0.07370	0.07320	0.06650	0.06150	0.05510	0.05540	0.04890
$\gamma_2 = -0.3$	$T=250$	0.34620	0.31530	0.29860	0.27730	0.26480	0.24910	0.23940	0.22490	0.21530	0.20240	0.19800	0.19000
$\gamma_1 = -0.3$	$T=100$	0.13240	0.11850	0.10780	0.09990	0.09670	0.08820	0.08610	0.07930	0.07660	0.06860	0.06700	0.06100
$\gamma_2 = -0.7$	$T=250$	0.55120	0.49670	0.46210	0.42880	0.40490	0.38240	0.36240	0.34230	0.32840	0.31050	0.29560	0.28160
$\gamma_1 = -0.3$	$T=100$	0.15500	0.13620	0.12650	0.11700	0.11190	0.10800	0.09920	0.09000	0.08740	0.07720	0.07500	0.06880
$\gamma_2 = -1$	$T=250$	0.64710	0.58760	0.54930	0.51030	0.48590	0.45460	0.43140	0.40620	0.39200	0.37100	0.35650	0.33980
$\gamma_1 = -0.7$	$T=100$	0.13310	0.11850	0.10660	0.09810	0.09340	0.08650	0.08540	0.07880	0.07730	0.06670	0.06660	0.05920
$\gamma_2 = -0.3$	$T=250$	0.55230	0.49490	0.46210	0.43010	0.40830	0.38030	0.36220	0.34060	0.32750	0.30900	0.29580	0.27970
$\gamma_1 = -0.7$	$T=100$	0.17920	0.15630	0.14370	0.12960	0.12500	0.11070	0.10860	0.09860	0.09530	0.08330	0.08120	0.07480
$\gamma_2 = -0.7$	$T=250$	0.78710	0.72120	0.66650	0.61990	0.58720	0.55080	0.51890	0.49070	0.47080	0.44420	0.42680	0.40420
$\gamma_1 = -0.7$	$T=100$	0.20990	0.18610	0.16720	0.15110	0.14230	0.12830	0.12450	0.11320	0.10750	0.09650	0.08960	0.08230
$\gamma_2 = -1$	$T=250$	0.86090	0.80530	0.75230	0.70040	0.66730	0.62930	0.59770	0.56600	0.53970	0.51500	0.49290	0.47040

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

Tablo 7. Trendden Arındırılmış Sollis Birim Kök Testinin Güçlülük Özelliği

		$\rho = 0.5$											
		$\theta_1 = 1.0$						$\theta_2 = 1.0$					
k		1	2	3	4	5	6	7	8	9	10	11	12
$\gamma_1 = -0.1$	$T=100$	0.29880	0.23530	0.20050	0.16140	0.14460	0.11970	0.10680	0.09030	0.07980	0.06830	0.06170	0.05420
$\gamma_2 = -0.3$	$T=250$	0.88760	0.81560	0.74030	0.69980	0.60480	0.54400	0.49010	0.43860	0.39800	0.35010	0.31960	0.28900
$\gamma_1 = -0.1$	$T=100$	0.62250	0.51320	0.42990	0.36060	0.31580	0.27080	0.23760	0.20270	0.18190	0.15820	0.14340	0.12890
$\gamma_2 = -0.7$	$T=250$	0.99390	0.98320	0.96360	0.93870	0.91000	0.88100	0.84560	0.80330	0.76820	0.73140	0.69970	0.66500
$\gamma_1 = -0.1$	$T=100$	0.79980	0.70180	0.62050	0.53820	0.48620	0.43100	0.38800	0.34390	0.31370	0.25790	0.26250	0.23720
$\gamma_2 = -1$	$T=250$	0.99830	0.99540	0.99120	0.98200	0.97320	0.96130	0.94940	0.93130	0.91580	0.89530	0.87790	0.85760
$\gamma_1 = -0.3$	$T=100$	0.55650	0.41240	0.31860	0.23550	0.18980	0.14870	0.12320	0.09420	0.08270	0.06500	0.05780	0.04820
$\gamma_2 = -0.3$	$T=250$	0.99270	0.96410	0.90480	0.81870	0.73490	0.64340	0.55350	0.47720	0.41690	0.35150	0.30310	0.26150
$\gamma_1 = -0.3$	$T=100$	0.84100	0.67650	0.53740	0.41220	0.32950	0.26480	0.21620	0.17480	0.14760	0.12420	0.10620	0.09140
$\gamma_2 = -0.7$	$T=250$	0.99990	0.99710	0.98560	0.95860	0.91400	0.85670	0.79230	0.72540	0.65840	0.59720	0.54430	0.49500
$\gamma_1 = -0.3$	$T=100$	0.93620	0.81870	0.69520	0.56930	0.48000	0.40150	0.34430	0.29080	0.25060	0.21520	0.19130	0.16920
$\gamma_2 = -1$	$T=250$	1.00000	0.99970	0.99800	0.98790	0.97110	0.94680	0.91630	0.87870	0.84150	0.79640	0.76460	0.72380
$\gamma_1 = -0.7$	$T=100$	0.84060	0.67140	0.53050	0.40930	0.33610	0.26320	0.21990	0.17610	0.15040	0.12560	0.10490	0.09320
$\gamma_2 = -0.3$	$T=250$	0.99980	0.99770	0.98660	0.95800	0.91040	0.85410	0.78700	0.72350	0.66340	0.60220	0.54620	0.49410
$\gamma_1 = -0.7$	$T=100$	0.95370	0.81180	0.64340	0.48590	0.38280	0.29810	0.24110	0.19110	0.15940	0.13600	0.11800	0.10210
$\gamma_2 = -0.7$	$T=250$	1.00000	0.99940	0.99360	0.97180	0.92900	0.86860	0.79940	0.72180	0.65660	0.58690	0.53180	0.47420
$\gamma_1 = -0.7$	$T=100$	0.98220	0.87900	0.73590	0.58300	0.47330	0.37750	0.31520	0.25950	0.21850	0.18430	0.16170	0.14010
$\gamma_2 = -1$	$T=250$	1.00000	0.99970	0.99700	0.98620	0.96150	0.92040	0.87350	0.81660	0.76230	0.70760	0.66160	0.61060

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

Tablo 8. Trendden Arındırılmış SOLLIS Birim Kök Testinin Güçlülük Özelliği

		$\rho = 0.9$											
		$\theta_1 = 0.01$						$\theta_2 = 1.0$					
	k	1	2	3	4	5	6	7	8	9	10	11	12
$\gamma_1 = -0.1$	$T=100$	0.08430	0.07800	0.07210	0.06570	0.06310	0.06000	0.06160	0.05640	0.05340	0.04820	0.04640	0.04190
$\gamma_2 = -0.3$	$T=250$	0.19640	0.18300	0.17450	0.16270	0.16000	0.14890	0.14370	0.13530	0.13420	0.12680	0.12540	0.11740
$\gamma_1 = -0.1$	$T=100$	0.10770	0.09660	0.08940	0.08010	0.07810	0.07550	0.07320	0.06650	0.06540	0.05750	0.05540	0.05000
$\gamma_2 = -0.7$	$T=250$	0.30690	0.28260	0.26970	0.25270	0.23890	0.22790	0.21970	0.21260	0.20360	0.19570	0.18800	0.17620
$\gamma_1 = -0.1$	$T=100$	0.12380	0.10940	0.10120	0.09230	0.08860	0.08420	0.08250	0.07320	0.07180	0.06440	0.06180	0.05640
$\gamma_2 = -1$	$T=250$	0.37200	0.33860	0.32570	0.30360	0.29270	0.27440	0.26690	0.25420	0.24450	0.23960	0.22580	0.21620
$\gamma_1 = -0.3$	$T=100$	0.10140	0.09260	0.08380	0.07800	0.07460	0.07370	0.07320	0.06650	0.06150	0.05510	0.05540	0.04890
$\gamma_2 = -0.3$	$T=250$	0.34620	0.31530	0.29860	0.27730	0.26480	0.24910	0.23940	0.22490	0.21530	0.20240	0.19800	0.19000
$\gamma_1 = -0.3$	$T=100$	0.13240	0.11850	0.10780	0.09990	0.09670	0.08820	0.08610	0.07930	0.07660	0.06860	0.06700	0.06100
$\gamma_2 = -0.7$	$T=250$	0.55120	0.49670	0.46210	0.42880	0.40490	0.38240	0.36240	0.34230	0.32840	0.31050	0.29560	0.28160
$\gamma_1 = -0.3$	$T=100$	0.15500	0.13620	0.12650	0.11700	0.11190	0.10800	0.09920	0.09000	0.08740	0.07720	0.07500	0.06880
$\gamma_2 = -1$	$T=250$	0.64710	0.58760	0.54930	0.51030	0.48590	0.45460	0.43140	0.40620	0.39200	0.37100	0.35650	0.33980
$\gamma_1 = -0.7$	$T=100$	0.13310	0.11850	0.10660	0.09810	0.09340	0.08650	0.08540	0.07880	0.07730	0.06670	0.06660	0.05920
$\gamma_2 = -0.3$	$T=250$	0.55230	0.49490	0.46210	0.43010	0.40830	0.38030	0.36220	0.34060	0.32750	0.30900	0.29580	0.27970
$\gamma_1 = -0.7$	$T=100$	0.17920	0.15630	0.14370	0.12960	0.12500	0.11070	0.10860	0.09860	0.09530	0.08330	0.08120	0.07480
$\gamma_2 = -0.7$	$T=250$	0.78710	0.72120	0.66650	0.61990	0.58720	0.55080	0.51890	0.49070	0.47080	0.44420	0.42680	0.40420
$\gamma_1 = -0.7$	$T=100$	0.20990	0.18610	0.16720	0.15110	0.14230	0.12830	0.12450	0.11320	0.10750	0.09650	0.08960	0.08230
$\gamma_2 = -1$	$T=250$	0.86090	0.80530	0.75230	0.70040	0.66730	0.62930	0.59770	0.56600	0.53970	0.51500	0.49290	0.47040

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

Tablo 9. Trendden Arındırılmış SOLLIS Birim Kök Testinin Güçlülük Özelliği

		$\rho = 0.9$											
		$\theta_1 = 1.0$						$\theta_2 = 1.0$					
	k	1	2	3	4	5	6	7	8	9	10	11	12
$\gamma_1 = -0.1$	$T=100$	0.29880	0.23530	0.20050	0.16140	0.14460	0.11970	0.10680	0.09030	0.07980	0.06830	0.06170	0.05420
$\gamma_2 = -0.3$	$T=250$	0.88760	0.81560	0.74030	0.69980	0.60480	0.54400	0.49010	0.43860	0.39800	0.35010	0.31960	0.28900
$\gamma_1 = -0.1$	$T=100$	0.62250	0.51320	0.42990	0.36060	0.31580	0.27080	0.23760	0.20270	0.18190	0.15820	0.14340	0.12890
$\gamma_2 = -0.7$	$T=250$	0.99390	0.98320	0.96360	0.93870	0.91000	0.88100	0.84560	0.80330	0.76820	0.73140	0.69970	0.66500
$\gamma_1 = -0.1$	$T=100$	0.79980	0.70180	0.62050	0.53820	0.48620	0.43100	0.38800	0.34390	0.31370	0.25790	0.26250	0.23720
$\gamma_2 = -1$	$T=250$	0.99830	0.99540	0.99120	0.98200	0.97320	0.96130	0.94940	0.93130	0.91580	0.89530	0.87790	0.85760
$\gamma_1 = -0.3$	$T=100$	0.55650	0.41240	0.31860	0.23550	0.18980	0.14870	0.12320	0.09420	0.08270	0.06500	0.05780	0.04820
$\gamma_2 = -0.3$	$T=250$	0.99270	0.96410	0.90480	0.81870	0.73490	0.64340	0.55350	0.47720	0.41690	0.35150	0.30310	0.26150
$\gamma_1 = -0.3$	$T=100$	0.84100	0.67650	0.53740	0.41220	0.32950	0.26480	0.21620	0.17480	0.14760	0.12420	0.10620	0.09140
$\gamma_2 = -0.7$	$T=250$	0.99990	0.99710	0.98560	0.95860	0.91400	0.85670	0.79230	0.72540	0.65840	0.59720	0.54430	0.49500
$\gamma_1 = -0.3$	$T=100$	0.93620	0.81870	0.69520	0.56930	0.48000	0.40150	0.34430	0.29080	0.25060	0.21520	0.19130	0.16920
$\gamma_2 = -1$	$T=250$	1.00000	0.99970	0.99800	0.98790	0.97110	0.94680	0.91630	0.87870	0.84150	0.79640	0.76460	0.72380
$\gamma_1 = -0.7$	$T=100$	0.84060	0.67140	0.53050	0.40930	0.33610	0.26320	0.21990	0.17610	0.15040	0.12560	0.10490	0.09320
$\gamma_2 = -0.3$	$T=250$	0.99980	0.99770	0.98660	0.95800	0.91040	0.85410	0.78700	0.72350	0.66340	0.60220	0.54620	0.49410
$\gamma_1 = -0.7$	$T=100$	0.95370	0.81180	0.64340	0.48590	0.38280	0.29810	0.24110	0.19110	0.15940	0.13600	0.11800	0.10210
$\gamma_2 = -0.7$	$T=250$	1.00000	0.99940	0.99360	0.97180	0.92900	0.86860	0.79940	0.72180	0.65660	0.58690	0.53180	0.47420
$\gamma_1 = -0.7$	$T=100$	0.98220	0.87900	0.73590	0.58300	0.47330	0.37750	0.31520	0.25950	0.21850	0.18430	0.16170	0.14010
$\gamma_2 = -1$	$T=250$	1.00000	0.99970	0.99700	0.98620	0.96150	0.92040	0.87350	0.81660	0.76230	0.70760	0.66160	0.61060

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

Tablo 6, Tablo 7, Tablo 8 ve Tablo 9’da ifade edilen Sollis Birim Kök testinin, Sollis (2009) çalışmasında yer alan γ değerlerinden bazıları ele alınarak simülasyon çalışması yapılmıştır. γ_1 için -0.1, -0.3 ve -0.7 değerleri, γ_2 için -0.3, -0.7 ve -1 değerleri ele alınmıştır. γ_1 ’in -0.3 ve γ_2 ’nin -0.3 ile γ_1 ’in -0.7 ve γ_2 ’nin -0.7 olması Sollis (2009)’da ifade edilen simetrik doğrusal olmayan bir düzenlemenin yani KSS tipi olduğu ifade edilmektedir. γ_1 ve γ_2 ’nin belirlenen değerler içinde farklı değerler alması ile asimetrik doğrusal olmayan düzenlemeyi göstermektedir.

Tablo 6 ile Tablo 7, Trendden Arındırılmış Sollis Birim kök testi güçlülük özelliği simülasyon sonuçlarına göre gecikme uzunluğu arttıkça değerlerde azalma meydana gelerek birim kök testinin güç kaybına neden olduğunu ifade etmektedir. Tablo 8 ve Tablo 9’da otokorelasyon katsayısı, Tablo 6 ile Tablo 7’deki otokorelasyon katsayısından farklı olsa bile simülasyon sonuçları aynı değerleri vermektedir. Böylece Trendden Arındırılmış Sollis Birim kök testi güçlülük özelliğinin otokorelasyon katsayısından etkilenmediğini karşılaştırmalı olarak Tablo 6 ile Tablo 7, Tablo 8 ile Tablo 9’da net bir şekilde görülebilmektedir.

Dickey-Fuller Birim kök testi, KSS Birim kök testi, Kruse Birim kök testi, Sollis Birim kök testi ve tüm bu birim kök testlerin güçlülük özelliği ile yapılan simülasyon çalışmalarında, gecikme uzunluğu arttıkça birim kök testinde ciddi boyut bozulması ve testin gücünde önemli ölçüde güç kaybına neden olduğu sonucuna ulaşılmaktadır.

Tablo 10. Trendden Arındırılmış Kruse Birim Kök Testi

k	$\rho = 0.5$ için		$\rho = 0.9$ için	
	$T = 100$	$T = 250$	$T = 100$	$T = 250$
1	0.04880	0.05080	0.04960	0.05310
2	0.04430	0.05020	0.04500	0.05000
3	0.04880	0.04890	0.04500	0.05190
4	0.04390	0.04580	0.04400	0.04850
5	0.04530	0.04580	0.04260	0.05000
6	0.04250	0.04350	0.03880	0.04570
7	0.04010	0.04260	0.03780	0.04720
8	0.03870	0.04460	0.03580	0.04850
9	0.03740	0.04490	0.03570	0.04910
10	0.03430	0.04220	0.03230	0.04790
11	0.03270	0.04510	0.03170	0.04670
12	0.03070	0.04710	0.03030	0.04710

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

Tablo 10’da simülasyon çalışması ile 100 ve 250 gözlem sayısı için Trendden arındırılmış Kruse birim kök testi, 1. gecikmeden 12. gecikmeye kadar hesaplanan değerlerinde önemli bir şekilde azalma ile boyut bozulmalarına neden olmaktadır. Otokorelasyon katsayılarının teste etkisinin olduğu gözlemlenmiştir, çünkü 0.5 ile 0.9 değerleri için farklı sonuçlar elde edilmiştir.

Tablo 11. Trendden Arındırılmış Kruse Birim Kök Testinin Güçlülük Özelliği

k	$\rho = 0.5$ için		$\rho = 0.9$ için	
	$T = 100$	$T = 250$	$T = 100$	$T = 250$
1	0.16530	0.73970	0.22820	0.88690
2	0.14260	0.67470	0.19570	0.84010
3	0.13220	0.63130	0.18180	0.80010
4	0.11730	0.59150	0.16480	0.76040
5	0.11390	0.56200	0.15350	0.73050
6	0.09960	0.52850	0.14220	0.69800
7	0.09670	0.50890	0.13280	0.67380
8	0.09040	0.48630	0.11990	0.64620
9	0.08370	0.47260	0.11630	0.62740
10	0.07920	0.44880	0.10670	0.60080
11	0.07600	0.43760	0.10330	0.57940
12	0.07270	0.41760	0.09350	0.55920

Not: Tablodaki değerler %5 anlamlılık düzeyinde birim köklü sıfır hipotezinin reddedilme olasılıklarını ifade etmektedir.

Tablo 11’de T=100 ve 250 gözlem sayısı ve 0.5 ile 0.9 otokorelasyon katsayıları için gecikme sayısı arttıkça tüm değerlerde azalış olduğu ifade edilmektedir. T=100 ise $\rho = 0.5$ için k=1’de %17 iken k=12’de %7 değeri ile gecikme sayısı arttıkça değerdeki azalış net bir şekilde görülmektedir. $\rho = 0.9$ için k=1’de %23 iken k=12’de %9 değerini alarak $\rho = 0.5$ ile aynı durumu paylaşmaktadır yani burada da gecikme sayısı arttıkça değerde azalış zıt bir durumun olduğunu ifade etmektedir. T=250 ise $\rho = 0.5$ için k=1’de %74 iken k=12’de %42 değeri ile $\rho = 0.9$ için k=1’de %89 iken k=12’de %56 değeri ile de 100 gözlem ile ulaşılan sonuçlara varılmıştır.

4. Sonuç

Doğrusal zaman serileri pek çok ekonometri çalışmasında mevcut olmuştur. Önemli ölçüde rassal zaman serisi, trendli zaman serisi, mevsimsel zaman serisi ve konjonktürel zaman serisi kalıpları birçok çalışmada kullanılmaktadır. Doğrusal zaman serilerinin çalışmalarda fazla tercih edilme sebebi yorumlanmasının daha basit ve anlaşılır olmasıdır.

Bir piyasadaki işsizlik oranlarındaki azalışlar yerine artışlar ihmal edilirse ciddi boyutta sorunlar ortaya çıkabilmektedir. Doğrusal olmayan seriler doğrusal zaman serileri ile analiz edilirse doğrusal olmayan zaman serileri sonuçlarına göre daha zayıf tahminlere neden olabilmektedir. Bu sebeple çalışmada doğrusal birim kök testlerinden Dickey-Fuller birim kök testi ve doğrusal olmayan birim kök testlerinden de KSS, Sollis ve Kruse birim kök testleri ele alınarak simülasyon çalışması ile gecikme uzunluğuna bağlı olarak boyut ve güç arasında oluşan farklılıklar ifade edilmeye çalışılmıştır.

WinRATS programı kullanılarak gözlem sayısı 100 ve 250 ile otokorelasyon katsayısı 0.5 ve 0.9 için gecikme uzunluğu 1 ila 12 olarak hesaplanarak 10.000 tekrara dayalı simülasyon çalışması yapılmıştır. Doğrusal birim kök testlerinden Dickey-Fuller birim kök testi ve Dickey-Fuller birim kök testi güçlülük özelliği artan her bir gecikme sayısı ile birim kök testinde boyut çarpıklığı ve gücünde azalış olduğu gözlemlenmiştir. Doğrusal olmayan birim kök testlerinden KSS birim kök testi, Sollis birim kök testi ve Kruse birim kök testlerinde de gecikme uzunlukları arttıkça birim kök testlerinde ciddi derecede boyut bozulmaları ve güç kaybı olduğu görülmektedir.

Doğrusal ve doğrusal olmayan birim kök testlerinin birkaçı ile küçük ve büyük gecikmelerin boyut ve güç üzerindeki etkilerinin araştırılması amaçlanmıştır. Doğrusal birim kök testlerinden Dickey-Fuller Birim Kök testi ve doğrusal olmayan birim kök testlerinden Kruse, KSS, Sollis birim kök testleri ile yapılan simülasyon çalışmasında gecikme uzunluğu arttıkça birim kök testlerinin gücünde azalış ve boyut çarpıklığının olduğu ispat edilmiştir. Bu sorunların üstesinden gelenebilmesi için uygulamacılara önerimiz, gecikme uzunluğuna bağlı düzeltilmiş kritik değerler ile kıyaslanma yapılmasıdır. Aynı zamanda bu makalenin birçok ekonometrik çalışmaya temel bir bakış açısı kazandıracığı öngörülmektedir.

Kaynakça

- Abadir, K.M., Distaso, W. (2007). Testing Joint Hypotheses When One of the Alternatives is One-Sided. *Journal of Econometrics*, No:140, p. 695-718.
- Dickey, D. A., Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, C:74, No:366, p.427-431.
- Gujarati, D. N., Porter, D. C. (2012). Temel Ekonometri. Çev. Ümit Şenesen ve Gülay G. Şenesen, 5. bs., *İstanbul, Literatür Yayıncılık*.
- Hepsağ, A., Akçalı, B. Y. (2015). Zayıf Formda Piyasa Etkinliğinin Asimetrik Doğrusal Olmayan Birim Kök Testi ile Analizi: G-7 ve E-7 Ülkeleri Örneği. *BDDK Bankacılık ve Finansal Piyasalar*, C.9, No:2, s.73-90.
- Kapetanios, G., Shin, Y., Snell, A. (2003). Testing for a Unit Root in the Nonlinear STAR Framework. *Journal of Econometrics*, 112(2), p.359–379.
- Kruse, R., (2011). A New Unit Root Test Against ESTAR Based on a Class of Modified Statistics. *Statistical Papers*, No:52, p.71-75.
- Sollis, R., (2009). A Simple Unit Root Test Against Asymmetric STAR Nonlinearity with an Application Real Exchange Rates in Nordic Countries. *Economic Modelling*, No:26, p.118-119.



Cinsiyete Dayalı Ücret Ayrımcılığının ISCO Meslek Gruplarına Göre Analizi

Ebubekir ERDOĞAN

Türkiye İstatistik Kurumu / Anketör

ebubekir_erdogan@hotmail.com

Orcid No: 0000-0001-8906-3005

Cavit YEŞİLYURT

Atatürk Üniversitesi / Prof. Dr.

cavitiesilyurt@atauni.edu.tr

Orcid No: 0000-0001-9814-4085

Özet

Bu çalışmada Türkiye işgücü piyasalarında kadın ve erkek çalışanlar arasında cinsiyetten kaynaklı ücret ayrımcılığının olup olmadığı tespit edilmeye çalışılmıştır. Her yıl düzenli olarak Türkiye İstatistik Kurumu (TÜİK) tarafından uygulanan Hanehalkı işgücü anketininin 2019 yılı verileri kullanılarak analiz yapılmıştır. Yapılan çalışmada kamu çalışanları veri setinden çıkarılarak sadece özel sektörde ücretli maaşlı ve yevmiyeli olarak tam zamanlı çalışanlar dâhil edilmiştir. Uluslararası Standart Meslek Sınıflaması (ISCO)'nda toplamda 10 ana meslek grubu vardır fakat daha reel sonuçlara ulaşabilmek için 1. Grupta bulunan Silahlı Kuvvetlerle ilgili meslekler veri setinden çıkarılmıştır. Diğer 9 meslek grubu için ayrı ayrı analiz yapılmıştır. Oaxaca-Blinder ayrıştırma yöntemi kullanılarak yapılan analizlerde 2 model oluşturulmuştur. Birinci modelde yaş ve kıdem değişkenleri, ikinci modelde ise yaş, kıdem ve eğitim değişkenleri kullanılmıştır. Ücret ayrımcılığının ortaya çıkış sebepleri beşeri sermayeden veya cinsiyetten kaynaklandığı görülmüştür. Analiz sonuçlarına göre bütün meslek guruplarında ücret ayrımcılığı tespit edilmiştir. Tespit edilen ücret ayrımcılığının büyük bir kısmının cinsiyetten kaynaklandığı görülmüştür.

Anahtar Sözcükler: Cinsiyet, ISCO Meslek Kodları, Ücret, Gelir Eşitsizliği

Sorumlu Yazar / Corresponding Author: 1-Ebubekir ERDOĞAN, Türkiye İstatistik Kurumu, Erzurum Bölge Müdürlüğü.

2-Prof. Dr. Cavit YEŞİLYURT, Atatürk Üniversitesi, İktisadi ve İdari Bilimler Fakültesi.

Atf / Citation: ERDOĞAN E., YEŞİLYURT C. (2022). Cinsiyete Dayalı Ücret Ayrımcılığının ISCO Meslek Gruplarına Göre Analizi. İstatistik Araştırma Dergisi, 12 (1), 16-38.

Gender Wage Discrimination Analysis by ISCO Occupational Groups

Abstract

In this study, it has been tried to determine whether there is a gender-based wage discrimination between male and female employees in the Turkish labor market. The analysis was carried out using the 2019 data of the Household labor force survey, which is regularly applied every year by the Turkish Statistical Institute. In the study, public employees were excluded from the data set, and only full-time salaried and casual employees in the private sector were included. There are 10 main occupational groups in the International Standard Classification of Occupations (ISCO), but in order to achieve more realistic results, the occupations related to the Armed Forces in the 1st group have been excluded from the data set. Separate analyzes were made for the other 9 occupational groups. Two models were created in the analyzes using the Oaxaca-Blinder decomposition method. Age and seniority variables were used in the first model, and age, seniority, and education variables were used in the second model. It has been seen that the reasons for the emergence of wage discrimination stem from human capital or gender. According to the results of the analysis, wage discrimination has been determined in all occupational groups. It has been observed that most of the wage discrimination identified is due to gender.

Keywords: Gender, ISCO Occupational Codes, Wages, Income Inequality

1. Giriş

Tam istihdam, bir ekonomideki bütün kaynakların tam kapasite ile üretime dâhil edilme durumudur. Gelişmekte olan ülkeler veya gelişmiş ülkeler iş gücü piyasalarında bu düzeye ulaşmak için çaba sarf ederek, gerekli ekonomik politikaları hayata geçirmeye çalışmaktadırlar. Toplumların zamanla değişen yapılarından dolayı günümüzde kadın istihdamı artmış çalışma hayatına kadınlar da girmiştir. Kadınların dâhil olduğu iş gücü piyasalarında maruz kaldıkları sorunlar da ortaya çıkmaya başlamıştır. Beşeri sermayeden kaynaklanan sorunlar anlaşılabilir ve zamanla çözümü bulunabilir, fakat tüm beşeri özelliklere sahip çalışanlar arasında ortaya çıkan ücret farkı anlaşılabilen yani cinsiyetten kaynaklı ücret ayrımcılığı olarak yorumlanmıştır. Küresel düzeyde sorun haline gelen cinsiyete dayalı ücret eşitsizliğinin her geçen gün uluslararası kamuoyunun ilgisini çekmiştir ve sorunların çözümü için gerekli adımlar atılmaya başlanmıştır. Türkiye Cumhuriyetinin kadın istihdamını artırıcı ve ücretler konusunda eşitlik sağlamak amacı ile hayata geçirdiği, geçirmeyi planladığı politikaları ve uluslararası anlamda desteklediği, taraf olduğu sözleşmeler konunun önemi açısından özetle aşağıda verilmiştir.

Uluslararası çalışma örgütünün 100 no'lu Ücret Sözleşmesi resmi olarak Haziran 1951 tarihinde yürürlüğe girmiştir ve cinsiyet ayrımı yapılmadan tüm sektörler için eşit işlerde çalışan, eşit iş yapanların maddi ve sosyal haklarının eşit olmasını istemiştir. Türkiye'de 1967 yılında kabul edilen ve yürürlüğe giren bu sözleşmenin 1. Maddesi "Eşit değerde iş için erkek ve kadın işçiler arasında ücret eşitliği" şeklindedir. ILO Sözleşmesi (1967). (https://www.ilo.org/ankara/conventions-ratified-by-turkey/WCMS_377269/lang--tr/index.htm)

Birleşmiş Milletler 1979 yılında kadınların aleyhinde olan her türlü ayrımcılığın ortadan kaldırılmasını ve önlenmesinin CEDAW sözleşmesi ile tüm üye devletlerde uygulanmasını istemiştir. Türkiye'nin de 1985 yılında kabul ettiği bu sözleşmenin 11/1 maddesinde "tazminatlar da dâhil eşit ücret alma ve eşit değerde yapılan işe karşı eşit muamele görme ile birlikte işin niteliğinin değerlendirilmesinde eşit muamele görme hakkı" tanımlanmıştır (Kadın İnsan Hakları Yeni Çözümler Derneği/Savunuculuk/Uluslararası sözleşmeler ve mekanizmalar/Cedaw 2017).

Tüm devletler bu sorunu çözmek için farklı uygulamalar hayata geçirmiş, pozitif ayrımcılığı eşitsizlik olarak görmemişlerdir. Uluslararası Çalışma Örgütü'ne üye tüm devletlerin onayladığı 183 sayılı Sözleşme, "hangi meslekte ne tür bir iş yapıyorlarsa yapsınlar tüm kadınları kapsamaktadır. Bu sözleşme, ekonomik güvenlik, iş güvencesi, anne ve bebek sağlığı ile toplumsal cinsiyet eşitliği konularında kadına pozitif ayrımcılık sunan bir içeriktir" şeklindedir (Tüzünkan, 2016).

2. Literatür

Ülkelerin gelecekteki iş gücü piyasaları hakkında bilgi sahibi olmak ve gelecek hakkında plan yapabilmek için zaman zaman durum tespit çalışmaları yapılmıştır. Yapılan çalışmalar gelişmekte olan ülkelerle beraber gelişmiş olan ülkeleri de kapsamaktadır. Tüm bu çalışmalarda işgücü ve iş hayatında kadın çalışanlara doğru bir negatif ücret ayrımcılığı olduğu saptanmıştır (Wright, Ermisch, 1991). Kadınların karşılaştıkları bu ayrımcılık sadece kazanç seviyesiyle yeterli kalmamış liyakatli yükselmenin de önüne geçmiştir. Türkiye'deki cinsiyetten kaynaklı ücret ayrımcılığı konusu Türkiye'nin Avrupa Birliğine girmesinde de engel teşkil etmektedir ve bu konu sık sık gündeme gelerek konu hakkında raporlar düzenlenmiştir (Eraslan, 2012).

Ücret gelir eşitsizliğinin tespitinde farklı teoriler ortaya atılmıştır. Bu teorilerin en çok kullanılanı ve en yaygın olanı Neo-Klasik yaklaşım ve Kurumsal yaklaşım teorileridir. Bahsi geçen yaklaşıma göre beşeri sermaye özellikleri aynı olan kadın ve erkek çalışanlar arasında eğer kadın daha düşük ücret alıyorsa işverene kadın çalışanın maliyeti daha az olmalıdır, diğer türlü verimliliği eşit olan çalışanlar arasında adalet sağlanmalı çalışma barışı bozulmamalıdır (Palaz, 2002). Diğer bir yaklaşım türü olan Kurumsal yaklaşım, işverenler ile tekeli kuruluşların sisteminden kaynaklı olarak bazı grupları diğer gruplara göre daha dezavantajlı duruma gelmelerini sağlamaktadır. Neo-Klasik yaklaşım da kadın çalışanların erkek çalışanlara nazaran daha düşük gelir elde etmelerinin temel nedeni yukarıda bahsedilen durumdur (Palaz, 2002).

Blinder (1973) yaptığı çalışmada ayrımcılığın altında yatan nedenleri bulmaya çalışmıştır ve ücret ayrımcılığının bileşenlere ayrılması gerektiğini savunmuştur. Bu çalışmayı iki türden oluşturmuştur. Birincisi erkekler arasında (beyaz-zenci) ücret farkını ele almıştır ve farklılığın detaylarını incelemiştir. İkinci olarak da cinsiyet ayrımı yaparak beyaz kadınlar ve beyaz erkekler arasında analiz yapmıştır. Birinci çalışmasında beyaz erkekler %45.6 oranında daha fazla saatlik kazanç elde etmektedir. Aradaki farkın (%33) beşeri özellik olarak kadınların erkeklerden daha düşük olduğundan kaynakladığını söylemektedir. Aradaki beşeri özellikleri yok sayarsak ücretlerin %67'si erkeklerin lehine gerçekleştiğini söylemiştir (Blinder, 1973).

Oaxaca ve Neuman (2003) cinsiyet ayrımcılığında Blinder-Oaxaca'ya ek olarak Heckman düzeltme modelini de kullanmışlardır. İsrail iş gücünü analiz ederek erkekler ile kadınlar arasındaki ücret ayrımcılığın yanında etnik ayrımcılığı yani doğu ve batı bölgelerini de incelemişlerdir. Çalışma sonucunda ücret ayrımcılığı oranı etnik ayrımcılık oranından fazla olduğunu tespit etmişlerdir. Öyle ki erkekler kadınlardan daha fazla ücret alırken batıda yaşayanlar da doğuda yaşayanlardan daha fazla ücret aldığı ortaya konulmuştur.

Blinder-Oaxaca ayrıştırma yöntemini kullanarak Yunan iş gücü piyasalarını inceleyen Cholezas ve Tsakloglou (2006), kamu sektöründe erkekler lehine çok az bir fark bulmuşlar. Fakat aynı çalışma özel sektör için erkekler ile kadınlar arasındaki farkı, kadınlar için erkeklerin saatlik ücretinin dörtte biri kadar bulmuştur. Genel olarak çalışmada özellikle özel sektörde gelir ücret eşitsizliğinin kadınların aleyhinde olduğu sonucuna varılmıştır (Cholezas, Tsakloglou, 2006).

Avusturya iş gücü piyasalarını ele alan Böheim vd.(2007) kadın erkek ücretlerinde kadınların maruz kaldığı ayrımcılığı ortaya çıkarmıştır. Gelir ücret eşitsizliği erkekler lehine %14 oranında gerçekleştiği saptamıştır. Avusturya'da en son çalışmalarda kadın erkek ücret eşitsizliği daha önceki çalışmalara nazaran azaldığı tespit edilmiştir, bunun nedeni ise kadınların beşeri sermaye özelliklerinin giderek artması olduğu vurgulanmıştır (Böheim, Hofer, Zulehner, 2007).

Emek konusunda yapılan çalışmaların birçoğu kadın çalışanların emeklerinde uğradığı olumsuz ayrımcılığı konu edinmiştir, bu durum emek piyasalarında gözlemlenen en önemli aksaklıklardan biridir. Yapılan analizde ayrımcılık konusunda en net gözlemlenen ve açık bir şekilde tespit edilen konu ücret ayrımcılığıdır (Topbaş, 2003). 1994 TÜİK Hanehalkı Gelir Dağılım anketi olan şimdiki ismi Hanehalkı Bütçe Anketi verilerini kullanarak özel sektörde kadın erkek ücret ayrımcılığı derecesini tespit etmek için yapılan çalışmada %62 oranında ayrımcılık tespit edilmiştir. Bu çalışmada Oaxaca ve Cotton ayrıştırma modeli kullanılmıştır her iki yöneteme göre de sonuçlar aynı çıkmıştır (Yamak, Topbaş, 2004).

Her alanda olduğu gibi toplumsal cinsiyet eşitsizliğinin görüldüğü alanlardan biri de çalışma alanıdır. Cinsiyete göre çalışma farklılıkları devam etmekte ve kadın çalışanlar belli başlı sektörlerde ve de mesleklerde toplanırlar. Çalışma hayatında kadınların aile içinde rolleri ve toplumun kadının yeri evidir anlayışı ile kadınlara çalışma hayatında gerekli değer verilmemiş ve kadınlar genelde düşük ücret alarak çalışma hayatlarına devam etmişlerdir. Bu durum kadın erkek ücretlerindeki eşitsizliği artırmıştır (Etiler, 2017). Çalışmada ücretsiz aile işçiliğine de yer verilirken bu çalışma şekli kadınlar arasında oldukça yaygın olduğu görülmüştür. Kadın çalışanların toplumsal cinsiyet rolünde en temel görevleri aile üyelerinin bakımı ve ev işleridir, ancak bunun karşılığı olarak ödenmeyen ücret, belirlenemeyen bir emektir.

Kadın çalışanlar erkeklere göre Türkiye ortalamasından daha düşük ücretlerde çalıştırılmakta hatta aynı eğitim seviyesinde ve aynı meslekte olsalar dahi bu değişmemektedir. Ayrıca daha güvencesiz ve esnek işlerde

çalıştırılmaktadırlar ve bazı durumlarda da yönetici konumuna da getirilmemektedirler. Kadınlar hane içindeki rolleri gereği ev işi, çocuk bakımı ve hanenin diğer işlerine zaman ayırdıkları için bu işlerde yoğunlaşmaktadırlar. Kreş, anaokulu gibi çocuk bakımı hizmetlerinin daha ulaşılabilir ve yaygın hale getirilmesi, yaşlı ve hasta bakımına yönelik kamusal hizmetlerin yaygınlaştırılması kadınların bakım emeğine harcadığı vakti azaltacak ve onların işgücüne katılımlarını kolaylaştıracaktır (Kaymaz, Şentürk, 2016).

Antalya ilinde organize sanayi bölgesinde çalışan 91 işveren/yönetici ile yapılan araştırmada işveren ve yönetici pozisyonu açısından kadın çalışanların karşılaştıkları ayrımcılığı ele alan çalışmada, kadın çalışanlar sadece işten atılma korkusu ile çalıştıklarını ve sadece böyle bir ayrımcılığa maruz kaldıklarını ifade etmişlerdir. Araştırmaya katılan kadın çalışanların iş yerinde karşılarına çıkan en büyük problemin iş yerinde kreşin olmamasıdır ancak ücretli veya ücretsiz fark etmez doğum izinlerinin olduğunu ifade etmişlerdir (Alparslan, Bozkurt, Özgöz, 2015). Genel olarak katılımcılar görevlerin hangisi erkek işi hangisi kadın işi bunun ayrımının net bir şekilde yapılmasını düşünmekte çünkü işe alımlarda cinsiyet farklılığının belirleyici bir unsur olduğunu düşünmekte. Katılımcılar ücret konusunda herhangi bir ayrımcılığa maruz kalmadıklarını fakat hamile olan çalışanların işe alınma sürecinde olumsuz karşılandığını ve işe girişini olumsuz yönde etkilediğini düşünmekte (Alparslan, Bozkurt, Özgöz, 2015).

3. Uygulama

Bu araştırmada cinsiyete dayalı ücret ayrımcılığı meslek guruplarına göre ayrı ayrı analiz yapılacaktır. Analiz yönteminde Oaxaca-Blinder ayrıştırma yöntemi kullanılacaktır. Oaxaca-Blinder ayrıştırma yönteminde kullanılan temel denklem aşağıdadır.

$$\ln(W_E) - \ln(W_K) = (X_E - X_K) \beta_E + X_K (\beta_E - \beta_K)$$

Oluşturulan denklemde $\ln(W_E)$ erkeklerin ortalama kazançlarını, $\ln(W_K)$ kadınların ortalama kazançlarını X_E erkeklerin beşeri sermaye özelliklerini, X_K kadınların beşeri sermaye özelliklerini göstermektedir. Modelde beşeri sermayeden kaynaklı fark $(X_E - X_K) \beta_E$ şeklinde ifade edilirken cinsiyetten kaynaklı fark ise $X_K (\beta_E - \beta_K)$ şeklinde ifade edilmiştir.

Araştırma da iki farklı model oluşturulacaktır. Birinci model de yaş ve kıdem değişkenleri oluşturulmuştur ve aşağıda gösterilmiştir.

$$\ln(W_t) = \beta_0 + \beta_1 YAŞ_t + \beta_2 KIDEM_t \quad (1)$$

İkinci modelde ise yaş, kıdem, eğitim değişkenleri ile oluşturulmuş ve aşağıda gösterilmiştir.

$$\ln(W_t) = \beta_0 + \beta_1 YAŞ_t + \beta_2 KIDEM_t + \beta_3 EĞİTİM \quad (2)$$

Modeller oluşturulurken kukla değişken (Dummy Variables) kullanılmıştır. Kukla değişken guruplar arasındaki farkın, matematiksel olarak yorumlanabilmesi için verilerin sınıflandırılarak modele dâhil edilmesini sağlamaktadır. Kukla değişken nitel özellikteki verileri nicel hale getirerek modelin analiz edilmesini sağlamaktadır. Genelde kukla değişken oluşturulurken 0-1 kodlarını alır, bunun sebebi yorumlama yaparken zorlanılmamasıdır. Kukla değişkenin daha iyi anlaşılması için örnek vermek gerekirse, eğitim seviyeleriyle alakalı kukla değişken oluşturulurken bir okul bitirmeyenler 1 değerini alırsa diğer tüm eğitim seviyeleri 0 değerinin alır ve modele bu şekilde dâhil edilir. Her bir eğitim seviyesi için bu şekilde ayrı ayrı kukla değişken oluşturulur.

Tüm eğitim seviyeleri için kukla değişken oluşturulurken referans eğitim seviyesi Yükseköğretim veya Fakülte mezunları alınmıştır.

3.1. Başkanlar Üst Düzey Yöneticiler ve Diğer İdari Müdürler

ISCO meslek grubunda birinci sırada yer alan meslekler aşağıda verilmiştir, yapılan analizler sadece bu meslek grubu içindir. Eğitim seviyesinde referans grup olarak Yükseköğretim veya Fakülte mezunları alınmıştır (Başkanlar, üst düzey yöneticiler ve kanun yapıcılar - Ticari ve idari müdürler - Üretim ve uzmanlaşmış hizmet müdürleri - Ağırhama, perakende ve diğer hizmet müdürleri).

Tablo 1. Tanımlayıcı İstatistikler

Cinsiyet	Değişkenler	Ortalama	Standart Sapma	Katılımcı Sayısı
Erkek	Aylık Kazanç (Ln)	8.529	7.993	1485
	Yaş	39.91	8.645	
	Kıdem	7.75	7.041	
	Bir Okul Bitirmeyenler	0.007	0.259	
	İlkokul	0.559	0.496	
	Meslek Lisesi	0.741	0.261	
	Yüksek Lisans veya Doktora	0.125	0.331	
Kadın	Aylık Kazanç	8.505	7.999	503
	Yaş	37.14	7.477	
	Kıdem	6.91	6.835	
	Bir Okul Bitirmeyenler	0.000	0.000	
	İlkokul	0.652	0.476	
	Meslek Lisesi	0.151	0.358	
	Yüksek Lisans veya Doktora	0.398	0.195	

Bu meslek grubunda toplam 1485 (%74.70) erkek çalışan ve 503 (%25.30) kadın çalışan olduğu görülmüştür (Tablo 1). Aylık kazanç erkeklerde ortalama 5063.04 (Ln 8.529) TL iken kadınlarda ortalama 4941.34 (Ln 8.505) TL olmuştur. Burada aylık kazançlar arasındaki fark daha iyi anlaşılın diye reel gelirler de verilmiştir. Erkekler için yaş kıdem ve eğitim sonuçları ise Tablo 2’de verilmiştir.

Tablo 2. Blinder - Oaxaca Ayrıştırma Modeli Sonuçları

	Ücret Farkı			
	Model 1		Model 2	
	Katsayı	t	Katsayı	t
Erkek	8.384	615.08	8.384	615.08
Kadın	8.347	340.76	8.347	340.76
Ln (W _e) – Ln (W _k)	0.036	0.28	0.037	0.28
Beşeri Sermaye	0.041	0.857	-0.047	0.148
Ayrımcılık	-0.005	0.264	0.084	0.249
Bir Okul Bitirmeyenler	-	-	-0.420	-2.137
İlkokul	-	-	-0.188	-9.192
Genel Ortaokul	-	-	-0.133	-6.682
Genel Lise	-	-	-0.197	-9.724
Meslek Lisesi	-	-	-0.168	-8.277
Yüksek Lisans veya Doktora	-	-	0.212	10.445

Blinder-Oaxaca ayrıştırma modeline göre yapılan analiz çalışmasında sonuçlar Tablo 2’de verilmiştir. Model 1’de (yaş ve kıdem) cinsiyete dayalı ücret ayrımcılığı 0.036 oranıyla erkekler lehine bulunmuştur. Bulunan bu ayrımcılığın (0.041) %114’ü beşeri sermayeden kaynaklanırken, (-0.005) -%14 ise ayrımcılıktan kaynaklandığı gözlemlenmiştir. Model 1’e göre farklılığın beşeri sermayeden kaynaklandığı görülmüştür ve ayrımcılık negatif yönde gözlemlenmiştir (Yüzdesel hesaplamalar ilgili modelin toplam ücret farkına bölünmesiyle elde edilmiştir).

Model 2’de (yaş-kıdem-eğitim) de cinsiyete dayalı ücret ayrımcılığı 0.037 oranıyla erkekler lehine bulunmuştur. Ücret ayrımcılığında beşeri sermayeden kaynaklı fark (-0.047) -127 oranıyla gerçekleşirken ayrımcılık (0.084) % 227 oranıyla tespit edilmiştir. Model 2’de eğitim seviyelerinde Yüksek Lisans veya Doktora mezunları 0.212 oranıyla pozitif yönde çıkarken diğer tüm eğitim seviyeleri farklı oranlarda negatif yönde sonuçlanmıştır. Bu şekilde negatif çıkan sonuçlar aylık kazançta olumlu yönde herhangi bir etkiye sahip olmadığı görülmüştür. Yapılan benzer çalışmalarda okuyazar olmayanlar kadınlarda -70.15 erkeklerde -44.12, İlkokul mezunu olanlar kadınlarda -59.71 erkeklerde -31.98, ortaokul mezun olanlar kadınlarda -63.99 erkeklerde -22.44 oranlarında negatif yönde çıkarken üniversite, Yüksek Lisans veya Doktora mezunları ise pozitif yönde kadınlarda 77.60 erkeklerde ise 29.05 oranında gerçekleşmiştir (Gürler, Üçdoğruk, 2007). Bu meslek gurubu için eğitimin önemsiz olduğu zamanda ayrımcılık beşeri sermayeye dayanırken eğitim değişkeniyle beraber ücret farkının ayrımcılığa yani cinsiyete dayandığı görülmüştür.

Ayrıca Model 2’ye göre eğitim değişkeni beşeri sermaye özelliklerinden önemli sayılmıştır yani bir çalışanın eğitim seviyesi, yaşından ve kıdeminden işveren açısından daha değerli görülmüştür. Bu meslek gurubunun pozisyon bakımından üst düzey yöneticiler olduğundan Yüksek Lisans veya Doktora eğitim seviyesinin önemli olması reel durumla da örtüşmektedir. Eğitim değişkeninin negatif yönde değer alması kadınların erkeklere göre geçen zaman içerisinde maaşına katkısı daha az olduğunu göstermektedir. Yani 25 yıl çalışan kadın ve erkek personeller arasında eğitim değişkeni kadına daha az getiri sağlamaktadır (Kaya, Selim, 2018). Bu durum diğer analiz çeşitleri için de geçerlidir.

3.2. Sağlık, Hukuk, Eğitim ve Diğer Profesyonel Meslek Mensupları

ISCO meslek gurubunda ikinci sırada yer alan meslekler aşağıda verilmiştir, yapılan analizler sadece bu meslek gurubu içindir. Eğitim seviyesinde referans gurup olarak “Yüksekokul veya Fakülte” mezunları alınmıştır. (Bilim ve mühendislik alanlarındaki profesyonel meslek mensupları - Sağlık profesyonelleri - Eğitim ile ilgili profesyonel meslek mensupları - İş ve yönetim ile ilgili profesyonel meslek mensupları - Bilgi ve iletişim teknolojisi ile ilgili profesyonel meslek mensupları - Hukuk, sosyal ve kültür ile ilgili profesyonel meslek mensupları).

Tablo 3. Tanımlayıcı İstatistikler

Cinsiyet	Değişkenler	Ortalama	Standart Sapma	Katılımcı Sayısı
Erkek	Aylık Kazanç (Ln)	8.437	7.864	1783
	Yaş	35.20	9.134	
	Kıdem	4.82	5.652	
	Bir Okul Bitirmeyenler	0.563	0.405	
	İlkokul	0.753	0.431	
	Meslek Lisesi	0.178	0.383	
	Yüksek Lisans veya Doktora	0.292	0.168	
Kadın	Aylık Kazanç	8.168	7.559	1639
	Yaş	32.09	8.405	
	Kıdem	3.92	5.104	
	Bir Okul Bitirmeyenler	0.000	0.000	
	İlkokul	0.791	0.406	
	Meslek Lisesi	0.151	0.358	
	Yüksek Lisans veya Doktora	0.451	0.207	

Bu meslek gurubunda toplamda 1783 (%52.10) erkek çalışan, 1639 (%47.90) kadın çalışan olduğu görülmüştür (Tablo 3). Aylık kazanç erkeklerde ortalama 4619.04 (Ln W_e 8.437) TL iken kadın çalışanlarda ortalama 3527.56 (Ln W_k 8.168) TL olmuştur. Bu meslek gurunda erkek çalışanların yaş ve kıdem ortalamaları kadın çalışanlara göre oldukça yüksek görülmüştür. Erkekler için yaş kıdem ve eğitim sonuçları ise Tablo 4’de verilmiştir.

Tablo 4. Blinder-Oaxaca Ayrıştırma Modeli

	Ücret Farkı			
	Model 1		Model 2	
	Katsayı	t	Katsayı	t
Erkek	8.307	711.08	8.307	711.07
Kadın	8.066	778.32	8.066	778.3
$\ln(W_e) - \ln(W_k)$	0.241	15.43	0.241	15.43
Beşeri Sermaye	0.056	8.66	0.055	6.45
Ayrımcılık	0.185	12.52	0.186	13.58
Bir Okul Bitirmeyenler	-	-	-0.036	-2.508
İlkokul	-	-	-0.047	-3.143
Genel Ortaokul	-	-	-0.147	-9.882
Genel Lise	-	-	-0.140	-9.143
Meslek Lisesi	-	-	-0.096	-6.458
Yüksek Lisans veya Doktora	-	-	0.311	20.399

Blinder-Oaxaca ayrıştırma modeline göre yapılan analiz çalışması sonuçları Tablo 4’de verilmiştir. Model 1’de (yaş ve kıdem) cinsiyete dayalı ücret ayrımcılığı 0.241 oranıyla erkeklerin lehine olacak şekilde gerçekleşmiştir. Beşeri sermayeden kaynaklı fark (0.056) %23 iken ayrımcılıktan kaynaklı fark (0.185) %77 oranında gerçekleşmiştir. Ücret ayrımcılığında farkın büyük bir kısmının cinsiyete dayandığı görülmüştür. Model 2 (yaş-kıdem-eğitim) de ayrımcılık Model 1 ile paralel sonuçlar elde edilmiştir. İstatistiksel olarak anlamlı bir fark oluşturmamakla beraber beşeri sermayeden kaynaklı fark (0.055) %22.5 oranıyla gerçekleşirken cinsiyetten kaynaklı fark (0.186) %77.5 oranıyla gerçekleşmiştir.

Analiz çalışmasında profesyonel meslek mensuplarının kazançlarında eğitim seviyeleri bakımından referans eğitim gurubuna göre negatif yönde sonuçlanan eğitim seviyeleri bir okul bitirmeyenler (-0.36), İlkokul (-0.47), genel ortaokul (-0.147), Genel Lise (-0.140) ve Meslek Lisesi (-0.96) olarak gerçekleşmiştir. Referans eğitim seviyesine göre Yüksek Lisans veya Doktora eğitim seviyesi 311 oranıyla pozitif yönde etkilemiştir. Yapılan çalışmalarda eğitim katsayılarının negatif yönde değer almaları normal görülmüş ve bu durum şu şekilde yorumlanmıştır. Eğitim değişkeninin geçen zaman içerisinde aylık kazançta getirisi yıl geçtikçe düşmektedir ve bu durumdan dolayı negatif kat sayılar çıkmaktadır. Yapılan analiz sonuçlarına göre eğitimin katsayısı, kadınlarda erkeklere göre daha az olduğunu söylenmiştir. Yani kadınların eğitim getirisi erkeklere göre daha azdır (Kaya, Selim, 2018). Bu durum diğer analiz çeşitleri için de geçerlidir.

3.3. Sağlık, Hukuk, Eğitim ve Diğer Yardımcı Profesyonel Meslek Mensupları

ISCO meslek gurubunda üçüncü sırada yer alan meslekler aşağıda verilmiştir, yapılan analizler sadece bu meslek gurubu içindir. Eğitim seviyesinde referans grup olarak “Yüksekokul veya Fakülte” mezunları alınmıştır (Bilim ve mühendislik ile ilgili yardımcı profesyonel meslek mensupları - Yardımcı sağlık profesyonelleri - İş ve idare ile ilgili yardımcı profesyonel meslek mensupları - Hukuk, sosyal, kültür ve benzeri alanlar ile ilgili yardımcı profesyonel meslek mensupları - Bilgi ve iletişim teknisyenleri).

Tablo 5. Tanımlayıcı İstatistikler

Cinsiyet	Değişkenler	Ortalama	Standart Sapma	Katılımcı Sayısı
Erkek	Aylık Kazanç	3277.97	1500.113	3094
	Yaş	36.26	9.009	
	Kıdem	6.9	7.062	
	Bir Okul Bitirmeyenler	0.390	0.621	
	İlkokul	0.384	0.486	
	Meslek Lisesi	0.152	0.122	
	Yüksek Lisans veya Doktora	0.292	0.455	
Kadın	Aylık Kazanç	2761.07	1154.211	1053
	Yaş	32.08	8.178	
	Kıdem	4.92	5.548	
	Bir Okul Bitirmeyenler	0.002	0.533	
	İlkokul	0.570	0.495	
	Meslek Lisesi	0.323	0.176	
	Yüksek Lisans veya Doktora	0.200	0.400	

Bu meslek gurubunda toplamda 3094 (%74.61) erkek çalışan, 1053 (%25.39) kadın çalışan olduğu görülmüştür (Tablo 5). Aylık kazanç erkeklerde ortalama 3277.97 (Ln W_e 8.094) TL iken kadın çalışanlarda ortalama 2761.07 (Ln W_k 7.923) TL olmuştur. Bu meslek gurunda erkek çalışanların yaş ve kıdem ortalamaları kadın çalışanlara göre oldukça yüksek görülmüştür. Erkekler için yaş kıdem ve eğitim sonuçları ise Tablo 6'da verilmiştir.

Tablo 6. Blinder-Oaxaca Ayrıştırma Modeli

	Ücret Farkı			
	Model 1		Model 2	
	Katsayı	t	Katsayı	t
Erkek	8.02	1234.91	8.02	1234.91
Kadın	7.864	808.94	7.864	808.91
Ln (W_e) – Ln (W_k)	0.156	13.35	0.156	13.35
Beşeri Sermaye	0.043	10.12	0.011	1.74
Ayrımcılık	0.113	9.88	0.145	13.25
Bir Okul Bitirmeyenler	-	-	-0.36	-2.508
İlkokul	-	-	-0.14	-9.143
Genel Ortaokul	-	-	-0.114	-7.889
Genel Lise	-	-	-0.100	-6.779
Meslek Lisesi	-	-	-0.136	-8.701
Yüksek Lisans veya Doktora	-	-	0.101	7.006

Analiz sonuçlarına göre Model 1'de ve Model 2'de cinsiyete dayalı ücret farkı 0.156 oranıyla gerçekleşmiştir ve bu farkın büyük bir kısmının ayrımcılıktan dolayı ortaya çıktığı görülmüştür. Model 1 için beşeri sermayeden kaynaklı fark (0.043) %28 oranıyla gerçekleşirken Model 2 için beşeri sermayeden kaynaklı fark (0.011) %7 oranında gerçekleşmiştir. Yine Model 1 için ayrımcılıktan kaynaklı fark (0.113) %72 oranıyla gerçekleşirken Model 2 için ayrımcılıktan kaynaklı fark (0.145) %93 oranında gerçekleşmiştir.

Bu meslek grubu için bakıldığı zaman beşeri sermaye özelliklerinden olan eğitim değişkeni Model 1 ve Model 2’de önemli sayılacak derece de artışa neden olmadığı görülmüştür. Analiz çalışmasında Profesyonel meslek mensuplarının kazançlarında eğitim seviyeleri bakımından referans eğitim gurubuna göre negatif yönde etkileyen eğitim seviyeleri Bir Okul Bitirmeyenler (-0.36), İlkokul (-0.14), Genel Ortaokul (-0.114), Genel Lise (-0,100) ve Meslek Lisesi (-0.136) olarak gerçekleşmiştir. Yüksek Lisans veya Doktora (101) eğitim seviyesi ise pozitif yönde etkilediği görülmüştür.

3.4. Klavye Kullanarak Çalışan Tüm Büro Elemanları

ISCO meslek gurubunda dördüncü sırada yer alan meslekler aşağıda verilmiştir, yapılan analizler sadece bu meslek gurubu içindir. Eğitim seviyesinde referans grup olarak “Yüksekokul veya Fakülte” mezunları alınmıştır (Genel büro elemanları ile klavye kullanan büro elemanları - Müşteri hizmetlerinde çalışan elemanlar - Sayısal işlemler yapan ve malzeme kayıtları tutan büro elemanları - Diğer büro hizmetlerinde çalışan elemanlar).

Tablo 7. Tanımlayıcı İstatistikler

Cinsiyet	Değişkenler	Ortalama	Standart Sapma	Katılımcı Sayısı
Erkek	Aylık Kazanç	2753.64	1013.512	2784
	Yaş	35.27	9.309	
	Kıdem	5.87	6.284	
	Bir Okul Bitirmeyenler	0.006	0.801	
	İlkokul	0.380	0.485	
	Meslek Lisesi	0.115	0.106	
	Yüksek Lisans veya Doktora	0.206	0.404	
	Kadın	Aylık Kazanç	2568.91	
Yaş		32.54	7.987	
Kıdem		4.58	5.166	
Bir Okul Bitirmeyenler		0.002	0.522	
İlkokul		0.538	0.498	
Meslek Lisesi		0.173	0.130	
Yüksek Lisans veya Doktora		0.172	0.378	

Bu meslek gurubunda toplamda 2784 (%55.95) erkek çalışan, 2192 (%44.05) kadın çalışan olduğu görülmüştür (Tablo 7). Aylık kazanç erkeklerde ortalama 2753.64 (Ln W_e 7.920) TL iken kadın çalışanlarda ortalama 2568.91 (Ln W_k 7.851) TL olmuştur. Bu meslek gurunda erkek çalışanların yaş ve kıdem ortalamaları kadın çalışanlara göre oldukça yüksek görülmüştür. Erkekler için yaş kıdem ve eğitim sonuçları ise Tablo 8’de verilmiştir.

Tablo 8. Blinder-Oaxaca Ayrıştırma Modeli

	Ücret Farkı			
	Model 1		Model 2	
	Katsayı	t	Katsayı	t
Erkek	7.872	1427.33	7.872	1427.32
Kadın	7.804	1302.54	7.804	1302.53
$\ln(W_e) - \ln(W_k)$	0.068	8.36	0.068	8.36
Beşeri Sermaye	0.028	8.21	0.004	0.92
Ayrımcılık	0.040	5.26	0.064	8.75
Bir Okul Bitirmeyenler	-	-	-0.44	-3.545
İlkokul	-	-	-0.192	-14.111
Genel Ortaokul	-	-	-0.128	-9.754
Genel Lise	-	-	-0.141	-10.367
Meslek Lisesi	-	-	-0.140	-10.398
Yüksek Lisans veya Doktora	-	-	0.77	6.158

Model 1’de ve Model 2’de cinsiyete dayalı ücret farkının büyük bir kısmının ayrımcılıktan dolayı ortaya çıktığı görülmüştür. Beşeri sermayeden kaynaklı fark Model 1 için (0.028) % 41 oranıyla gerçekleşirken Model 2 için (0.004) %6 oranında gerçekleşmiştir. Cinsiyetten kaynaklı ayrımcılık Model 1 için (0.040) %59 oranında gerçekleşirken Model 2 için bu oran (0.064) % 94 olarak gerçekleşmiştir.

Bu meslek grubu için eğitim değişkeni beşeri sermayeden kaynaklı ayrımcılığı Model 2’de azaltsa da genel olarak cinsiyetten kaynaklı ayrımcılık her iki modelde ortaya çıkmıştır. Analiz çalışmasında Profesyonel meslek mensup yardımcılarının kazançlarında eğitim seviyeleri bakımından referans eğitim gurubuna göre negatif yönde etkileyen eğitim seviyeleri Bir Okul Bitirmeyenler (-0.44), İlkokul (-0.192), Genel Ortaokul (-0.128), Genel Lise (-0.141) ve Meslek Lisesi (-0.140) olarak gerçekleşmiştir. Yüksek Lisans veya Doktora (0.77) eğitim seviyesi ise pozitif yönde etkilediği görülmüştür.

3.5. Kişisel Hizmet Bakım Veren Satış Koruma Hizmeti Çalışanları

ISCO meslek gurubunda beşinci sırada yer alan meslekler aşağıda verilmiştir, yapılan analizler sadece bu meslek gurubu içindir. Eğitim seviyesinde referans gurup olarak “Yüksekokul veya Fakülte” mezunları alınmıştır (Kişisel hizmetler veren elemanlar - Satış hizmetleri veren elemanlar - Kişisel bakım hizmetleri veren elemanlar - Koruma hizmetleri veren elemanlar).

Tablo 9. Tanımlayıcı İstatistikler

Cinsiyet	Değişkenler	Ortalama	Standart Sapma	Katılımcı Sayısı
Erkek	Aylık Kazanç	2403.94	686.649	7775
	Yaş	35.61	10.747	
	Kıdem	4.80	5.789	
	Bir Okul Bitirmeyenler	0.018	0.134	
	İlkokul	0.142	0.349	
	Meslek Lisesi	0.002	0.046	
	Yüksek Lisans veya Doktora	0.183	0.386	
Kadın	Aylık Kazanç	2240.58	589.085	3191
	Yaş	32.93	10.251	
	Kıdem	3.17	3.964	
	Bir Okul Bitirmeyenler	0.185	0.134	
	İlkokul	0.210	0.407	
	Meslek Lisesi	0.022	0.467	
	Yüksek Lisans veya Doktora	0.194	0.395	

Bu meslek gurubunda toplamda 7775 (%70.90) erkek çalışan, 3191 (%29.10) kadın çalışan olduğu görülmüştür (Tablo 9). Aylık kazanç erkeklerde ortalama 2403.94 (Ln W_e 7.784) TL iken kadın çalışanlarda ortalama 2240.58 (Ln W_k 7.714) TL olmuştur. Bu meslek gurubu için yaş ortalamaları yakınsa kıdem ortalamaları arasındaki fark oldukça yüksektir. Erkekler için yaş kıdem ve eğitim sonuçları ise Tablo 10'da verilmiştir.

Tablo 10. Blinder-Oaxaca Ayrıştırma Modeli

	Ücret Farkı			
	Model 1		Model 2	
	Katsayı	t	Katsayı	t
Erkek	7.756	3120.83	7.756	3120.83
Kadın	7.693	2394.54	7.693	2394.54
Ln (W_e) – Ln (W_k)	0.062	15.49	0.062	15.49
Beşeri Sermaye	0.009	10.44	0.003	2.17
Ayrımcılık	0.053	13.07	0.059	15.27
Bir Okul Bitirmeyenler	-	-	-0.920	-9.552
İlkokul	-	-	-0.269	-19.871
Genel Ortaokul	-	-	-0.197	-15.717
Genel Lise	-	-	-0.112	-9.085
Meslek Lisesi	-	-	-0.120	-9.552
Yüksek Lisans veya Doktora	-	-	0.88	9.573

Model 1'de ve Model 2'de cinsiyete dayalı ücret farkının büyük bir kısmının yine ayrımcılıktan dolayı ortaya çıktığı görülmüştür. Bu meslek gurubu için bakıldığı zaman her iki modelde de cinsiyete dayalı ücret ayrımcılığı 0.062 oranıyla erkeklerin lehine olacak şekilde gerçekleşmiştir. Beşeri sermayeden kaynaklı ücret farkı Model 1 için (0.009) %15 oranında gerçekleşirken Model 2 için (0.003) %5 oranında gerçekleşmiştir. Cinsiyetten kaynaklı fark ise Model 1 için (0.053) %85 oranında gerçekleşmişken Model 2 için (0.059) %95 oranında gerçekleşmiştir. Analiz çalışmasında kişisel hizmet veren elemanlar kazançlarında eğitim seviyeleri bakımından referans eğitim

gurubuna göre negatif yönde etkileyen eğitim seviyeleri Bir okul bitirmeyenler (-0.92) , İlkokul (-0.269), Genel Ortaokul (-0.197), Genel Lise (-0.112) ve Meslek Lisesi (-0.120) olarak gerçekleşmiştir. Yüksek Lisans veya Doktora (0.88) eğitim seviyesi ise pozitif yönde etkilediği görülmüştür. Bu meslek gurubunda ayrıştırma modelinde %5 anlamlılık seviyesinde model anlamlı çıkmıştır. Erkekler modelinde analiz çalışmalarında model %5 anlamlılık seviyesinde anlamsız çıkmıştır. Aynı konu üzerinde yapılan çalışmaların bazılarında veriler anlamlı çıkmamıştır (Çağlayan Akay, Saçaklı Saçıldı, Güriş, 2010).

3.6. Nitelikli Tarım, Orman ve Su Ürünleri Çalışanları

ISCO meslek gurubunda altıncı sırada yer alan meslekler aşağıda verilmiştir, yapılan analizler sadece bu meslek gurubu içindir. Eğitim seviyesinde referans gurup olarak “Yüksekokul veya Fakülte” mezunları alınmıştır (Pazara yönelik nitelikli tarım çalışanları- Pazara yönelik nitelikli ormancılık, su ürünleri ve avcılık çalışanları - Kendi geçimine yönelik çiftçiler, balıkçılar, avcılar ve toplayıcılar).

Tablo 11. Tanımlayıcı İstatistikler

Cinsiyet	Değişkenler	Ortalama	Standart Sapma	Katılımcı Sayısı
Erkek	Aylık Kazanç	2289.36	601.679	389
	Yaş	40.10	11.439	
	Kıdem	4.59	6.598	
	Bir Okul Bitirmeyenler	0.133	0.3407	
	İlkokul	0.030	0.173	
	Meslek Lisesi	0.020	0.1095	
	Yüksek Lisans veya Doktora	0.054	0.226	
Kadın	Aylık Kazanç	2067.50	136.218	48
	Yaş	38.71	10.264	
	Kıdem	4.08	4.937	
	Bir Okul Bitirmeyenler	0.187	0.394	
	İlkokul	0.062	0.244	
	Meslek Lisesi	0.087	0.269	
	Yüksek Lisans veya Doktora	0.000	0.000	

Bu meslek gurubunda toplamda 389 (%89.02) erkek çalışan, 48 (%10.98) kadın çalışan olduğu görülmüştür (Tablo 11). Aylık kazanç erkeklerde ortalama 2289.36 (Ln W_e 7.736) TL iken kadın çalışanlarda ortalama 2067.50 (Ln W_k 7.634) TL olmuştur. Bu meslek gurubunda çalışan kişi sayısı oldukça azdır. Buna rağmen yaş gurubu ve kıdem değerleri birbirine oldukça yakındır aynı şekilde aylık kazanç değerleri de yakın değerler almıştır. Erkekler için yaş kıdem ve eğitim sonuçları ise Tablo 12’de verilmiştir.

Tablo 12. Blinder-Oaxaca Ayrıştırma Modeli

	Ücret Farkı			
	Model 1		Model 2	
	Katsayı	t	Katsayı	t
Erkek	7.713	794.47	7.713	794.47
Kadın	7.632	878.79	7.632	878.79
$\ln(W_e) - \ln(W_k)$	0.081	6.26	0.081	6.26
Beşeri Sermaye	0.001	1.32	0.014	1.63
Ayrımcılık	0.080	6.22	0.067	4.97
Bir Okul Bitirmeyenler	-	-	-0.760	-0.758
İlkokul	-	-	-0.017	-0.215
Genel Ortaokul	-	-	-0.020	-0.245
Genel Lise	-	-	-0.035	-0.297
Meslek Lisesi	-	-	0.095	1.323
Yüksek Lisans veya Doktora	-	-	0.025	1.101

Model 1’de ve Model 2’de cinsiyete dayalı ücret farkının büyük bir kısmının ayrımcılıktan dolayı ortaya çıktığı görülmüştür. Bu meslek gurubu için bakıldığı zaman cinsiyete dayalı ücret ayrımcılığı 0.081 oranıyla erkeklerin lehine olacak şekilde gerçekleşmiştir. Model 1’de beşeri sermayeden kaynaklı ücret farkı %1 iken ayrımcılıktan kaynaklı %99 oranın da gerçekleştiği görülmüştür. Model 2’de ise beşeri sermayeden kaynaklı ücret farkı %17 olarak gerçekleşirken ayrımcılıktan dolayı gerçekleşen ücret farkı %83 olmuştur. Analiz çalışmasında pazara yönelik tarım çalışanları kazançlarında eğitim seviyeleri bakımından referans eğitim gurubuna göre negatif yönde etkileyen eğitim seviyeleri Bir okul bitirmeyenler (-0.076), İlkokul (-0.017), Genel Ortaokul (-0.020) ve Genel Lise (-0.035) olarak gerçekleşmiştir. Meslek Lisesi (0.095) ve Yüksek Lisans veya Doktora (0.025) oranıyla pozitif yönde etkilediği görülmüştür. Katılımcıların eğitim seviyelerinde mezun oldukları okul bakımından Genel ortaokul, Genel Lise ve Yüksek Lisans veya Doktora bu eğitim seviyelerinde çok az kişi olduğu için analizlerde değerleri çıkmamıştır. Bu meslek gurubunda ayrıştırma modelinde %5 anlamlılık seviyesinde model anlamlı çıkmıştır. Erkekler modelinde analiz çalışmalarında model %5 anlamlılık seviyesinde anlamsız çıkmıştır.

3.7. İnşaat Metal İşleme İle Diğer Sanatkâr Çalışanları

ISCO meslek gurubunda yedinci sırada yer alan meslekler aşağıda verilmiştir, yapılan analizler sadece bu meslek gurubu içindir. Eğitim seviyesinde referans gurup olarak “Yüksekokul veya Fakülte” mezunları alınmıştır (İnşaat ve ilgili işlerde çalışan sanatkârlar (elektrikçiler hariç) - Metal işleme, makine ve ilgili işlerde çalışan sanatkârlar - El sanatları ve basım ile ilgili işlerde çalışanlar - Elektrik ve elektronik işlerde çalışan sanatkârlar - Gıda işleme, ağaç işleri, giyim eşyası ve diğer sanatkârlar ve ilgili işlerde çalışanlar).

Tablo 13. Tanımlayıcı İstatistikler

Cinsiyet	Değişkenler	Ortalama	Standart Sapma	Katılımcı Sayısı
Erkek	Aylık Kazanç	2526.04	763.808	8062
	Yaş	36.30	10.081	
	Kıdem	5.35	6.451	
	Bir Okul Bitirmeyenler	0.029	0.169	
	İlkokul	0.066	0.249	
	Meslek Lisesi	0.005	0.222	
	Yüksek Lisans veya Doktora	0.189	0.391	
Kadın	Aylık Kazanç	2209.18	440.118	862
	Yaş	36.85	9.889	
	Kıdem	3.66	4.303	
	Bir Okul Bitirmeyenler	0.047	0.0212	
	İlkokul	0.116	0.320	
	Meslek Lisesi	0.002	0.340	
	Yüksek Lisans veya Doktora	0.128	0.335	

Bu meslek gurubunda toplamda 8062 (%90.33) erkek çalışan, 862 (%9.66) kadın çalışan olduğu görülmüştür (Tablo 13). Aylık kazanç erkeklerde ortalama 2526.04 (Ln W_e 7.834) TL iken kadın çalışanlarda ortalama 2209.18 (Ln W_k 7.700) TL olmuştur. Diğer meslek guruplarından farklı olarak bu meslek gurubunda ortalama yaş değerleri erkeklerden yüksek çıkmıştır. Erkekler için yaş kıdem ve eğitim sonuçları ise Tablo 14’de verilmiştir.

Tablo 14. Blinder-Oaxaca Ayrıştırma Modeli

	Ücret Farkı			
	Model 1		Model 2	
	Katsayı	t	Katsayı	t
Erkek	7.802	2975.13	7.802	2975.13
Kadın	7.686	1480.18	7.686	1480.18
Ln (W_e) – Ln (W_k)	0.115	19.91	0.115	19.91
Beşeri Sermaye	0.011	6.53	0.011	19.91
Ayrımcılık	0.104	18.11	0.104	18.35
Bir Okul Bitirmeyenler	-	-	-0.132	-11.256
İlkokul	-	-	-0.351	-17.183
Genel Ortaokul	-	-	-0.222	-11.816
Genel Lise	-	-	-0.117	-8.392
Meslek Lisesi	-	-	-0.062	-3.641
Yüksek Lisans veya Doktora	-	-	0.071	7.100

Model 1’de ve Model 2’de cinsiyete dayalı ücret farkının büyük bir kısmının ayrımcılıktan dolayı ortaya çıktığı görülmüştür. Bu meslek gurubu için bakıldığı zaman cinsiyete dayalı ücret ayrımcılığı Model 1 ve Model 2 için 0.115 oranıyla erkeklerin lehine olacak şekilde gerçekleşmiştir. Model 1 ve Model 2’de beşeri sermayeden kaynaklı ücret farkı (0.011) %10 oranında gerçekleşirken ayrımcılıktan kaynaklı fark (0.104) %90 oranında gerçekleşmiştir. Analiz çalışmasında İnşaat işlerinde çalışan sanatkarlar kazançlarında eğitim seviyeleri bakımından referans eğitim gurubuna göre negatif yönde etkileyen eğitim seviyeleri Bir Okul Bitirmeyenler

(-0.132), İlkokul (-0.351), Genel Ortaokul (-0.222), Genel Lise (-0.117) ve Meslek Lisesi (0.095) oranıyla negatif yönde etkilediği görülmüştür. Yüksek Lisans veya Doktora eğitim seviyesi ise pozitif yönde etkilemiştir. Bu meslek gurubunda ayrıştırma modelinde %5 anlamlılık seviyesinde model anlamlı çıkmıştır. Erkekler modelinde Model 1 analiz çalışmalarında %5 anlamlılık seviyesinde anlamsız çıkmıştır.

3.8. Sabit Tesis ve Diğer Teçhizat Operatörleri

ISCO meslek gurubunda sekizinci sırada yer alan meslekler aşağıda verilmiştir, yapılan analizler sadece bu meslek gurubu içindir. Eğitim seviyesinde referans gurup olarak “Yüksekokul veya Fakülte” mezunları alınmıştır (Sabit tesis ve makine operatörleri - Montajcılar - Sürücüler ve hareketli makine ve teçhizat operatörleri).

Tablo 15. Tanımlayıcı İstatistikler

Cinsiyet	Değişkenler	Ortalama	Standart Sapma	Katılımcı Sayısı
Erkek	Aylık Kazanç	2506.98	706.776	6993
	Yaş	38.62	10.126	
	Kıdem	5.11	6.000	
	Bir Okul Bitirmeyenler	0.180	0.133	
	İlkokul	0.045	0.208	
	Meslek Lisesi	0.095	0.286	
	Yüksek Lisans veya Doktora	0.158	0.365	
	Aylık Kazanç	2172.11	316.141	
Kadın	Yaş	36.65	9.311	
	Kıdem	3.82	4.096	
	Bir Okul Bitirmeyenler	0.066	0.2496	
	İlkokul	0.039	0.193	
	Meslek Lisesi	0.023	0.126	
	Yüksek Lisans veya Doktora	0.099	0.299	

Bu meslek gurubunda toplamda 6993 (%86.96) erkek çalışan, 1049 (%13.04) kadın çalışan olduğu görülmüştür (Tablo 15). Aylık kazanç erkeklerde ortalama 2596.98 ($\ln W_e$ 7.862) TL iken kadın çalışanlarda ortalama 2172.11 ($\ln W_k$ 7.683) TL olmuştur. Erkekler için yaş kıdem ve eğitim sonuçları ise Tablo 16’da verilmiştir.

Tablo 16. Blinder-Oaxaca Ayrıştırma Modeli

	Ücret Farkı			
	Model 1		Model 2	
	Katsayı	t	Katsayı	t
Erkek	7.797	2875.66	7.797	2875.66
Kadın	7.675	2035.22	7.675	2035.22
$\ln(W_e) - \ln(W_k)$	0.122	26.4	0.122	26.4
Beşeri Sermaye	0.010	7.65	0.021	10.50
Ayrımcılık	0.112	24.62	0.101	22.1
Bir Okul Bitirmeyenler	-	-	-0.010	-7.662
İlkokul	-	-	-0.224	-8.382
Genel Ortaokul	-	-	-0.135	-5.543
Genel Lise	-	-	-0.046	-2.435
Meslek Lisesi	-	-	0.012	0.571
Yüksek Lisans veya Doktora	-	-	0.015	0.678

Model 1’de ve Model 2’de cinsiyete dayalı ücret farkının büyük bir kısmının ayrımcılıktan dolayı ortaya çıktığı görülmüştür. Bu meslek gurubu için bakıldığı zaman Model 1 ve Model 2 için cinsiyete dayalı ücret ayrımcılığı 0.122 oranıyla erkeklerin lehine olacak şekilde gerçekleşmiştir. Model 1’de beşeri sermayeden kaynaklı ücret farkı (0.010) %8 iken Model 2’de (0.021) %17 oranında gerçekleşmiştir. Cinsiyetten kaynaklı ücret ayrımcılığı Model 1 için (0.112) %92 oranında gerçekleşirken Model 2 için (0.101) %83 oranında gerçekleşmiştir. Analiz çalışmasında sabit tesis ve makine operatörleri kazançlarında eğitim seviyeleri bakımından referans eğitim gurubuna göre negatif yönde etkileyen eğitim seviyeleri Bir okul bitirmeyenler (-0.010), İlkokul (-0.224), Genel Ortaokul (-0.135) ve Genel Lise (-0.046) oranıyla negatif yönde etkilediği görülmüştür. Yüksek Lisans veya Doktora (0.015) ile Meslek Lisesi (0.095) eğitim seviyeleri ise pozitif yönde etkilemiştir. Bu meslek gurubunda ayrıştırma modelinde %5 anlamlılık seviyesinde model anlamlı çıkmıştır. Model 1 ve Model 2 analiz çalışmalarında kadınlar ve erkekler %5 anlamlılık seviyesinde anlamsız çıkmıştır.

3.9. Nitelik Gerektirmeyen Her Türlü İşlerde Çalışanlar

ISCO meslek gurubunda dokuzuncu sırada yer alan meslekler aşağıda verilmiştir, yapılan analizler sadece bu meslek gurubu içindir. Eğitim seviyesinde referans gurup olarak “Yüksekokul veya Fakülte “ mezunları alınmıştır (Temizlikçiler ve yardımcılar - Tarım, ormancılık ve balıkçılık sektörlerinde nitelik gerektirmeyen işlerde çalışanlar - Madencilik, inşaat, imalat ve ulaştırma - sektörlerinde nitelik gerektirmeyen işlerde çalışanlar - Yiyecek hazırlama yardımcıları - Cadde ve sokaklarda satış ve hizmetlerinde çalışanlar - Çöpçüler, atık toplayıcılar ve diğer nitelik gerektirmeyen işlerde çalışanlar).

Tablo 17. Tanımlayıcı İstatistikler

Cinsiyet	Değişkenler	Ortalama	Standart Sapma	Katılımcı Sayısı
Erkek	Aylık Kazanç	2288.31	485.617	4607
	Yaş	36.55	11.132	
	Kıdem	3.79	5.280	
	Bir Okul Bitirmeyenler	0.063	0.244	
	İlkokul	0.052	0.222	
	Meslek Lisesi	0.004	0.020	
	Yüksek Lisans veya Doktora	0.123	0.328	
Kadın	Aylık Kazanç	2099.57	224.886	2313
	Yaş	39.58	9.568	
	Kıdem	3.20	4.004	
	Bir Okul Bitirmeyenler	0.067	0.250	
	İlkokul	0.342	0.181	
	Meslek Lisesi	0.000	0.000	
	Yüksek Lisans veya Doktora	0.080	0.271	

Bu meslek gurubunda toplamda 4607 (%66.58) erkek çalışan, 2313 (%33.42) kadın çalışan olduğu görülmüştür (Tablo 17). Aylık kazanç erkeklerde ortalama 2288.31 (Ln W_e 7.735) TL iken kadın çalışanlarda ortalama 2099.57 (Ln W_k 7.649) TL olmuştur. Bu meslek gurubunda da kadınların ortalama yaş değerleri erkeklerden yüksek çıkmıştır. Erkekler için yaş kıdem ve eğitim sonuçları ise Tablo 18’ de verilmiştir.

Tablo 18. Blinder-Oaxaca Ayrıştırma Modeli

	Ücret Farkı			
	Model 1		Model 2	
	Katsayı	t	Katsayı	t
Erkek	7.718	3001.86	7.718	3001.86
Kadın	7.644	4104.10	7.644	4104.09
Ln (W_e) – Ln (W_k)	0.073	23.11	0.073	23.11
Beşeri Sermaye	0.005	5.76	0.008	7.04
Ayrımcılık	0.068	21.79	0.065	20.84
Bir Okul Bitirmeyenler	-	-	-0.068	-3.815
İlkokul	-	-	-0.127	-4.311
Genel Ortaokul	-	-	-0.046	-1.748
Genel Lise	-	-	-0.011	-0.550
Meslek Lisesi	-	-	0.023	1.149
Yüksek Lisans veya Doktora	-	-	0.002	0.156

Model 1’de ve Model 2’de cinsiyete dayalı ücret farkının büyük bir kısmının ayrımcılıktan dolayı ortaya çıktığı görülmüştür. Bu meslek gurubu için bakıldığı zaman Model 1 ve Model 2 için cinsiyete dayalı ücret ayrımcılığı 0.073 oranıyla erkeklerin lehine olacak şekilde gerçekleşmiştir. Model 1 de beşeri sermayeden kaynaklı ücret farkı (0.005) %7 iken Model 2’de (0.008) %11 oranında gerçekleşmiştir. Cinsiyetten kaynaklı ayrımcılık ise Model 1 için (0.068) %93 oranında gerçekleşmişken Model 2 için (0.065) %89 oranında gerçekleşmiştir. Analiz çalışmasında temizlikçiler ve yardımcıları kazançlarında eğitim seviyeleri bakımından referans eğitim gurubuna

göre negatif yönde etkileyen eğitim seviyeleri Bir okul bitirmeyenler (-0.068), İlkokul (-0.127), Genel Ortaokul (-0.046) ve Genel Lise (-0.011) oranıyla negatif yönde etkilediği görülmüştür. Yüksek Lisans veya Doktora (0.002) ile Meslek Lisesi (0.023) eğitim seviyeleri ise pozitif yönde etkilemiştir. Bu meslek gurubunda ayrıştırma modelinde %5 anlamlılık seviyesinde model anlamlı çıkmıştır.

3.10. Tüm Meslek Guruplarının Analiz Sonuçlarının Bir Arada Gösterilmesi

Yapılan analiz sonuçlarında tüm meslek guruplarında cinsiyete dayalı ücret ayrımcılığı tespit edilmiştir. Aşağıdaki tabloda tespit edilen ücret ayrımcılığının beşeri sermaye ve ayrımcılık oranı her bir meslek gurubu için ayrı ayrı tablo halinde sunulmuştur.

Tablo 19. Meslek Guruplarına Göre Sonuçların Karşılaştırılması

Meslek Grubu	Cinsiyete Dayalı Ücret Eşitsizliği	Beşeri Sermaye (Model 1)	Ayrımcılık (Model 2)
Başkanlar Üst Düzey Yöneticiler	0.036	-0.047	0.084
Sağlık Hukuk Eğitim ve Diğer Profesyoneller	0.241	0.055	0.186
Sağlık Hukuk Eğitim ve Diğer Yardımcı Profesyoneller	0.156	0.011	0.145
Klavye Kullanarak Çalışan Tüm Büro Elemanları	0.068	0.004	0.064
Kişisel Bakım Satış ve Koruma Hizmeti Çalışanları	0.062	0.003	0.059
Nitelikli Tarım Orman ve Su Ürünleri Çalışanları	0.081	0.014	0.067
İnşaat Metal İşleme ile Diğer Sanatkâr Çalışanları	0.115	0.011	0.104
Sabit Tesis ve Diğer Teçhizat Operatörleri	0.122	0.021	0.101
Nitelik Gerektirmeyen Her Türlü İşlerde Çalışanlar	0.073	0.008	0.065

Analiz çalışmasıyla ilgili paralel olarak örtüşen sonuçlar ve yorumlardan birkaç tanesi Tablo 20’de verilmiştir.

Tablo 20. Diğer Çalışmalarda Bulunan Ücret Farkları ve Yorumları

Diğer Çalışmalarda Bulunan Ücret Farkları	
Yazar	Oran/Yorum
Toksöz, Memiş, 2020	Cinsiyete dayalı ücret farklarını imalat sektöründe analiz eden çalışmada Türkiye imalat sektöründe cinsiyetten kaynakla kazanç farkının olduğunu ve bu farkın sebeplerinden birinin de bu sektör için kayıt dışı istihdamın çok fazla olmasından kaynakladığını söylemiştir.
Yalçın, Arlı, Aytaç, Başol, Aydemir, 2019	%262.8
Tunç,2018	Az gelişmiş Ülkelerde cinsiyetten kaynaklı ücret artışı daha fazla iken gelişmiş ülkeler de bu durum daha azdır.
Duman, 2018	%8.8
Acun, 2018	Analiz sonuçlarına bakıldığı zaman kadınların aleyhinde bir sonuca ulaşıldığı ve erkeklere göre daha düşük kazanç aldığını ortaya çıkarmaktadır.
Çelik Uğuz, Topbaş, 2016	%0.57
Cergibozan, Özcan, 2012	İşgücü piyasalarında kadın ve erkek çalışanlar arasında incelemeler yapan çalışmalarda genel sonuç kadın çalışanlar erkek çalışanlardan daha az gelir elde ettiği yönündedir.
Kiren Gürler, Üçdoğruk, 2007	Çalışmanın sonuçlarına göre beşeri sermaye özelliklerine istinaden insan sermayesi ve meslek bilgileri gelir farklılığına katkısı diğer kazanç getirilerine göre temel insan sermayesinden daha fazladır. Her iki model sonucuna göre Türkiye de cinsiyetler arası önemli boyutlarda ayrımcılığın olduğu saptanmıştır.
Hisarcıklılar, Ercan, 2005	%120.4
Yamak, Topbaş, 2004	Oaxaca % 78 Cotton % 80
Palaz, 2002	Yapılan teorik ve ampirik çalışmalar sonucunda Türkiye’de işgücü piyasalarında kadınlara yönelik cinsiyetten dolayı ücret ayrımcılığıyla sürekli olarak karşı karşıya kalındığını söylemiştir.

4. Tartışma

Blinder-Oaxaca ayrıştırma modelinin kullanıldığı 2015 yılı Türkiye İstatistik Kurumu İşgücü anket veri seti ile üç farklı model oluşturularak yapılan, Türkiye’de Emek Piyasalarında Cinsiyete Dayalı Ücret Ayrımcılığı başlıklı çalışma da bölgesel analiz yapılmıştır. Bu çalışmadaki amaç, ücret ayrımcılığının kadın erkek ayrımının yanında bölgesel farklılıkları da tespit etmektir. Araştırmada üç farklı model kurulmuş ve ücret farkının en çok cinsiyete dayalı ayrımcılıktan kaynaklandığı tespit edilmiştir. İBBS Düzey 1’de bulunan 12 bölge üzerinde yapılan çalışmada ücret eşitsizliği en fazla Doğu’da, en az ise Ege bölgesinde olduğu görülmüştür. Doğuda diğer bölgelere göre daha yüksek çıkmasının kadınların işgücüne katılım oranının düşük olmasından kaynaklandığı düşünülmüştür. Orta Anadolu’dan batıya gidildikçe kadın erkek ücret eşitsizliğinin azaldığı görülmüştür, buna rağmen Batı bölgesinde erkekler kadınlara göre yüksek ücret aldığı tespit edilmiştir (Yalçın, Arlı, Aytaç, Başol, Aydemir, 2019).

Türkiye’de kadın erkek çalışanların iş gücü piyasalarındaki durumunu araştıran TÜİK 2010 iş gücü anketi verilerini kullanarak hem Türkiye geneli hem de bölgesel olarak yapılan analiz çalışmasında, Türkiye geneli için yapılan analizde istatistiksel olarak anlamlı bir ücret farkı bulunmadığı ortaya çıkmıştır. Oaxaca-Blinder ve Reimers ayrıştırma analiz yöntemlerini kullanarak yaptıkları çalışmada İstanbul ve İç Anadolu bölgesinde kadınlar lehine sonuçlara ulaşılmıştır. Marmara ve Karadeniz bölgelerinde ise kadınlar aleyhinde sonuçlara ulaşılmıştır. Ege ve Doğu Anadolu bölgelerinin de ise anlamlı derece ayrımcılık bulunmamıştır (Cergibozan, Özcan, 2012).

Ucuz işgücü kaynağı olarak görülen kadın emeğine olan artan talep, ihracata dayalı sanayileşme modelinin 1970’lerin sonlarında benimsenmesinden bu yana devam etmektedir (Toksöz, Memiş, 2020). Emek piyasalarında kadınların düşük ücretli istihdam edildikleri kayıtlı istihdam edilmelerine rağmen cinsiyet ücret açığı ile karşı karşıya olduklarından kaynaklı, toplumsal cinsiyet eşitsizliğinin en temel meselelerinden biri olmuştur. Türkiye sanayi üretiminde de diğer ülkeler gibi kadın erkek istihdamında kadınlar aleyhinde ücret açığına maruz kalmıştır. 2004-2016 Hanehalkı İşgücü Anketi verilerine göre yapılan çalışmada imalat sanayi alt sektörlerinde kayıt dışı istihdamda saatlik ücretlerdeki cinsiyet ücret açığı analiz edilmiştir. İhracat ve kadın istihdam oranı seçili sektörlerde (gıda, tekstil ve hazır giyim imalatı) nispeten yüksek çıkmış, kayıt dışılık ve cinsiyet ücret farkı ilişkisi, araştırmanın odak sorunu olmuştur (Toksöz, Memiş, 2020). Elde edilen sonuçlar ihracat yönelimi yüksek olan sektörlerde kadın ücret düzeyinin düşük olduğu ve cinsiyet eşitsizliğinin yüksek düzeyde olduğunu ve bu sorunun varlığını sürdürdüğü görülmüştür.

Türkiye genelinde kadın ve erkelerin iş gücüne katılımını belirleyen etmenleri Mincer’ in İnsan Sermayesi Modeli’ ne göre inceleyen ve 2002 Hanehalkı bütçe anketi verilerini kullanarak yapılan çalışmada önemli boyutlarda ayrımcılık olduğu saptanmıştır. Türkiye geneli ekonomik ve sosyal sebeplerden kaynaklı olarak İşgücüne katılım oranı çok düşük görülmüştür. Oaxaca (1973) ayrıştırma modelinin kullanıldığı çalışmada gelir farklılığı ve donanım farklılığı olmak üzere iki kısımda cinsiyetler arası ücret geliri incelenmiştir. Yapılan analiz sonuçlarında meslek bilgileri yani tecrübesi ve beşeri sermaye özellikleri temel insan sermayesi modelinden katkı olarak daha fazladır. Araştırma sonuçlarına göre incelenen modellerle yapılan analizde ayrımcılık erkekler lehine gelişmiştir. İşgücü piyasalarında kadın istihdamının artması ve ücret ayrımcılığı ile ilgili sorunların kalıcı çözümü üniversite eğitiminin yaygınlaştırılması ile giderilebileceği de ifade edilmiştir (Gürler, Üçdoğruk, 2007).

Diğer ülkelerle kıyasla Türkiye’ de kadınların iş hayatına katılma oranları oldukça düşüktür ve genel olarak kadınların kayıt dışı istihdamda daha çok çalıştıkları bilinmektedir. Ekonomik anlamda son yıllarda kayıtlı çalışmayanlar ile mücadele edilmesinin sonuçları olarak kayıt dışı istihdamda oldukça azalma vardır. Bu kadın çalışanlar açısından olumlu bir durumdur fakat tam anlamıyla bu konu çözülmüş değildir. Kadın çalışanlarda eğitim seviyesi arttıkça sigorta güvencesine sahip işlerde çalışma isteği de artmaktadır, eğitim seviyesi azaldıkça sigortasız işlerde çalışma durumu da artmaktadır. Kentleşmenin yüksek olduğu illerde kadın istihdamının arttığı ve sigortalı işlere önem verildiği gözlemlenmiştir. Son yıllarda bu konuyla alakalı çalışmalar da oldukça fazladır (Öztürk, Başar, 2018). TÜİK 2015 yılı Hanehalkı Bütçe anketi verileri kullanılarak yapılan çalışmada toplumun genel ekonomik durumu erkeklere göre kadınların çalışma hayatında ve sosyal güvenlik yani kayıt dışı çalışma tercihinde önemli bir etken olduğu ortaya çıkmıştır.

Türkiye’de cinsiyetten kaynaklı ücret ayrımcılığı konusunda yapılan güncel analiz çalışmalarına bakıldığı zaman 7 bölge üzerinde analiz çalışması yapan Eraslan (2012). Diğer bölgelerde (Türkiye Geneli – İstanbul – Batı Doğu Marmara – Ege Akdeniz – Orta Batı Anadolu – Batı Karadeniz K. Doğu Anadolu) erkekler lehine sonuçlara ulaşmış fakat Orta- Güneydoğu Anadolu bölgesinde, erkeklerin aleyhinde sonuçlara ulaşarak kadınların saatlik ücretlerinin daha yüksek olduğunu tespit etmiştir. Bunun sebebini ise işgücüne katılan kadın çalışan sayısının az olmasına bağlamıştır. Diğer bölgeler için analiz sonuçlarında kadınlarda saatlik ücretin az olmasına yani eşit işe eşit ücret verilmemesinden kaynaklandığını öne sürmüştür (Eraslan, 2012).

5. Sonuç

Gelişmiş ve gelişmekte olan ülkelerin işgücü piyasalarında çok sıklıkla karşılaştıkları sorunların başında ücret eşitsizliği gelmektedir. Yapılan çalışmalarda ücret eşitsizliğinin temelde iki sebebinin olduğu ortaya çıkmıştır. Bu sebeplerden birisi beşeri sermaye özelliklerinin farklı olması diğer sebep ise ücret eşitsizliğinin cinsiyetten kaynaklanmasıdır (Kaya, Selim, 2018). Beşeri sermayeden kaynaklı ücret farklılığı gerçek ekonomik durumla örtüşmektedir fakat ücret eşitsizliğinin cinsiyete dayanması eşitsizlik konusunun tekrar ele alınmasına neden olmuştur.

Türkiye de 1999 yılında cinsiyet ayrımcılığı konusunda ilk çalışmalardan biri olan “Türkiye de kamu-özel sektör ilişkisi ve cinsiyet” başlıklı çalışmada Kamu sektöründe eşitsizliğin olmadığını fakat özel sektörde kadınların gelir eşitsizliğine maruz kaldıkları tespit edilmiştir (Tansel, 1999). Türkiye de cinsiyetten kaynaklı ücret ayrımcılığı konusunda yapılan çalışmaların birçoğunda istatistiki bölge birimleri sınıflandırmasına göre analiz çalışmaları yapılmıştır. Oaxaca-Blinder ayrıştırma yöntemini kullanarak Türkiye’de 13 bölge ile yapılan analiz sonuçları da

tüm bölgeler de cinsiyetten kaynaklı ücret ayrımcılığı tespit edilmiştir (Kaya, Selim, 2018). Farklı bir çalışma da ise 7 bölge arasında ücret ayrımcılığı analizi yapılmış ve Karadeniz bölgesi en yüksek ayrımcılığın olduğu bölge çıkmıştır (Cergibozan, Özcan, 2012). İnsan sermayesi modelini kullanarak analiz çalışması yapan Gürler ve Üçdoğruk meslek guruplarının gelir üzerindeki etkisini ölçmüştür. Kanun yapıcı, üst düzey yönetici, profesyonel meslek mensuplarının kadın ve erkekler de diğer meslek guruplarına göre gelir etkisinin (ortalama gelirin) daha fazla olduğu görülmüştür. Oaxaca ayrıştırma modeli ile Cotton ayrıştırma modelini ayrı ayrı analiz çalışmasında kullanarak Türkiye de cinsiyetten kaynaklı ücret ayrımcılığını araştıran Yamak ve Topbaş, Oaxaca modeline göre %78, Cotton modeline göre %80 oranında ayrımcılık tespit etmişlerdir.

Bu çalışmanın amacı diğer çalışmalardan farklı olarak Türkiye’de Uluslararası Standart Meslek Kodlarında bulunan meslek guruplarının cinsiyetten kaynaklı ücret ayrımcılığının var olup olmadığının tespit edilmesidir. Bu kapsamda ilk önce SPSS paket programında çoklu doğrusal regresyon analizi yapılarak geliri etkileyen değişkenler ve hangi oranda etkilediği tespit edilmiş, daha sonra Stata paket programında Blinder- Oaxaca ayrıştırma modeli kullanılarak ücret ayrımcılığının olup olmadığı saptanmıştır. Yapılan çoklu doğrusal regresyon analizin de bağımlı değişken olan ücretin kıdem, yaş ve eğitim ile ilişkili olduğu ortaya çıkmıştır. İlk başlarda yaş ve kıdem değişkeni ücret üzerinde olumlu etki yaparken belirli bir zaman sonra (25 yıl) bu etki negatife döndüğü gözlemlenmiştir (Kaya, Selim, 2018). Eğitim değişkeni ücret üzerinde meslek guruplarına göre değişkenlik göstermiş, profesyonel meslek mensuplarında eğitimin ücret üzerindeki etkisine diğer meslek guruplarına göre daha yüksek oranda olumlu yönde etkilediği ortaya çıkmıştır.

Yapılan analiz sonuçlarına göre cinsiyetten kaynaklı ücret ayrımcılığı bütün sektörler de tespit edilmiştir. Daha önceki yapılan sonuçlarla paralel çıkan analiz sonuçlarında en çok ücret ayrımcılığı %0.241 oranıyla Sağlık, hukuk ve eğitim profesyonel meslek mensuplarında görülmüştür. Diğer meslek gurubu analizlerine bakıldığı zaman Sağlık, hukuk ve eğitim yardımcı profesyonel meslek mensupları %0.156 oranıyla ikinci sırada yer almaktadır. Sabit tesis ve diğer teçhizat operatörleri olarak adlandırılan meslek grubu %0.122 oranıyla üçüncü sırada yer alırken onu %0.115 oranıyla İnşaat metal işleme ile diğer sanatkâr çalışanları adlandırılan meslek gurubu takip etmiştir. Diğer Analiz gruplarında %0.081 oranıyla Nitelikli tarım, orman ve su ürünleri meslek gurubu altıncı sırada yer alırken onu %0.068 oranıyla Klavye kullanarak çalışan tüm büro elemanları olarak adlandırılan meslek gurubu takip etmektedir. Düşük meslek gurupları ise 0.073 oranıyla Nitelikli gerektirmeyen her türlü işlerde çalışanlar diye adlandırılan meslek gurubu iken en düşük oran ise %0.036 oranıyla Başkanlar üst düzey yöneticiler ve diğer idari müdürler olarak adlandırılan meslek gurubu olmuştur.

Oluşan bu ücret ayrımcılığında beşeri sermayeden kaynaklı fark en çoktan en aza sırasıyla; Sağlık, hukuk eğitim ve diğer profesyonel meslek mensupları, sabit tesis ve diğer teçhizat operatörleri, nitelikli tarım orman ve su ürünleri meslek mensupları, inşaat metal işleme ile diğer sanatkârlar, sağlık hukuk eğitim ve diğer yardımcı profesyonel meslek mensupları, nitelik gerektirmeyen her türlü işlerde çalışanlar, klavye kullanarak çalışan tüm büro elemanları, kişisel bakım satış ve koruma hizmeti çalışanları, başkanlar üst düzey yöneticiler ve diğer idari müdürler olarak adlandırılan meslek guruplarında gerçekleşmiştir (Sağlık hukuk eğitim yardımcı profesyonel meslek mensupları ile İnşaat metal işleme ve diğer sanatkârlar aynı orana sahiptir). Cinsiyetten kaynaklı ayrımcılık ise en çoktan en aza sırasıyla; Sağlık, hukuk, eğitim ve diğer profesyonel meslek mensupları, sağlık, hukuk, eğitim diğer yardımcı profesyonel meslek mensupları, inşaat metal işleme ile diğer sanatkâr meslek mensupları, sabit tesis ve diğer teçhizat operatörleri, başkanlar üst düzey yöneticiler ve diğer idari müdürler, nitelikli tarım orman ve su ürünleri meslek mensupları, nitelik gerektirmeyen her türlü işlerde çalışanlar, klavye kullanarak çalışan tüm büro elemanları, kişisel bakım satış ve koruma hizmeti çalışanları olarak adlandırılan meslek guruplarında gerçekleşmiştir.

2019 yılında kadınların işgücüne katılım oranı %34.4 seviyesinde gerçekleşmiştir (TÜİK, 2019). Türkiye’de son 26 yılda kadınların işgücüne katılım oranı diğer OECD ülkelerinin artış oranının altında kalmıştır (Yalçın, Arlı, Aytaç, Başol, Aydemir, 2019). İşgücü piyasalarında kadın istihdamı ve kadınların işgücüne katılım oranlarının artması yaşanan ayrımcılıkların önüne geçebileceği düşünülmektedir. Yapılan analizlerde görülüyor ki kadınların işgücüne katılımının az olduğu bölgeler de ayrımcılık daha fazladır. Ayrıca eğitim seviyesinin artması ve üniversite eğitiminin yaygınlaştırılması cinsiyete dayalı ücret ayrımcılığının da azalmasına neden olmaktadır. Günümüzde kanun koyucular kadın çalışanların korunması için gerekli yasalar düzenlemeleri hayata geçirmişlerdir ve sorunun kalıcı olarak çözülmesine çaba sarf etmişlerdir. Fakat işgücü piyasalarında yapılan güncel analiz çalışmalarında görülüyor ki özel sektörde tam anlamıyla ayrımcılık önlenmiş değildir.

Blinder - Oaxaca ayrıştırma modeline göre yapılan analiz çalışmasında her meslek gurubu için ayrı ayrı analiz yapılarak meslek gurupları açısından daha detaylı inceleme ve karşılaştırma fırsatı sunulduğu düşünülerek yapılan çalışmanın literatüre katkı sunacağı beklenmektedir.

Kaynakça

- Acun, S. (2018). Türkiye de Ücret Eşitsizliğinin Analizi. *Bulletion of Economic Theory and Analysis*, 3(4), 237-260.
- Alparslan, A., Bozkurt, Ö., Özgöz, A. (2015). İşletmelerde Cinsiyet Ayrımcılığı ve Kadın Çalışanların Sorunları, Ş.Seyahat Acentalarına Yönelik Bir Araştırma. *Seyahat ve Otel İşletmecilliği Dergisi*, 14(1), 90-107.
- Blinder, Alan S., (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources*, Vol. 8, No. 4, 1973, pp. 436-455.
- Böheim, R., Hofer H., Zulehner C. (2007). Wage Differences Between Austrian Men and Women: Semper Diem?. *Emprica*, vol.34, pp.213-299.
- Cergibozan, R., Özcan, Y. (2012). Türkiye için Bölgelere Göre Ücret Ayrıştırma Analizi: Ekonometrik Yaklaşım. *Kırklareli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 1(1). 27-48.
- Cholezas, I., Tsakoglou, P. (2006). Gender earnings differentials in the Greek labour market. *Economic Policy Studies*, forthcoming.
- Çelik Uğuz, S., Topbaş, F. (2016). Turizmde Kadın İstihdamı ve Ücret Ayrımcılığı: Karşılaştırmalı Bir Analiz. *Anatolia: Turizm Araştırmaları Dergisi*, 27(1). 62-78.
- Duman, A. (2018). Education Mismatches in the Labor Markets and Their Impact on Wages across Sectors: Evidence from Turkey. *Sosyoekonomi*, 26(37). 227-242.
- Eraslan, V. (2012). Türkiye İşgücü Piyasasında Ücret Seviyesinde Cinsiyet Ayrımcılığı: Blinder-Oaxaca Ayrıştırma Yöntemi. *İktisat Fakültesi Mecmuası*, 62(1), 231-248.
- Etiler, N. (2017). Kesişmeyen Kümeler: Kadın Emegi Ve Ücret. *Ttb Mesleki Sağlık Ve Güvenlik Dergisi*, 17(66), 19-27.
- Güriş, S., Çağlayan Akay, E., Saçaklı Saçıldı, İ. (2010). Schooling and wage distribution for male and female employees in the public and private sector. *International Research Journal of Finance and Economics*, 40, 146-154.
- Hisarcıklılar, M., Ercan, H. (2005). Gender based wage differentials in Turkey. *Bilgi Sosyal Bilimler Dergisi*, (1), 45-62.
- Kaya, G., Selim, R. (2018). The gender wage inequality in Turkey. *Press Academia Procedia (PAP)*, 7, 408-413.
- Kaymaz, Y., Şentürk, B. (2016). Lojistik Sektöründe Kadın Çalışanlara Yönelik Bir Analiz: İzmir Örneği. *Uluslararası İktisadi Ve İdari İncelemeler Dergisi*, 469-480.
- Kiren Gürler, Ö., Üçdoğruk, Ş. (2007). Türkiye’de Cinsiyete Göre Farklılığının Ayrıştırma Yöntemiyle Uygulanması. *Yaşar Üniversitesi*, 2(6), 571-589.
- Neuman S., Oaxaca R., (2003). Gender versus Ethnic Wage Differentials Among Professionals: Evidence from Israel. *Annals of Economics and Statistics*, pp.267-292.
- Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*. 14(3). 693-709.
- Öztürk, S., Başar, D. (2018). Türkiye’de Kadınların İşgücü Piyasasına Yönelik Tercihleri: Kayıt Dışı Sektör Özelinde Bir Analiz. *Sosyal Güvenlik Dergisi*, 8(2), 41-58
- Palaz, S. (2002). Türkiye’de Cinsiyet Ayrımcılığı Analizinde NeoKlasik Yaklaşımına Karşı Kurumcu Yaklaşım: Eşitliği Sağlayıcı Politika Önerileri. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 90-106.
- Tansel, A. (1999). Public-Private Employment Choice, Wage Differentials and Gender In Turkey. *Bonn: Iza Discussion Paper*, No: 1262.
- Toksöz, G., Memiş, E. (2020). Kayıt Dışı İstihdamda Toplumsal Cinsiyet Eşitsizlikleri: İmalat Sanayinde Cinsiyet Ücret Açığı. *Ekonomik Yaklaşım Derneği*, 31(116), 1-33.
- Topbaş, F. (2003). Emek Piyasalarında Cinsiyete Dayalı Ücret Ayrımcılığı ve Türkiye Örneği. *Karadeniz Teknik Üniversitesi. Sosyal Bilimler Enstitüsü. Yüksek Lisans Tezi*.
- TÜİK, ILO (2020). Cinsiyete Dayalı Ücret Farkının Ölçümü Türkiye Uygulaması. *Ankara: ILO*.
- Türkiye İstatistik Kurumu (2019). (<https://www.tuik.gov.tr/>)

Atf / Citation: ERDOĞAN E., YEŞİLYURT C. (2022). Cinsiyete Dayalı Ücret Ayrımcılığının ISCO Meslek Gruplarına Göre Analizi. *İstatistik Araştırma Dergisi*, 12 (1), 16-38.

Tüzünkan, D. (2016). Kadın İşgücüne Yönelik ILO Sözleşmeleri ve Uygulanırlığı: Mevzuatlar Açısından Türkiye-İtalya Karşılaştırması. *Kastamonu Üniversitesi İİBF Dergisi*, 12, 313-324.

Tunç, M. (2018). Kalkınmada Kadın Ayrımcılığı ve Toplumsal Cinsiyet Eşitsizliğinin Rolü: Ülkelerarası Farklı Gelir Gruplarına Göre Yatay Kesit Analiz. *Sosyoekonomi*, 26(38), 221-251.

Wright Robert, E., Ermisch John, F., (1991). Gender Discrimination in the British Labour Market: A Reassessment. *Economic Journal, Royal Economic Society*, vol. 101(406), pages 508-522, May.

Yalçın, E. C., Arlı, N. B., Aytaç, S., Başol, O., Aydemir, M. (2019). Türkiye'de Emek Piyasalarında Cinsiyete Dayalı Ücret Ayrımcılığı: İBBS Düzey 1'e Göre Farklılıklar. *Kocaeli Üniversitesi Sosyal Bilimler Dergisi*, 2(38), 293-314.

Yamak, N., Topbaş, F. (2004). Kadın emeği ve cinsiyete dayalı ücret ayrımcılığı. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 18(3-4), 143-156.



The Impact of Income Level on Household Greenhouse Gas Emissions: A Case Study for Turkey

Ersin ERCAN

Turkish Standards Institution / Dr.

ersinercan@yahoo.com

Orcid No: 0000-0003-2956-4398

Mehmet Mustafa YATARKALKMAZ

Turkish Aerospace / Dr.

mustafa.yatarkalkmaz@tai.com.tr

Orcid No: 0000-0002-0321-144X

Onur Fatih BULUT

Turkish Standards Institution

onurfatihb@tse.org.tr

Orcid No: 0000-0003-3559-2926

Abstract

This study examines the effects of income levels on household greenhouse gas (GHG) emissions. For this aim, the impact of income level on household carbon footprints was analysed. Based on a largescale household survey in Ankara, direct and some indirect GHG emissions at the household level were estimated. Furthermore, the results were examined with the Environmental Kuznets Curve (EKC) Hypothesis which was studied in Turkey. Results showed that there is a significant relationship between household GHG emissions and income level groups. Depending on the level of income, consumption growth, and per capita, energy usage results in an increase in greenhouse gas emissions. Household average emission is found as 6.934 tons of CO_{2e} per capita. Household carbon footprint varied across Ankara's district according to income level. Wealthy districts usually have a higher per capita carbon footprint than poor districts. Besides the wealthiest socio-economic level was found to emit 1.87 times as much CO_{2e} as the lowest. It is identified that emissions from heat production take the larger share of household carbon footprints. It is also determined that the only emission that decreases with higher income levels is indirect emissions from water consumption.

Keywords: ANOVA Analysis, Greenhouse Gas Emissions, Household Income

Corresponding Author / Sorumlu Yazar: 1-Ersin ERCAN, Turkish Standards Institution

2-Mehmet Mustafa YATARKALKMAZ, Turkish Aerospace

3-Onur Fatih BULUT, Turkish Standards Institution

Citation / Atıf: ERCAN E., YATARKALKMAZ M. M., BULUT O. F. (2022). The Impact of Income Level on Household Greenhouse Gas Emissions: A Case Study for Turkey. *İstatistik Araştırma Dergisi*, 12 (1), 39-55.

Gelir Düzeyinin Hanehalkı Sera Gazı Emisyonlarına Etkisi: Türkiye için Bir Vaka Çalışması

Özet

Bu çalışma, gelir düzeylerinin hane halkı sera gazı emisyonları (SGE) üzerindeki etkilerini incelemektedir. Bu amaçla gelir düzeyinin hane halkı karbon ayak izleri üzerindeki etkisi analiz edilmiştir. Ankara'da gerçekleştirilen geniş ölçekli bir hane halkı araştırmasına dayalı olarak, hane düzeyinde doğrudan ve bazı dolaylı sera gazı emisyonları tahmin edilmiştir. Diğer taraftan sonuçlar Türkiye için çalışılan Çevresel Kuznets Eğrisi (EKC) Hipotezi ile incelenmiştir. Buna göre hane halkı sera gazı emisyonları ile gelir düzeyi arasında önemli bir ilişki olduğu tespit edilmiştir. Gelir düzeyine, tüketim artışına ve kişi başına düşen enerji kullanımına bağlı olarak, enerji kullanımı sera gazı emisyonlarında artışa neden olmaktadır. Hane halkı ortalama emisyonu 6.934 ton CO_{2e} olarak tespit edilmiştir. Hane halkı karbon ayak izi, Ankara ili genelinde gelir düzeyine göre farklılık göstermiştir. Genellikle yüksek gelire sahip ilçeler, düşük gelire sahip ilçelere göre daha yüksek kişi başına karbon ayak izine sahiptir. Ayrıca en yüksek sosyo-ekonomik düzeyin, en düşük sosyo-ekonomik düzeye göre 1.87 kat daha fazla CO_{2e} saldığı tespit edilmiştir. Isı üretiminden kaynaklanan emisyonların hane halkı karbon ayak izi içerisinde en büyük paya sahip olduğu belirlenmiştir. Diğer taraftan gelir düzeyi yükseldikçe azalan tek emisyonun su tüketiminden kaynaklanan dolaylı emisyonlar olduğu da tespit edilmiştir.

Anahtar kelimeler: ANOVA Analizi, Hane Halkı Geliri, Sera Gazı Emisyonları

1. Introduction

There are several articles about Greenhouse Gas (GHG) emissions published for Turkey, from cement manufacturing to fossil fuel consumption. However, there has not been a comprehensive study in the literature that calculates actual household emissions and examines the factors affecting them. Global warming caused by GHG emissions gets more and more important day by day. Environmental concerns have started to put significant pressure on people in recent years (Plassmann, Norton, Attarzadeh, Jensen, Brenton & Edwards-Jones, 2010). The United States Environmental Protection Agency (EPA) defines climate change as a significant change in the measures of climate, such as temperature, rainfall or wind, lasting for an extended period of time – decades or longer. The most important reason for global warming is the increase in the amount of carbon dioxide in the atmosphere. S. Arrhenius is the first to predict climate change as a result of CO₂ accumulation (Türkeş, 2001). As reported by the European Environment Agency (EEA), CO_{2e} emissions from the residential sector accounted for about 27% of the global total CO_{2e} emissions in 2016 (European Environment Agency, 2015). According to the Turkish Statistical Institute (TUIK), Turkey's total emissions in 2018 were 520.9 million tons. 71.6% of the total emissions arise from energy production, followed by industrial processes and product use with 12.5%, agricultural activities with 12.5% and waste, including landfills and wastewater, with 3.4% (Turkish Statistical Institute, 2020a). It is stated that among the largest emitters of greenhouse gases in Europe are Germany, United Kingdom, France, Turkey, Poland, Italy, and Spain (Kijewska & Bluszcz, 2016). Wang et al. analysed the carbon emissions of 170 countries. They reported that energy consumption and urbanization contributed most to carbon emissions (Wang, Shi, 2019). It is determined that non-renewable energy sources had twice the effect on carbon emissions in the European Union region than renewable energy sources (Mert & Bölük, 2016). Cong et al. have demonstrated household CO_{2e} emissions at a community level for Japan. They have prepared map-based emissions by combining the statistical data on households with detailed emission intensities (Cong, Saito, Hirata, Ito, 2019). Barrett et al. have revealed consumption-based emissions results for the United Kingdom and suggested consumption-based emissions as a complementary indicator to the current approach of measuring territorial emissions (Barrett, Peters, Wiedmann, Scott, Lenzen, Roelich & Le Quéré, 2013). It has been explored the interaction effects of income, income inequality, and democracy on Canadian households' emissions using quantile technique with the EKC

hypothesis (AbdelHady, 2019). Adriana and Pratama have calculated the GHG emissions associated with household water and proposed ways to reduce emissions (Adriana & Pratama, 2020). Li et. al. have designated household CO_{2e} emissions in Northwest China. They have analysed the relationship between per capita emissions and explanatory factors using spatial econometric models (Li, Huang, Yang, Chuai, Li, Qu d, Zhang, 2016). Mi et al. have studied an environmentally extended multiregional input–output approach to estimate household carbon footprints for 12 different income groups in China’s 30 regions. They found that China’s households contributed 34% of the national carbon footprint in 2012 (Mi, Zheng & Meng, 2020). Soneja et al. have conducted research to develop field-based estimates of the amount of black carbon and particulate matter that is emitted from households due to the use of traditional cook stoves in Nepal (Soneja, Breysse & Tielsch, 2013). Vardopoulos and Konstantinou have studied to explore the possible relationship between unemployment rates and their effect on climate change through CO_{2e} emissions from human economic activity (Vardopoulos & Konstantinou, 2017). In the case of China, empirical research was done with the time series data between 1995 and 2010. Consequently, it is found that when increasing 1% of the gross domestic product (GDP) growth index and technology, household carbon emission intensity is lowered by 0.3604% and 0.0412% (Liu, Zeng Qu, Wang Q.H. & Wang L, 2013). It is indicated that gross domestic product, population, and normalized difference vegetation index are all positive driving forces for urban CO_{2e} emissions in China (Wang & Fang, 2018). Intertemporal lifestyle effects on carbon emissions have been studied in China. It was indicated that income and demographic effects contributed only 25.1% of the total household emissions. The remaining 74.9%, is defined as the effect of intertemporal lifestyle changes (Zhang, Shi, Wang, Xue, Song & Sun, 2020). The greenhouse gas emissions of 10 cities were examined by Kennedy et al. Fuel combustion, waste and electricity consumption were included in the calculation (Kennedy, Steinberger, Gasson, Hillman, Havránek, Hansen, Pataki, Phdungsilp, Ramaswami & Mendez, 2010). A similar study was conducted by Long and Fang in Tokyo. According to the results, the first and second emissions sources were found to be electricity and gas consumption, accounting for 22% and 23% of total household emissions, respectively. The four most significant impact factors on household emissions were found to be household residents, household income, house size, and preferred temperature (Zhu, Huang, Chen, 2019). In a study investigating the regional carbon footprint of Germany, it was found that the highest-income households emit 4.25 times as much as the lowest household group (Miehe, Scheumann, Jones, Kammen & Finkbeiner, 2016). In this study, there are theoretical background, methodology, and empirical results. The final section summarizes the conclusions.

2. Methodology

2.1 Study Area and Data Sources

We investigated the relationship between household GHG emissions and socio-cultural level in our study. We used consumption data to calculate household emissions. The survey method was chosen to obtain data, which we put into the calculation of emission. First, we determined the common emission sources for households. According to studies, consumption of fuel, electricity, heat, water, and waste disposal have an important place in the distribution of domestic CO_{2e} emissions (Long, Yoshida, Fang, Zhang & Dhondt, 2019; Hargreaves, Preston, White & Thumim, 2013). In this context, we categorized these sources in accordance with the EN ISO 14064-1 standard. Thus, direct emissions, indirect emissions, and other indirect emissions are included in the content of the study. At this point, as the products consumed or purchased by the households for each part would be varied and would be difficult to identify through a survey; we excluded upstream and downstream emissions from furnishing, equipment/device use, housing, food, clothing, and services such as entertainment, education, and health. Furthermore, the literature on carbon footprint was not adequate to calculate Turkey’s total household emissions. We formed questions which contain socio-cultural parameters and then carried out the survey to cover the entire city of Ankara (capital city of Turkey), including urban and rural regions. The survey consists of two parts. The first part includes the questions aimed at obtaining general information and determining the socio-cultural level of the household. In this part, the number of households, their age, and education levels were queried. Moreover, individuals’ jobs with their income were also questioned. In the second part, there are questions about the sources of emissions and the factors affecting them, which we considered significant. For example, heating, an important household emission source, could be varying according to the house properties and conditions. Therefore, in the

second part, questions about the size, facades, type, and age of houses were included. Furthermore, the types of energy sources used for heating, hot water supply, and cooking were queried. After all, we aimed to reach consumption data in order to calculate the total household emissions. However, as expected, people did not know the quantity of fuel, electricity, water, etc. consumed in a month. At this point, we obtained the average monthly bills of the consumers. Then we determined the household consumption data by dividing the bill total to the current unit price. On the other hand, fuel consumption was identified by questions about the number of cars belonging to the household, the fuel types of the cars, and the annual distances made with these cars. Finally, it was questioned whether the households separated domestic waste and how often they threw it out. Households in Ankara are allocated into five income groups; rich, upper-middle, middle, lower-middle, poor. For this classification, current income level research was made for Ankara (Endeksa, 2020). Survey research was carried out in 25 districts of Ankara.

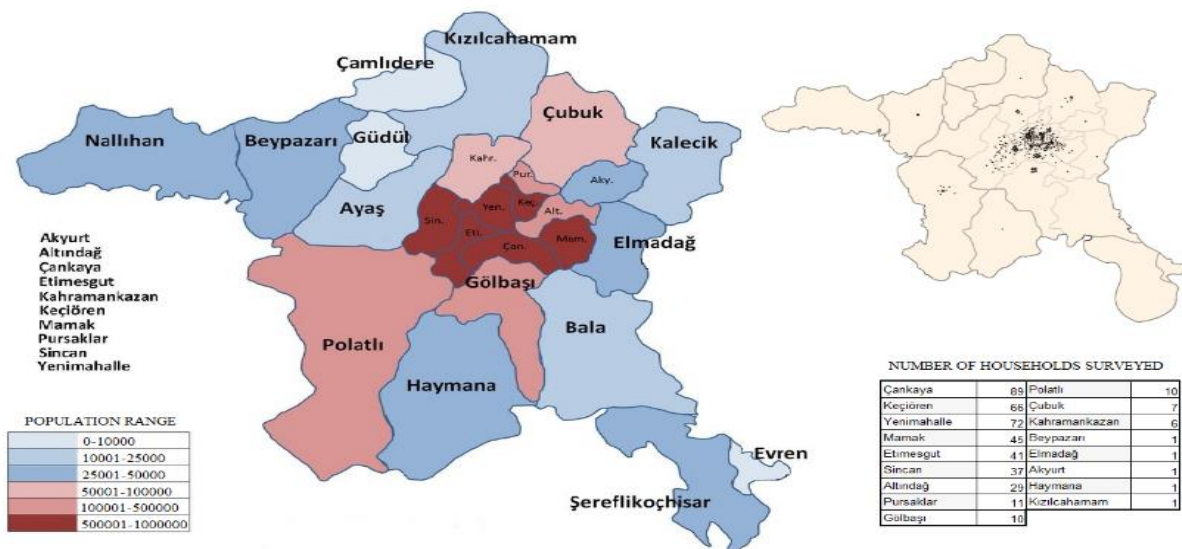


Figure 1. The distribution maps of population density and survey conducted

The total number of surveys according to the district populations of Ankara was distributed according to the stratified sampling method. The numerical representation of the surveys is shown in Figure 1 at the bottom right. The distribution of the surveys by addresses is shown in Figure 1 at the top right. As can be seen, the population density map of Ankara and the survey distribution map are compatible. The surveys conducted in each district were carried out with the simple random sampling method after stratifying according to the population. Then a survey is sent to the relevant households according to the population of the district where the research is conducted. The households were surveyed to ask questions about their socio-cultural level and resources for calculating GHG emissions, respectively. According to the results of the surveys, GHG emissions are calculated and the relationship between the emissions and the socio-cultural level was examined conforming to the statistical methods. Besides the survey research, we searched for the Gini Coefficient for Turkey. As it is known, the Gini coefficient, a widespread measure of income inequality, is a useful indicator for climate change. According to the data (Mi, Zheng & Meng, 2020) the disparity in GHG emissions could be diminished synchronously by increasing the incomes of the poor and changing the lifestyles of the wealthy to decrease the carbon intensity of their consumption patterns. The coefficient takes values between 0 and 1, where higher values correspond to greater inequality. In 2018, the Gini coefficient for Turkey was 0.408 (Turkish Statistical Institute, 2020b). The Environmental Kuznets Curve (EKC) hypothesis tries to explain the relationship between environmental degradation and economic growth, which has been the subject of long-term controversy. The EKC hypothesis states that there is an Inverse-U shaped relationship between environmental degradation and per capita income. A relationship between GHG emissions and per capita income is valid for Turkey (Uzar & Eyuboglu, 2019).

2.2. Calculation of GHG Emission

2.2.1 Heating and Transportation

Greenhouse gas emissions from transportation and heating primarily come from the burning of fossil fuels. According to the EN ISO 14064-1 standard, these emission sources are considered as direct emissions, owned or controlled by a household. In general, emissions of each greenhouse gas from the sources are calculated by multiplying fuel consumption by the corresponding emission factor. In the study, the IPCC 2006 Tier 1 method was applied to calculate direct emissions (Mi, Zheng & Meng, 2020). In the Tier 1 method, emissions of CO₂, CH₄ and N₂O are calculated by multiplying the amount of fuel consumed and the emission factor (IPCC, 2006). The following equation is used:

$$\text{Emissions}_{\text{GHG, fuel}} = \text{Fuel Consumption} \times \text{Emission Factor}_{\text{GHG, fuel}} \quad (1)$$

where;

Emissions_{GHG fuel} is emissions of a given GHG by type of fuel (kg GHG),

Fuel Consumption is amount of fuel combusted (TJ),

Emission Factor_{GHG fuel} is default emission factor of a given GHG by type of fuel (kg gas/TJ).

For CO₂, it includes the carbon oxidation factor, assumed to be 1.

Moreover, CO₂, CH₄, N₂O and other gases are converted to tons CO_{2e} by using the last updated global warming potential values. To calculate the total emissions by gas from the source, the emissions as calculated in Equation 2 are summed over all fuels: (IPCC, 2006)

$$\text{Emissions}_{\text{GHG}} = \sum_{\text{fuels}} \text{Emissions}_{\text{GHG, fuel}}$$

2.2.2 Electricity

Emissions from electricity consumption occur where electricity is generated (EN ISO 14064-1, 2019). The amount of CO₂ emissions resulting from fossil fuels in electric power generation varies depending on the type of fossil fuel. A significant portion of the greenhouse gas emissions in Turkey stem from electricity generation (Özcan & Öztürk, 2015). We preferred Tier 2 for electricity emissions. In the Tier 2 approach, default emission factors (Tier 1) are replaced by country-specific emission factors. There are some grid electricity emission factors for countries documented in some international publications. When analysing data for Turkey, we found big differences, up to 50%, between values. This would cause inconsistency. Therefore, we used emission factors from our previous study for this paper (Yatarkalkmaz & Özdemir, 2019). Electricity emission factors were developed using country-specific data such as the carbon content of the fossil fuels used in national electricity production, the quality of the fossil fuels, and the state of technological development. Equation 3 is used to calculate electric emission factor.

$$\text{Emissions}_{\text{Electricity}} = \text{Electricity consumption (kwh)} \times \text{Emission Factor}_{\left(\frac{\text{kgCO}_2}{\text{kwh}}\right)} \quad (2)$$

where;

Emissions_{Electricity} is emissions of a given GHG (kg GHG),

Electricity consumption is amount of electricity (kwh),

Emission Factor_{Electricity} is country-specific emission factor of a given GHG (kg CO₂ /kwh).

2.2.3 Domestic Waste and Water Consumption

An effective greenhouse gas inventory requires detailed understanding of a household's greenhouse gas emissions. Therefore, we have focused on other indirect emissions as well as direct emissions and energy-indirect emissions. At this point, we delved into waste volume and water consumption because they have a significant effect on

household GHG emissions. Different types of waste generate various quantities of greenhouse gases. Depending on the type of waste, CO₂ is formed from the degradation of fossil fuels while CH₄ is formed from the decomposition of biogenic materials in landfill (GHG Protocol, 2001). We used the average-data method to calculate emissions from domestic waste. This method involves estimating emissions based on total waste according to the landfill method and using average emission factors (GHG Protocol, 2021). Equation 3 is used to calculate waste generated in housing.

$$\text{Emissions}_{\text{Waste}} = \text{Waste generated (kg)} \times \text{Emission Factor} \left(\frac{\text{kgCO}_2\text{e}}{\text{tonne}} \right) \quad (3)$$

where;

Emission_{Waste} = Emissions of a given GHG by type of waste (kg GHG),

Waste_{Generated} = Amount of waste (kg)

Emission factor = Default emission factor of a given GHG by waste (kgCO_{2-Eq}/tonne).

Equation 4 is used to calculate emissions from water consumption:

$$\text{Emissions}_{\text{Waste}} = \text{Water consumed (kg)} \times \text{Emission Factor} \left(\frac{\text{kgCO}_2\text{e}}{\text{tonne}} \right) \quad (4)$$

2.3. Statistical Analysis

To apply ANOVA analysis, data must be normally distributed in each of the groups. This assumption is checked before statistical analysis. The following hypotheses are used for this purpose.

Normality assumption hypothesis group 1:

H₀: In the poor income group, GHG emissions are normally distributed.

H₁: In the poor income group, GHG emissions are not normally distributed.

Normality assumption hypothesis group 2:

H₀: In the lower-middle income group, GHG emissions are normally distributed.

H₁: In the lower-middle income group, GHG emissions are not normally distributed.

Normality assumption hypothesis group 3:

H₀: In the middle income group, GHG emissions are normally distributed.

H₁: In the middle income group, GHG emissions are not normally distributed.

Normality assumption hypothesis group 4:

H₀: In the upper-middle income group, GHG emissions are normally distributed.

H₁: In the upper-middle income group, GHG emissions are not normally distributed.

Normality assumption hypothesis group 5:

H₀: In the rich income group, GHG emissions are normally distributed.

H₁: In the rich income group, GHG emissions are not normally distributed.

After obtaining all p values, since the smallest p value is greater than 0.05, the null hypotheses for normality is accepted for each income group. After confirming that the depended variable representing greenhouse gas emissions showed a normal distribution, whether the mean consumption differs between the average emissions between 5 income groups was investigated. For this purpose, one-way analysis of variance (ANOVA) was used to reveal any significant differences between income groups. During these analysis, the SPSS 22.0 program was used. The following hypotheses were used for the ANOVA analysis:

H₀: There is no difference in average emissions between income groups.

H₁: At least two of the average emissions between income groups differ.

These hypotheses can also be expressed mathematically as:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_1: \text{At least one } \mu_k \neq \mu_l \quad (k \neq l \text{ and } k, l=1,2,3,4,5)$$

The ANOVA model used is given below.

$$y_{ij} = \mu + \mu_i + \epsilon_{ij} \quad i=1,2,3,4,5 \text{ (income levels)} \quad (5)$$

This notation represents the independent one-way ANOVA analysis on 5 income level group. While applying separate one-way ANOVA analysis, we have assumed that there is no interaction between explanatory variables and we have just focused on one categorical variable to explain the dependent variable, ignoring all other explanatory categorical variables at each step. In general, the ANOVA table for the one-way case is indicated in Table 1:

Table 1. ANOVA table for the one-way case

Source	Sum of Squares	df	MS	F
Factor	$SS_F = J \sum (\bar{y}_i - \bar{y}_{...})^2$	I-1	$MSF = SS_F / (I - 1)$	MSF / MSE
Residual	$SS_E = \sum \sum (y_{ij} - \bar{y}_{...})^2$	I(J-1)	$MSE = SS_E / (I(J - 1))$	
Corr. Total	$SST = \sum \sum (y_{ij} - \bar{y}_{...})^2$	IJ-1		

3. Results And Discussion

3.1. Data Analysis

447 people participated in the survey. Data was examined according to normal distribution analysis. In the study, since the relationship between the GHG emission of the household and the socio-cultural level was examined, the average of the survey results was calculated on household basis. When the results of the first part of the 447 questionnaires were analysed, there were 1413 people in these households.

The most important assumption of ANOVA is that the dependent variable is normally distributed in each subgroup of the explanatory variable. The explanatory variable is a 5-level categorical variable. Income group with 5 levels was used as the explanatory variable. From this point of view, analysis of normality was performed in each of the 5 groups. Since $p > 0.05$ was found in all 5 tests, the H_0 hypothesis was accepted. That is, the dependent variable was normally distributed in the subgroup ($j=1,2,3,4,5$). Normal distribution test results for p values are 0.1783, 0.3153, 0.4855, 0.5458, 0.6581, respectively, from poor to rich according to the income groups.

Table 2. The ANOVA test results

Source of variance	SS	df	MS	F	p-value
Between groups	4E+10	4	9E+09	1384	<0.001
In groups	1E+10	2135	6E+06		
Total	5E+10	2139			

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_1: \text{At least one } \mu_k \neq \mu_l \quad (k \neq l \text{ and } k, l=1,2,3,4,5)$$

The consistency of the remaining data within itself was checked. According to the ANOVA test results (Table 2), the p value between groups was found very closed to zero. Since the p-value was less than 0.05, the H_0 hypothesis was rejected.

At least one of the groups has been inferred to have different mean at the %0.05 significance level.

3.2 Carbon Footprint

We calculated the household carbon footprints of five groups based on their income distributions. When the emissions were examined, it was found that the emissions were high for upper income groups. Thereafter, household emission averages (tCO_{2e}) were determined as 5.07 for the poor group, 4.82 for the lower-middle group, 5.96 for the middle group, 7.96 for the upper-middle group, and 9.49 for the rich group. Contrary to expectations, the emissions of the lower income group were not the lowest. A closer look would reveal that at lower income level, the lowest emissions were observed (Figure 2). This decrease has been caused by the heavy use of non-clean fossil fuels (coal and fuel oil). At a poor level, 37.5% of the respondents get heated using non-clean fossil fuels. While this rate is 11% in the lower middle level, it has decreased to 0.5% in the middle. No non-clean fossil fuel use has been found in the upper middle and rich levels.

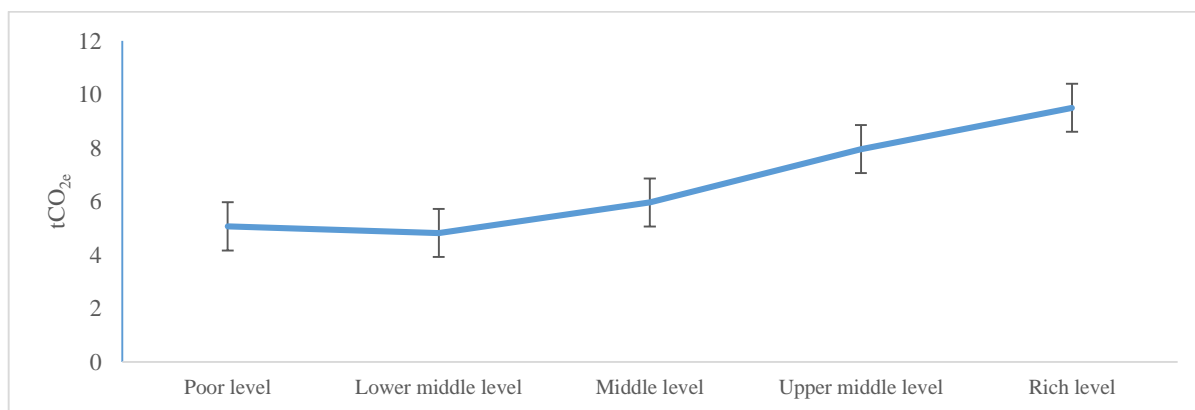


Figure 2. Average emissions (tCO_{2e}) according to income levels.

Whereas emissions have considerably increased between the lower middle income and rich income levels.

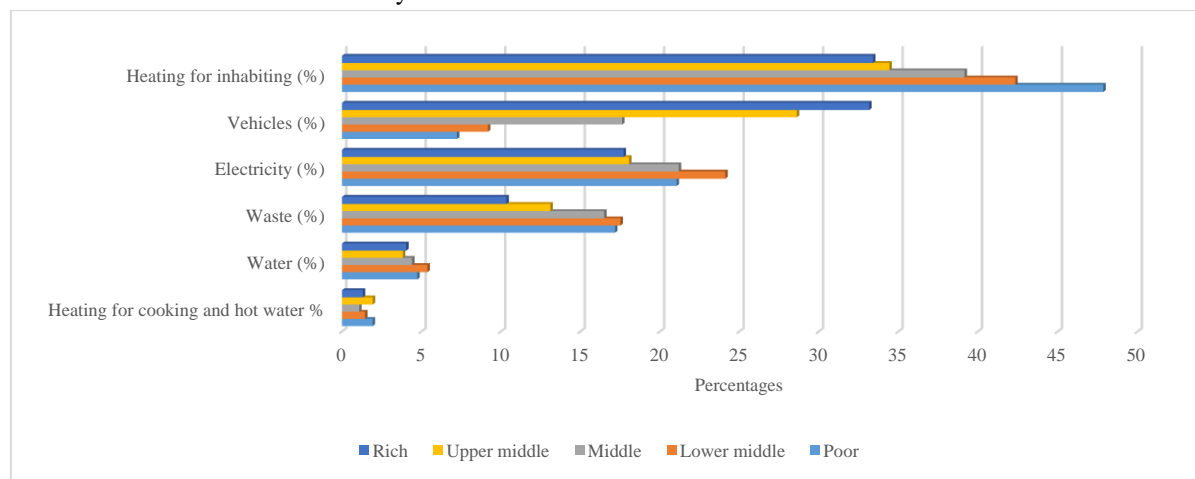


Figure 3. Percentages of household emissions sources

Heating accounts for 39% of emissions (heating for living + hot water supply and cooking) in the middle level. This is because the largest share of household emissions for all income levels belongs to heating. While this share is higher in the poor level (47.385%), it decreases by 33% in the rich level due to low carbon fuel use and vehicle emissions occupying an important place. This shows us the importance of the facade of houses and insulation. It can be said that, in total, 61% of emissions are direct emissions. On the other hand, 20% of emissions come from electricity use, which is an energy-indirect emission. In addition, 19% of emissions are from waste disposal and water supply, which is an indirect emission (Figure 3).

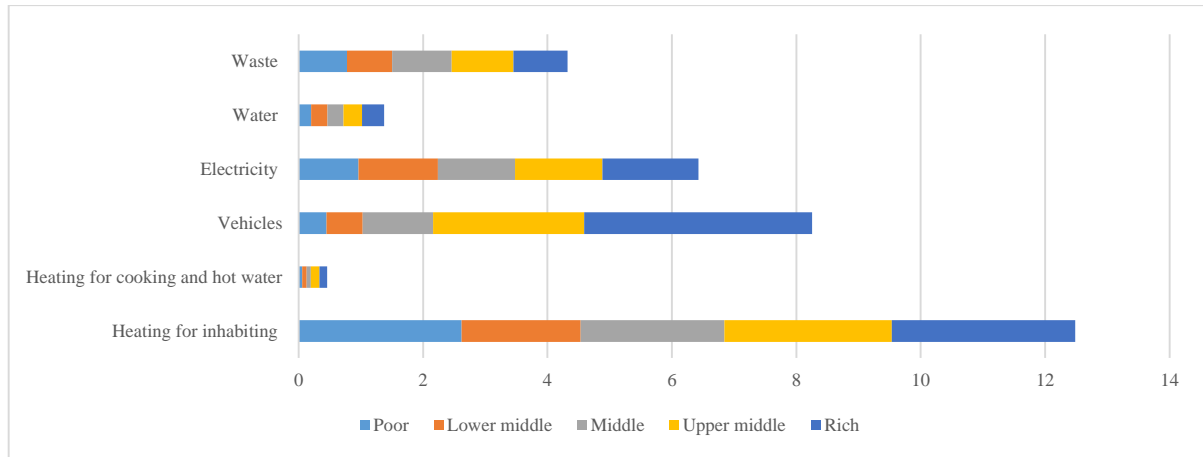


Figure 4. Average values (tCO_{2e}) of household emission sources

It is seen in Figure 4 that emissions from heating for habitation take the larger share of household carbon footprints, independent of income level. The city of Ankara is located in a continental climate. While the average temperature is 24 °C in the summer months, it decreases to -1 °C in the winter months (Climate Data, 2022). So, heating of the buildings in Ankara comes with a huge carbon footprint. We have stated factors affecting emissions from heating, that is the facade and the condition of the house. As can be predicted, much, little or no exposure to the sun will affect the amount of energy needed for heating. Accordingly, the least energy need for heating was seen in houses facing south, and the highest energy need was seen in houses facing north (Figure 5). Emissions due to occupied heating were highest in north and northwest facing houses. Values (kg CO_{2e}/per m²) were found to be 21.91 and 19.01, respectively. West and south-facing houses had the lowest emissions, with 17.45 and 17.36 emissions, respectively.

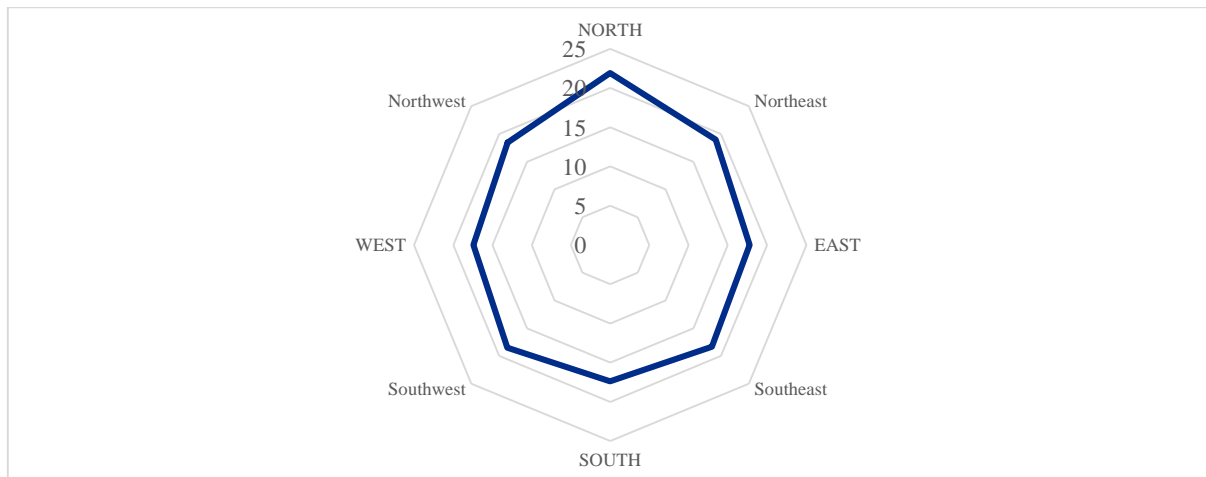


Figure 5. Facade and average heating emissions (kg CO_{2e}/per m²) of houses using natural gas

Another factor, when we compare the age of the houses with the emissions from heating (Figure 6), the emission values of the older houses come out higher. In the research, 44% of the houses were between 0–10 years old, and the lowest average heating emission value (kg CO_{2e}/per m²) was found to be 17.14 in this age group. While the average energy (Tj/kg) required for heating houses aged 0–10 year was 0.19, this value was determined as 0.30 for houses aged 40–50 years. It is obvious that as the ages of the houses increase, so does the energy required to heat the building.

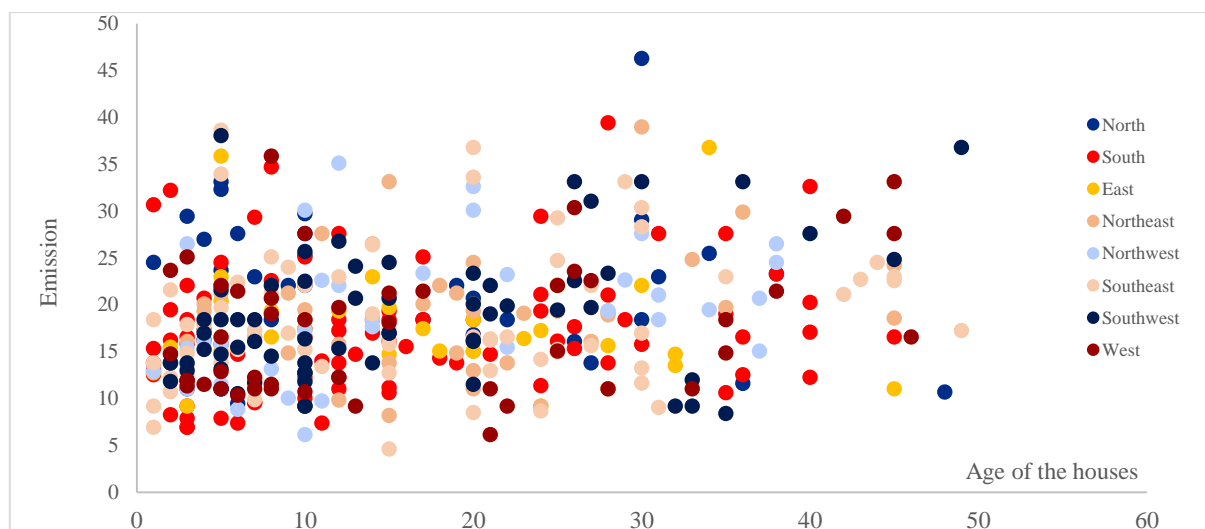


Figure 6. House age - average heating emissions (kg CO_{2e}/per m²) of houses using natural gas

Almost every household's emissions have been increased from a poor level to a rich level. There have been notable increases in two sources: vehicle use and meeting hot water and cooking needs. Because electricity is commonly used instead of natural gas for cooking and hot water in low-income households, emissions from this source were found to be lower than in higher-income households. When emissions from private vehicles are examined closely, not only the number of vehicles but also the distance travelled by these vehicles show a big difference (Figure 7). Vehicle owners in low-income areas travel an average of 3000 km per year. This figure was 19600 km at the rich level. It was seen that the use of vehicles has made a significant contribution to these increases. It was discovered that using private vehicles for transportation was very rare at the poor level (only 25% of respondents have their own vehicle). At the richest level (with the highest emissions), 95.45% of respondents own a vehicle, with a second owned by 25.75%.

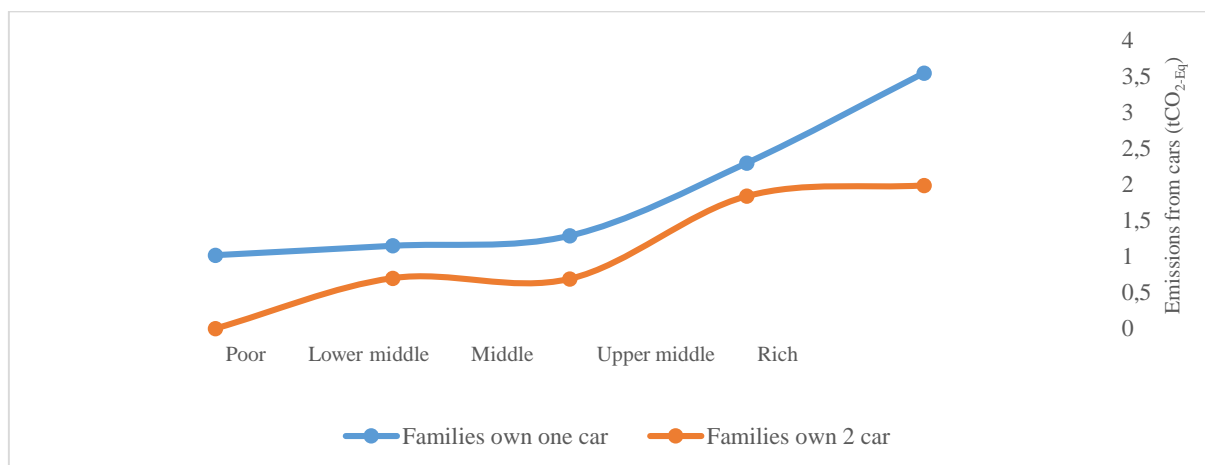


Figure 7. House age - average heating emissions (kg CO_{2e}/per m²) of houses using natural gas

3.3 Per Capita Carbon Footprint

One of the important parameters according to per capita carbon footprint is how many people the family consists of. As expected, when the family population is asymmetric across income levels, the per capita carbon footprint could be higher or lower than it needs to be. Therefore, it is important to determine the family population so that the survey results are not misleading. This situation did not affect per capita emissions as the average number of people per family varied between 3.5 and 3.7 (Figure 8) among income groups.

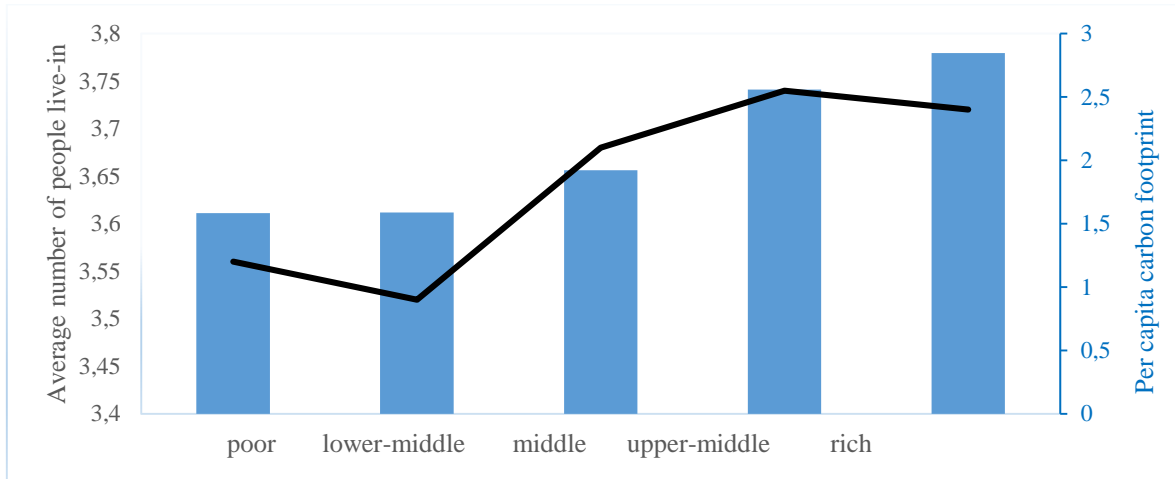


Figure 8. Per capita carbon footprint tCO_{2e} - average family population

The difference between the income groups was found to be high (up to almost 10 times). While the per capita carbon footprint (ton/CO_{2e}) in the poor level was 1.585, it was 2.104 in the upper-middle level, 1.922 in the middle level and 2.848 in the rich level. The ability to use more energy with an increasing income level reveals that there is a direct proportion between per capita emissions and income level. Although the per capita emissions in the middle group were found to be lower than the lower middle level, the values of the two levels were found to be close to each other. This decrease can be explained by the distribution of the survey data, its inherent uncertainty and other environmental factors. As a result (Figure 9), it has been clearly determined that the higher the income, the higher the per capita carbon footprint.

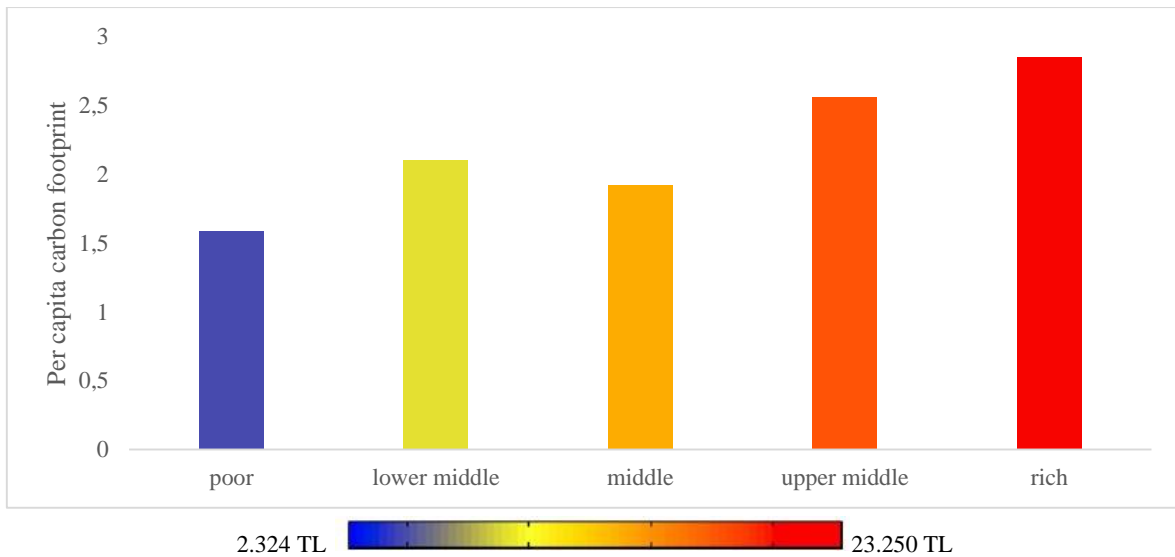


Figure 9. Per capita footprint (tCO_{2e}) according to amount of income

When the emissions related to electricity consumption are examined, a considerable increase (60.41% from poor to rich) has been observed. Grid electricity is being used within the boundaries of Ankara province, where the study was conducted, and all types of electricity are the same (no use of renewable energy). Keeping in mind family populations, which are close to each other, this increase in electricity emissions can be explained by lifestyle. Per capita electricity consumption has increased with the improvement in living conditions and welfare paralleled with industrialization. The fact that families with a higher income have more electrical devices means

they increase their electricity consumption and emissions. On the other hand, regardless of their income, families with a high population did not have emissions from electricity consumption. The number of toddlers in the family can explain this situation.

There has been a significant increase (up to 75%) in water consumption at lower income levels. This indicates that water consumption for personal use has increased. As the education level increases, it was seen that water consumption decreases in line with the awareness of individuals caring about the sustainability of natural resources.

Within the scope of the research, the origin of the domestic waste was examined. A study conducted in Turkey looking at the components of municipal solid wastes figured out that kitchen wastes have a significant place with 48.70%. Paper waste constitutes the second place with a rate of 19.10%. Plastic and glass waste are the third and fourth highest components, with 14.80% and 11.50%, respectively (Gökpur, Zıba & Dolaz, 2019). Accordingly, the reason for the high emissions from waste in the higher income group can be explained by the fact that it is related to more food consumption. When we evaluate waste generation and water consumption according to education levels regardless of the amount of income (Figure 10), we can say that the average amount of waste emissions did not differ from each other (low correlation; 0.033 stated in Table 2). In other respects, higher education levels come with lower water consumption. This situation (together with those described above in water consumption) can be explained by the increase in education levels contributing to water awareness.

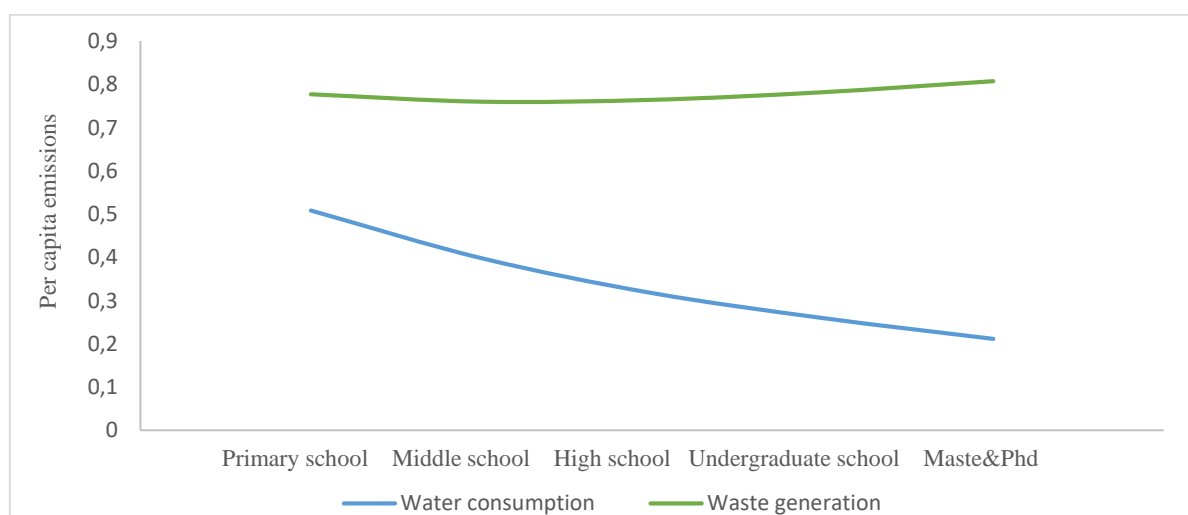


Figure 10. Per capita emissions (tCO_{2e}) from water consumption and waste generation

Another significant factor influencing emissions is whether the family lives in an apartment or a single-family home. While the average emission (tCO_{2e}) of those who live in a single house is 9.973, this number falls to 6.779 for those who live in apartments. The most important factor causing this difference is the heating for the inhabitants. Heating emissions (tCO_{2e}) constituted 2.426 (37.33% of total) for apartment residents and 3.775 (42.54%) for families living in a single house. The largest share of household emissions is sourced from heating. It was determined that as the income level increases, the energy used for heating also increases. Parallel to economic growth, the use of natural gas has scaled up rapidly in recent years. It is used extensively owing to its being a clean resource and ease of use. Natural gas is widely used in the districts where the survey was conducted. Therefore, there can be dramatic results for other regions, especially underdeveloped and rural ones, using non-clean sources. It has been determined that electricity consumption has increased at higher income levels. Lighting and refrigerators account for the majority of Turkey's electricity consumption (gazelektrik.com, 2020). As the amount of income increases, people want to live in larger houses. Big houses with more rooms need more enlightenment.

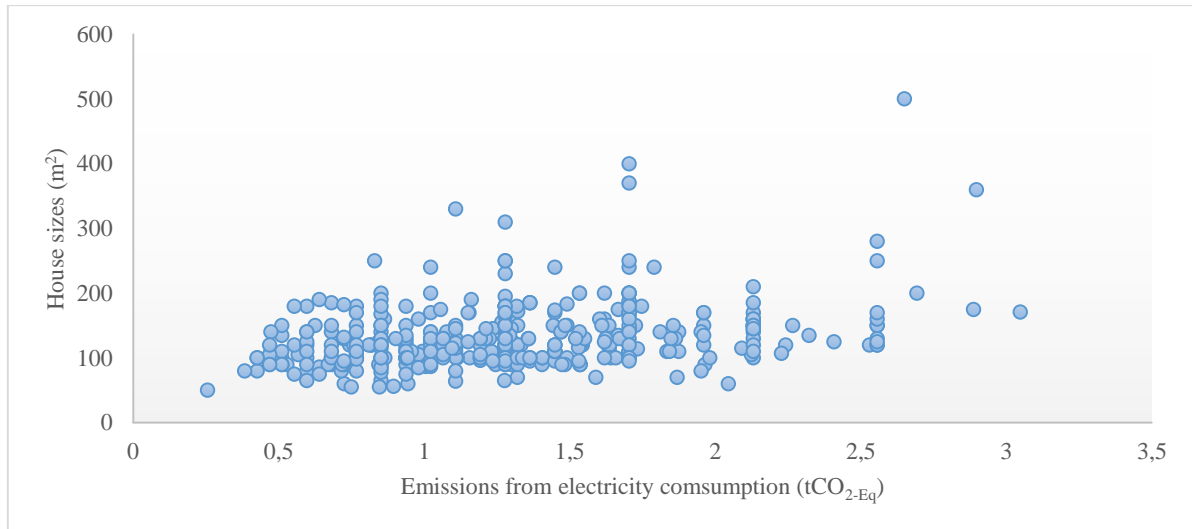


Figure 11. Emissions according to house sizes

Devices that work non-stop due to the constant need for electricity are used more in high-income families as seen in Figure 11. For example, some families may use deep freezers to store more food together with their refrigerator. Television has the second largest place in electricity consumption. In recent years, the number of televisions in families has increased due to the changes in watching habits and the fact that people have more private spaces. This increase is reflected in the amount of electricity-related emissions. Another important source of emissions was seen in vehicle use. As the income group increases, people do not prefer public transportation and prefer to travel in their own personal vehicles. Cars with LPG conversion systems and gasoline-powered cars are mainly used in lower income groups. These cars have lower emission values when compared with diesel cars. Diesel vehicles (lower fuel consumption) are mostly used at higher income levels. Since the prices of diesel vehicles are higher, these cannot be purchased by those with lower income levels. Food generates greenhouse gases while reaching to the plates and then for disposal. Approximately one-third of the food purchased from the market in developed countries is wasted as household waste (Premanandh, 2011). So, consuming as much as needed will reduce emissions from waste generation. The only emission source that decreases with the higher income level is emissions due to water consumption. Increasing income levels affect water consumption in two different ways. One of them is improving the living standards (spending more time outside, using a dishwasher, being able to pay for some cleaning tasks such as dry cleaning, carpet washing, etc.); and the other is conscious use thanks to education. While the emission from water consumption per capita in the poor level was 0.35 tons of CO_{2e}, it was 0.20 tons of CO_{2e} in the rich level. According to a study conducted in Balıkesir, Turkey, it has been observed that as the income level decreases, water consumption increases (Aliağaoğlu & Mirioğlu, 2019).

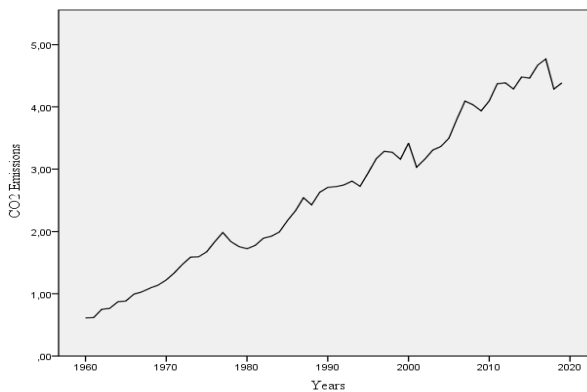


Figure 12. CO₂ emissions (ton per capita) of Turkey between 1960-2020 (World Bank Data, 2020)

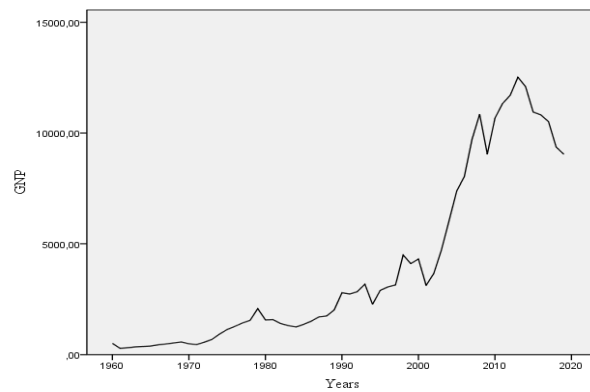


Figure 13. Gross national product (\$) of Turkey between 1960-2020 (World Bank Data, 2020)

Turkeys' last 60 years' development (Figure 12 and Figure 13) has meant that more growth and more enrichment comes with more energy use. Need for energy has continuously gone up due to the increasing urban population, developing technology, high level of industrialization and the rise in the information sector. A large part of the energy obtained from the use of fossil fuels has negatively affected the environment. We compared our result that economic growth affects positively the level of CO_{2e} emissions with the previous Environmental Kuznets Curve (EKC) hypotheses studied for Turkey. According to papers, with the increasing income level, a positive and significant role of energy consumption on CO_{2e} emissions has figured in Turkey (Lebe, 2016; Kemal & Hizarci, 2017; Dam, Karakaya & Bulut, 2014; Turkish Statistical Institute, 2020c). A study done by Turkish Statistical Institute (TUIK) shows that annual emission of a person in Turkey in 2018 is calculated as 6.4 tons CO_{2e} (Turkish Statistical Institute, 2020d).

4. Conclusion

In this paper parameters on household emissions are studied. By reducing the use of fossil fuels in energy needs, fore and most household GHG emissions can be reduced. For this purpose, much attention can be paid to social policies.

In this context, public transportation can be encouraged and energy management of buildings may be supported. Policies that neglect fuel efficiency considerations by reducing the fuel consumption of vehicles are no longer sustainable. Also some obligations could be adopted to electricity generating companies such as certain proportion of total production, given to grid, should be supplied from renewable sources rather than using fossil fuels. That's because these plants provide consistent electricity over long periods of time and are relatively inexpensive to construct. Carbon-based fuels, on the other hand, emit large amounts of CO_{2e}, which contributes to climate change. Other pollutants, such as sulphur and nitrogen oxides, are produced by these plants, resulting in acid rain. Giving more importance to recycling will have a huge recovery on amount of emissions. As we gain more technical and management competencies, we can build an increased capacity to use cycles in the end to end material production. Recycled resources will replace newly mined or manufactured materials in an increasing number of new products. These energy-intensive recycling procedures will have a decreasing impact on greenhouse gas emissions as our economy decarbonizes. To make this happen, we'll need technology, organizational capacity, human innovation, and political will. Additional government regulation, as well as financial incentives and disincentives, are needed to hardwire sustainability management into organizational activity in order to steer a large economy away from behaviours that degrade environmental quality. Besides it would be more accurate to say that, consuming as much as needed will strengthen the efforts done against climate change.

References

- AbdelHady, M. (2019). Income, inequality, and households' emissions: an assessment of the environmental kuznets curve in Canada. *Concordia University Department of Economics*, Research paper, (12-50).
- Adriana P, R., Pratama, A. (2020). Greenhouse gas emission estimations for Depok's (West Java, Indonesia) middle-class household water end-uses. *IOP Conference series earth and environmental science*.
- Aliağaoğlu, A., Mirioğlu, G. (2019). Balıkesir Şehrinde Su Tüketimi: Coğrafi Bir Yaklaşım (Water Consumption in Balıkesir City: A Geographical Approach). *Coğrafi Bilimler Dergisi / Turkish Journal of Geographical Sciences*, 17(2), 260-280.
- Barrett, J., Peters, G., Wiedmann, T., Scott, K., Lenzen, M., Roelich, K., Le Quéré, C. (2013). Consumption-based GHG emission accounting: a UK case study. *Climate policy*.
- Climate Data, 2020. Ankara - Climate graph, temperature graph, weather by month. <https://tr.climate-data.org/asya/turkiye/ankara/ankara-172/> last access 06.04.2022.
- Cong, R., Saito, M., Hirata, R., Ito, A. (2019). Spatiotemporal analysis on CO₂ emissions from households in Japan. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLII-2/W16, 2019, PIA19+MRSS19 – Photogrammetric image analysis & Munich Remote Sensing Symposium, 18–20 September 2019, Munich, Germany.
- Dam, M. M., Karakaya E. & Bulut, Ş. (2014). Çevresel Kuznets Eğrisi ve Türkiye: Amprik Bir Analiz. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, Eyi Özel Sayısı. 85-96.
- EN ISO 14064-1 (2019). Greenhouse Gases-Part 1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals.
- Endeksa, 2020. Research on Ankara income level, <https://www.endeksa.com/tr/analiz/ankara/demografi>, last access 26.09.2020.
- European Environment Agency (2015). Greenhouse gas emissions induced by household consumption, per Euro spent of expenditure in 12 household consumption categories, 2000-2007. <https://www.eea.europa.eu/data-and-maps/figures/direct-and-indirect-green-house-1>
- GHG Protocol (2021). Technical Guidance for Calculating Scope 3 Emissions, 3-9
- Gökpür, H., Zıba, C. A., Dolaz, M. (2019). Kahramanmaraş İli Dulkadiroğlu Bölgesi Katı Atık Bileşenlerinin Araştırılması. *Mühendislik Bilimleri ve Tasarım Dergisi*, 7(2), 345-351.
- Hargreaves K, Preston I, White V, Thumim J. (2013). The distribution of Household CO₂ emissions in Great Britain. *JRF Programme Paper Climate Change and Social Justice*.
- IPCC (2006). In: Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T., Tanabe, K. (Eds.). Guidelines for National Green House Gas Inventories Volume 2 Chapter 2 Stationary Combustion, IGES, Japan.
- Kemal, M.B., Hizarci, B.B. (2017). The relationship between energy consumption, CO₂ emissions and GDP per capita: A revisit of the evidence from Turkey. *Alphanumeric Journal*, 5, 353-368.
- Kennedy, C, Steinberger, J, Gasson, B, Hillman, T., Havránek, M., Hansen, Y, Pataki, D, Phdungsilp, A., Ramaswami, A., Villalba Mendez, G. (2010). Methodology for inventorying greenhouse gas emissions from global cities. *Energy Policy* 37 (9).
- Kijewska A, Bluszcz A. (2016). Analysis of greenhouse gas emissions in the European Union member states with the use of an agglomeration algorithm, The Silesian University of Technology, Faculty of Mining and Geology, 44-100 Gliwice, ul. *Akademicka 2*, Poland.

- Lebe, F. (2016). Çevresel Kuznets Eğrisi Hipotezi: Türkiye için Eşbütünlük ve Nedensellik Analizi. *Doğuş Üniversitesi Dergisi*, 17(2), 177-194
- Li, J., Huang, X., Yang H., Chuai, X., Li a, Y., Qu d, J., Zhang, Z. (2016). Situation and determinants of household carbon emissions in Northwest China. *Habitat International*, (51) 178-187.
- Liu L.N., Zeng Qu, J.J. & Wang Q.H., Wang L. (2013). Analysis the influence factors of China's household carbon intensity. *Environment, Energy and Sustainable Development – Sung, Kao & Chen (eds)*, DOI: 10.1201/b16320-94.
- Long, Y., Yoshida, Y., Fang, K., Zhang, H. & Dhondt, M. (2019). City-level household carbon footprint from purchaser point of view by a modified input-output model. *Appl. Energy*, 236, 379–387.
- Mert, M., Bölük, G. (2016). Do foreign direct investment and renewable energy consumption affect the CO2 emissions? New evidence from a panel ARDL approach to Kyoto Annex countries. *Environ. Sci. Pollut. Res.* 23 (21), 21669–21681
- Mi, Z., Zheng, J., Meng, J. et al. (2020). Economic development and converging household carbon footprints in China. *Nat Sustain* 3, 529–537. <https://doi.org/10.1038/s41893-020-0504-y>
- Miehe, R., Scheumann, R., Jones, C. M., Kammen, D. M., & Finkbeiner, M. (2016). Regional carbon footprints of households: A German case study. *Environment, Development and Sustainability*, 18, 577–591. <https://doi.org/10.1007/s10668-015-9649-7>
- Özcan M, Öztürk S. (2015). Türkiye'nin Elektrik Enerjisi Üretimi Kaynaklı Sera Gazı Emisyonunda Beklenen Değişimler ve Karbon Vergisi Uygulaması. *EMO*.
- Per Capita Electricity Consumption Per Household in Turkey, <https://gazelektrik.com/faydali-bilgiler/elektrik-tuketimi> last access 22.08.2020.
- Plassmann, K., Norton, A., Attarzadeh, N., Jensen, M. P., Brenton, P., Edwards-Jones, G. (2010). Methodological complexities of product carbon footprinting: a sensitivity analysis of key variables in a developing country context. *Environmental Science & Policy*, 13(5), 393-404.
- Premanandh, J. (2011). Factors affecting food security and contribution of modern technologies in food sustainability. *J. Sci. Food Agric.*, 91: 2707-2714.
- Soneja, S., Breysse, P., Tielsch, J. (2013). Assessment of household emissions due to cookstoves in southern Nepal, conference paper.
- Turkish Statistical Institute (2020a), Greenhouse Gas Emissions Statistics. Available at: <http://www.tuik.gov.tr/PreHaberBultenleri.do?id=24588> (accessed 10 August 2020)
- Turkish Statistical Institute, (2020b). Income distribution and living conditions statistics for Turkey, from http://www.tuik.gov.tr/PreTablo.do?alt_id=1011, last access 06.08.2020.
- Turkish Statistical Institute (2020c). Income. Life. Consumption and Poverty. Inflation and Price. Ankara. Retrieved July 13. 2020. from <http://www.tuik.gov.tr/UstMenu.do?metod=kategorist>
- Turkish Statistical Institute (2020d). Environment with Statistics, 2018. Ankara. Retrieved August 4, 2020 from <http://www.tuik.gov.tr/PreHaberBultenleri.do?id=33675>
- Türkeş, M. (2001). Küresel iklimin korunması iklim değişikliği çerçeve sözleşmesi ve Türkiye, Tesisat Mühendisliği. *TMMOB Makina Mühendisleri Odası, Süreli Teknik Yayın*, 61: 14-29.
- Uzar, U.; Eyuboglu, K. (2019). The nexus between income inequality and CO2 emissions in Turkey. *J. Cleaner Production*, 227, 149–157

Vardopoulos, I., Konstantinou, Z. (2017). Study of the possible links between CO₂ emissions and employment status. *Sustainable Development, Culture, Traditions Journal*, Volume 1b.

Wang, H., Liu, G., Shi, K. (2019). What are the driving forces of urban CO₂ emissions in China? A refined scale analysis between national and urban agglomeration levels. *Int. J. Environ. Res. Public Health*, 16, 3692

Wang, S., Li, G., Fang, C. (2018). Urbanization, economic growth, energy consumption, and CO₂ emissions: empirical evidence from countries with different income levels. *Renew. Sust. Energ. Rev.*, 81, 2144–2159.

World Bank Data, 2020. <https://data.worldbank.org/> last access:14 October 2020.

Yatarkalkmaz, M.M., Özdemir, M.B. (2019). The calculation of greenhouse gas emissions of a family and projections for emission reduction. *Journal of Energy Systems*, 3(3): 96-110, DOI: 10.30521/jes.566516

Zhang H.W., Shi X.P., Wang K.Y., Xue J.J., Song L.G., Sun Y.P. (2020). Intertemporal lifestyle changes and carbon emissions: evidence from a China household survey. *Energy Econ*, 86:104655 <https://doi.org/10.1016/j.eneco.2019.104655>.

Zhu L., Huang Y., Chen W. (2019). Household carbon emission characteristics and adaptive strategies for low carbon community planning. *E3S Web of Conferences*, 136, 04058 <https://doi.org/10.1051/e3sconf/2019136040> E 3S



Açık Veri ve Coğrafi Bilgi Sistemleri Kullanarak Türkiye için Sürdürülebilir Kalkınma Göstergeleri Tahmini

Selin TEKİN

Türkiye İstatistik Kurumu / TÜİK Uzman Yardımcısı

selin.tekin@tuik.gov.tr

Orcid No: 0000-0002-5382-9627

Birkan ERGÜÇ

Türkiye İstatistik Kurumu / TÜİK Uzmanı

birkan.erguc@tuik.gov.tr

Orcid No: 0000-0002-9470-0803

Serdar ÖZ

Türkiye İstatistik Kurumu / Başkan Yardımcısı

serdar.oz@tuik.gov.tr

Orcid No: 0000-0002-2933-2661

Özet

Tüm dünyada sürdürülebilir kalkınmaya duyulan ihtiyacın artması ve karşı karşıya kalınan çevresel, sosyal ve ekonomik krizlerle mücadele edebilme hedefi 2015 yılında Birleşmiş Milletler tarafından kabul edilen Sürdürülebilir Kalkınma Amaçlarının (SKA) temelini oluşturmaktadır. Şehirler ve yerleşim alanlarının sürdürülebilir kılınması da bu amaçların merkezinde yer alan unsurlardandır. Gelişen şehirlerin ve giderek büyüyen yerleşim alanlarının durumu insanın hayat kalitesini doğrudan etkilemektedir. Bu çalışmada Sürdürülebilir Şehirler ve Topluluklar Amacı kapsamında “Gösterge 11.2.1:Toplu taşıma araçlarına kolay erişimi olan nüfusun oranı” ve “Gösterge 11.7.1: Kamu kullanımına tamamen açık, yapılaşmış alanların şehirlerdeki ortalama payı” Coğrafi Bilgi Sistemleri (CBS) ve açık veri kaynakları kullanılarak hesaplanmıştır. Gösterge 11.2.1 Ankara, İstanbul ve İzmir illeri için, Gösterge 11.7.1 tüm iller için hesaplanmıştır. Şehirlerin toplu taşıma verileri kent bilgi sistemlerinden elde edilirken, kent alanı ve kamusal açık alan verileri uydu görüntüsü ve kitle kaynaklı açık veri kaynaklarından elde edilmiştir. Her iki göstergenin hesaplanmasında da Birleşmiş Milletlerin yönergeleri izlenmiştir. Çalışmada ulusal sürdürülebilir kalkınmanın izlenmesi ve değerlendirilmesi için yeni kaynak ve yöntemlerle gösterge üretilmesi amaçlanmıştır.

Anahtar Sözcükler: Coğrafi Bilgi Sistemleri, Kamusal Açık Alan, Kentsel Sürdürülebilirlik, Sürdürülebilir Kalkınma Amaçları, Toplu Taşıma, Türkiye

Sorumlu Yazar / Corresponding Author: 1-Selin TEKİN, Türkiye İstatistik Kurumu, Yazılım Daire Başkanlığı, Coğrafi Bilgi Sistemleri Grup Başkanlığı.

2-Birkan ERGÜÇ, Türkiye İstatistik Kurumu, Yazılım Daire Başkanlığı, Coğrafi Bilgi Sistemleri Grup Başkanlığı.

3-Serdar ÖZ, Türkiye İstatistik Kurumu, Başkan Yardımcısı.

Atf / Citation: TEKİN S., ERGÜÇ B., ÖZ S. (2022). Açık Veri ve Coğrafi Bilgi Sistemleri Kullanarak Türkiye için Sürdürülebilir Kalkınma Göstergeleri Tahmini. İstatistik Araştırma Dergisi, 12 (1), 56-67.

Estimating Sustainable Development Indicators Using Open Data and Geographic Information Systems for Turkey

Abstract

The basis of the Sustainable Development Goals (SDG) adopted by United Nations in 2015 is the increasing need for sustainable development all over the world and also the aim of fighting environmental, social and economic crises. Making cities and residential areas sustainable is one of the main elements of this goals. The status of developing cities and growing urban areas directly affects the quality of human life. In this study, two indicators under the goal of Sustainable Cities and Communities, "Indicator 11.2.1: "Proportion of population that has convenient access to public transport" and " Indicator 11.7.1: "Average share of the built-up area of cities that is open space for public use for all" were estimated using Geographic Information Systems (GIS) and open data sources. Indicator 11.2.1 was estimated for three major provinces namely Ankara, İstanbul and İzmir and Indicator 11.7.1 was estimated at the province level for whole Turkey. The public transportation data was obtained from the city information systems whereas urban area and public open space data was obtained from open data sources based on satellite imagery and crowd-sourcing. United Nations guidelines were followed for estimating both indicators. In this study it was aimed to produce indicators using novel sources and methods for monitoring and evaluating national sustainable development.

Keywords: Geographic Information Systems, Open Public Space, Public Transportation, Sustainable Development Goals, Turkey, Urban Sustainability

1. Giriş

İnsanlık, geçirdiği tüm gelişim süreçleri boyunca çevresindeki kaynakların ve süreçlerin hem etkisi altında kalmış hem de çevresini etkileyip dönüştürmüştür. Bu dönüşüm yaşanırken elde edilen pozitif kazanımların sürekli olabilmesi hedefi sürdürülebilirlik fikrini ortaya çıkartmıştır. Sürdürülebilirlik konusu, enerji, ekonomi ve çevre üçgeninin tam merkezinde bulunduğundan hükümetler ve hükümetler arası örgütlerin yanı sıra, sosyal bilimcilerden fen ve doğa bilimcilerine, politikacılardan yerel ve uluslararası çevre örgütlerine kadar oldukça geniş bir yelpazede tartışılmış, farklı tanımlar ve yaklaşımlar geliştirilmiştir. (Yeni, 2014). Sürdürülebilir kalkınma kavramı ise Dünya Çevre ve Kalkınma Komisyonu tarafından 1987 yılında gelecek nesillerin kendi ihtiyaçlarını karşılayabilme yeteneğini ortadan kaldırmaksızın şimdiki neslin ihtiyaçlarının karşılanması olarak tanımlanmıştır (UN-WCED, 1987). İnsanlığın günümüze kadar elde ettiği kalkınmanın sürekli kılınarak gelecek nesiller tarafından da faydalanılabilir olması adına kalkınmanın doğrudan ya da dolaylı olarak ortaya çıkarttığı çevresel, sosyal ve ekonomik krizleri önlemek ya da etkilerini minimuma indirmek temel prensiptir.

Tüm dünyada sürdürülebilir kalkınmaya duyulan acil ihtiyacın artması ile Sürdürülebilir Kalkınma Amaçları fikri hızla zemin kazanmıştır (Sachs, 2012). Sürdürülebilir Kalkınma Amaçları diğer önceliklerin yanı sıra iklim değişikliği, ekonomik eşitsizlik, inovasyon, sürdürülebilir tüketim, barış ve adalet gibi yeni alanları da dâhil ederek Binyıl Kalkınma Hedeflerinin üzerine inşa edilmiştir (Permai, Mukhaiyar, Satyaning, Soleh, & Aini, 2018). 25 Eylül 2015 tarihinde Birleşmiş Milletler'in (BM) Sürdürülebilir Kalkınma Zirvesi'nde, aralarında Türkiye'nin de bulunduğu 193 üye ülkenin imzası ile kabul edilen 2030 Sürdürülebilir Kalkınma Amaçları (SKA 2030), 17 temel amaç ve bunlara bağlı olan 169 adet hedef ve 247 tekrarlı (231 tekil) göstergeden oluşmaktadır. Bu yeni küresel gündem, 2030 yılına kadar herkes için insan hakları ilkesini öne çıkaran bir anlayışla ülkelerin kalkınma plan ve politikalarının şekillenmesini öngörmektedir (United Nations, 2015).

Sürdürülebilir Kalkınma Amaçlarının uygulanmasında en önemli unsur öncelikle ulusal hedeflerin belirlenmesi ve bu doğrultuda üretilen göstergeler ile hedeflere ne ölçüde ulaşılabildiğinin değerlendirilerek gereken tedbirlerin alınmasıdır. Küresel bir kalkınma stratejisi olan bu yaklaşımın küresel verimlilik hareketine dönüşebilmesi için hükümetlerin, özel sektörün, sivil toplum kuruluşlarının, bilim dünyasının ve ilgili tüm paydaşların ortak bir anlayışla işbirliği yapmaları gerekmektedir. (Peşkirioğlu, 2016).

Sürdürülebilir kalkınma mücadelesi için "ya şehirlerde kazanılacak ya da şehirlerde kaybedilecek" ifadesi sıklıkla dile getirilmektedir (Kharrazi, Qin, & Zhang, 2016). Şehirler toplumun sürdürülebilir kalkınmasını ileri taşıyan kilit unsurlar haline geldiğinden kentsel sürdürülebilir kalkınma göstergelerinin araştırılması ve uygulanması, kentsel gelişim stratejilerinin oluşturulmasına rehberlik edecek, kamu politikalarının etkisini izleyebilecek ve sosyal propaganda yolu ile bilgi şeffaflığını ve kamu denetimini teşvik edebilecektir (Xiang, 2017). Kentsel sürdürülebilirliği

temel alan ve Sürdürülebilir Kalkınma Amaçlarının 17 temel amacından olan Amaç 11: Sürdürülebilir şehirler ve topluluklar oluşturmayı, şehirleri ve insan yerleşimlerini kapsayıcı, güvenli, dayanıklı ve sürdürülebilir kılmayı hedeflemektedir. Ancak standartlaştırılmış, açık ve kıyaslanabilir veri eksikliği ve şehir ölçeğinde kentsel sürdürülebilirliği izleyecek güçlü veri toplama kurumlarının olmaması bu göstergeye dair temel problemlerdir (Klopp & Petretta, 2017). Ayrıca sürdürülebilir şehirler ve topluluklar hedefleri çoğunlukla kent alanlarına hitap etmesine rağmen; bu hedeflerin şehirlerde nasıl uygulanacağına dair net bir kılavuz olmadığına da dikkat çekilmektedir. (Koch & Krellenberg, 2018).

Büyüyen kentsel bölgelerin refahı, yeterli ve uygun ulaşım hizmetlerinin sağlanmasıyla yakından bağlantılıdır. Genişleyen bir kentsel nüfus, ticari faaliyetlere, eğitime, istihdama ve eğlence imkânlarına erişim gerektirir. Bu hizmetlerin konumu ve otoyollar, toplu taşıma ve otopark alanları gibi yeterli ulaşım altyapısının sağlanması, şehir planlamasının özüdür (Murray, Davis, Stimson, & Ferreira, 1998). Yeterli, uygun ve erişilebilir toplu taşıma, kent hayatının sürdürülebilirliği açısından oldukça önemlidir. Sürdürülebilir Kalkınma Hedefleri göstergelerinden olan ve bu çalışmada hesaplanan “Gösterge 11.2.1: Toplu taşıma araçlarına kolay erişimi olan nüfus oranı” bu temele dayanmaktadır. Bu gösterge ile 2030’a kadar özellikle kırılan durumda olan insanların, kadınların, çocukların, engellilerin ve yaşlıların ihtiyaçlarına özel önem gösterilerek, yol güvenliğinin artırılması, özellikle toplu taşıma sisteminin geliştirilmesiyle herkesin güvenli, uygun fiyatlı, erişilebilir ve sürdürülebilir ulaşım sistemlerine erişiminin sağlanması hedeflenmektedir. Tüm bunların yanında, Gösterge 11.2.1’in arz merkezli olması yani fiziksel toplu taşıma altyapısına dayanması bir çeşit zayıflık olarak değerlendirilmiştir (Brussel, Zuidgeest, Pfeffer, & Maarseveen, 2019). Bu durum bisikletle veya yürüyerek yapılabilecek kısa yolculukları teşvik eden sürdürülebilir ulaşım mantığına aykırı görülmektedir (Tiwari & Philip, 2021).

Sürdürülebilir şehirler aynı zamanda değişimi ve gelişimi sürdürmek için çevre ile dost tasarımlarla sosyoekonomik çıkarların ve doğal kaynakların korunmasını hedefleyen yaşanabilir şehirlerdir (Demircan & Sezen, 2018). Kentsel yeşil alanlar, insan ve doğa arasındaki bozulan ilişkileri dengeleyerek kentsel planlamaya bütünsel bir yaklaşım sağlar ve kentsel yaşam koşullarının iyileştirilmesinde de önemli bir rol oynar (Çetin, 2015). “Gösterge 11.7.1: Kamu kullanımına tamamen açık, yapılaşmış alanların şehirlerdeki ortalama payı” bu yaklaşımlara dayanmaktadır. Bu doğrultuda, 2030’a kadar özellikle kadınlar, çocuklar, yaşlılar ve engellilerin güvenli, kapsayıcı ve erişilebilir yeşil alanlara ve kamu alanlarına erişimlerinin güvence altına alınması hedeflenmektedir. Bu göstergeye dair temel zorluk ise kent içerisindeki yeşil alanlar uzaktan algılama yoluyla görece daha kolay bir şekilde elde edilebilirken “kamusal alan” kavramının arazi örtüsünden çok bir arazi kullanım kategorisi olduğundan elde etmesinin daha zor olmasıdır (Blaschke & Kovács-Györi, 2020).

Bu çalışmada, Gösterge 11.2.1 ve Gösterge 11.7.1 sözü edilen tüm beklenti ve zorlukları göz önünde bulundurularak coğrafi bilgi sistemleri aracılığıyla hesaplanmıştır. Sürdürülebilir Şehirler ve Topluluklar Amacı da dahil hesaplanan diğer Sürdürülebilir Kalkınma Amaçları göstergeleri “Sürdürülebilir Kalkınma Göstergeleri 2010-2019” haber bülteni olarak Türkiye İstatistik Kurumu tarafından yayımlanmıştır (TÜİK, 2021). Çalışma ile Türkiye’nin Sürdürülebilir Şehirler ve Topluluklar Amacı kapsamında mevcut durumunu ortaya koyarak kentlerdeki sürdürülebilir kalkınmanın planlanması ve izlenmesi için bir başlangıç yapmak hedeflenmiştir. Ayrıca uluslararası platformda sürdürülebilir kalkınma çalışmalarına Türkiye’nin verileri ile katkı sağlamak da amaçlamıştır. Bu amaca hizmet edecek iki gösterge yeni kaynak ve yöntemlerle üretilmiştir.

2. Çalışma Alanı

Çalışma alanı Türkiye’dir. Türkiye, nüfusları 81.910 ile 15.462.452, yüzölçümleri 798 km² ile 40.838 km² arasında değişen 81 ile sahiptir, toplam nüfusu 2020 yılı itibariyle 83.614.362’dir (TÜİK, 2021). En yüksek nüfusa sahip İstanbul, Ankara ve İzmir illeri toplam nüfusun %30,5’ini oluşturur. Gösterge 11.2.1 için gerekli olan toplu taşıma verileri, çalışmanın yapıldığı süreçte yalnızca Ankara, İstanbul ve İzmir illeri için elde edilebildiğinden gösterge bu iller için hesaplanmıştır. Gösterge 11.7.1 ise tüm Türkiye için il kırılımında hesaplanmıştır.

3. Veri Kaynakları

Her iki göstergenin hesaplanmasında kullanılan veriler ve verilerin temin edildiği kaynaklar sırasıyla Tablo 1 ve Tablo 2’de özetlenmiştir.

Tablo 1. Gösterge 11.2.1'in Hesaplanması için Kullanılan Veriler ve Veri Kaynakları

Veri Türü	Veri Kaynağı
Ankara İli Toplu Taşıma Verileri (Metro ve otobüs durakları, dolmuş güzergâhları)	Ankara Ego Ulaşım Bilgi Sistemi http://map.ego.gov.tr
İstanbul İli Toplu Taşıma Verileri (Metro, metrobüs, denizyolu ve otobüs durakları, dolmuş güzergâhları)	İstanbul Açık Veri Portalı https://data.ibb.gov.tr
İzmir İli Toplu Taşıma Verileri (Metro, denizyolu ve otobüs durakları)	İzmir Kent Rehberi http://kentrehberi.izmir.bel.tr/izmirkentrehberi
İl Sınırları	Harita Genel Müdürlüğü, 2019
Nüfus Yoğunluğu Gridleri (NYG)	Türkiye İstatistik Kurumu, 2019

Tablo 2. Gösterge 11.7.1'in Hesaplanması için Kullanılan Veriler ve Veri Kaynakları

Veri Türü	Veri Kaynağı
Kent Alanı ve Yerleşim Alanları	Landsat Küresel İnsan Yerleşimi (H-BASE) Yerleşim Kapsamı (Global Human Built-up And Settlement Extent), 2010
Meydan, Cadde, Sokak ve Yeşil Alan Verileri	Küresel İnsan Yerleşim Katmanı (Global Human Settlement Layer), 2018 Open Street Map, 2020
İl Sınırları	Harita Genel Müdürlüğü, 2019

4. Metodoloji

Göstergelerin hesaplanmasında Birleşmiş Milletler'in Sürdürülebilir Kalkınma Amaçları için yayımladığı yönergeler izlenmiştir (UN-Habitat, 2021). Göstergelerin hesaplanmasında ArcMap 10.8 ve QGIS 3.0 yazılımları kullanılmıştır.

4.1. Toplu Taşıma Araçlarına Kolay Erişimi Olan Nüfusun Oranı

Göstergenin hesaplanacağı illerin kent bilgi sistemlerinden elde edilen tüm otobüs, metro, metrobüs ve deniz ulaşımı durakları, dolmuş güzergâhları verileri WGS84 projeksiyon sistemine dönüştürülmüştür.

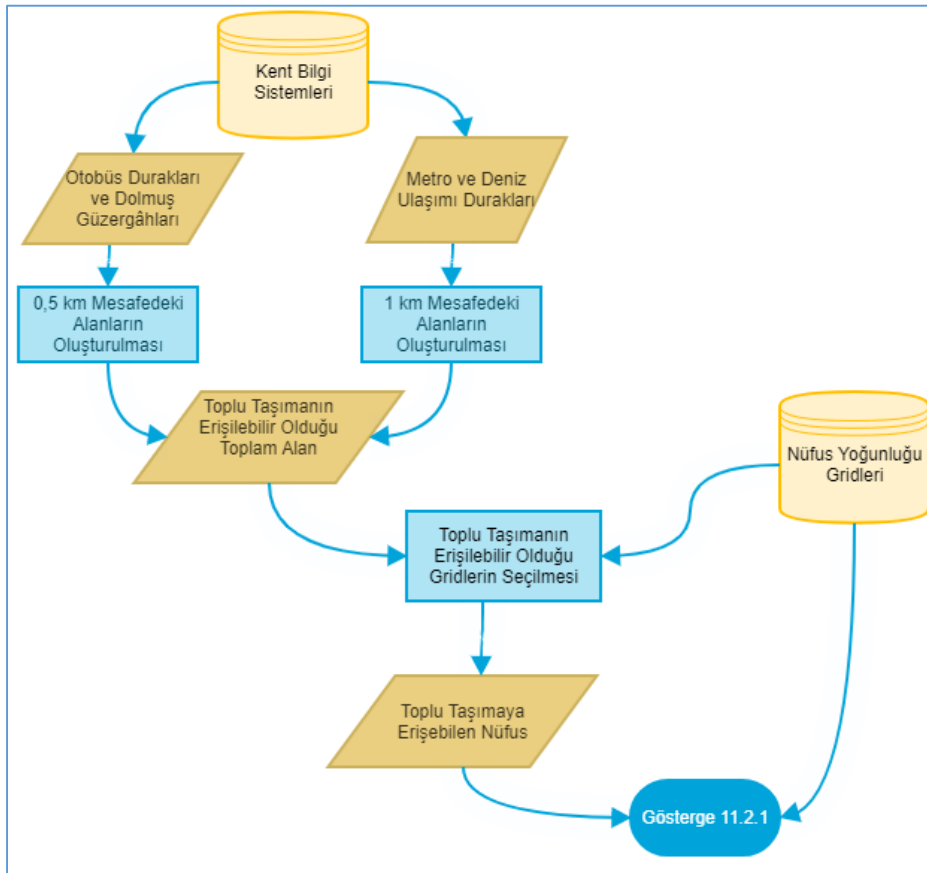
Birleşmiş Milletler'in tanımına göre toplu taşımaya erişilebilir kabul edilebilmesi için, bir referans noktasından toplu taşımaya 0,5 km mesafede ulaşabilmek gerekmektedir. Bu yaklaşımdan yola çıkılarak otobüs duraklarına, dolmuş güzergâhlarına 0,5 km, diğer toplu taşıma seçeneklerine oranla daha yüksek kapasitede olduklarından metro ve deniz ulaşımı duraklarına ise 1 km mesafede tampon bölge oluşturulmuştur. Elde edilen tampon bölge katmanları birleştirilerek şehir içerisindeki toplam ulaşım erişilebilir alan bulunmuştur (Şekil 1).

Toplu taşımaya erişebilen nüfusun bulunabilmesi için ise nüfus yoğunluğu gridleri kullanılmıştır. Nüfus yoğunluğu gridleri, veri mevcudiyeti, istatistiki gizlilik gibi faktörlere bağlı olarak toplulaştırma veya ayrıştırma metotlarıyla üretilebilmektedir. Toplulaştırma yöntemi, koordinatları bilinen adres noktalarında ikamet eden nüfusun içinde bulunduğu grid hücrelerine toplanmasıyla elde edilir ve yüksek doğruluktadır. Koordinatlı adres veri setinin bulunmadığı yerlerde ise idari sınırlar için üretilen nüfusun arazi kullanım haritası gibi bir yardımcı veri kullanılarak

gridlere ayrıştırılması yöntemi kullanılır ve toplulaştırma yöntemine göre daha düşük doğruluktur. Veri gizliliğinin ihlal edildiği durumlarda veya adres noktalarına dayalı kısmi veri bulunması halinde iki yöntemin birlikte (hibrit) kullanılması da söz konusudur (Eurostat, 2018). Adres ile eşleştirilmiş nüfus verisi henüz tüm Türkiye için mevcut olmadığından çalışmada ayrıştırma yöntemi kullanılarak elde edilen, 1 km² çözünürlüğe sahip nüfus yoğunluğu gridleri kullanılmıştır. TÜİK tarafından üretilen nüfus yoğunluğu gridleri Çevre ve Şehircilik Bakanlığı'nın coğrafi veri portalında yayımlanmaktadır (TUCBS Ulusal Coğrafi Bilgi Platformu, 2021).

Göstergenin hesaplanması için toplu taşımanın erişilebilir kabul edildiği alan ile kesişen nüfus yoğunluğu gridlerindeki nüfus, toplam nüfusa oranlanmıştır (Formül 1).

$$\text{Gösterge 11.2.1} = 100 \times \frac{\text{Toplu Taşımaya Kolay Erişebilen Nüfus}}{\text{Toplam Nüfus}} \quad (1)$$



Şekil 1. Gösterge 11.2.1'in Hesaplanması için Kullanılan Yöntem

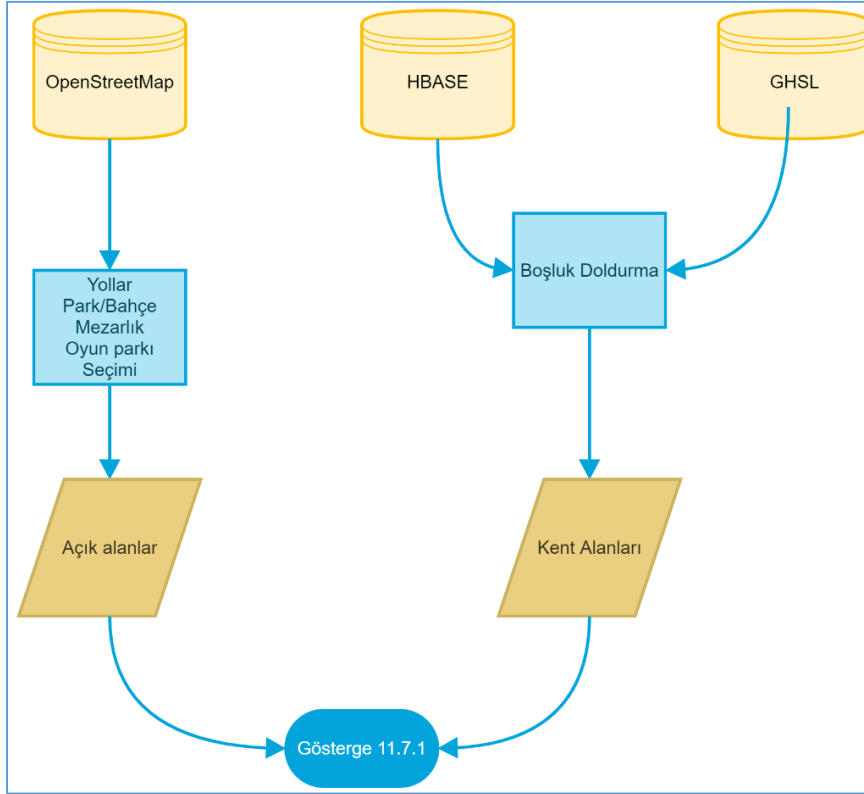
4.2. Kamu Kullanımına Tamamen Açık, Yapılaşmış Alanların Şehirlerdeki Ortalama Payı

Kamuya açık kentsel alan kavramı farklı kültürlerde farklı olarak değerlendirilse de bu çalışmada, açık kamusal alan kavramı, parklar ve mezarlıklar, rekreasyon alanları, çocuk parkları, cadde, sokak ve meydanlar olarak kabul edilmiştir. Kentsel alanın tespiti için ise Landsat uydu görüntüsü temelli kent alanı ve yerleşim alanları verisi - HBASE (Wang, Huang, Brown de Colstoun, Tilton, & Tan, 2017) kullanılarak kent alanları ortaya konmuş, bu verideki boşluklar Global Human Settlement Layer (Corbane, Florczyk, Pesaresi, Politis, & Syrris, 2018) verisi ile doldurulmuştur. Meydan, cadde ve sokak verileri ile yeşil alan verileri OpenStreetMap açık veri kaynağından (OpenStreetMap contributors, 2020) elde edilmiştir.

Tüm veriler WGS 1984 Web Mercator Auxiliary Sphere koordinatlarına dönüştürülmüştür. Cadde ve sokakların payını hesaplamak için otoban ve bağlantılarında 50 m, anayollarda 20 m ve diğer yollarda 6 m tampon analizi yapılarak yol çizgilerinden alanlar elde edilmiştir. Yeşil alanların tespitinde ise park, bahçe, çocuk parkı, ağaç parkı,

tema parkı, su parkı, mezarlık ve rekreasyon alanları filtrelenmiş ve kent alanları içerisinde bulunan veya kent alanına en fazla 100 m uzaklıkta bulunan yeşil alanlar çalışmada kullanılmıştır. (Şekil 2)

Elde edilen alan verileri tüm Türkiye için oranlanmış (Formül 2) ve gösterge değeri bulunmuştur ayrıca bu alanlar il sınırlarına göre kesilerek birbirine oranlanmış ve gösterge değeri il bazında elde edilmiştir. İl bazında elde edilen sonuçların il nüfuslarıyla olan korelasyonu hesaplanmıştır.



Şekil 2. Gösterge 11.7.1'in Hesaplanması için Kullanılan Yöntem

$$Gösterge\ 11.7.1 = 100 \times \frac{\text{kamu kullanımına açık alanlar} + \text{sokaklara ayrılmış alanlar}}{\text{toplam şehir kentsel yerleşim alanı}} \quad (2)$$

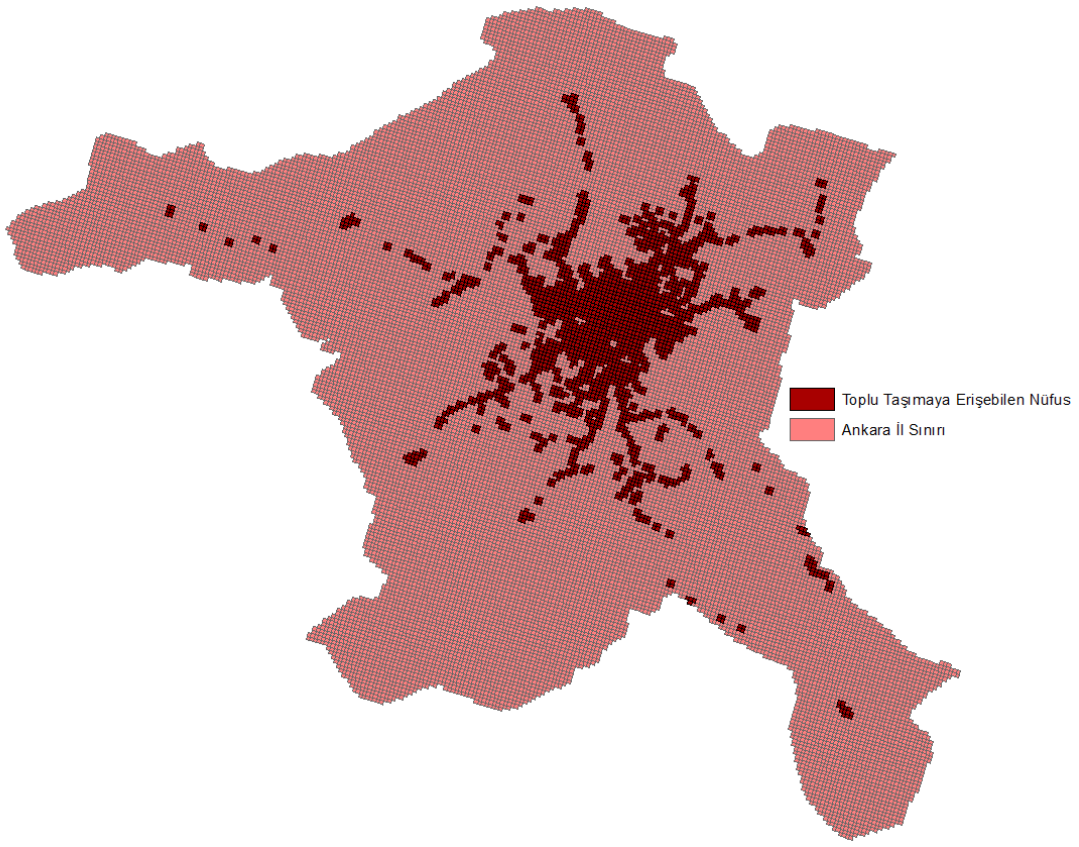
5. Tartışma ve Sonuç

Çalışma sonucunda Gösterge 11.2.1'in hesaplandığı şehirler için elde edilen sonuçlar Şekil 3, Şekil 4 ve Şekil 5'te gösterilmiştir. Toplu taşımaya kolay erişimi olan nüfusun toplam nüfusa oranı Ankara, İstanbul ve İzmir için sırasıyla %87.9, %98.1 ve %84 olarak hesaplanmış, sonuçlar Tablo 3'te sunulmuştur. İzmir için elde edilen değer diğer iki il ile gösterdiği farklılık kent bilgi sisteminde dolmuş güzergâhları verisi olmaması olarak açıklanabilir.

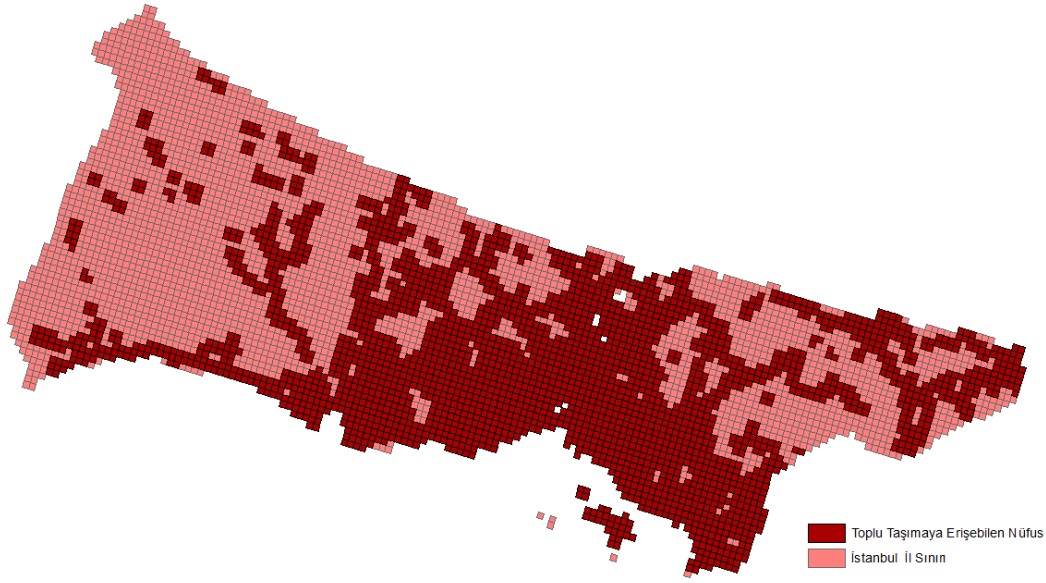
Gösterge 11.2.1 yayınlayan diğer ülkeler incelendiğinde gösterge hesaplanırken farklı yaklaşımların benimsendiği görülmüştür. Örneğin; Almanya iş günlerinde en az 28 seferi olan durakları belirleyerek bunlara 500 metre mesafede bulunan alanı erişilebilir kabul etmiş ve gösterge değerini %79.9 olarak hesaplamıştır (Almanya İstatistik Ofisi, 2020). Norveç mesafe parametresine ek olarak iş günlerinde 06:00- 20:00 saatleri arasında saatte en az bir sefer olan toplu taşıma hizmetini erişilebilir toplu taşıma olarak kabul etmiş ve gösterge değerini %92.3 olarak hesaplamıştır (Norveç İstatistik Kurumu, 2019). Danimarka ise gösterge değerini kent alanlarında %83.3, kırsal alanlarda %17.4 olarak hesaplamış ayrıca erişilebilirlik oranlarını kırsal ve kentsel alanda cinsiyet kırılımında da yayımlamıştır (Danimarka İstatistik Kurumu, 2019).

Tablo 3. Toplu Taşıma Araçlarına Kolay Erişimi Olan Nüfusun Oranı

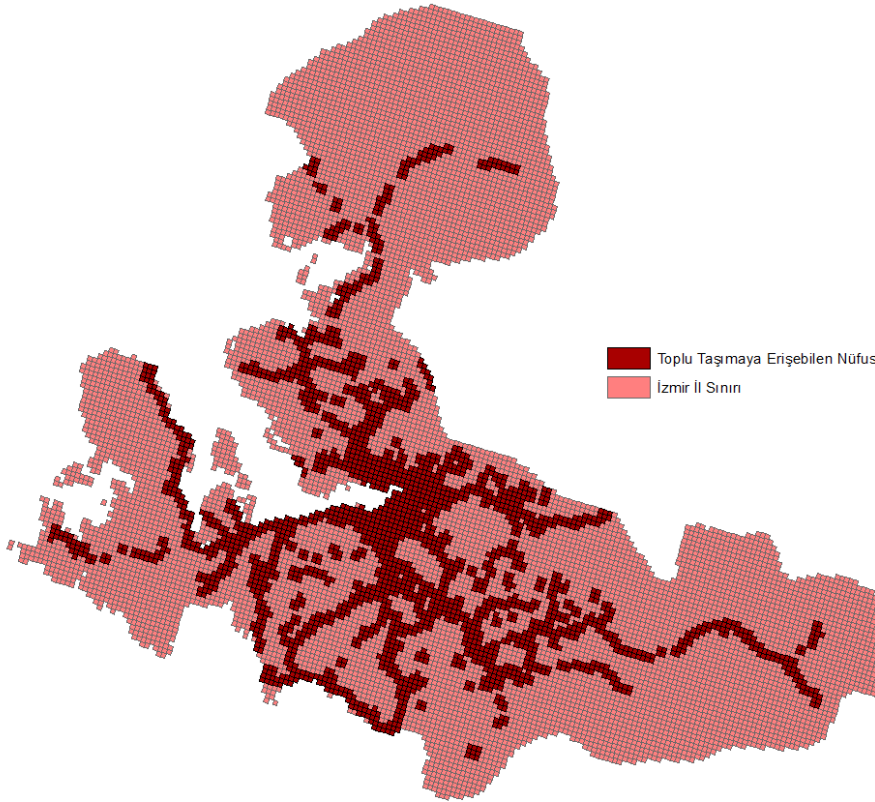
İl	Nüfus (%)
Ankara	87.9
İstanbul	98.1
İzmir	84



Şekil 3. Ankara İli için Toplu Taşıma Araçlarına Kolay Erişimi Olan Nüfus



Şekil 4. İstanbul İli için Toplu Taşıma Araçlarına Kolay Erişimi Olan Nüfus



Şekil 5. İzmir İli için Toplu Taşıma Araçlarına Kolay Erişimi Olan Nüfus

Gösterge 11.7.1 için elde edilen sonuçlar Şekil 6'da sunulmuştur. Ülke bazında elde edilen toplam kentsel alan ve kamu kullanımına açık alanlar oranlanarak yapılan hesapta Türkiye ortalaması %13.56 olarak bulunmuştur. Açık alan oranı en düşük olan iller sırasıyla Bingöl, Tunceli, Erzurum, en yüksek olan iller ise İstanbul, Trabzon, Ordu'dur. Gösterge değeri il kırılımında yapılan hesaplama sonuçları ile birlikte Tablo 4'te sunulmuştur. Gösterge hesaplanırken

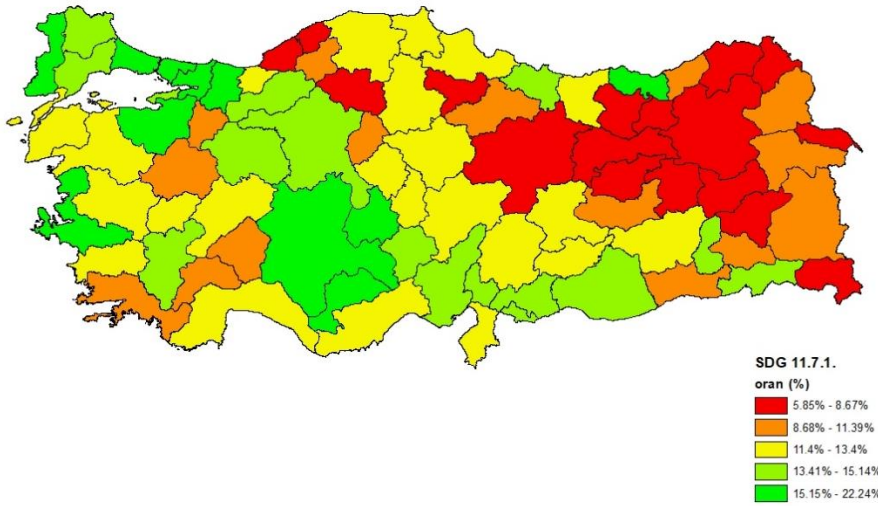
kent alanlarının ve kamusal alanların elde edildiği veri kaynaklarının zamansal değişkenlik göstermesi ve açık kaynak verileri doğruluklarının şehirlerin gelişmişlik düzeylerine göre farklılık göstermesi hesaplama sonucunda çıkan oranları etkileyebileceği değerlendirilmektedir. 2020 il bazında Adrese Dayalı Nüfus Kayıt Sistemi (ADNKS) yılsonu nüfusları ile Gösterge 11.7.1 sonuçları arasındaki ilişkinin gücü korelasyon analizi ile incelenmiştir. SPSS yazılımı kullanılarak Pearson korelasyon katsayısı (r) 0.42; p ise 0.0001 olarak hesaplanmıştır. Her iki değişken arasında güçlü olmasa da pozitif ve anlamlı bir ilişki tespit edilmiştir. Bu korelasyon, kitle-kaynaklı veri kalitesinin yüksek nüfuslu yerleşim yerlerinde artması şeklinde yorumlanabileceği gibi, düşük nüfuslu yerleşim yerlerinde, açık alanların içinde yer alan, otoyol, bulvar, cadde gibi öğelerin daha seyrek olmasıyla da açıklanabilir. Yazarlar, her iki durumda bu korelasyonun oluşmasında etkili olduğunu değerlendirmektedir.

Türkiye için elde edilen Gösterge 11.7.1'in hesaplandığı çeşitli ülkelerin sonuçları da incelenmiştir. Buna göre Almanya'nın %31.78 (Almanya İstatistik Ofisi, 2018); Norveç'in %59 (Norveç İstatistik Kurumu, 2018); İngiltere'nin %4.9 (Birleşik Krallık SKA, 2019); İskoçya'nın %5.3 (Birleşik Krallık SKA, 2019); Galler'in %3.4 (Birleşik Krallık SKA, 2019); Danimarka'nın %26.1 (Danimarka İstatistik Kurumu, 2020) olarak hesapladığı görülmüştür.

Tablo 4. Kamu Kullanımına Tamamen Açık, Yapılaşmış Alanların Şehirlerdeki Payı

İl Adı	%	İl Adı	%
ADANA	14.53	KONYA	14.55
ADIYAMAN	11.2	KÜTAHYA	9.3
AFYONKARAHİSAR	12.95	MALATYA	11.91
AĞRI	9.91	MANİSA	12.52
AMASYA	10.5	KAHRAMANMARAŞ	12.4
ANKARA	13.3	MARDİN	11.2
ANTALYA	12.66	MUĞLA	10.07
ARTVİN	8.27	MUŞ	7.61
AYDIN	11.86	NEVŞEHİR	13.7
BALIKESİR	13.07	NİĞDE	13.47
BİLECİK	10.69	ORDU	18.1
BİNGÖL	6.16	RİZE	17.87
BİTLİS	7.29	SAKARYA	15.76
BOLU	16.36	SAMSUN	13.77
BURDUR	9.64	SİİRT	10.49
BURSA	16.06	SİNOP	14.21
ÇANAKKALE	13.19	SİVAS	8.99
ÇANKIRI	8.72	TEKİRDAĞ	14.97
ÇORUM	11.37	TOKAT	10.03
DENİZLİ	13.5	TRABZON	20.32
DİYARBAKIR	12.99	TUNCELİ	6.25
EDİRNE	16.72	ŞANLIURFA	14.41
ELAZIĞ	11.13	UŞAK	12.09
ERZİNCAN	8.68	VAN	9.79
ERZURUM	6.9	YOZGAT	12.26
ESKİŞEHİR	14.51	ZONGULDAK	7.77
GAZİANTEP	14.12	AKSARAY	17.55
GİRESUN	17.53	BAYBURT	7.85
GÜMÜŞHANE	6.99	KARAMAN	15.17
HAKKÂRİ	6.94	KIRIKKALE	11.29

HATAY	12.04	BATMAN	13.2
ISPARTA	10.25	ŞIRNAK	12.92
MERSİN	13.05	BARTIN	7.5
İSTANBUL	21.05	ARDAHAN	8.3
İZMİR	16.34	IĞDIR	7.36
KARS	10.66	YALOVA	13.1
KASTAMONU	13.72	KARABÜK	12.4
KAYSERİ	12.76	KİLİS	13.56
KIRKLARELİ	14.32	OSMANİYE	14.28
KİRŞEHİR	12.57	DÜZCE	14.48
KOCAELİ	15.89	<i>Türkiye ortalaması</i>	<i>13.56</i>



Şekil 6. Kamu Kullanımına Tamamen Açık, Yapılmış Alanların Şehirlerdeki Payı

Bu çalışma ile Türkiye için daha önce üretilmemiş olan Sürdürülebilir Kalkınma Göstergelerinden ikisi üretilmiş ve kentlerdeki sürdürülebilir kalkınmanın izlenebilmesi ve planlanması yolunda bir başlangıç adımı atılmıştır. Hesaplanan göstergeleri diğer ülkeler ile karşılaştırmanın metodolojik farklılıklar nedeniyle yanıltıcı olabileceği hatırd tutularak değerlendirildiğinde Gösterge 11.2.1 için sonuçların daha yakın olduğu, Gösterge 11.7.1 için ise ülkeler arasında daha büyük farklar olduğu göze çarpmaktadır. Bunun başlıca nedeni, açık alan kavramının farklı şekillerde yorumlanmasıdır. Çalışmada üç büyük il için hesaplanan Gösterge 11.2.1'nin önümüzdeki süreçte tüm Türkiye için il kırılımında hesaplanması hedeflenmektedir. Ayrıca gösterge için yalnızca motorlu toplu taşıma alt yapısına dayandığı da dikkate alınarak bisikletle veya yürüyerek yapılabilecek kısa yolculukların elverişliliğinin de gelecek hesaplamalara dâhil edilmesi değerlendirilmelidir. Gösterge 11.7.1'in ise merkezi bir coğrafi veri kayıt sisteminden elde edilebilecek, standartlaştırılmış veriler ile hesaplanması güvenilirliğini arttıracaktır. Nihai hedef ise, ulusal düzeyde üretilecek geniş yelpazedeki gösterge setleri ile ülkemizde sürdürülebilir kalkınmanın izlenmesini ve değerlendirilmesini gerçekleştirerek hem karar alıcılar hem de toplum için katılımcı ve kapsayıcı planlama süreçlerini sağlamaktır.

Teşekkür

Yazarlar, Sn. Sevinç METİN BAŞOĞLU ve Sn. Hasan AZTOPAL'a çalışma boyunca verdikleri teknik ve manevi destek için teşekkür eder.

Kaynakça

- Almanya İstatistik Ofisi (2018). <https://sdg-indikatoren.de/en/11-7-1/> adresinden alındı
- Almanya İstatistik Ofisi (2020). <https://sdg-indikatoren.de/en/11-2-1/> adresinden alındı
- Birleşik Krallık SKA (2019). <https://sdgdata.gov.uk/11-7-1/> adresinden alındı
- Blaschke, T., & Kovács-Györi, A. (2020). Earth Observation to Substantiate the Sustainable Development Goal 11: Practical Considerations and Experiences from Austria. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLIII-B4-2020, 2020XXIV ISPRS Congress.
- Brussel, M., Zuidgeest, M., Pfeffer, K., & Maarseveen, M. v. (2019). Access or Accessibility? A Critique of the Urban Transport SDG Indicator. *ISPRS International Journal of Geo-Information* 8, 67.
- Corbane, C., Florczyk, A., Pesaresi, M., Politis, P., & Syrris, V. (2018). GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014). *R2018A. European Commission, Joint Research Centre (JRC)*, doi: 10.2905/jrc-ghsl-10007 PID: <http://data.europa.eu/89h/jrc-ghsl-10007>.
- Çetin, M. (2015). Using GIS analysis to assess urban green space in terms of accessibility: Case study in Kutahya. , *International Journal of Sustainable Development & World Ecology* Vol.22, No.5, 420-424.
- Danimarka İstatistik Kurumu (2019). <https://www.dst.dk/en/Statistik/Sdg/11-baeredygtige-byer-og-lokalsamfund/delmaal-02/indikator-1> adresinden alındı
- Danimarka İstatistik Kurumu (2020). <https://www.dst.dk/en/Statistik/Sdg/11-baeredygtige-byer-og-lokalsamfund/delmaal-07/indikator-1> adresinden alındı
- Demircan, N., & Sezen, I. (2018). Use of Green Spaces for Liveable and Sustainable Cities; Urban Allotment Gardens. *Iğdır Üni. Fen Bilimleri Enst. Der. / Iğdır Univ. J. Inst. Sci. & Tech.* 8(1), 263-270.
- Eurostat (2018). *Methodological Manual on Territorial Typologies* .
- Kharrazi, A., Qin, H., & Zhang, Y. (2016). Urban Big Data and Sustainable Development Goals: Challenges and Opportunities. *Sustainability* 8(12), 1293, DOI:10.3390/su8121293.
- Klopp, J. M., & Petretta, D. L. (2017). The urban sustainable development goal: Indicators, complexity and the politics of measuring cities. *Cities* 63 (2017), 92-97.
- Koch, F., & Krellenberg, K. (2018). How to Contextualize SDG 11? Looking at Indicators for Sustainable Urban Development in Germany. *International Journal of Geo-Information* 7, 464.
- Murray, A. T., Davis, R., Stimson, R., & Ferreira, L. (1998). Public Transportation Access. *Transpn Res.-D, Vol. 3, No. 5*, 319-328.
- Norveç İstatistik Kurumu (2018). <https://www.ssb.no/sdg-en> adresinden alındı
- Norveç İstatistik Kurumu (2019). <https://www.ssb.no/sdg-en> adresinden alındı
- OpenStreetMap contributors (2020). Türkiye verisi <https://geofabrik.de> adresinden indirildi. . <https://www.openstreetmap.org> adresinden alındı
- Permai, S., Mukhaiyar, U., Satyaning, N., Soleh, M., & Aini, Q. (2018). Spatial weighting approach in numerical method for disaggregation of MDGs indicators. *IOP Conf. Series: Materials Science and Engineering* 332 (2018) 012049.
- Peşkirioğlu, N. (2016). 2030 Sürdürülebilir Kalkınma Hedefleri: Küresel Verimlilik Hareketine Doğru. *Anahtar Dergisi*, Kasım 2016.
- Sachs, D. (2012). From Millennium Development Goals to Sustainable Development Goals. *Lancet*, 2012; 379:2206–11.

Tiwari, G., & Philip, C. (2021). Development of public transport systems in small cities: A roadmap for achieving sustainable development goal indicator 11.2. *IATSS Research* 45 (2021) 31–38.

TUCBS Ulusal Coğrafi Bilgi Platformu (2021, 09 10). *ATLAS*: <https://atlas.gov.tr> adresinden alındı

TÜİK (2021). Sürdürülebilir Kalkınma Göstergeleri 2010-2019 Haber Bülteni, <https://sdg.tuik.gov.tr>.

UN-Habitat (2021). United Nations Human Settlements Programme SDG11.7.1 Metadata, <https://unstats.un.org/sdgs/metadata/files/Metadata-11-07-01.pdf>.

UN-Habitat (2021). United Nations Human Settlements Programme SDG 11.2.1 Metadata, <https://unstats.un.org/sdgs/metadata/files/Metadata-11-02-01.pdf>.

United Nations (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development*.

UN-WCED (1987). Report of the World Commission on Environment and Development: Our Common Future. <https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf>.

Wang, P., Huang, C., Brown de Colstoun, E. C., Tilton, J. C., & Tan, B. (2017). Documentation for the global human built-up and settlement extent (HBASE) dataset from Landsat. *Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC)*, doi: <https://doi.org/10.7927/H4DN434S>.

Xiang, N. (2017). A Review on the research and practice of city sustainable development indicators and indices. *IOP Conf. Series: Earth and Environmental Science*, 88 (2017) 012024.

Yeni, O. (2014). Sürdürülebilirlik ve Sürdürülebilir Kalkınma: Bir Yazın Taraması. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* 16/3, 181-208.



ARAŞTIRMA MAKALESİ

RESEARCH ARTICLE

Tabaka Sınırlarının Belirlenmesinde, Kümeleme Analiz Yöntemleri ve Tabaka Sınırı Belirleme Yöntemlerinin Karşılaştırılması¹

Osman SERT

Türkiye İstatistik Kurumu / TÜİK Uzmanı

osman.sert@tuik.gov.tr

Orcid No: 0000-0002-0928-851X

Alpaslan AKÇORAOĞLU

Hacı Bayram Veli Üniversitesi / Prof. Dr.

alpaslan.akcoraoglu@hbv.edu.tr

Orcid No: 0000-0003-4368-9998

Özet

Tabakalı örneklemede örnekleme çerçevesi birbiriyle örtüşmeyen tabakalara bölünür. Bu bölünme çoğu pratik durumda coğrafi bölgeler, cinsiyet, yaş gibi doğal durumu yansıtacak şekilde kendiliğinden oluşur. Bu şekilde kendiliğinden oluşan tabakalar araştırma değişkenine göre içsel olarak homojen olmayabilir. Fakat araştırma değişkeninin tabaka sınırları katı bir şekilde önceden belirlenmemişse tabakaların içsel olarak homojenliği sağlanabilir. Bu yolla araştırma değişkenine göre içsel olarak homojen tabakalar oluşturulur ve tahmin hassasiyeti en üst düzeye çıkarılmış olur. Bu amaçla bu çalışmada tabaka-içi homojenliği sağlamak için Kümeleme Analiz Yöntemleri ve Tabaka Sınırı Belirleme Yöntemleri karşılaştırılmıştır. Bu karşılaştırma için beş farklı çarpıklık değerine sahip veri setleri türetilmiş ve her veri seti için ayrıştırma yöntemlerine göre tabaka sınırları belirlenmiştir. Tahminlerin güvenilirliğini arttırmak için her yöntemden 1000 kez bağımsız örnek seçilmiştir. Böylece her yöntemden elde edilen ortalama tahmin edicisine ilişkin Kök Hata Kareler Ortalamaları (KHKO) hesaplanmış ve en küçük KHKO değerini veren yöntemin tabaka sınırlarının optimum sınırlar olduğu sonucuna ulaşılmıştır. Analizler R programında yer alan “NbClust” ve “Stratification” paketleri ile yapılmıştır. Çalışmada elde edilen sonuçlara göre en küçük üç çarpıklık değerine sahip simülasyonlar için “Lavallee-Hidiroglou”, dördüncü simülasyon için “Ortalama Kümeleme” ve en büyük çarpıklık değerine sahip simülasyon için ise “K-Ortalamalar Kümeleme” yöntemleri ile elde edilen tabaka sınırları, optimum tabaka sınırları olarak belirlenmiştir.

Anahtar Sözcükler: Kümeleme Analizi, Optimum Tabaka Sınırları, Tabakalı Tesadüfi Örnekleme

¹ Bu çalışma ilk yazarın doktora tezinden türetilmiştir. Çalışmada ifade edilen görüşler tamamen yazarlara aittir ve Türkiye İstatistik Kurumu'nu bağlamaz.

Sorumlu Yazar / Corresponding Author: 1-Osman SERT, Türkiye İstatistik Kurumu, Yöntem Araştırmaları Daire Başkanlığı, Veri Analiz Teknikleri Grup Başkanlığı.

2-Alpaslan AKÇORAOĞLU, Hacı Bayram Veli Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü.

Atıf / Citation: SERT O., AKÇORAOĞLU A. (2022). Tabaka Sınırlarının Belirlenmesinde, Kümeleme Analiz Yöntemleri ve Tabaka Sınırı Belirleme Yöntemlerinin Karşılaştırılması. İstatistik Araştırma Dergisi, 12 (1), 68-81.

Comparison of Cluster Analysis Methods and Strata Boundary Determination Methods for Determination of Strata Boundaries

Abstract

In stratified sampling, the sampling frame is divided into non-overlapping strata. In most practical cases, this division occurs spontaneously, reflecting the natural state such as geographical regions, gender and age. Self-forming strata formed in this way may not be internally homogeneous according to the research variable. However, internal homogeneity of the strata can be achieved if the strata boundaries of the research variable are not strictly predetermined. Thus, internally homogeneous strata are formed according to the research variable and precision of an estimate is maximized. For this purpose, in this study, Cluster Analysis Methods and Optimum Strata Boundary Determination Methods were compared to ensure intra-stratum homogeneity. For this comparison, data sets with five different skewness values were derived. For each data set, strata boundaries were determined according to the decomposition methods. In order to increase the reliability of the estimations, independent samples were selected from each method for 1000 times. Thus, the Root Mean Squares Error (RMSE) of the mean estimator obtained from each method was calculated and it was concluded that the method with the smallest RMSE value had the optimum strata boundaries. Analyzes were made with the "NbClust" and "Stratification" packages in the R program. According to the results obtained in the study the strata boundaries obtained by the "Lavallee-Hidiroglou" Method for the simulations with the 3 lowest skewness values; "Average Clustering" Method for the fourth simulation; and "K-Means Clustering" Method for the simulation with the largest skewness value were determined as the optimum strata boundaries.

Keywords: Cluster Analysis, Optimum Strata Boundaries, Stratified Random Sampling

1. Giriş

Tabakalı Tesadüfi Örnekleme Yöntemi; işsizlik oranı, gelir eşitsizliği, genel mutluluk düzeyi, turizm istatistikleri, mali aracı kuruluş istatistikleri ve girişimlerde bilişim teknolojileri kullanım istatistiklerinin hesaplanması gibi ülkelerin istatistik kurumları tarafından yürütülen birçok parametre tahmininde kullanılan önemli bir örnekleme tekniğidir. Tabakalı örneklemede örnekleme çerçevesi, örtüşmeyen gruplara veya tabakalara bölünür. Oluşturulan tabakaların araştırma değişkenine göre içsel olarak homojen olması hedeflenir ve böylece tahminin hassasiyeti en üst düzeye çıkarılmış olur. Fakat çoğu pratik durumda, bu gibi optimum tabakaları oluşturmak çok zordur. Çünkü araştırmacılar coğrafi bölgeler (örn; Batı Karadeniz, Orta Anadolu, vb.), idari bölgeler (örn; iller, ilçeler, vb.), cinsiyet, yaş veya diğer doğal ölçütler (örn; kentsel-kırsal alan) gibi doğal durumu yansıtacak şekilde topluluğu tabakalaştırır (Khan, Reddy, & Rao, 2015). Fakat kullanılan değişkenin tabaka sınırları katı bir şekilde önceden belirlenmemişse, (örn; firmaların ciro değerleri ya da çalışan sayıları gibi) tabakaların içsel olarak homojenliği sağlanabilir. Bu nedenle, tahminlerin hassasiyetini arttırmak için Optimum Tabaka Sınırları (OTS) belirlenmelidir. OTS'nin belirlenmesinde temel husus tabakaların mümkün olduğunca içsel olarak homojen olması, yani tabaka varyanslarının belirli bir örnek dağılımı için mümkün olduğunca küçük olmasıdır.

Bu amaçla bu çalışmada sırasıyla Tabaka Sınırı Belirleme Yöntemleri ve Kümeleme Analiz Yöntemlerine ait literatür bilgisi verilmiş, sonrasında türetilen veri setleri üzerinde uygulama yapılmış ve sonuçlar sunulmuştur.

2. Tabaka Sınırı Belirleme Yöntemleri

Tabakalı Örnekleme Yönteminde oluşturulan tabakaların içsel olarak homojen olmaması sorunu birçok araştırmacı tarafından incelenmiş ve tahminlerin hassasiyetini arttıran Tabaka Sınırları Belirleme Yöntemleri geliştirilmiştir. Literatürde değişkenlerin tabaka sınırlarını belirlemek için birçok çalışma yapılmıştır. Genel olarak bu yöntemler yakınsama ve optimizasyon bazlı olmak üzere iki başlıkta toplanır (Hidiroglou & Kozak, 2017). Yakınsama bazlı yöntemlerden en yaygın olanları Dalenius ve Hodges (1957) tarafından önerilen Birikimli Kök Frekans Yöntemi ile Gunning ve Horgan (2004) tarafından önerilen Geometrik Tabakalama Yöntemidir. Mahalanobis (1952), Ekman (1959), Sethi (1963), Singh (1971) ve Thomsen (1976) çalışmaları ise diğer yakınsama örnekleri arasında gösterilebilir. Optimizasyon Bazlı Tabakalama Yöntemlerinin en yaygın kullanılanları ise Lavallee ve Hidiroglou (1988) ve Kozak (2004) tarafından ortaya atılan yöntemlerdir. Diğer yöntemler ise Sweet ve Sigman (1995), Rivest

(2002), Lednicki ve Wieczorkowski (2003), Kesintürk ve Er (2007), Verma, Kozak ve Zielinski (2007), Benedetti, Bee ve Espa (2010), Brito, Ochi, Montenegro ve Maculan (2010) ve Ballin ve Barcaroli (2013) tarafından önerilmiştir. Bu çalışmada literatürde en çok karşılaşılan ve karşılaştırılan aşağıdaki 3 yöntem kullanılmıştır.

2.1. Birikimli Kök Frekans Yöntemi

Birikimli Kök Frekans Yöntemi en popüler tabaka sınırı belirleme yöntemlerinden biridir. Bu yöntemde ilgilenilen değişkenin toplam tahmin varyansı en aza indirgenerek tabaka sınırları belirlenir. Bu işlemde tabakalama değişkeninin tabakalama noktaları arasında yaklaşık olarak eşit dağıldığı varsayılır. Sınırlar, frekansların karekökünün kümülatif fonksiyonu üzerinde eşit aralıklar alınarak elde edilir (Dalenius & Hodges, 1957). Sınırları belirlemek için, aşağıdaki denklemlerin yinelemeli olarak çözülmesi gerekir:

$$\frac{S_h^2 + (\mu_h - \bar{X}_h)^2}{S_h} = \frac{S_{h+1}^2 + (\mu_{h+1} - \bar{X}_h)^2}{S_{h+1}}, \quad h = 1, 2, \dots, L - 1 \quad (2.1)$$

Burada h tabaka indisi, L tabaka sayısı, μ_h kitle ortalaması, \bar{X}_h h tabakasının kitle ortalaması ve S_h^2 h tabakasındaki tabakalama değişkeninin kitle varyansdır.

Yöntemin işlem adımları aşağıda verilmiştir:

1. Veri küçükten büyüğe sıralanır,
2. Veri sınıflara ayrılır. Sınıf sayısı (J) tabaka sayısından fazla olmalıdır ($J > L$),
3. Her sınıftaki frekans hesaplanır ($f_i, i = 1, 2, \dots, J$),
4. Her sınıfın frekansının karekökü hesaplanır,
5. Frekansların kareköklerinin toplamı hesaplanır ($\sum_{i=1}^J \sqrt{f_i}$),
6. Frekansların kareköklerinin toplamı, tabaka sayısına bölünür ($Q = \frac{1}{L} \sum_{i=1}^J \sqrt{f_i}$),
7. Hesaplanan Q değeri her tabakanın üst sınırı olur ($Q, 2Q, \dots, (L - 1)Q, LQ$).

2.2. Geometrik Yöntem

Bu tabaka sınırı belirleme yöntemi, ilgilenilen yardımcı değişkenin değişim katsayısını her tabakada eşitleyerek tabaka sınırlarının belirlenmesi fikrine dayanmaktadır. Sadece ilgili değişken x ve tabaka sayısı (L) gerektiren sınırların hesaplanması, geometrik bir ilerlemeye dayanır (Gunning & Horgan, 2004).

$$b_h = m \left(\frac{M}{m} \right)^{\frac{h}{L}}, \quad h = 0, 1, \dots, L - 1 \quad (2.2)$$

Bu sınırlar, m ve M 'nin sırasıyla x 'lerin minimum ve maksimum değerleri olduğu yerlerdir. Ortaya çıkan tabakalar $[b_h; b_h + 1)$ için $h = 0, 1, \dots, L - 1$ şeklinde gösterilebilir. Örnek olarak; tabaka sayısı 4 ($L=4$), kitledeki en küçük değer 5 ($m=5$) ve en büyük değer 50000 ($M=50000$) ise tabaka sınırları;

$$b_h = m \left(\frac{M}{m} \right)^{\frac{h}{L}} = 5 \left(\left(\frac{50000}{5} \right)^{1/4} \right)^h = 5 * 10^h, \quad (h = 0, 1, 2, 3, 4)$$

ile hesaplanır. Buna göre tabaka sınırları sırasıyla 5-50;51-500;501-5000;5001-50000 olacaktır.

Gunning ve Horgan (2004) algoritmanın normal dağılımlar için çalışmadığına dikkat çekmiştir. Daha genel olarak, yöntem herhangi bir simetrik dağılım için çalışmamaktadır. Bununla birlikte sınırlar geometrik olarak arttığından, minimum değeri sıfıra yakın değişkenler için yöntemin iyi çalışmayacağını da belirtmişlerdir. Ayrıca en yüksek değerlerin nadir olduğu yüksek derecede sağa çarpık dağılımlar için en iyi sonuçları elde etmeyi beklemişlerdir.

2.3. Lavallee-Hidiroglou Yöntemi

Lavallee ve Hidiroglou (1988) çalışmasında önerilen tabakalama yöntemi, kitlenin tam-sayım ve örnekleme tabakalarına ayrıştırılması özelliğiyle daha önce tanımlanmış olan yöntemlerden farklılık göstermektedir. Tam-sayım tabakasında tabakadaki tüm birimler kesin olarak seçilirken, örnekleme tabakalarında birimler yerine

koymadan basit tesadüfi örnekleme yöntemi kullanılarak seçilir. Yöntemin amacı, tahmin edicinin verilen değişim katsayısına (CV) ve örnek dağıtım şekline göre örnek hacmini minimize edecek tabaka sınırlarını hesaplamaktır (Lavallee & Hidirolou, 1988). Buradaki problem, verilen örnek dağıtım yöntemini kullanarak istenen değişim katsayısı kısıtında minimum örnek hacmini elde edecek tabaka sınırlarının $y_1 < b_1 < b_2 < \dots < b_{L-1} < y_N$ hesaplanmasıdır (1. tabaka sınırı: $y_1 \dots b_1$, 2. tabaka sınırı: $b_1 + 1 \dots b_2$ olacak şekilde). Bu yöntemde toplam örnek hacmi (n) aşağıdaki formülle tabakalara dağıtılır.

$$n = N_L + \frac{N(\sum_{h=1}^{L-1} W_h^2 S_h^2 / \hat{A}_h)}{N(c\bar{X})^2 + \sum_{h=1}^{L-1} W_h S_h^2} \quad (2.3)$$

Burada;

\hat{A}_h ; örnek dağıtım yöntemi,

W_h ; örnekleme ağırlığı (N_h/N),

N_L ; tam-sayım tabakasına dağıtılan birimlerin sayısı,

c ; güven seviyesi,

\bar{X} ; kitle ortalaması,

S_h^2 ; h tabakasının kitle varyansı olarak ifade edilmiştir.

\hat{A}_h formülü ise aşağıdaki gibidir:

$$\hat{A}_h = \frac{N_h^{2q_1} \bar{X}_h^{2q_2} S_h^{2q_3}}{\sum_{h=1}^L N_h^{2q_1} \bar{X}_h^{2q_2} S_h^{2q_3}} \quad (2.4)$$

Burada, $0 \leq 2q_1 \leq 1$, $0 \leq 2q_2 \leq 1$ ve $0 \leq 2q_3 \leq 1$ 'tür. Bu genel formül bilinen bütün dağıtım yöntemlerini göstermektedir. Aşağıdaki tabloda hangi kombinasyonun hangi dağıtım yöntemini temsil ettiği ve istenen dağıtım yöntemini seçmek için gerekli q değerleri gösterilmiştir (Lavallee & Hidirolou, 1988).

Tablo 1. \hat{A}_h ile Dağıtım Yöntemleri Arasındaki İlişki

Dağıtım Yöntemi	$2q_1$	$2q_2$	$2q_3$
Neyman Dağıtım	1	0	1
X-Oransal Dağıtım	1	1	0
N-Oransal Dağıtım	1	0	0
Güç Dağıtım	p	p	0

Buna göre algoritmanın işleyişi aşağıdaki adımlarda verilmiştir:

1. Kitle küçükten büyüğe sıralanır,
2. Öncelikle geçici tabaka sınırları belirlenir $b_0 < b'_1 < b'_2 < \dots < b'_{L-1} < b_L$,
3. Bu tabaka sınırları için tek tek tabaka ağırlığı, ortalaması ve varyansı hesaplanır,
4. Başlangıç tabaka sınırları aşağıdaki formülle yeni hesaplananlarla değiştirilir (b''_1, \dots, b''_{L-1}),

$$b''_h = \frac{\beta'_{h+1} + \sqrt{\beta'_{h+1}{}^2 - 4\alpha'_h \gamma'_h}}{2\alpha'_h}, h = 1, 2, \dots, L-1 \quad (2.5)$$

Burada yer alan α, β, γ terimleri, örnek hacmi hesabı için oluşturulan formüldeki 2. dereceden reel kök hesabındaki katsayılardır. Bu hesap $\alpha_h b_h^2 + \beta_h b_h + \gamma_h = 0$ şeklinde gösterilebilir,

5. Ardışık iki hesaplama aynı olana veya ihmal edilebilir miktarlarda farklılık gösterene kadar 3. ve 4. adım tekrarlanır.

3. Kümeleme Analiz Yöntemleri

Kümeleme Analizi temel olarak benzer birimlerin bir araya getirilmesi yaklaşımını kullanmaktadır. Kümeleme Analizi, gruplanmamış verileri benzerliklerine göre gruplandırarak araştırmacıya özetleyici bilgiler sunar. Kümeleme analizindeki amaç, benzer olarak bir araya getirilen birimlerden oluşan kümelerde; küme içi değişimin minimum, kümeler arası değişimin ise maksimum olmasını sağlamaktır. Literatürde birçok kümeleme analiz yöntemi bulunmaktadır ve genel olarak hiyerarşik (aşamalı) ve hiyerarşik olmayan (aşamalı olmayan) yöntemler biçiminde ikiye ayrılır.

3.1. Hiyerarşik Kümeleme Yöntemleri

Hiyerarşik Kümeleme Yönteminde verideki n sayıda gözlemin her biri bir küme olarak düşünülür. Bu kümelere birleştirme işlemleri uygulanır ve sonuçta tek bir küme oluşana kadar bu işlemler devam eder. Hiyerarşik kümeleme yönteminin genel adımları aşağıda verilmiştir (Servi, 2009):

1. Küme sayısı n olarak alınır. K_1, K_2, \dots, K_n ile ifade edilen bu n sayıda küme için benzerlik matrisi hesaplanır.
2. Benzerlik matrisinde n küme sayısı, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, n$ ve $i \neq j$ olmak üzere tüm $\min(d(K_i, K_j))$ uzaklığına sahip iki küme belirlenir. Bu iki küme birleştirilerek yeni bir küme oluşturulur.
3. Benzerlik matrisi yeni oluşan küme de dikkate alınarak güncellenir.
4. Tek bir küme elde edilinceye kadar 2. ve 3. adımlar tekrarlanır.

Hiyerarşik kümelemede benzerlik matrisinin hesaplanması tanımlanan uzaklık fonksiyonuna bağlıdır. Literatürde birçok uzaklık fonksiyonu tanımı vardır. Uzaklık ölçüleri arasındaki fark bu çalışmanın konusu olmadığı için bu çalışmada uzaklık ölçüsü olarak sıklıkla kullanılan “Öklid Uzaklığı” kullanılmıştır. Öklid uzaklığı, gözlem vektörleri arasındaki farkların kareleri toplamının karekökünün alınması ile $d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$ formülü ile hesaplanır (Erişoğlu, 2011). Bu formülde p değişken sayısıdır.

Bu çalışmada hiyerarşik kümeleme analiz yöntemlerinden aşağıdaki 6 tanesi kullanılacaktır:

i. Tek Bağlantı Kümeleme Yöntemi: Florek, Lukaszewicz, Steinhaus ve Zubrzycki (1951) çalışmasında ilk kez önerilen bu yöntemde iki farklı kümedeki en yakın gözlemlere göre belirlenen uzaklık değeri, iki küme arasındaki uzaklık olarak tanımlanır. Tek bağlantı kümeleme aynı zamanda “en yakın komşuluk” olarak da adlandırılır. Bu yöntemde x biriminin y ve z birimlerinden oluşan yeni kümeye olan uzaklığı,

$$d(x, \{y, z\}) = \min(d(x, y), d(x, z)) \quad (3.1)$$

eşitliği ile hesaplanır.

ii. Tam Bağlantı Kümeleme Yöntemi: ilk olarak Sorensen (1948) tarafından önerilen bu yöntemde iki farklı kümedeki birbirine en uzak gözlem çifti arasındaki uzaklık değeri, iki küme arasındaki uzaklık olarak alınır. Tam bağlantı kümeleme yöntemi aynı zamanda “en uzak komşuluk” olarak da tanımlanır. Bu yöntemde x biriminin y ve z birimlerinden oluşan yeni kümeye olan uzaklığı,

$$d(x, \{y, z\}) = \max(d(x, y), d(x, z)) \quad (3.2)$$

eşitliği ile hesaplanır.

iii. Ortalama Bağlantı Kümeleme Yöntemi: Ayrı gruplarda yer alan gözlem çiftleri arasındaki ortalama uzaklık iki küme arasındaki uzaklık olarak alınır. Her biri sırasıyla n ve m adet birimden oluşan K_i ve K_j kümeleri arasındaki uzaklık

$$(K_i, K_j) = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m d(x_i, x_j) \quad (3.3)$$

eşitliği ile tanımlanır.

iv. Merkezi Bağlantı Kümeleme Yöntemi: İki küme arasındaki uzaklık kümelerin kendi merkezleri arasındaki uzaklık olarak alınır. K_i kümesinin merkezi \bar{x} ve K_j kümesinin merkezi \bar{y} olarak alındığında K_i kümesinin K_j kümesine olan uzaklığı merkez bağlantı yöntemine göre

$$d(K_i, K_j) = d(\bar{x}, \bar{y}) \quad (3.4)$$

şeklinde hesaplanır.

v. Medyan Bağlantı Kümeleme Yöntemi: Merkez bağlantı kümeleme tekniğinde iki küme bağlandığında yeni oluşan kümenin merkezi, bağlanan kümelerden birim sayısı fazla olan kümenin merkezine yakın olmaktadır (Erişoğlu, 2011). Bu durumu ortadan kaldırmak için medyan bağlantı tekniği önerilmiştir. Bu yönteme göre K_i kümesinin, K_j ve K_l kümelerinin bağlanması ile elde edilen $\{K_j \cup K_l\}$ kümesine olan uzaklığı

$$d(K_i, (K_j \cup K_l)) = \frac{1}{2}d(K_i, K_j) + \frac{1}{2}d(K_i, K_l) - \frac{1}{4}d(K_j, K_l) \quad (3.5)$$

eşitliği ile hesaplanır.

vi. Ward Bağlantı Kümeleme Yöntemi: Artırmalı (incremental) hata kareler toplamı yöntemi olarak da adlandırılan Ward'ın hiyerarşik kümeleme yönteminde amaç, artan sınıf içi hata kareler toplamının minimum yapılmasıdır. Bu yönteme göre K_i kümesinin, K_j ve K_l kümelerinin bağlanması ile elde edilen $\{K_j \cup K_l\}$ kümesine olan uzaklığı

$$d(K_i, (K_j \cup K_l)) = \frac{(n_i+n_j)d(K_i,K_j)+(n_i+n_l)d(K_i,K_l)-n_jd(K_j,K_l)}{n_i+n_j+n_l} \quad (3.6)$$

şeklinde hesaplanır (Ward, 1963).

3.2. Hiyerarşik Olmayan Kümeleme Yöntemi (K-Ortalamalar Kümeleme Yöntemi)

K-Ortalamalar Kümeleme Yöntemi kümeleme analizinde en yaygın olarak kullanılan bir kümeleme algoritmasıdır. K-Ortalamalar Kümeleme Yöntemi, hata kareler toplamının minimum yapılmasına dayalı olarak verideki en uygun parçalanmayı bulmayı amaçlayan adimsal optimizasyon yöntemidir (Erişoğlu, 2011). Bu yöntem, Hiyerarşik Kümeleme Yöntemlerinin aksine ilk aşamada oluşturulacak küme sayısının belirlenmesi ile başlar. Bu sebeple verinin parçalanması ve bu parçalanmaya göre küme merkezlerinin oluşturulması kümeleme sonucuna oldukça etki eder. Bu etkinin azaltılması için başlangıçta verinin nasıl parçalanacağına ilişkin genel bir yöntem yoktur. Bu amaçla Hiyerarşik Kümeleme Yöntemleri ile küme sayısına karar verip, K-Ortalamalar Kümeleme Yöntemi daha sonra uygulanabilir. Bu yöntemdeki aşamalar aşağıdaki gibi sıralanabilir:

1. Veri hakkında bilgi varsa bu bilgi ışığında yoksa rassal olarak başlangıç küme merkezleri seçilir. Temelde bu merkezler birbirinden uzak olan bir dizi gözlemdir.
2. Her birim merkeze olan uzaklığına göre tanımlanmış olan en yakın kümeye atanır.
3. Bu atama sonucu yeni oluşan küme merkezleri bulunur.
4. Her birimden her bir merkeze olan uzaklık tekrar hesaplanır ve kümede olmayan gözlemler en yakın kümelere taşınır.
5. Küme merkezleri nispeten kararlı kalana kadar bu süreç devam ettirilir.

4. Bulgular

Ekonomik ve işyeri araştırmalarının çoğunda, hedef değişkenlerin dağılımları (örnek: İşletmelerin cirosu, hane halkı gelirleri, vb.) yaygın olarak birçok küçük ve birkaç büyük birime sahip çarpık dağılımlara benzemektedir (Khan, Reddy, & Rao, 2015). Bu sebeple bu çalışmada gerçek bir duruma örnek olması açısından farklı çarpıklık değerlerine sahip beş farklı kitle türetilmiş ve tabaka sınırlarını belirlemek ve tahminleri en iyileştirmek için yukarıda açıklanan Kümeleme Analiz Yöntemleri ve Tabaka Sınırı Belirleme Yöntemleri karşılaştırılmıştır. Bu karşılaştırma yapılırken 7 farklı kümeleme analizi ve 3 farklı tabaka sınırı belirleme yöntemi olmak üzere, 10 farklı ayrıştırma yöntemi kullanılmıştır. Bu ayrıştırma yöntemleriyle elde edilen tabakalar belirlendikten sonra oransal dağıtım yöntemiyle örnek birimler tabakalara dağıtılmıştır. Tahminlerin güvenilirliğini arttırmak için her yöntemden 1000 defa örnek seçilmiştir. Böylece her yöntemden elde edilen ortalama tahmin edicisine ilişkin Kök Hata Kareler Ortalamaları (KHKO) elde edilmiş ve en küçük KHKO değerini veren yöntemin tabaka sınırlarının optimum sınırlar olduğu sonucuna ulaşılmıştır.

Bu bölümde aşağıdaki notasyonlar kullanılmıştır:

$$k = 1, \dots, 10$$

$$k : \text{Ayrıştırma Yöntem İndisi}$$

$$j = 1, \dots, 1000$$

$$j : \text{Örnek Sayısı}$$

$$h = 1, \dots, H$$

$$h : \text{Tabaka Sayısı}$$

$$i = 1, \dots, n_{kh}$$

$$n_{kh} : k. \text{ Yöntem ve } h. \text{ tabaka için örnek hacmi}$$

x_{kjhi} : k. ayırıştırma yöntemi, j. tekrar ve h. tabakadaki i. birimin değeri

$$\bar{x}_{kj} = \frac{1}{N} \sum_{h=1}^{H_k} N_{kh} \frac{\sum_{i=1}^{n_{kh}} x_{kjhi}}{n_{kh}} \quad : k. Yöntem ve j. tekrar için ortalama tahmin edicisi (TE)$$

$$\bar{x}_k = \frac{\sum_{j=1}^{1000} \bar{x}_{kj}}{1000} \quad : k. Yöntem için ortalama tahmin edicisi$$

$$V(\bar{x}_k) = \frac{\sum_{j=1}^{1000} (\bar{x}_{kj} - \bar{x}_k)^2}{1000} \quad : k. Yöntem için ortalamanın varyansı$$

$$\sqrt{V(\bar{x}_k)} = \sqrt{\frac{\sum_{j=1}^J (\bar{x}_{kj} - \bar{x}_k)^2}{J}} \quad : k. Yöntem için ort. TE standart hatası$$

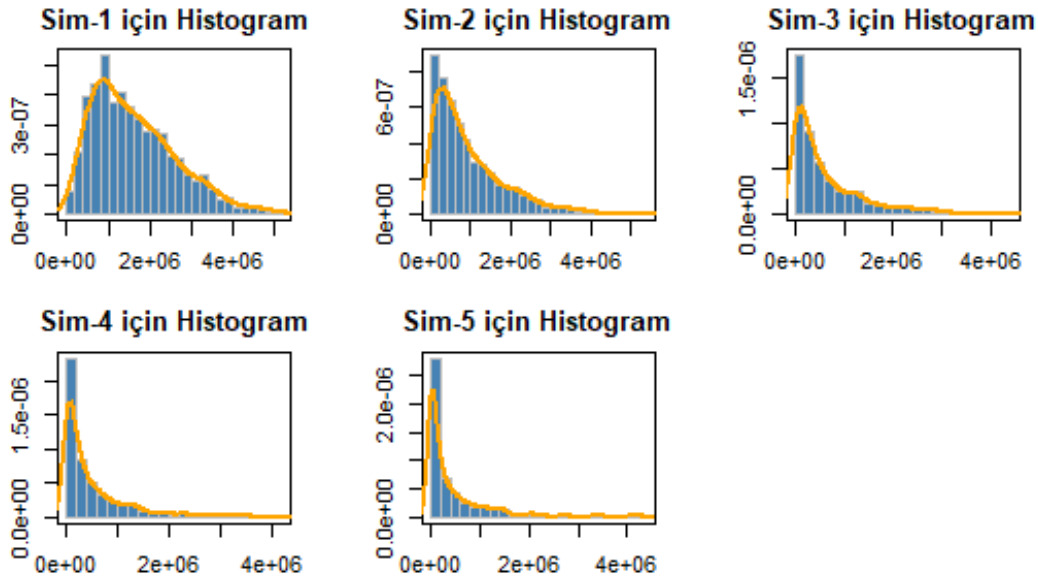
$$KHKO_k = \sqrt{V(\bar{x}_k) + (\bar{x}_k - \bar{X})^2} \quad : k. Yöntem için KHKO$$

KHKO formülünde yer alan \bar{X} kitle ortalamasıdır. Uygulamada k indisi ile gösterilen 10 ayırıştırma yönteminin 7'si kümeleme analizi 3'ü tabaka sınırı belirleme yönteminden oluşmaktadır ve aşağıdaki listede sıralanmıştır:

1. K-Ortalamalar Kümeleme Yöntemi (K-Ort)
2. Medyan Bağlantı Kümeleme Yöntemi (Medyan)
3. Merkezi Bağlantı Kümeleme Yöntemi (Merkezi)
4. Ortalama Bağlantı Kümeleme Yöntemi (Ortalama)
5. Tam Bağlantı Kümeleme Yöntemi (Tam)
6. Tek Bağlantı Kümeleme Yöntemi (Tek)
7. Ward Bağlantı Kümeleme Yöntemi (Ward)
8. Birikimli Kök Frekans Yöntemi (Cumrootf)
9. Geometrik Yöntem (Geo)
10. Lavalley-Hidiroglou Yöntemi (LH)

Uygulamada j indisi ile gösterilen tekrar sayısı örnek tahminlerinin hassasiyetini arttırmak için önsel olarak 1000 ile sınırlandırılmıştır. 100, 500 ve 1000 tekrarlar denemeler yapılmış ve 1000 tekrarın yeterli olduğuna karar verilmiştir.

Veri setleri R paket programında standart tanımı $rbeta(n, shape1, shape2, ncp=0)$ şeklinde olan $rbeta$ formülü ile yaratılmıştır. Bu formül beta dağılımına sahip n birimden oluşan rassal veri yaratmak için kullanılır. Beta dağılımı $[0,1]$ aralığında iki tane pozitif şekil parametresi ($shape1$ ve $shape2$) ile ifade edilmiş bir sürekli olasılık dağılımı ailesidir (Kennedy, 1988). Bu formülde gözlem sayısı " n ", beta dağılım parametreleri " $shape1$ ", " $shape2$ " ve opsiyonel olan merkeziyetsizlik parametresi (non-centrality parameter) " ncp " ile gösterilmektedir. Uygulamadaki amacımız farklı çarpıklığa sahip veriler üretmek olduğu için, $shape1$ parametresine sırasıyla 2, 1, 0.5, 0.25 ve 0.2 değerleri; $shape2$ parametresine ise bütün veri setlerinde sabit olacak şekilde 10 değeri atanmıştır. Ayrıca tabaka sınırı belirleme yöntemlerinden olan Geometrik Yöntem minimum değeri sıfıra yakın olan verilerde iyi sonuçlar vermediği için beta dağılımı ile türetilen veriler 10 000 000 değeri ile çarpılmıştır. Bu şekilde elde edilmiş 5 ayrı veri setine ait histogram görselleri ve özet istatistik bilgiler Şekil 1 ve Tablo 2'de özetlenmiştir.



Şekil 1. Oluşturulan Veri Setlerine Ait Histogram Görselleri

Tablo 2. Oluşturulan Veri Setlerine Ait Temel İstatistiki Göstergeler

	Sim-1	Sim-2	Sim-3	Sim-4	Sim-5
Birim Sayısı	1000	1000	1000	1000	1000
Medyan	1 455 129.95	671 381.80	362 454.47	231 387.71	156 643.89
Ortalama	1 642 935.46	955 930.56	628 164.23	484 826.27	397 109.25
Min	86 777.06	960.38	19.84	0.46	0.01
Maks	5 160 506.73	5 291 443.90	4 232 040.98	4 136 807.14	4 282 776.29
Standart Sapma	1 005 070.57	889 196.69	710 045.05	651 775.55	601 677.32
Çarpıklık	0.82	1.48	1.70	2.24	2.86

Yukarıda histogram görsellerinden ve tablodaki çarpıklık değerlerinden de görüldüğü gibi çarpıklık değeri simülasyon birden beşe doğru artmaktadır. Bunun amacı, farklı çarpıklık değerleri için yöntemlerin performanslarını karşılaştırmaktır.

Literatürde hem türetilmiş hem de gerçek veri setleri kullanılarak yapılan çalışmalar vardır. Gunning & Horgan, 2004 çalışmasında farklı çarpıklık değerine sahip 4 veri seti üzerinden uygulamalarını yapmışlardır. Bu veri setleri; bir İrlanda firmasının borçlarına ait muhasebe kayıtları, Amerika Birleşik Devletleri'ndeki şehir nüfusları, 4 yıllık Amerika Birleşik Devletleri kolejlerindeki öğrenci sayıları ve Amerika Birleşik Devletleri'ndeki büyük ticari bankaların kaynaklarıdır. Khan, Reddy, & Rao, 2015 ise kendi çalışmalarında Fiji'deki çiftçilerin şeker kamışı üretim verilerini elde edebilmek için yardımcı değişken olarak hektar cinsinden şeker kamışı ekim alanı verisini kullanmıştır. Hidiroglou & Kozak, 2017 çalışmasında hem gerçek veri seti üzerinden bir uygulama yapmış hem de Slanta & Krenzke, 1996 çalışmasına atıfta bulunarak, bu çalışma bulgularını paylaşmıştır. Slanta & Krenzke, 1996 ise kendi çalışmalarında 5 farklı örnek hacmine sahip $F(x)=(j-1/2)/N$ formülü ile türetilmiş veri seti üzerinde, Lavallee ve Hidiroglou yöntemini performansını sınımlarıdır (Slanta & Krenzke, 1996).

Kümeleme analizi için R paket programında bulunan “*NbClust*” paketi kullanılmıştır (Charrad, Ghazzali, Boiteau, & Niknafs, 2014). Kümeleme analizi için kullanılan formülün standart formu aşağıdaki gibidir:

```
NbClust(data, diss= NULL, distance= "euclidean", min.nc= 2, max.nc=15, method = "NULL", index = "all", alphaBeale = 0.1) (4.1)
```

Burada “data” kullanılacak veri setini, “diss” benzerlik matrisini, “distance” uzaklık ölçüsünü, “min.nc” oluşturulabilecek en küçük küme sayısını, “max.nc” oluşturulabilecek en büyük küme sayısını, “method” kullanılacak analiz yöntemini, “index” kullanılacak kümeleme sayısı belirleme kriterini ve “alphaBeale” ise küme sayısı belirleme kriterlerinden biri olan Beale endeksi için önem düzeyini ifade etmektedir, varsayılan ayarında değeri 0.1’dir. Benzerlik matrisi tanımlanmamışsa, uzaklık ölçüsü olarak “euclidean”, “maximum”, “manhattan”, “canberra”, “binary”, “minkowski” uzaklık ölçülerinden biri seçilmelidir, varsayılan ayar Öklid uzaklık ölçüsüdür. Uygulanacak kümeleme analiz yöntemleri ise “ward.D”, “ward.D2”, “single”, “complete”, “average”, “mcquitty”, “median”, “centroid”, “kmeans” seçenekleri arasında seçilmelidir. Küme sayısı belirleme kriterleri ise Charrad, Ghazzali, Boiteau ve Niknafs (2014)’de tanımlanmış 30 farklı yöntemden oluşmaktadır. Bu yöntemlerin içinden bir tane ya da tamamı belirleme yöntemi olarak seçilebilir. Örneğin “index=CCC” yazılırsa kübik kümeleme kriterine (Cubic Clustering Criterion) göre en uygun küme sayısı seçilir. “index=all” seçeneğinde 30 yöntem için de uygun küme sayısı belirlenir, bu durumda ise en çok yöntem tarafından seçilen küme sayısı, en uygun küme sayısı olarak belirlenmiş olur.

Bu çalışmada kümeleme analizi için gerçekleştirilen uygulamada, uzaklık ölçüsü olarak Öklid Uzaklığı, küme büyüklükleri için minimum değer olarak 2, maksimum değer olarak 10 değerleri, küme sayısı belirleme kuralları ise “index=all” seçeneği seçilmiştir. Böylece bu küme değerleri arasında uygun küme sayıları daha önceden tanıtılan 7 kümeleme analiz yöntemi için hesaplanmıştır.

Tabaka sınırı belirleme yöntemleri için ise R paket programında bulunan “*stratification*” paketi ile uygulama gerçekleştirilmiştir. Bu pakette yer alan formüllerin varsayılan ayarları aşağıda gösterilmiştir. Birikimli Kök Frekans Yöntemi için 4.2, Geometrik Yöntem için 4.3 ve Laval-Hidiroglou Yöntemi için 4.4 formül geçerlidir.

```
strata.cumrootf(x, n = NULL, CV = NULL, Ls = 3, certain = NULL, alloc = list(q1 = 0.5, q2 = 0, q3 = 0.5), rh = rep(1, Ls), model = c("none", "loglinear", "linear", "random"), model.control = list(), nclass = NULL) (4.2)
```

```
strata.geo(x, n = NULL, CV = NULL, Ls = 3, certain=NULL, alloc = list(q1 = 0.5, q2 = 0, q3 = 0.5), rh = rep(1, Ls), model = c("none", "loglinear", "linear", "random"), model.control = list()) (4.3)
```

```
strata.LH(x, n = NULL, CV = NULL, Ls = 3, certain = NULL, alloc = list(q1 = 0.5, q2 = 0, q3 = 0.5), takenone = 0, bias.penalty = 1, takeall = 0, rh = rep(1, Ls), model = c("none", "loglinear", "linear", "random"), model.control = list(), initbh = NULL, algo = c("Kozak", "Sethi"), algo.control = list()) (4.4)
```

Burada “x” tabakalama değişkeni vektörünü, “n” örnek hacmini, “CV” değişim katsayı değerini, “Ls” tabaka sayısını, “certain” vektör x’te yer alan ve örnekte seçilmesi kesin olan birimleri, “alloc” örnek birimlerin tabakalara dağıtım yöntemini belirleyen katsayıları, “rh” her tabakadaki beklenen cevaplama oranını, “model” tabaklama değişkeni X ile araştırma değişkeni Y arasındaki farklılığı tanımlayacak modeli, “model.control” model parametrelerini, “nclass” *Birikimli Kök Frekans Yöntemi* için sınıf sayısını, “takenone” örnek birim seçilmeyecek tabaka sayısını, “takeall” tabakadaki bütün birimlerin örnek seçileceği tabaka sayısını, “bias.penalty” “takenone=1” olduğunda araştırma tahmin edicisinin beklenen hata kareler ortalamasının hesabında kullanılan 0 ile 1 arasında değer alan ceza parametresini, “initbh” başlangıç tabaka sınırlarını, “algo” seçilecek optimizasyon algoritmasını ve “algo.control” ise optimizasyon algoritmasını kontrol edecek parametreleri göstermektedir. Bu formüllerde yer alan birçok seçenek, tabakalama değişkeni ile araştırma değişkeninin farklı olduğu durumlar için sunulmuştur. Bu çalışmada böyle bir ayırım yapılmadığı için sadece kullanılan alternatiflere ait bilgiler verilmiştir. Formüllerde “n” ya da “CV” den bir tanesi seçilmelidir. Örnek hacmi seçilmişse değişim katsayı değeri, değişim katsayısı seçilmişse örnek hacmi seçilmez. Bu çalışmada CV değer 0,01 alınmıştır. Örnek birimlerin tabakalara dağıtımını için Hidiroglou ve Srinath (1993)’ta açıklanan 3 sayısal değer yer almalıdır. Buna göre;

- $q_1=0,5$ ve $q_2=q_3=0$ ise oransal dağıtım,
- $q_1=q_2=p/2$ ve $q_3=0$ ise güç dağıtım,
- $q_1=q_3=0,5$ ve $q_2=0$ ise Neyman dağıtım yöntemini ifade etmektedir.

Tabaka sınırı belirleme yöntemlerinin dezavantajı tabaka sayısının önceden belirlenmesidir. “Ls” yani tabaka sayısı bu sebeple dolu olmalıdır. Bu çalışmada yukarıda adı geçen 3 tabaka sınırı belirleme yöntemi için 3’ten 7’ye kadar tabaka sayıları önsel olarak belirlenmiş ve veri setleri 5 farklı tabaka sayısı için tabakalara ayrılmıştır. Bu işlemler sonucunda 10 ayrıştırma yöntemi için (7’si kümeleme 3’ü tabaka sınırı belirleme yöntemi) tabaka büyüklükleri belirlenmiştir. Bu aşamadan sonra örnek hacmi belirlenip, örnekler tabakalara dağıtılıp veri setlerine ilişkin ortalama tahminlerinin KHKO’ları karşılaştırılmıştır. Bu uygulama için örnek hacim büyüklüğü 200 olarak belirlenmiştir. Örneklerin tabakalara dağıtımını oransal dağıtım yöntemiyle yapılmıştır. Yukarıdaki yöntemler için örneklerin tabakalara dağılımı belirlendikten sonra örnek seçim işlemi Tabakalı Basit Tesadüfî Örnekleme Yöntemiyle yapılmış, daha önce de değinildiği üzere örnek tekrar sayısı 1000 olarak belirlenmiştir. Sonuç olarak her bir yöntemde 1000 defa örnek seçilmiş ve her bir yönteme ait ortalama tahmini ve ortalamanın varyans tahminlerine ulaşılmıştır. Bu 10 yöntemi birbiri ile kıyaslamak için hesaplanan KHKO değerleri Tablo 3’te verilmiştir:

Tablo 3. Yöntemlerin KHKO Değerleri

Yöntem	Sim-1	Sim-2	Sim-3	Sim-4	Sim-5
Ortalama	31 113.80	24 162.47	18 705.67	5 905.44	8 239.33
Merkezi	30 329.63	16 160.24	18 818.21	8 800.44	12 586.51
Tam	25 994.18	19 412.70	8 911.38	16 376.18	6 634.93
K-Ortalamalar	34 985.42	22 557.16	16 424.74	15 175.73	5 970.40
Medyan	22 273.20	15 056.90	18 941.25	21 072.09	15 034.48
Tek	61 580.44	45 483.98	24 152.47	28 521.64	28 028.16
Ward	25 279.13	13 669.21	16 237.64	11 772.29	9 231.71
cumrootf_3	25 585.63	22 662.02	17 576.92	17 047.95	17 460.59
cumrootf_4	19 762.29	18 006.49	12 772.06	13 338.31	15 434.84
cumrootf_5	16 310.49	14 401.40	10 439.38	10 386.40	11 249.25
cumrootf_6	13 582.09	11 572.44	9 356.08	9 329.73	10 908.58
cumrootf_7	11 756.70	11 040.84	7 486.96	8 180.40	8 854.51
Geo_3	40 128.71	46 157.35	40 203.60	38 610.95	35 606.67
Geo_4	31 710.79	38 756.59	36 688.03	35 514.85	35 714.71
Geo_5	25 678.77	32 962.36	32 613.52	33 475.64	33 305.12
Geo_6	21 735.49	27 582.34	27 279.27	30 168.55	29 867.17
Geo_7	18 690.07	24 479.32	24 781.22	27 248.03	26 280.88
LH_3	24 814.91	22 628.00	16 660.09	15 004.08	14 628.92
LH_4	18 502.59	16 950.56	12 282.63	11 805.17	10 903.23
LH_5	15 736.37	13 489.36	9 720.74	8 738.27	8 754.91
LH_6	13 444.24	11 853.11	8 934.04	7 579.56	6 909.30
LH_7	11 240.72	10 335.59	7 252.25	6 604.77	6 311.94

Tablo 3'te oransal dağıtım ile örnek dağıtım yapılan, kümeleme ve tabaka sınırı belirleme yöntemlerine göre tabaka sınırı belirlenen beş farklı çarpıklığa sahip veri seti için elde edilen ortalama tahminine ilişkin KHKO'ları verilmiştir. Kırmızı ile boyanmış hücreler her veri seti için en düşük %10'luk değerleri göstermektedir. Buna göre;

- Çarpıklık derecesi 0.82 olan 1. veri setinde; Lavallee-Hidiroglou Yöntemi 11240.72 değeri ile en küçük KHKO değerine sahiptir. Bu sebeple 1. veri seti için optimum tabaka sayısı 7 olarak belirlenmiştir. Buna göre 1. Veri seti için optimum tabaka sınırları; Lavallee-Hidiroglou Yönteminde 7 tabaka için hesaplanan sınırlardır. Bu sınırlar; 0-700000, 700000-1200000, 1200000-1750000, 1750000-2400000, 2400000-3100000, 3100000-3900000, 3900000-ve üzeri şeklinde hesaplanmıştır.

- Çarpıklık derecesi 1.48 olan 2. veri setinde; Lavallee-Hidiroglou Yöntemi 10335.59 değeri ile en küçük KHKO değerine sahiptir. Bu sebeple 2. veri seti için optimum tabaka sayısı 7 olarak belirlenmiştir. Buna göre 2. Veri seti için optimum tabaka sınırları; Lavallee-Hidiroglou Yönteminde 7 tabaka için hesaplanan sınırlardır. Bu sınırlar; 0-350000, 350000-750000, 750000-1200000, 1200000-1750000, 1750000-2400000, 2400000-3300000, 3300000-ve üzeri şeklinde hesaplanmıştır.
- Çarpıklık derecesi 1.7 olan 3. veri setinde; Lavallee-Hidiroglou Yöntemi 7252.25 değeri ile en küçük KHKO değerine sahiptir. Bu sebeple 3. veri seti için optimum tabaka sayısı 7 olarak belirlenmiştir. Buna göre 3. Veri seti için optimum tabaka sınırları; Lavallee-Hidiroglou Yönteminde 7 tabaka için hesaplanan sınırlardır. Bu sınırlar; 0-220000, 220000-520000, 520000-850000, 850000-1300000, 1300000-1850000, 1850000-2600000, 2600000-ve üzeri şeklinde hesaplanmıştır.
- Çarpıklık derecesi 2.23 olan 4. veri setinde; Ortalama Bağlantı Kümeleme Yöntemi 5905.44 değeri ile en küçük KHKO değerine sahiptir. Bu sebeple 4. veri seti için optimum tabaka sayısı 9 olarak belirlenmiştir. Buna göre 4. Veri seti için optimum tabaka sınırları; Ortalama Bağlantı Kümeleme Yöntemi ile hesaplanan sınırlardır. Bu sınırlar; 0-220000, 220000-620000, 620000-1000000, 1000000-1500000, 1500000-2100000, 2100000-2400000, 2400000-3100000, 3100000-4100000, 4100000-ve üzeri şeklinde hesaplanmıştır.
- Çarpıklık derecesi 2.86 olan 5. veri setinde ise; K-Ortalamlar Kümeleme Yöntemi 5970.4 değeri ile en küçük KHKO değerine sahiptir. Bu sebeple 5. veri seti için optimum tabaka sayısı 7 olarak belirlenmiştir. Buna göre 5. Veri seti için optimum tabaka sınırları; K-Ortalamlar Kümeleme Yöntemi ile hesaplanan sınırlardır. Bu sınırlar; 0-150000, 150000-400000, 400000-700000, 700000-1100000, 1100000-1800000, 1850000-2800000, 2800000-ve üzeri şeklinde hesaplanmıştır.

5. Tartışma ve Sonuç

Bu çalışmada 1000 birimden oluşan farklı çarpıklık değerlerine sahip beş farklı kitle türetilmiştir. Tabakalı Tesadüfi Örneklemde kullanılmak üzere bu kitlelerin tabaka sınırlarını belirlemek ve tahminleri en iyileştirmek için Kümeleme Analiz Yöntemleri ve Tabaka Sınırı Belirleme Yöntemleri karşılaştırılmıştır. Bu karşılaştırma yapılırken 7 farklı kümeleme analizi, 3 farklı tabaka sınırı belirleme yöntemi olmak üzere 10 farklı ayırıştırma yöntemi kullanılmıştır. Tabakalar belirlendikten sonra oransal dağıtım yöntemiyle örnek birimler tabakalara dağıtılmıştır. Tahminlerin güvenilirliğini arttırmak için her yöntemden 1000 defa örnek seçilmiş, bu örneklerin ortalama tahminlerine ilişkin KHKO değerleri karşılaştırılmış ve sonuçları sunulmuştur.

Lavallee-Hidiroglou Yöntemi bir optimizasyon yöntemi olduğu için 5 veri setinin 3'ünde en başarılı ayırıştırma sağlamıştır. En düşük çarpıklığa sahip ilk üç veri setinde Lavallee-Hidiroglou Yöntemi, optimum tabakalamayı sağlamışken, 4. veri setinde Ortalama Bağlantı Kümeleme Yöntemi, en çarpık veri setinde ise K-Ortalamlar Kümeleme Yöntemi en iyi tabakalamayı yapmıştır. Çarpıklık değerleri arttıkça kümeleme analiz yöntemlerinin başarısının artması, çarpık dağılıma sahip veri setlerinde, kümeleme analiz yöntemlerinin de tabaka sınırı belirleme çalışmalarında kullanılabileceğini göstermektedir. Ayrıca özellikle hiyerarşik kümeleme yöntemlerinde küme sayısının önsel olarak belirlenmemesi, tabaka sınırı belirleme yöntemlerine kıyasla bir üstünlük olarak değerlendirilebilir. Buna rağmen hiyerarşik kümeleme yöntemlerinin uygulama mantığı gereği büyük veri setleri için maliyetli olmaları bir dezavantaj olarak kabul edilmelidir.

Sonuç olarak bu çalışma, çeşitli çarpıklık dağılımlarına göre tabakalı tesadüfi örneklemede kullanılan tabaka sınırlarının belirlenmesini karşılaştıran öncü bir çalışmadır. Bu kapsamda özellikle hiyerarşik olmayan K-ortalamlar yöntemine ait bulgular dikkat çekicidir. Çalışmanın bundan sonraki aşamalarında, hanehalkı/iş yeri düzeyinde farklı tahmin edicilere göre model denemeleri ve performans karşılaştırmaları yapılabilir. Bu makale yazarın doktora tez çalışmasının bir kısmını oluşturmaktadır ve tez çalışmasında hem gerçek veri seti üzerinde hem de diğer tabaka dağıtım yöntemleri için çalışma gerçekleştirilmiştir.

Kaynakça

- Ballin , M., & Barcaroli, G. (2013). Joint determination of Optimal Stratification and Sample Allocation Using Genetic Algorithm. *Survey Methodology*, 369-393.
- Benedetti, R., Bee, M., & Espa, G. (2010). A Framework for Cut-Off Sampling in Business Survey Design. *Journal of Official Statistics*, 651-671.
- Brito, J., Ochi, L., Montenegro, F., & Maculan, N. (2010). An Iterative Local Search Approach Applied to the Optimal Stratification Problem. *Int. Trans. Oper. Res.*, 753-764.
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 1-36.
- Dalenius, T., & Hodges, J. (1957). The Choice of Stratification Points. *Scand. Actuar. Journal*, 198-203.
- Ekman, G. (1959). An Approximation Useful in Univariate Stratification. *Ann. Math. Stat.*, 219-229.
- Erişoğlu, M. (2011). Uzaklık Ölçülerinin Kümeleme Analizine Olan Etkilerinin İncelenmesi ve Geliştirilmesi.
- Florek, K., Lukaszewicz, J., Steinhaus, H., & Zubrzycki, S. (1951). Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicum*, 282-285.
- Gunning, P., & Horgan, J. (2004). A New Algorithm for the Construction of Stratum Boundaries in Skewed Populations. *Survey Methodology*, 159-166.
- Hidiroglou, M., & Kozak, M. (2017). Stratification of Skewed Populations: A Comparison of Optimisation-based Versus Approximate Methods. *International Statistics Review*, 1-19.
- Hidiroglou, M., & Srinath, K. (1993). Problems Associated with Designing Subannual Business Surveys. *Journal of Business & Economic Statistics*, 397-405.
- Kennedy, D. (1988). A Note On Stochastic Search Methods For Global Optimization. *Advances in Applied Probability*, 476-478.
- Keskintürk, T., & Er, S. (2007). A Genetic Algorithm Approach to Determine Stratum Boundaries and Sample Sizes of Each Stratum in Stratified Sampling. *Comput. Stat. Data Anal.*, 53-67.
- Khan, M. G., Reddy, K. G., & Rao, D. (2015). Designing Stratified Sampling in Economic and Business Surveys. *Journal of Applied Statistics*, 1-20.
- Kozak, M. (2004). Optimal Stratification Using Random Search Method in Agricultural Surveys. *Stat. Transition*, 797-806.
- Lavallee, P., & Hidiroglou, M. A. (1988). On the Stratification of Skewed Populations. *Survey Methodology*, 33-43.
- Lednicki, B., & Wieczorkowski, R. (2003). Optimal Stratification and Sample Allocation Between Subpopulations and Strata. *Stat. Transition*, 287-306.
- Mahalanobis, P. (1952). Some Aspects of the Design of Sample Surveys. *The Indian Journal of Statistics*, 1-7.
- Rivest, L. (2002). A Generalization of the Lavallee and Hidiroglou Algorithm for Stratification in Business Surveys. *Survey Methodology*, 191-198.
- Servi, T. (2009). Çok Değişkenli Karma Dağılım Modeline Dayalı Kümeleme Analizi.

Atf / Citation: SERT O., AKÇORAOĞLU A. (2022). Tabaka Sınırlarının Belirlenmesinde, Kümeleme Analiz Yöntemleri ve Tabaka Sınırı Belirleme Yöntemlerinin Karşılaştırılması. *İstatistik Araştırma Dergisi*, 12 (1), 68-81.

Sethi, V. (1963). A Note on the Optimum Stratification of Populations for Estimating the Population Means. *Australian J. Stat.*, 20-33.

Singh, R. (1971). Approximately Optimum Stratification on the Auxiliary Variable. *J. Am. Stat. Assoc.*, 829-833.

Slanta, J., & Krenzke, T. (1996). Applying the Lavallée and Hidiroglou Method to Obtain Stratification Boundaries for the Census Bureau's Annual Capital Expenditure Survey. *Survey Methodology*, 65-75.

Sorensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter*, 1-34.

Sweet, E., & Sigman, R. (1995). Evaluation of Model-Assisted Procedures for Stratifying Skewed Populations Using Auxiliary Data. *U.S. Bureau of the Census*, 1-9.

Thomsen, I. (1976). A Comparison of Approximately Optimal Stratification Given Proportional Allocation with Other Methods of Stratification and Allocation. *Metrika*, 15-25.

Verma, M. R., Kozak, M., & Zielinski, A. (2007). Modern Approach to Optimum Stratification. *Review and Perspectives*, 223-250.

Ward, J. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 236-244.