

# İSTATİSTİK ARAŞTIRMA DERGİSİ

JOURNAL OF STATISTICAL  
RESEARCH

JSRTR

ISSN 2791-7614

Cilt  
Volume **11**

Sayı  
Issue **02**

Yıl  
Year **2021**

**İSTATİSTİK ARAŞTIRMA DERGİSİ  
ULUSLARARASI HAKEMLİ DERGİ**

JOURNAL OF STATISTICAL RESEARCH  
INTERNATIONAL PEER-REVIEWED JOURNAL

**ISSN: 2791-7614  
CİLT/VOL. 11 • SAYI / ISSUE 2 • 2021**

**Türkiye İstatistik Kurumu Adına Sahibi / Owner on Behalf of Turkish Statistical Institute**  
Prof. Dr. Sait Erdal DİNÇER

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

**Editör Yardımcısı / Assistant Editor**  
Prof. Dr. Nurcan METİN

**Sorumlu Müdür / Responsible Manager**  
Organizasyonel Gelişim 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.

**©Türkiye İstatistik Kurumu Yayını  
©Turkish Statistical Institute Publication**

## **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.

Dergimizin Temmuz ve Aralık sayılarında çalışmaların daha nitelikli hale gelmesinde çok değerli öneri, eleştiri ve katkılarını esirgemeyen ve aşağıda adı geçen sayın hakemlere şükranlarımı iletmek isterim.

Bu sayıda makaleleri ile destek veren yazarlarımıza, dergimizin Editörler Kuruluna üye olmayı kabul ederek bizi onurlandıran değerli bilim insanlarına, desteklerinden ötürü TÜİK Başkanı Sayın Prof. Dr. S. Erdal DİNÇER'e ve emeği geçen herkese teşekkürlerimi sunarım.

Dergimizin Aralık 2021 (Cilt 11, Sayı 2) 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, Çag 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, Poland  
Nikolai KOLEV, University of Sao Paulo, Brezilya  
Debasis KUNDU, Indian Institute of Technology, Hindistan  
W. Y. Wendy LOU, University of Toronto, Kanada  
Dr. Ojonugwa USMAN, Federal College of Education (Technical), Nijerya

**2021 Yılında ‘İstatistik Araştırma Dergisi’ne Bilimsel Katkı Sağlayan Hakemler**  
**Referees Who Provided Scientific Contributions to The Journal of Statistical Research in 2021**

Prof. Dr.	Ebru ÇAĞLAYAN AKAY, Marmara Üniversitesi, İstanbul Mehmet Ali CENGİZ, Ondokuz Mayıs Üniversitesi, Samsun Burak GÜRİŞ, İstanbul Üniversitesi, İstanbul İsmet KOÇ, Hacettepe Üniversitesi, Ankara Ünal Halit ÖZDEN, İstanbul Ticaret Üniversitesi, İstanbul Mustafa SEVÜKTEKİN, Uludağ Üniversitesi, Bursa 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 Metehan YILGÖR, Bandırma Onyedi Eylül Üniversitesi, Balıkesir
Doç. Dr.	Hasan BULUT, Ege Üniversitesi, İzmir Aycan HEPSAĞ, İstanbul Üniversitesi, İstanbul Mehmet Kenan TERZİOĞLU, Trakya Üniversitesi, Edirne
Dr.	Yasemin KOLDERE AKIN, Trakya Üniversitesi, Edirne Melek ASTAR, Fatih Sultan Mehmet Vakıf Üniversitesi, İstanbul Seda KARAKAŞ GEYİK, İstanbul Üniversitesi, İstanbul N. Tuba YILMAZ SOYDAN, Marmara Üniversitesi, İstanbul Nazan ŞAK, Osmaniye Korkut Ata Üniversitesi, Osmaniye Turgut ÜN, Marmara Üniversitesi, İstanbul Nurbanu BURSA, Hacettepe Üniversitesi, Ankara

**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

- 1-13 Şaban KIZILARSLAN  
*Kantil Regresyon Modellerinde Uyum İyiliği Ölçüleri ve Model Seçimi* / Goodness of Fit Measures and Model Selection in Quantile Regression
- 14-29 Fulden KÖMÜRYAKAN  
*Marital Status Differences in Women's Labor Force Participation in Turkey: Nonlinear Decomposition Approach* / Türkiye'de Kadınların İşgücüne Katılımında Medeni Durum Farklılıkları: Doğrusal Olmayan Ayrıştırma Yaklaşımı
- 30-55 Cemil İSKENDER  
*Econometric Analysis of Population Increase and Population Projections in Turkey* / Türkiye'deki Nüfus Artışı ve Nüfus Projeksiyonlarının Ekonometrik Analizi
- 56-70 Onur ŞENTÜRK, Nuri ÇELİK  
*Türkiye'de Hanehalkı Yoksulluğunu Etkileyen Faktörlerin Belirlenmesi* / Determination of Factors Affecting Household Poverty in Turkey
- 71-82 Asila KOÇAK  
*Türkiye' de Toplumsal Cinsiyetin Çok Boyutlu Ölçekleme İle İncelenmesi* / Examining Gender in Turkey with Multidimensional Scaling



## Kantil Regresyon Modellerinde Uyum İyiliği Ölçüleri ve Model Seçimi

Şaban KIZILARSLAN

Van Yüzüncü Yıl Üniversitesi / Araştırma Görevlisi

saban.kizilarслан@marmara.edu.tr

Orcid No: 0000-0003-1545-9597

### Özet

Bu çalışmada kantil regresyon modellerinde uyum iyiliği ve model seçimi için kullanılan kriterlerden öne çıkanlar bir arada sunulmuştur. Bu kriterler belirlilik oranına ( $R^2$ ) benzer bir ölçü olan  $R^1$  oranı,  $C^1$  testi, asimetrik ağırlıklandırılmış ortalama mutlak hata (ATWE) ile Akaike (AIC) ve Schwarz (BIC) bilgi kriterleridir. Ayrıca etkinliklerinin artırılması amacıyla uyarlanmış AIC ve BIC kriterlerine de yer verilmiştir. Bu kriterlerin uygulamada kullanımını göstermek amacıyla, Mincer ücret denklemi desiller düzeyinde yuvalanmış modeller olarak tahmin edilmiştir. Potansiyel tecrübenin karesine kısıt konarak oluşturulan kısıtlı ve kısıtsız modeller için, açıklanan kriterler yardımıyla uyum iyiliği incelenmiş ve model seçimi yapılmıştır. Sonuçta genel olarak kısıtsız model tercih edilmiş ve potansiyel tecrübe ile logaritmik ücret ilişkisinin tüm kantillerde teoride varsayıldığı gibi karesel olması gerektiği belirlenmiştir. Yalnızca uyarlanmış bilgi kriterlerine göre en yüksek kantillerde kısıtlı model tercih edilmiş, yani potansiyel tecrübenin karesinin modelin uyum iyiliğini yüksek kantillerde anlamlı düzeyde arttırmadığı tespit edilmiştir. Kantil regresyon modelleri için topluca sunulan ve uygulamada kullanımı gösterilen uyum iyiliği ve model seçimi kriterlerinin, literatürde daha sonra yapılacak çalışmalar için destekleyici nitelikte olacağı düşünülmektedir.

*Anahtar Sözcükler: Bilgi Kriterleri, Kantil Regresyon, Mincer Ücret Denklemi, Model Seçimi, Uyum İyiliği.*

**Sorumlu Yazar / Corresponding Author:** 1-Şaban KIZILARSLAN, Van Yüzüncü Yıl Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü.

**Atf / Citation:** KIZILARSLAN Ş. Kantil Regresyon Modellerinde Uyum İyiliği Ölçüleri ve Model Seçimi. İstatistik Araştırma Dergisi, 2021, 11 (2), 1-13.

## Goodness of Fit Measures and Model Selection in Quantile Regression

### Abstract

In this study, the prominent criteria used for goodness of fit and model selection in quantile regression models are presented together. These criteria are the  $R^1$  ratio, which is a measure similar to the coefficient of determination ( $R^2$ ), the  $C^1$  test, the asymmetric weighted mean absolute error (ATWE), and the Akaike (AIC) and Schwarz (BIC) information criteria. In addition, more efficient AIC and BIC criteria, which were adapted to increase their effectiveness, are also included. To illustrate the practical use of these criteria, the Mincer wage equation was estimated as nested models at the decile level. For the constrained and unconstrained models created by placing constraint on the square of the potential experience, the goodness of fit was examined and the model selection was made with the help of the explained criteria. As a result, the unconstrained model was generally preferred and it was determined that the relationship between potential experience and logarithmic wage should be quadratic in all quantiles as assumed in theory. However, according to the adapted information criteria, the restricted model was preferred in the highest quantiles, that is, it was determined that the square of potential experience did not significantly increase the goodness of fit of the model in high quantiles. It is thought that the goodness of fit and model selection criteria, which are presented collectively for quantile regression models and used in practice, will be supportive for future studies in the literature.

*Keywords: Goodness of Fit, Information Criteria, Mincer Wage Equation, Model Selection, Quantile Regression.*

### 1. Giriş

Regresyon modellerinde kurulan modelin uyum iyiliğini incelemek ve alternatif regresyon modelleri arasında optimal model seçimi yapmak önem arz etmektedir. Uyum iyiliği kavramı genel olarak modelde yer alan bağımsız değişkenlerin bağımlı değişkendeki değişmeyi açıklayabilme gücüdür. Uyum iyiliği modelin tahmin başarısı üzerinden analiz edilebildiği gibi öngörü başarısı üzerinden de incelenebilmektedir. Model seçimi ise genellikle aynı bağımlı değişkeni açıklamak üzere farklı açıklayıcı değişkenler veya farklı fonksiyonel biçimler kullanılarak elde edilen regresyon modelleri arasında uyum iyiliği en yüksek modelin belirlenmesi sürecidir. Bazı durumlarda bir modeldeki bazı değişkenlere kısıt veya kısıtlar konarak kısıtlı bir alt model elde edilir ve yuvalanmış modeller olarak isimlendirilen bu kısıtlı ve kısıtsız modeller arasında model seçimi yapılır. Yuvalanmış modeller arasında seçim yapma işlemi, aslında kısıt konan değişken veya değişkenlerin modelin uyum iyiliğine anlamlı bir katkısının olup olmadığının incelenmesi işlemidir.

Regresyon modellerinin uyum iyiliğini incelemek ve model seçimi yapmak amacıyla kullanılan pek çok kriter mevcuttur. Bunlar arasında öne çıkan bazıları belirlilik katsayısı ( $R^2$ ); F, Benzerlik Oranı (LR), Langrange Çarpanı (LM) ve Wald testleri ile Akaike ve Schwarz (Bayesyen) bilgi kriterleridir (AIC, BIC). Bunların yanı sıra, modellerin tahmin ve öngörü başarılarını dikkate alan ve genellikle model hatalarının birer fonksiyonu olarak hesaplanan Ortalama Hata Kare (MSE) ve Ortalama Mutlak Hata (MAE) gibi ölçüler ve bu ölçülerin çeşitlendirilmeleri bulunmaktadır (Gürüş ve Çağlayan-Akay, 2018; Pham, 2019).

Kantil regresyon modelleri, bağımlı değişken dağılımının ortalamasının ele alındığı klasik regresyon modellerine alternatif olarak önerilmiştir. Dağılımın kantillerinden faydalanarak istenen her noktada inceleme yapma avantajına sahip olan bu modeller, aşırı değerlerin varlığı ve değişen varyans gibi durumlarda klasik modellere göre daha dirençlidir. Ayrıca klasik modelde yer alan kısıtlayıcı bir varsayım olan hata terimlerinin normal dağılması varsayımı kantil modellerinde söz konusu değildir. Bağımlı değişkenin medyanının ele alındığı regresyon modeli, kantil regresyonun özel bir durumudur ve medyan regresyon olarak adlandırılmaktadır (Koenker, 2005).

Kantil regresyon modellerinde uyum iyiliğini incelemek ve model seçimi yapabilmek için çeşitli kriterler geliştirilmiştir. Klasik regresyon modellerinde parametre tahminleri için kullanılan hata kereleri minimizasyonu, kantil regresyon modellerinde ağırlıklandırılmış mutlak hatalarla değiştirilir. Kantil regresyon modellerinde uyum iyiliği ve model seçimi kriterleri, genel olarak bu değişikliğin dikkate alınarak klasik modeller için kullanılan kriterlerin uyarlanması ile elde edilmektedir. Literatürde kantil regresyon modelleri için çeşitli uyum iyiliği ve model seçimi kriterleri yer almaktadır. Hurvich ve Tsai (1990) medyan regresyon için AIC tipi bilgi kriteri önermişlerdir. Machado (1993), kantil regresyonu durumunu da içeren M-tipi robust tahminçiler için Schwarz bilgi



kriterini (BIC) ortaya koymuştur. Koenker ve Machado (1999), çalışmalarında kantil regresyon için klasik modellerde kullanılan belirlilik katsayısına benzer bir uyum iyiliği ölçüsü olan  $R^1$  oranını önermiş ve bu kriterin model seçimi için kullanımını açıklamışlardır. Ayrıca yine aynı çalışmada kantil regresyon modelleri için LR ve Wald tipi testler önerilmiştir. He ve Zhu (2003), doğrusal veya doğrusal olmayan kantil regresyon modelleri için gradyan vektörünün cusum sürecine dayanan bir uyum eksikliği testi önermişlerdir. Koenker (2005) kantil regresyon modelleri için AIC kriterinin yapısını sunmuştur. Chernozhukov ve Fernández-Val (2005), kantil regresyon modellerinde çeşitli temel hipotezleri test etmek için kullanılabilir, alt örneklemeyle dayalı bir test yapısı önermişlerdir. Furno (2011), kantil regresyon modelleri için klasik modellerde kullanılan F ve LR testlerinin yapısına benzer bir test olan  $C^1$  uyum iyiliği testini sunmuştur. Haupt vd. (2011), çeşitli kantil regresyon modelleri için bir çapraz doğrulama (cross-validation) yöntemi önerdikleri çalışmalarında, model karşılaştırması için  $R^1$  ile birlikte ortalama ağırlıklandırılmış mutlak hata (ATWE) ölçüsünden faydalanmışlardır. Dong vd. (2019), düşük ve yüksek boyutlu parametrik kantil regresyon modelleri için bir uyum eksikliği testi geliştirmişlerdir. Shin vd. (2021), kantil modeller için kullanılan AIC ve BIC kriterlerinin etkinliğini arttırmak amacıyla, alternatif bir kayıp fonksiyonu kullanarak daha etkin olduğunu ifade ettikleri uyarlanmış AIC ve BIC kriterleri geliştirmişlerdir.

Bu çalışmada, kantil regresyon modellerinde uyum iyiliğini analiz etmek ve model seçimi yapmak amacıyla kullanılan bazı kriterler bir arada sunulmuştur. Çalışmada öne çıkan kriterler seçilirken, klasik regresyon modellerindeki kriterlere benzer yapıda olan genel kriterler tercih edilmiştir. Çalışmada ele alınan kriterler  $R^1$ ,  $C^1$  testi, ATWE ile bilgi kriterleri AIC ve BIC'dir. Bu kriterlere ek olarak etkinliği artırılmış iki bilgi kriterine de çalışmada yer verilmiştir. Sunulan kriterlerin uygulamada kullanımını örneklemek amacıyla Mincer ücret denklemi, yuvalanmış modeller olarak desiller düzeyinde tahmin edilmiş ve kısıtlı ile kısıtsız modeller arasında açıklanan kriterler ile model seçimi yapılmıştır. Toplu olarak sunulan ve uygulama örneği verilen bu kriterlerin, yerli literatürde daha sonra yapılacak kantil regresyon çalışmalarında kullanılabilirliği göz önüne alındığında, bu çalışmanın yerli literatüre katkı sağlayacağı öngörülmektedir.

Çalışmanın bir sonraki bölümünde, ele alınan kriterlerin teorik yapıları tanıtılmıştır. Üçüncü bölümde uygulamanın detaylarına ve kullanılan değişkenlere ait bilgilere yer verilmiştir. Bir sonraki bölümde uygulamanın bulguları sunulmuş ve son bölümde çalışmadan elde edilen sonuçlar toparlanmıştır.

## 2. Kantil Regresyonda Uyum İyiliği ve Model Seçimi Ölçüleri

Klasik regresyon modelinde bağımlı değişkenin koşullu ortalaması modellenmektedir. Koenker ve Basset (1978) tarafından öncülük edilen kantil regresyon modelinde ise bağımlı değişkenin koşullu kantillerinin modellenmesi önerilmektedir. Böylece bağımlı değişkenin yalnızca ortalaması için değil, dağılımının herhangi bir noktasında incelenmesi mümkün olmaktadır. Ayrıca klasik regresyonda yer alan kısıtlayıcı varsayımlardan bir tanesi olan hata terimlerinin normal dağılması varsayımı kantil regresyon modelleri için söz konusu değildir. Bu nedenle bu modeller aşırı değerlere karşı klasik modellere karşı daha dirençlidir.

Kantil regresyon modellerinde genellikle bağımlı değişkenin bağımsız değişkenlere koşullu kantil fonksiyonu kullanılmaktadır.  $y$  tesadüfi değişkeninin  $x$  tesadüfi değişkeni üzerine koşullu  $\theta$ . kantili aşağıdaki gibi tanımlanmaktadır:

$$q_{\theta}(x) = F^{-1}(\theta|x) = \inf \{y : F(y|x) \geq \theta\} \quad (1)$$

Burada  $F$  kümülatif dağılım fonksiyonudur ve değişkenin ilgili kantilinin hesaplandığı kantil fonksiyonu  $Q(\theta)$ , bu kümülatif dağılım fonksiyonunun tersi olarak tanımlanmaktadır. Bu kantil fonksiyonu yardımıyla  $\theta$ . kantil için doğrusal kantil regresyon modeli aşağıdaki gibi gösterilebilir:

$$Q_{\theta}(y|x) = x'\beta_{\theta} \quad (2)$$

Burada  $x'$  açıklayıcı değişkenler matrisinin transpozunu,  $\beta_{\theta}$  ise ilgili kantil için parametreler vektörünü ifade etmektedir. Kantil regresyonda parametre tahminleri, klasik regresyonda kullanılan hata kareleri toplamı yerine, hataların asimetrik ağırlıklandırılmış mutlak değerleri toplamının minimizasyonu ile elde edilmektedir. Parametre tahminleri için kullanılan minimizasyon fonksiyonu aşağıdaki gibidir:

$$\hat{\beta}_{\theta} = \min_{\beta} \sum \rho_{\theta}(y - x'\beta) \quad (3)$$

Burada kullanılan  $\rho_\theta$  kayıp fonksiyonu,

$$\rho_\theta(z) = z(\theta - I(z < 0)) \quad (4)$$

olarak tanımlanmaktadır ve  $I$  gösterge fonksiyonudur. Bu kayıp fonksiyonu  $\theta = 0.5$  değeri yani medyan için hesaplandığında mutlak değer fonksiyonuna dönüşmektedir ve bu fonksiyonla medyan için kurulan regresyon modeli özel olarak medyan regresyon olarak adlandırılmaktadır (Koenker ve Basset, 1978; Koenker, 2005; Davino vd., 2014).

Kantil regresyon modellerinde modelin uyum iyiliğini ölçmek ve birden fazla kantil regresyon modeli arasında seçim yapmak için kullanılan bazı ölçüler geliştirilmiştir. Bu bölümde, bu ölçülerden öne çıkan ve yeni önerilen birkaç tanesi tanıtılacaktır.

## 2.1. $R^1$ Uyum İyiliği Kriteri

Kantil regresyon modellerinde uyum iyiliğini ölçmek ve model seçimi yapmak amacıyla önerilen  $R^1$  uyum iyiliği kriteri, klasik regresyon modelleri için kullanılan belirlilik (determinasyon) katsayısı  $R^2$ 'ye benzer bir ölçüdür. Belirlilik katsayısı hesaplanırken modelin toplam hata karelerinden faydalanılırken,  $R^1$  kriterinde benzer bir yaklaşımla, parametre tahmininde kullanılan asimetric ağırlıklandırılmış mutlak değerleri toplamından faydalanılmaktadır. İncelenen  $\theta$ . kantil için, kantil regresyon modeline ait  $R^1$  değeri aşağıdaki gibi hesaplanmaktadır:

$$R^1(\theta) = 1 - \frac{\sum \rho_\theta(y - x' \hat{\beta}_\theta)}{\sum \rho_\theta(y - y_\theta)} \quad (5)$$

Burada  $y_\theta$ , bağımlı değişkeninin örnek dağılımının  $\theta$ . kantilidir. Hesaplanan değer 1'e yaklaşması modelin uyum iyiliğinin yüksek, 0'a yaklaşması ise düşük olduğunu ifade etmektedir.  $R^1$  kriteri birden fazla modelin karşılaştırılması için de kullanılabilir. Yuvalanmış kantil regresyon modellerinde, kısıtlı ve kısıtsız modellerin karşılaştırılması için bu kriter aşağıdaki gibi hesaplanabilir:

$$R^1(\theta) = 1 - \frac{\sum \rho_\theta(y - x' \hat{\beta}_{\theta, \text{kısıtsız}})}{\sum \rho_\theta(y - x' \tilde{\beta}_{\theta, \text{kısıtlı}})} \quad (6)$$

Burada kısıtsız ve kısıtlı modellerin ağırlıklandırılmış mutlak hata toplamları kullanılmaktadır. Elde edilen değer, kısıt konan değişken veya değişkenlerin modele ne kadar katkı yaptığını göstermektedir. Örneğin kısıtsız bir model ile bu modelden bir değişken çıkartılarak elde edilen kısıtlı model için yukarıda verilen kriter hesaplandığında, hesaplanan değer, kısıt konan değişkenin modelin uyum iyiliğine katkısını ortaya koymaktadır. Klasik regresyon modelleri için global bir ölçü olan  $R^2$  ölçüsünün aksine,  $R^1$  kriteri ilgili kantil için hesaplanan yerel bir ölçüdür (Koenker ve Machado, 1999).

## 2.2. $C^1$ Testi

Klasik regresyon modellerinde, bağımsız değişkenlerin bağımlı değişkenin koşullu ortalamasını açıklama başarısını bir arada test etmek amacıyla F Testi kullanılmaktadır. F Testi, aynı zamanda yuvalanmış modellerde kısıt konan değişkenlerin modelin açıklama gücünde anlamlı bir fark oluşturup oluşturmadığını test etmek amacıyla kullanılabilir. Ancak bu testin güvenilirliği normallik varsayımının sağlanmasına bağlıdır. Furno (2011) tarafından kantil regresyon modelleri için F ve Benzerlik Oranı (LR) testlerine benzer bir yaklaşımla genelleştirilen  $C^1$  testi önerilmiştir.  $C^1$  testi, Chow (1960) tarafından önerilen LR testi ile yakından ilişkilidir. Bu test, normal dağılımlı hata terimlerine sahip kısıtlı ve kısıtsız modellerin karşılaştırılmasında kullanılmaktadır ve toplam hata karelerinden hesaplanmaktadır.  $C^1$  testi, kantil regresyon modelleri için benzer bir yaklaşım takip etmektedir. Ancak LR testinden farklı olarak, toplam hata kare yerine ağırlıklandırılmış mutlak hatalardan faydalanılmaktadır ve normal dağılım koşulu söz konusu değildir. İlgili kantil için kısıtlı ve kısıtsız modeller arasında tercih yapmak amacıyla  $C^1$  test istatistiği aşağıdaki gibi hesaplanabilir:

$$C^1(\theta) = \frac{\left[ \sum \rho_\theta(y - x' \tilde{\beta}_{\theta, \text{kısıtlı}}) - \sum \rho_\theta(y - x' \hat{\beta}_{\theta, \text{kısıtsız}}) \right] / d_1}{\sum \rho_\theta(y - x' \hat{\beta}_{\theta, \text{kısıtsız}}) / d_2} \quad (7)$$

Burada  $d_1$  kısıt sayısıdır.  $d_2$  ise,

$$d_2 = \text{gözlem sayısı} - \text{kısıtsız modelde tahmin edilen katsayı sayısı}$$

olarak hesaplanmaktadır. Bu test istatistiği asimptotik olarak  $F_{d_1, d_2}$  dağılımına sahiptir. Ayrıca klasik regresyon modellerindeki F testi ile  $R^2$  ilişkisine benzer bir ilişki,  $C^1$  testi ile daha önce kısıtlı ve kısıtsız modeller arasındaki uyum iyiliği farkını incelemek amacıyla verilen  $R^1$  kriteri için de söz konusudur.  $C^1$  test istatistiği, yuvalanmış modellere ait  $R^1$  kriteri yardımıyla aşağıdaki gibi hesaplanabilir:

$$C^1(\theta) = \frac{R^1(\theta)}{1 - R^1(\theta)} \frac{d_2}{d_1} \quad (8)$$

Bu testin sıfır hipotezi kısıtın geçerli olduğunu ifade etmektedir. Bu durumda sıfır hipotezi reddedilirse kısıtsız model tercih edilirken; sıfır hipotezi reddedilemezse, kısıt konan değişkenlerin modelin uyum iyiliğinde anlamlı bir fark oluşturmadığı sonucuna varılmakta ve kısıtlı model tercih edilmektedir. (Furno, 2011).

### 2.3. Ağırlıklandırılmış Ortalama Mutlak Hata (ATWE)

Klasik regresyon modellerinde, modelin tahmin ve öngörü başarısını ölçmek ve modelleri bu açıdan karşılaştırmak için kullanılan kriterlerden öne çıkanları Ortalama Hata Kare (MSE) ve Ortalama Mutlak Hata (MAE) kriterleridir. Ağırlıklandırılmış Ortalama Mutlak Hata (ATWE) kriteri, MSE ve MAE kriterlerine benzer bir yaklaşımla, kantil regresyon modelleri için genellenmiş bir kriterdir. Kantil regresyon modelinde ilgili  $\theta$ . kantil için ATWE kriteri,

$$ATWE(\theta) = \frac{1}{n} \sum \rho_\theta(y - x' \hat{\beta}_\theta) \quad (9)$$

olarak hesaplanabilir. Burada asimetrik ağırlıklandırılmış mutlak hata ortalamaları hesaplanmaktadır ve MSE ve MAE kriterlerine benzer olarak ATWE değeri daha düşük olan kantil modelinin uyum iyiliğinin daha yüksek olduğu ve alternatif modellere tercih edilebileceği ifade edilmektedir. Bu kriter, aynı zamanda veri setinin bir kısmı bölünerek, alternatif kantil modellerin öngörü başarılarının karşılaştırılmasında kullanılabilir (Haupt vd., 2011).

### 2.4. Bilgi Kriterleri

Regresyon modellerinde model seçimi kullanılan temel seçeneklerden bir tanesi bilgi kriterleridir. Bilgi kriterlerinden öne çıkan ve en sık kullanılan iki tanesi Akaike Bilgi Kriteri (AIC) ve Schwarz, diğer adıyla Bayesyen Bilgi Kriteri (BIC)'dir. Klasik regresyon modellerinde toplam hata karelerinden faydalanarak hesaplanan bilgi kriterleri, kantil regresyon modellerinde ağırlıklandırılmış mutlak hata toplamları kullanılarak hesaplanmaktadır. Kantil regresyon modellerinde ilgili  $\theta$ . kantil için bilgi kriterlerinin genel ifadesi;

$$IC(\theta) = n \ln \left[ \frac{1}{n} \sum \rho_\theta(y - x' \hat{\beta}_\theta) \right] + \alpha(n, k) / 2 \quad (10)$$

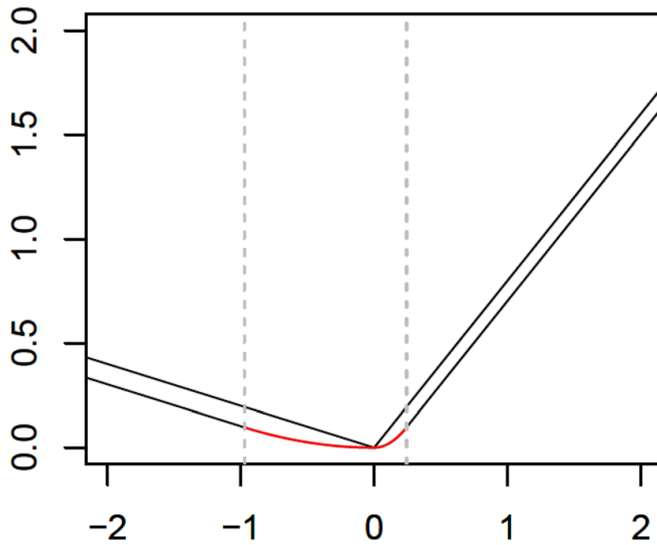
olarak verilebilir. Burada  $n$  gözlem sayısı,  $k$  ise modelin boyutu yani modeldeki parametre sayısıdır. Bu genel ifadede  $\alpha(n, k)$  fonksiyonunun farklı seçimleri ile AIC ve BIC kriterleri elde edilmektedir.  $\alpha(n, k) = 2k$  olarak seçildiğinde AIC kriteri elde edilirken,  $\alpha(n, k) = k \ln(n)$  olarak seçildiğinde ise BIC kriteri elde edilmektedir. Her iki kriter için de geçerli olmak üzere, genel olarak daha küçük değere sahip olan modelin optimal model olduğu sonucuna varılmaktadır (Koenker, 2005; Shin vd., 2021).

## 2.5. Uyarlanmış Bilgi Kriterleri

Shin vd. (2021), kantil regresyon modellerinde kullanılan AIC ve BIC kriterlerinin etkinliklerini arttırmak amacıyla bu kriterler için düzeltme önermişlerdir. Burada önerilen düzeltme, bilgi kriterlerinin değeri hesaplanırken kullanılan  $\rho_\theta$  kayıp fonksiyonu yerine alternatif bir kayıp fonksiyonu kullanmaktır. Daha önce verilen  $\rho_\theta$  kayıp fonksiyonu yerine önerilen modifiye edilmiş kayıp fonksiyonu aşağıdaki gibidir:

$$\rho_\theta^M(z) = \begin{cases} \theta z - \frac{\theta(1-\theta)}{2\lambda_\theta}, & \frac{(1-\theta)}{\lambda_\theta} \leq z \\ \frac{\lambda_\theta}{2} \frac{\theta}{(1-\theta)} z^2, & 0 \leq z < \frac{(1-\theta)}{\lambda_\theta} \\ \frac{\lambda_\theta}{2} \frac{(1-\theta)}{\theta} z^2, & -\frac{\theta}{\lambda_\theta} \leq z < 0 \\ -(1-\theta)z - \frac{\theta(1-\theta)}{2\lambda_\theta}, & z < -\frac{\theta}{\lambda_\theta} \end{cases} \quad (11)$$

Burada  $\lambda_\theta = 0.5e^{(-2.118-1.097 \times \min(\theta, 1-\theta))} n^\alpha / \hat{\sigma}$  olarak hesaplanmaktadır. Bu ifadede  $n$  örneklem büyüklüğü ve  $\alpha$  pozitif bir sabit değerdir, uygulamada bu değer 0.3 olarak alınmıştır.  $\hat{\sigma}$  ise hata dağılımının robust bir ölçek tahminidir. Burada  $\hat{\sigma}$  olarak tüm değişkenlerle tahmin edilen medyan regresyon modeline ait artıkların standart hatası kullanılmıştır. Şekil 1'de  $\rho_\theta$  ile modifiye edilmiş  $\rho_\theta^M$  kayıp fonksiyonlarına ait grafikler yer almaktadır.



Şekil 1.  $\rho_\theta$  ve  $\rho_\theta^M$  Kayıp Fonksiyonları (Kaynak: Shin vd., 2021)

Grafik incelendiğinde,  $\rho_\theta$  kayıp fonksiyonunda 0 değerinde yer alan sivri ucun asimetric bir kuadratik eğri ile değiştirildiği görülmektedir. Bu değişiklik, sıfır etrafında bir ortalama etkisi yaparak etkinliğin artmasını sağlamaktadır. Bu modifiye edilmiş kayıp fonksiyonu yardımıyla kantil regresyon modeli için daha etkin bilgi kriterleri aşağıdaki gibi hesaplanabilir:

$$EIC(\theta) = n \ln \left[ \frac{1}{n} \sum \rho_\theta^M \left( y - x' \hat{\beta}_\theta \right) \right] + \alpha(n, k) \quad (12)$$

Burada daha etkin AIC (E\_AIC) ve daha etkin BIC (E\_BIC) kriterlerinin hesaplanması için  $\alpha(n, k)$  fonksiyonu sırasıyla  $2k$  ve  $k \ln(n)$  olarak seçilmektedir. Yine bu bilgi kriterleri ile model seçimi yapılırken, daha düşük değere sahip modelin optimal model olduğu sonucuna varılmaktadır (Shin vd., 2021).

### 3. Uygulama: Mincer Ücret Denklemi

Bu çalışmada, Türkiye’de Mincer ücret denkleminin kantil düzeyinde geçerliliği incelenmiştir. Mincer (1974) tarafından önerilen Mincer ücret denklemi, belirli bir ücret düzeyine sahip bireyler için, beşeri sermaye değişkenleri ile bireylerin ücret düzeyleri arasındaki ilişkiyi analiz etmektedir. Burada ele alınan beşeri sermaye değişkenleri eğitim yılı ve potansiyel tecrübedir. Logaritmik olarak ele alınan ücretler ile eğitim arasındaki ilişki doğrusal olarak incelenirken, potansiyel tecrübe ile ücret ilişkisinin ise doğrusal olmadığı kabul edilmiş ve potansiyel tecrübe değişkeninin karesi de eklenerek ilişki karesel olarak ele alınmıştır. Standart Mincer ücret denklemi:

$$\text{LnUcret}_i = \alpha_0 + \alpha_1 \text{Eğitim}_i + \alpha_2 \text{Tecrübe}_i + \alpha_3 \text{Tecrübe}_i^2 + \varepsilon_i \quad (13)$$

şeklinde gösterilebilir Standart denkleme ücret düzeyi üzerinde etkili olduğu düşünülen farklı değişkenler eklenerek genişletilebilir (Çağlayan-Akay, Kömürçyan, 2021). Bu çalışmada Mincer ücret denklemi kantil düzeyinde ele alınmıştır. Çalışmada kantil regresyona ait uyum iyiliği ölçülerinden faydalanarak, kısıtlı ve kısıtsız kantil regresyon modelleri arasında model seçimi yapılmıştır. Bu amaçla Mincer ücret denkleminde yer alan potansiyel tecrübe değişkeninin fonksiyonel yapısı ele alınmıştır. Potansiyel tecrübe değişkeni teoride önerildiği gibi doğrusal olmayan bir biçimde, yani değişkenin kendisi ve karesi birlikte modele eklenerek kısıtsız kantil regresyon modeli oluşturulmuştur. Ardından tecrübe değişkeninin karesi modelden çıkartılarak, doğrusal biçimde yer aldığı kısıtlı kantil regresyon modeli elde edilmiştir. Böylece potansiyel tecrübenin ücretle ilişkisinin doğrusal mı yoksa karesel mi olduğu kantil düzeyinde test edilmektedir. Kısıtsız kantil regresyon modelinin genel gösterimi aşağıdaki gibidir:

$$Q_\lambda(\text{LnUcret}_i) = \beta_0 + \beta_1 \text{Eğitim}_i + \beta_2 \text{Tecrübe}_i + \beta_3 \text{Tecrübe}_i^2 + \beta_4 \text{Cinsiyet}_i + \beta_5 \text{Çalışma}_i \text{ Şekli}_i + \varepsilon_{ii} \quad (14)$$

Kısıtlı regresyon modelinde, kısıtsız modeldeki tecrübenin karesi değişkenine kısıt konmuştur:  $\beta_3 = 0$ . Kısıtlı regresyon modelinin genel gösterimi:

$$Q_\lambda(\text{LnUcret}_i) = \alpha_0 + \alpha_1 \text{Eğitim}_i + \alpha_2 \text{Tecrübe}_i + \alpha_3 \text{Cinsiyet}_i + \alpha_4 \text{Çalışma}_i \text{ Şekli}_i + \varepsilon_{2i} \quad (15)$$

şeklinde dir. Modellerde standart Mincer ücret denkleminde yer alan eğitim, tecrübe ve tecrübenin karesi değişkenlerinin yanı sıra kontrol değişkeni olarak cinsiyet ve çalışma şekli değişkenleri de eklenmiştir Çalışmada kullanılan değişkenler ve açıklamaları Tablo 1’de verilmiştir.

Tablo 1. Değişkenler ve Tanımları

Değişken	Tanım
LnUcret	Aylık ücretin doğal logaritması
Eğitim	Eğitim alınan yıl sayısı
Tecrübe	Yaş – Eğitim – 6 olarak hesaplanan potansiyel tecrübe
Tecrübe <sup>2</sup>	Potansiyel tecrübenin karesi
Cinsiyet	1 Erkek; 2 Kadın
Çalışma Şekli	1 Tam zamanlı; 2 Kısmi Zamanlı

Tüm değişkenler Türkiye İstatistik Kurumu (TÜİK) tarafından uygulanan 2016 yılına ait Hanehalkı İşgücü Anketi verilerinden alınmış ve hesaplanmıştır. Bu anket, 15 yaş ve üstündeki bireylere uygulanan ve işgücüne dair soruları kapsayan bir ankettir. TÜİK tarafından yapılan açıklamada, üç binden daha az gözlem ile yapılan analizlerden güvenilir tahminler elde edilemeyeceği ifade edilmiştir. Bu nedenle analizlerin hesaplama maliyetini düşürmek amacıyla veri setinden tesadüfi olarak 3000 gözlem seçilmiş ve analizler için bu alt örnek kullanılmıştır. Veri setinde yer alan bireyler, TÜİK tarafından “çalışma çağı” olarak adlandırılan 15-64 yaş aralığında yer alan ve pozitif bir gelir miktarına sahip olan bireylerdir. Ayrıca potansiyel tecrübe değişkeni en az 1 değerini alacak şekilde kodlanmıştır ve bu değer “bir yıla kadar potansiyel tecrübesi olan bireyleri” ifade etmektedir.

#### 4. Bulgular

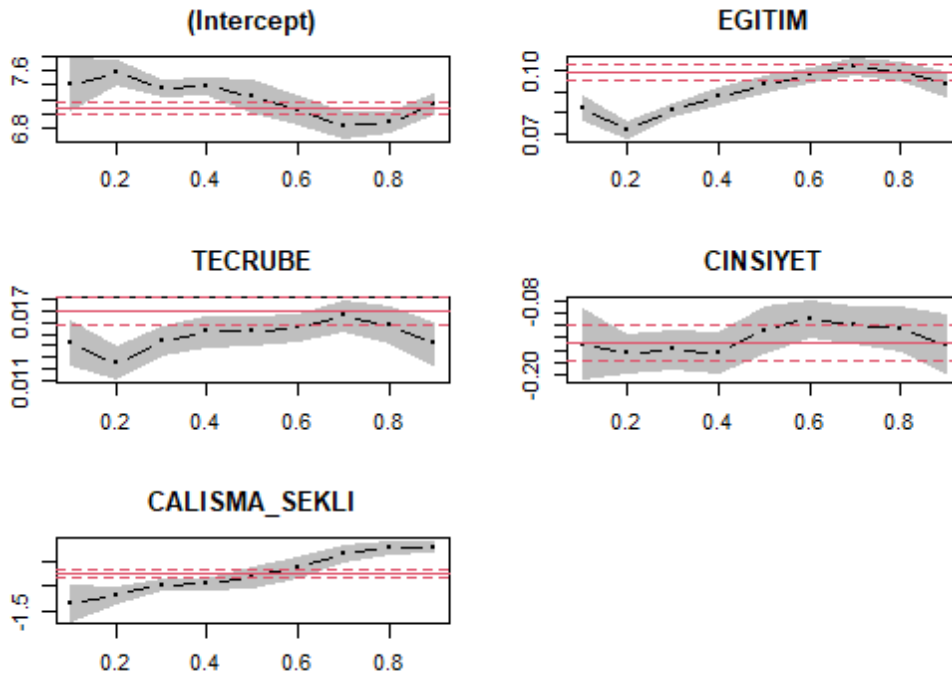
Kantil regresyon modellerinde model seçimi kriterlerini incelemek amacıyla, öncelikle kısıtlı ve kısıtsız Mincer ücret denklemi modelleri tahmin edilmiştir. Daha sonra açıklanan kriterler kullanılarak modeller arasında seçim yapılmıştır. Hem model seçiminin hem de ilgili kriterlerin dağılımı boyunca değişimini inceleyebilmek amacıyla desillerin, yani ondabirlik kantillerin hepsi için model tahmini yapılmıştır. Tecrübenin doğrusal olarak eklendiği kısıtlı kantil regresyon modelinin tahmin sonuçları Tablo 2’de verilmiştir.

Tablo 2. Kısıtlı Model Tahmin Sonuçları

Değişken	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Sabit	7.406	7.578	7.362	7.384	7.246	7.054	6.834	6.875	7.134
Eğitim	0.082	0.072	0.081	0.088	0.094	0.098	0.102	0.100	0.094
Tecrübe	0.014	0.013	0.014	0.015	0.015	0.016	0.017	0.016	0.014
Cinsiyet	-0.151	-0.166	-0.159	-0.164	-0.128	-0.109	-0.119	-0.126	-0.150
Çalışma Şekli	-1.330	-1.167	-0.968	-0.953	-0.821	-0.622	-0.366	-0.250	-0.227

\*\*\* Tüm katsayılar % 1 hata payı ile istatistiksel olarak anlamlıdır.

Tahmin sonuçları incelendiğinde, modelde yer alan değişkenlerin incelenen tüm kantillerde ücret düzeyi üzerinde istatistiksel olarak anlamlı etkilerinin olduğu görülmektedir. Mincer ücret denkleminin temel değişkenleri olan eğitim ve tecrübenin ücret düzeyi üzerinde pozitif etkisi olduğu belirlenmiştir. Kontrol değişkenlerinin katsayıları ise incelenen tüm kantillerde negatif olarak elde edilmiştir. Buna göre ücret dağılımının tamamında kadınların ücreti erkekle göre anlamlı düzeyde daha düşüktür. Benzer şekilde ücret dağılımının incelenen her noktasında kısmi zamanlı çalışanların ücreti tam zamanlı çalışanlara göre daha düşüktür. Bu modelde tecrübe doğrusal olarak ele alındığından, bireylerin potansiyel tecrübesinin artmasının, ücret düzeylerinin devamlı olarak artmasına neden olacağı şeklinde yorumlanmaktadır. Modelde yer alan değişkenlerin katsayılarının kantillere göre değişimi Şekil 2’de verilmiştir.



Şekil 2. Kısıtlı Model Katsayılarının Kantillere Göre Değişimi

Bu şekilde yatay eksen kantilleri ifade ederken, dikey eksen ilgili değişkenin katsayısını ifade etmektedir. Noktalı ve kesikli olarak verilen eğriler ile etrafında yer alan taralı alanlar, kantillerdeki değişken katsayıları ile bu katsayıya ait güven aralıklarını temsil etmektedir. Şekilde yer alan tam ve kesikli doğrular ise modelin EKK ile tahmin edilmesi durumunda ilgili değişken için elde edilecek katsayı ile bu katsayıya ait güven aralığını temsil etmektedir. Eğitim değişkenine ait grafik incelendiğinde, 20. kantilde eğitimin ücret üzerindeki etkisi daha düşükken, yüksek kantillere doğru gidildikçe eğitim değişkeninin ücret üzerindeki etkisi artmaktadır. Oysa EKK tahmin sonucuna bakıldığında, eğitim değişkeninin etkisinin, kantillere göre en yüksek çıkan değer civarında tahmin edildiği görülmektedir. Buradan, eğitimin ücretin farklı düzeyleri için farklılaşan etkisini görebilmek için kantil model kullanmanın gerekliliği anlaşılmaktadır. Cinsiyet değişkeni için bakıldığında, medyana kadarki kantillerde etki daha yüksek iken, daha yüksek kantillerde etkinin azalmaya başladığı görülmektedir. Bir başka deyişle, düşük ücret düzeyleri için erkekler ile kadınlar arasındaki ücret farkı daha fazla iken, yüksek ücret düzeylerinde bu fark azalmaktadır. Ancak çok yüksek ücret düzeylerinde aradaki farkın yeniden artmaya başladığı görülmektedir. Çalışma şekline göre bakıldığında ise düşük ücret düzeyinden yüksek ücret düzeyine doğru gidildikçe tam zamanlı çalışanlar ile kısmi zamanlı çalışanlar arasındaki ücret farkının azaldığı görülmektedir. Temel olarak ele alınan tecrübe değişkeninin ise etkisinin kantiller boyunca pek farklılaşmadığı, ancak artan görünümde olduğu görülmektedir.

Bir sonraki aşamada potansiyel tecrübenin ücret üzerindeki etkisinin doğrusal olmadığı teorisi dikkate alınarak, tecrübenin karesi de modele eklenmiş ve kısıtsız model tahmin edilmiştir. Tecrübenin karesel olarak eklendiği kısıtsız kantil regresyon modeli sonuçları Tablo 3’de verilmiştir.

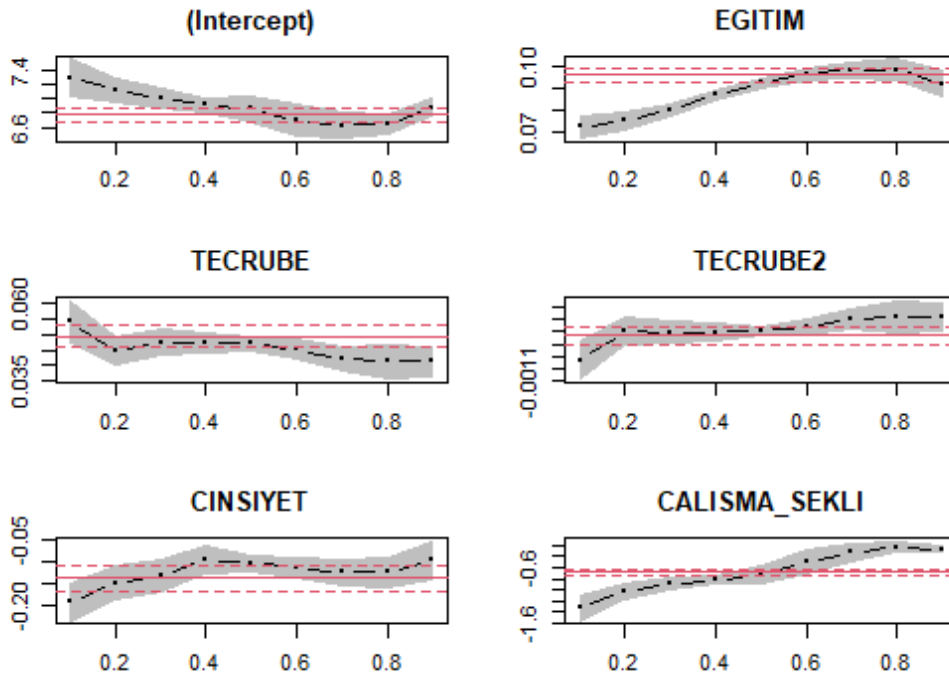
Model tahmin sonuçları incelendiğinde, kontrol değişkenleri olan cinsiyet ve çalışma şekli değişkenlerinin kısıtlı modelde olduğu gibi tüm kantillerde negatif katsayılı bulunduğu görülmektedir. Buna göre ücret dağılımının incelenen her noktasında kadınların ücret düzeyi erkeklerden ve yarı zamanlı çalışanların ücret düzeyi tam zamanlılardan daha düşük bulunmuştur. Eğitimin etkisi ise kısıtlı modeldekine benzer şekilde pozitifdir. Tecrübe değişkeni incelendiğinde, tüm kantillerde değişkenin kendisi pozitif katsayılı bulunurken, karesinin negatif katsayılı olduğu görülmektedir. Buna göre tecrübenin ücret üzerindeki etkisinin ters – U şeklinde olduğu söylenebilir. Bir başka deyişle, belirli bir düzeye kadar tecrübe ücreti arttırırken, bir düzeyden sonra bu etkinin terse döndüğü söylenebilir. Dolayısıyla Mincer ücret denkleminde ortaya konan tecrübe – ücret ilişkisinin geçerli olduğu tespit edilmiştir.

Tablo 3. Kısıtsız Model Tahmin Sonuçları

Değişken	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Sabit	7.291	7.121	7.004	6.911	6.857	6.699	6.631	6.654	6.881
Eğitim	0.072	0.075	0.080	0.087	0.093	0.097	0.098	0.099	0.092
Tecrübe	0.054	0.045	0.048	0.048	0.047	0.045	0.042	0.041	0.041
Tecrübe <sup>2</sup>	-0.0009	-0.0007	-0.0007	-0.0007	-0.0007	-0.0007	-0.0006	-0.0006	-0.0006
Cinsiyet	-0.193	-0.150	-0.133	-0.098	-0.105	-0.116	-0.127	-0.126	-0.099
Çalışma Şekli	-1.327	-1.014	-0.894	-0.824	-0.731	-0.506	-0.326	-0.231	-0.269

\*\*\* Tüm katsayılar % 1 hata payı ile istatistiksel olarak anlamlıdır.

Kısıtsız modelde değişken katsayılarının kantillere göre değişimi Şekil 3'de verilmiştir.



Şekil 3. Kısıtsız Model Katsayılarının Kantillere Göre Değişimi

Şekil incelendiğinde, eğitim, cinsiyet ve çalışma şeklinin kantillere göre etkisi kısıtlı modele benzer yapıdadır. Eğitim değişkeni, kısıtlı modelden farklı olarak en düşük kantilde en düşük etkiye sahipken, yüksek kantillere doğru gidildikçe etkisi artmakta, en yüksek kantillerde ise sabitlenmeye başlamaktadır. Cinsiyetin etkisi düşük kantillerde daha belirginken, yüksek kantillerde daha az ve yakın düzeyde kalmaktadır. Yani düşük ücret düzeylerinde erkekler ile kadınlar arasındaki ücret farklılıkları daha barizdir. Yarı zamanlılar ile tam zamanlılar arasındaki ücret farklılığı, düşük ücret düzeylerinden yüksek ücret düzeylerine doğru gidildikçe azalmaktadır.



Tecrübe değişkeninin dağılımı incelendiğinde ise, kısıtlı modeldekine oranla daha yüksek etkisinin olduğu ancak kısıtlı modelden farklı olarak kantillere göre azalan yapıda olduğu görülmektedir. Tecrübenin karesinin de katsayılarının negatif olduğu dikkate alınrsa, etkisinin yüksek kantillere doğru azaldığı görülmektedir. Buna göre Mincer ücret denkleminde belirtilen doğrusal olmayan ilişki yapısının tüm kantillerde geçerli olduğu ancak yüksek kantillerde bu doğrusal olmayan etkinin azalmaya başladığı söylenebilir. Bu noktada tecrübenin karesinin modele eklenmesiyle, kantiller boyunca modelin uyumunda istatistiksel olarak anlamlı bir gelişme olup olmadığını açıklanan kriterler yardımıyla incelemek mümkündür. Bu kriterler yardımıyla kısıtlı ve kısıtsız modeller arasında seçim yapılabilir ve tecrübenin karesinin modele eklenmesinin anlamlı bir fark oluşturup oluşturmadığı tespit edilebilir. Kısıtlı ve kısıtsız modeller için uygulanan uyum iyiliği ve model seçim kriterlerinin sonuçları Tablo 4’de verilmiştir.

Tablo 4. Kısıtlı – Kısıtsız Modeller İçin Uyum İyiliği ve Model Seçimi Sonuçları

KANTİL	YÖNTEM	R <sup>1</sup> (%)	C <sup>1</sup>	ATWE	AIC	BIC	E_AIC	E_BIC
0.10	KISITLI	5.03	158.52	0.078	-7651.5	-7639.5	-8176.8	-8164.8
	KISITSIZ			0.074	-7805.2	-7790.2	-8423.6	-8408.6
0.20	KISITLI	3.93	122.50	0.117	-6423.3	-6411.3	-6976.7	-6964.7
	KISITSIZ			0.113	-6542.6	-6527.6	-7357.5	-7342.5
0.30	KISITLI	5.24	165.53	0.143	-5826.5	-5814.5	-6609.6	-6597.6
	KISITSIZ			0.136	-5986.1	-5971.1	-7178.9	-7163.9
0.40	KISITLI	5.30	167.53	0.157	-5561.2	-5549.2	-6772.9	-6760.9
	KISITSIZ			0.148	-5723.5	-5708.5	-7869.7	-7854.7
0.50	KISITLI	4.86	153.07	0.159	-5505.4	-5493.4	-8076.7	-8064.6
	KISITSIZ			0.152	-5654.0	-5638.1	-8950.8	-8935.8
0.60	KISITLI	4.02	125.27	0.153	-5635.6	-5623.6	-10131.4	-10119.4
	KISITSIZ			0.147	-5757.5	-5742.5	-10556.7	-10541.7
0.70	KISITLI	3.33	103.00	0.137	-5962.4	-5950.3	-15140.9	-15128.9
	KISITSIZ			0.132	-6062.8	-6047.8	-19974.8	-19959.8
0.80	KISITLI	2.44	74.91	0.110	-6608.8	-6596.8	-16998.1	-16986.0
	KISITSIZ			0.108	-6681.9	-6666.9	-15272.3	-15257.3
0.90	KISITLI	2.24	68.55	0.071	-7945.7	-7933.7	-12755.5	-12743.5
	KISITSIZ			0.069	-8012.6	-7997.6	-11864.5	-11849.5

Tablo değerleri sırasıyla değerlendirilirse,  $R^1$  uyum iyiliği ölçüsüne göre, tecrübenin karesinin modele eklenmesi düşük kantillerde modelin uyumunu %5 civarında geliştirirken, yüksek kantillere doğru gidildikçe bu gelişim %2 civarına kadar düşmektedir. Bir başka deyişle, düşük ücret düzeylerinde tecrübenin karesinin eklenmesi modeli daha belirgin olarak geliştirirken, yüksek ücret düzeylerinde daha az geliştirmektedir.

$C^1$  testi sonuçlarına göre tüm kantillerde test istatistiği değerlerinin oldukça yüksek ve tablo kritik değeri olan  $F_{1,2994} = 3.84$  değerinden daha büyük bulunduğu görülmektedir. Buna göre tüm kantiller için, tecrübenin karesinin modele eklenmesiyle uyum iyiliğinde anlamlı bir gelişme olmadığı sıfır hipotezi reddedilmiştir. Tüm kantillerde kısıtsız modelin kısıtlı modele göre daha yüksek uyum iyiliğine sahip olduğu sonucuna varılmıştır. Bununla birlikte test istatistiği değerlerinin yüksek kantillere doğru gidildikçe düşük kantillere göre oldukça azaldığı da dikkat çekmektedir.

ATWE değerleri incelendiğinde, kısıtsız modele ait değerlerin tüm kantillerde kısıtlı modele göre daha düşük olduğu, yani kısıtsız modelin kısıtlı modele tercih edilebileceği görülmektedir. Ağırlıklandırılmış kantil hata ortalamaları kısıtsız modellerde kısıtlı modele daha düşük olmakla birlikte, bu farkın en yüksek kantillerde daha düşük olduğu görülmektedir.

Bilgi kriterleri incelendiğinde, AIC ve BIC kriteri sonuçlarının daha önceki ölçüler ile uyumlu olduğu görülmektedir. İncelenen tüm kantillerde kısıtsız modele ait AIC ve BIC değerleri kısıtlı modele göre daha düşüktür, dolayısıyla kısıtsız modelin tercih edilmesi uygundur. Etkin bilgi kriterlerinde ise sonuçlarda farklılıklar söz konusudur. E\_AIC ve E\_BIC kriterleri incelendiğinde, düşük kantillerden yüksek kantillere kadar, önceki sonuçlara benzer olarak kısıtsız modelin değerleri kısıtlı modele göre daha düşük bulunmuştur, dolayısıyla kısıtsız modelin tercih edilmesi uygundur. Ancak en yüksek iki kantilde, diğer kriterlerden farklı olarak, kısıtlı modelin değerleri daha düşük bulunmuştur ve kısıtlı modelin tercih edilmesi sonucuna varılmıştır. Bir başka deyişle, bu iki kriter gereği yüksek ücret düzeylerine kadar tecrübenin ücret üzerindeki etkisi doğrusal olmayan şekilde iken, en yüksek ücret düzeylerinde bu doğrusal olmayan etki anlamsızlaşmaktadır ve doğrusal olarak ele alınabilir. Bu sonuç, daha önceki ölçülerde yüksek kantillerde model gelişiminin azalması ile uyumludur ve klasik bilgi kriterlerinden farklı olarak tecrübenin karesinin modele eklenmesinin en yüksek kantillerde anlamlı bir katkısı olmadığını ifade etmektedir.

## 5. Sonuç

Klasik regresyon modellerinde uyum iyiliğini incelemek ve model seçimi yapmak amacıyla kullanılan bazı temel ölçüler, benzer şekilde kantil regresyon için uyarlanmıştır. Bu çalışmada, kantil regresyon modellerinde uyum iyiliği ve model seçimi için kullanılacak  $R^1$  kriteri,  $C^1$  testi, ATWE ile bilgi kriterleri AIC ve BIC tanıtılmıştır. Ayrıca, bilgi kriterlerinin etkinliğini arttırmak amacıyla uyarlanmış bir kayıp fonksiyonu yardımıyla yeniden hesaplanan E\_AIC ve E\_BIC kriterlerine de çalışmada yer verilmiştir. Açıklanan kriterlerin uygulamada kullanımlarını örneklemek amacıyla Mincer ücret denklemi uygulaması yapılmıştır. Mincer ücret denkleminde bireylerin potansiyel tecrübeleri ile ücretleri arasında doğrusal olmayan bir ilişkinin var olduğu öngörülmektedir. Çalışmada, Mincer denkleminde yer alan potansiyel tecrübenin karesi değişkenine kısıt konarak, potansiyel tecrübenin ücret üzerindeki etkisinin doğrusal olup olmadığı kantil düzeyinde incelenmiştir. Yuvalanmış olarak ele alınan kısıtlı ve kısıtsız kantil modelleri desiller için tahmin edilmiş ve açıklanan kriterler yardımıyla modellerin uyum iyiliği incelenip model seçimi yapılmıştır. Sonuçta, ilk dört kriter gereği incelenen tüm kantillerde kısıtsız model kısıtlı modele tercih edilmiştir. Bir başka deyişle, tüm kantillerde potansiyel tecrübenin karesinin modele eklenmesinin modelin uyum iyiliğine anlamlı bir katkı sağladığı ve dolayısıyla incelenen ilişkinin doğrusal olmadığı sonucuna ulaşılmıştır. Ancak bu katkı yüksek kantillere doğru gidildikçe azalan bir seyir göstermektedir. Nitekim daha etkin iki bilgi kriterine göre, diğer kriterlerden farklı olarak en yüksek iki kantilde kısıtlı model tercih edilmiştir. Buna göre, en yüksek kantillerde potansiyel tecrübenin ücretle ilişkisi doğrusal hale gelmekte ve değişkenin karesinin modele eklenmesinin modelin uyum iyiliği üzerinde belirleyici bir katkısı olmamaktadır. Burada toplu olarak verilen kriterlerin kantil regresyon uygulamalarında kullanılacak olması nedeniyle, çalışmanın yerli literatürde daha sonraki çalışmaları destekleyici nitelikte olduğuna inanılmaktadır.

## **Kaynakça**

- Chernozhukov, V. ve Fernández-Val, I. (2005). Subsampling Inference on Quantile Regression Processes. *Sankhyā: The Indian Journal of Statistics*, 67(2): 253-276.
- Chow, G. (1960). Tests of Equality between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28: 591-605.
- Çağlayan-Akay, E. ve Kömürbakan, F. (2021). Koşullu ve Koşulsuz Kantil Regresyon Modelleri Türkiye’de Ücret Eşitsizliği Hakkında Farklı Ne Söylüyor ? *Journal of Economy Culture and Society*, 64: 1-21.
- Davino, C., Furno, M. ve Vistocco, D. (2014). Quantile Regression: Theory and Applications. West Sussex: *John Wiley & Sons, Ltd.*
- Dong, C., Li, G. ve Feng, X. (2019). Lack-of-Fit Tests for Quantile Regression Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(3): 629-648.
- Furno, M. (2011). Goodness of Fit and Misspecification in Quantile Regressions. *Journal of Educational and Behavioral Statistics*, 36(1): 105-131.
- Güriş, S. ve Çağlayan-Akay, E. (2018). Ekonometri Temel Kavramlar. İstanbul: *DER Yayınları (5. Baskı)*.
- Haupt, H., Kagerer, K. ve Schnurbus, J. (2011). Cross-Validating Fit and Predictive Accuracy of Nonlinear Quantile Regressions. *Journal of Applied Statistics*, 38(12): 2939-2954.
- He, X. ve Zhu, L.-X. (2003) A Lack-of-Fit Test for Quantile Regression. *Journal of the American Statistical Association*, 98(464):1013-1022.
- Hurvich, C. M. ve Tsai, C. L. (1990). Model Selection for Least Absolute Deviations Regression in Small Samples. *Statistics & Probability Letters*, 9(3): 259-265.
- Koenker, R. (2005). Quantile Regression. New York: *Cambridge University Press.*
- Koenker, R. ve Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1): 33–50.
- Koenker, R. ve Machado, J.A.F. (1999). Goodness of Fit and Related Inference Processes for Quantile Regression. *J. Amer. Statist. Assoc.* 94(448): 1296–1310.
- Machado, J. A. F. (1993). Robust Model Selection and M-Estimation. *Econometric Theory*, 9: 478–493.
- Mincer, J. (1974). Schooling, Experience and Earnings. New York: *National Bureau of Economic Research.*
- Pham, H. (2019). A New Criterion for Model Selection. *Mathematics*, 7(12), 1215: 1-12.
- Shin, W., Kim, M. ve Jung, Y. (2021). Efficient Information-Based Criteria for Model Selection in Quantile Regression. *Journal of the Korean Statistical Society*, 1-37.



## Marital Status Differences in Women's Labor Force Participation in Turkey: Nonlinear Decomposition Approach

Fulden KÖMÜRYAKAN

Bandırma Onyedi Eylül University / Research Assistant

[fkomuryakan@bandirma.edu.tr](mailto:fkomuryakan@bandirma.edu.tr)

Orcid No: 0000-0002-4034-513X

### Abstract

The present study investigates the differences in the likelihood of prime working-age married and unmarried women's labor force participation and examines the labor force participation gap between them. Using Turkish Household Budget Survey data of 2018, this study estimates the probit model and employs a nonlinear decomposition method. To the best of my knowledge, this is the first study to address the labor force participation gap via the nonlinear decomposition method. The findings suggest that there is a 14% labor force participation gap between married and unmarried women in favor of unmarried women. Furthermore, nonlinear decomposition findings reveal that if the married and unmarried women had the same characteristics in terms of age, education, household economic status, and household demographic characteristics, 96.92% of the labor force participation gap between them would have disappeared. By understanding the driving forces behind the labor force participation gap, this study may help to develop policies to reduce the gap and increase married women's involvement in the economy.

*Keywords: Marital Status, Nonlinear Decomposition, Probit Model, Turkey, Women's Labor Force Participation.*

---

**Corresponding Author / Sorumlu Yazar:** 1-Fulden KÖMÜRYAKAN, Bandırma Onyedi Eylül Üniversitesi, Faculty of Economics and Administrative Sciences, Econometrics

**Citation / Atıf:** KÖMÜRYAKAN F. Marital Status Differences in Women's Labor Force Participation in Turkey: Nonlinear Decomposition Approach. İstatistik Araştırma Dergisi, 2021, 11 (2), 14-29.

## Türkiye'de Kadınların İşgücüne Katılımında Medeni Durum Farklılıkları: Doğrusal Olmayan Ayırıştırma Yaklaşımı

### Özet

Bu çalışma, çalışma çağındaki evli ve evli olmayan kadınların işgücüne katılım olasılıkları arasındaki farklılıkları ve aralarındaki işgücüne katılım farkını incelemektedir. 2018 yılı Türkiye Hanehalkı Bütçe Anketi verilerini kullanan bu çalışmada, probit modeli tahmin edilmekte ve doğrusal olmayan ayırıştırma yöntemi kullanılmaktadır. Bilgim çerçevesinde bu çalışma, işgücüne katılım farkını doğrusal olmayan ayırıştırma yöntemiyle ele alan ilk çalışmadır. Çalışmanın bulguları, evli ve evli olmayan kadınlar arasında, evli olmayan kadınlar lehine, %14'lük bir işgücüne katılım farkı olduğunu ortaya koymaktadır. Doğrusal olmayan ayırıştırma bulguları ise evli ve evli olmayan kadınların yaş, eğitim, hanenin ekonomik durumu ve hanehalkı demografik özellikleri bakımından aynı koşullara sahip olmaları durumunda, aralarındaki işgücüne katılım farkının %96,92'sinin ortadan kalkacağını ortaya koymaktadır. Bu çalışma, işgücüne katılım farkının arttıran belirleyicileri tespit ederek, işgücüne katılım farkını azaltmak ve evli kadınların ekonomiye katılımını artırmak için politikalar geliştirmeye yardımcı olabilir.

*Anahtar Sözcükler: Doğrusal Olmayan Ayırıştırma, Kadınların İşgücüne Katılımı, Medeni Durum, Probit Modeli, Türkiye.*

### 1. Introduction

Women participating in the labor force is an important driver of sustainable economic growth and development (Verick, 2014). Although an increase in women's labor force participation especially after the late 1940s in some labor markets, low labor force participation of women is still a serious problem especially for the emerging market economies. Therefore, low women's labor force participation is a crucial and current phenomenon that economies need to pay special attention to. Women's labor force participation has been of interest to many economists and a subject of many studies over the years across the countries, and most of them point out a low women labor supply. The main barrier towards women's labor force participation originates from the conventional social norms and cultural values of the society. In fact, the papers examining the women's labor force participation with social and cultural values (e.g., (Atasoy, 2016; Contreras & Plaza, 2010; Diwan & Vartanova, 2017; Gedikli, 2014; Goksel, 2013)) indicate that social norms, conservatism, traditionalism, machismo, patriarchy, and conventional cultural values influence the women's decision to participate in the labor force. Along with the social norms and cultural values, the institutional settings not covering family and child care leads to fewer women in the labor force with the idea of women should stay home, engage in chores, and take care of the children. In this case, the structure of the household, social, economic, and demographic features of women can be considered as the major determinants of women's labor force participation. In fact, many studies in the literature find a significant relationship between the likelihood of labor force participation and marital status, age, education, number of children, etc.

Most of the studies on labor force participation prove that there is a significant gap in labor force participation between men and women. However, this is not the only gap; there is also an important gap in labor force participation between married and unmarried women due to different reasons, other than the reasons every woman faces regardless of marital statutes such as social norms and conventional cultural values of the societies. One of the pioneer studies by Mincer (1962) brings forward married women do not only choose between participating in the labor force or leisure time but also have to consider the unpaid work in the home such as chores and childcaring. In fact, the existing literature on the labor force participation of women by marital status is considerably wide and the studies examining different labor markets documents a significant labor force participation gap between married and unmarried women. For instance, Lee, Jang, & Sarkar (2008) analyze the relationship between Korean women's marital status and labor force participation and find that married women are less likely to participate in the labor force than their unmarried counterparts. Contreras & Plaza (2010) examine the women's labor force participation determinants in the Chilean labor market with both classical and cultural values. They provide several important findings in terms of labor force participation determinants and findings related to the marital status indicate that the likelihood of labor force participation significantly reduces in case of women having a partner. A recent paper by Xiao & Asadullah (2020) analyzes the gender gap in labor force participation between Chinese men and women and the results indicate that being married reduces the probability of working for Chinese women.

Findings of another recent paper by Petrakis (2021) for the Greece labor market suggest that marital status has a significant and reducing effect on women's labor force participation decision.

Since Turkey has a conventional household structure in which individuals get married and have children in the early stages of their lives, marriage is one of the most significant constraints on women's labor supply. In fact, Ilkkaracan (2012) reveals that marriage is a more significant pattern on women's labor force participation than the presence of small children as a reason of gender roles. As Dayioglu & Kirdar (2010) indicate, since being married is considered a universal thing in Turkey, examining married women's labor force participation decision carries importance in terms of the labor market. Given the evidence that women's marital status is highly correlated to labor force participation, this study has two main objectives. First, it examines the differences in the likelihood of labor force participation of married and unmarried<sup>1</sup> women and diagnoses the patterns that could affect labor force participation decisions with a special focus on marital status. Second, it analyzes the labor force participation gap between them and reveals the determinants that could wider or narrow it. In order to achieve these objectives, this study adopts a probit model and nonlinear decomposition method. Although there is a wide literature on women's labor force participation in Turkey, to the best of my knowledge, this is the first study analyzing the labor force participation gap between married and unmarried women via the nonlinear decomposition method.

The next section covers women's labor force participation in Turkey and builds the background. Section 3 explains the dataset and sample used in this study. Section 4 contains the methodological issues and Section 5 reveals the empirical findings. Lastly, Section 6 provides further discussion and concludes remarks.

## 2. Women in Turkish Labor Force

This section builds an institutional background via national statistics and existing literature related to women's labor force participation in Turkey. Figure 1 illustrates the prime working-age<sup>2</sup> women's labor force participation and unemployment rates compiled by the Turkish Statistical Institute (TURKSTAT) in 2018.

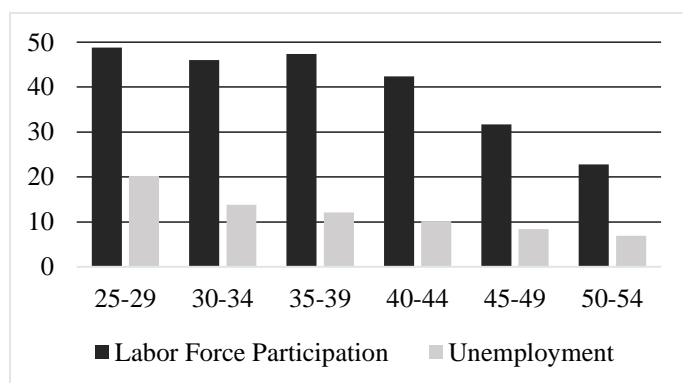


Figure 1. Women's labor force participation and unemployment rates

Source: TURKSTAT labor force statistics.

According to the statistics, younger women are more likely to participate in the labor force. Since the younger women who are especially in their mid-twenties seek jobs more, the unemployment rate is higher for them. The likelihood of labor force participation and therefore the possibility of being unemployed decreases as women get older. Turkey has a young and dynamic population, however, the average of the recent rates in Figure 1 reveals that almost 60 percent of the prime working-age women do not join the labor force and only 12 percent of them are unemployed. This suggests that 48 percent of women do not join the labor force in 2018 for several reasons other than unemployment even though they are in their prime working-age period. Moreover, the labor force participation rates of women are still considerably below most of the economies. World Bank's World Development Indicators reveal that women's labor force participation rates<sup>3</sup> in the European Union, the United States, the average of the world are 67.51, 67.23, and 52.54 percent, respectively. Surprisingly, despite the strong

<sup>1</sup> Unmarried women are single, divorced, separated, or widows.

<sup>2</sup> The prime working-age interval determined by the International Labour Organization (ILO) is 25 and 54.

<sup>3</sup> Calculated for the women aged between 15 and 64 by the ILO.

workforce with the young and dynamic population of Turkey, women's labor force participation rate is considerably below both developed economies and the average of the world.

The possible several reasons for women not participating in the labor force are maybe conventional cultural values, social norms influencing marriage in the early stages of their lifetime, higher fertility rates, lower educational attainment, institutional settings, household, and spouse characteristics (Verick, 2014). According to the marriage statistics of Turkey obtained from TURKSTAT, women's average age at first marriage is 24.8 in 2018 which can be considered as the beginning of the prime working-age period. Marriage is a reducing pattern on labor force participation of women, therefore, if marriage reduces the labor force participation of women, Turkish women get married instead of working in their mid-twenties. A similar finding from Palaz (2010) indicates that women join the labor market after graduation in their early twenties, however, the labor force participation decreases in their 24 and 34 due to marriage and childcare duties. Furthermore, Kasnakoglu & Dayioglu (1997) investigate the female labor force participation by exploiting the 1987 survey data and reveal that being married and having children tend to reduce the likelihood of labor force participation of Turkish women. Ozar & Gunluk-Senesen (1998) bring out that the role of wife and mother in the household are the important determinants of Turkish women's non-participation in the labor force. Tansel (2002) examines the labor force participation rates of both men and women by regions of Turkey. The rates of labor force participation by marital status indicate that unmarried women are more likely to participate than married women between 1998-2000 period. According to the findings of Ince & Demir (2006), marriage negatively affects Turkish women's labor force participation between 1988 and 2002 and causes women to be economically inactive and unpaid family workers. Dayioglu & Kirdar (2010) perform detailed analyses for the women's labor force participation in Turkey between 1988 and 2006. The findings indicate that married women are less likely to participate in the labor force than unmarried women. By examining the women's labor force participation for 1998, 2000, and 2008 Turkish data, Ilkcaracan (2012) defines marriage as an obstacle to women's labor force participation. However, marriage has become relatively less binding to labor force participation over the years. Varol (2017) examines the World Values Survey of 2007 for the Turkish labor force and suggests that since being married causes more responsibility to women, marriage harms the decision of women's labor force participation. Baslevant & Onaran (2003) adopt a different viewpoint and analyze the effect of the employment status of husbands on wives' labor force participation decisions in Turkey. The findings suggest that the two decisions are negatively correlated and if the husband is employed, women do not join the labor force. Similar to the international studies on women's labor force participation, national studies support that there is a significant gap between married and unmarried women in labor force participation.

Besides relatively early marriages in Turkey, demographic statistics compiled by TURKSTAT show that the fertility rate is 15.4 percent and the average total fertility in terms of the number of children is 2 per woman in 2018. Since fertility is a crucial reducing factor on women's labor force participation decisions with long maternity leaves and expensive and inaccessible childcaring, these relatively high fertility rates may lead women not to work. If we look at the education rates as another significant pattern on labor force participation, the rates indicate that only 23 percent of the prime working-age women have bachelor's or postgraduate degrees in 2018. 77 percent of the women of their prime working-age do not have a qualified degree in a particular labor force and consequently, they may have to decide between either obtaining unqualified jobs or being an unpaid family worker.

### **3. Data Description and Sample**

Household Budget Survey (HBS) conducted by the TURKSTAT includes a rich data covering information about social, economic, demographic patterns of the households in Turkey. Therefore, HBS data are convenient for this study which analyzes women's decisions to take place in the labor force who have different characteristics. HBS has been applied since 2002 by TURKSTAT and this study exploits more recent data that was administered in 2018. A stratified two-stage cluster sampling method for the households in Turkey was employed to collect HBS data and the survey was administered to 15,551 households in 2018. By following the previous studies on women's labor force participation, to receive more relevant results, the sample of this study was restricted to the prime working-age women. The main reason for this is to obtain more accurate results by excluding women who do not take part in the labor force due to preferences such as education or retirement and prevent the misleading effect of age. After applying the restrictions, the sample of this study contains 8,582 women aged between 25 and 54. The HBS data used in this study are licensed and can only be used with permission from TURKSTAT.

### **4. Methodology**

This study adopts a probit model and nonlinear decomposition methods. The following sub-sections contain methodological explanations of these methods.

#### 4.1. Probit Labor Force Participation Model

As discussed above, one of the objectives of this study is to determine the probability of women participating in the labor force, diagnose the patterns that could affect this probability, and examine the differences between married and unmarried women in terms of labor force participation. Therefore, this study estimates the probit model of labor force participation for married and unmarried women separately. The dependent variable of the model is a categorical variable that takes the value of 1 if the woman currently works or does not work but the connection with the working place is still on<sup>4</sup> in the last week of the survey. The labor force participation variable takes the value of 0 if the woman does not join the labor force due to several reasons. The likelihood of labor force participation for married and unmarried women is estimated by  $\Phi(X\beta)$ , where  $X$  is the  $N \times K$  matrix of explanatory variables,  $\Phi(\cdot)$  is the cumulative standard normal distribution probability density function, and  $\beta$  is the  $K \times 1$  vector of coefficients. The probit model constructed in this study can be expressed as below.

$$p_i = \Phi \left( \beta_0 + \sum_{j=1}^5 \delta_j Age_{ij} + \sum_{j=1}^5 \gamma_j Education_{ij} + \beta_1 HouseholdWage_i + \beta_2 NonlaborIncome_i + \beta_3 Children(0 - 5)_i + \beta_4 Children(6 - 14)_i + \beta_5 ReferencePerson_i \right) \quad (1)$$

The left side of the equation ( $p_i$ ) indicates the conditional probability of participation in the labor force on explanatory variables. The probit model constructed in the Equation 1 contains age categories, education levels, household economic status, and household demographic variables as the determinants that could affect labor force participation decisions. Women may have different preferences obtaining jobs in different periods of their lives. To capture this specific relationship between the age and the likelihood of joining the labor force age categories are divided into six categorical variables as 25-29, 30-34, 35-39, 40-44, 45-49, and 50-54. Education increases the cost of not participating in the labor force, thus, it raises the likelihood of seeking a job (Chin, 1995). Therefore, since education can be considered as an investment in terms of human capital, labor force participation is highly determined by education. Insomuch that educational attainments have more influence on women's decision to labor force participation than men (Komuryakan & Yilgor, 2021). In order to receive more detailed results about the influence of women's education on their likelihood of labor force participation, schooling is divided into categorical variables as no schooling, compulsory education, high school education, associate degree, bachelor's degree, and postgraduate degree.

Household economic structures along with the conservative and patriarchal values of society may cause women to stay out of the labor force. In most conventional households, women are encouraged to stay home, do household chores, and take care of children. This structure may cause if the household annual wage earned from the other people in the household is high, especially by men, the likelihood of labor force participation of the women may be lower. Therefore, the natural logarithm of the annual household wage excluding woman is added to the labor force participation model to analyze the probability of labor force participation of the woman in the case of a higher wage of the household other than woman. Another income determinant that could lower the possibility of women's working is the nonlabor income of the household. If the nonlabor income such as scholarship, interest, yield, fund, etc. other than wages is higher, women may not prefer to work. Hence, the labor force participation model contains the natural logarithm of the annual household nonlabor income including women's income. Household demographic features such as the number of children are one of the most significant determinants of the women's decision to participate in the labor market. The presence and the age of the children in the household are some of the major determinants of the women's decision to work. Since the presence of the preschooler and the older children have different impacts on the labor force participation, the model contains two continuous variables as the number of children aged between 0-5 and 6-14. Lastly, the model contains a categorical variable that denotes if the woman is the reference person in the household. The reference person is the person who is responsible for most of the expenditure of the household and may be referred to as the breadwinner and the head of the household. Being the head of the household is a significant factor in the decision of labor force participation.

It is highly expected that the impacts of the determinants that could affect the labor force participation decision vary by the marital status of the women. Age, education, household economic status, and household demographic structure may differ for married and unmarried women. For instance, education may be more significant for unmarried women whereas the number of children in the household may be more crucial for married women since childcaring may be considered her duty. Therefore, the probit model is constructed for separately married and

<sup>4</sup> The women who are on maternity leave, suspended, do not work due to such reasons as slowdown, on vacation, etc. are considered as working.



unmarried women to bring out the differences in the labor force participation decisions of married and unmarried women.

## 4.2. Nonlinear Decomposition

Decomposition is one of the most important statistical tools to analyze the differences between categories such as gender, occupation, age, marital status, etc. Decomposition was firstly introduced to the literature for the standard linear regression by Oaxaca (1973) and Blinder (1973). Therefore, this decomposition technique is mostly known as Oaxaca-Blinder (OB) decomposition. Over the years, the OB decomposition method was also developed for nonlinear regression models such as logit, probit, complementary log, etc. In order to decompose the differences in labor force participation between married and unmarried women after the estimation of the probit model aforementioned above, this study adopts the nonlinear decomposition method based on Yun (2004, 2005a, 2005b, 2008). This nonlinear decomposition method has two contributions to the literature. First, it handles the path dependency problem which originates because of a sensitivity of the nonlinear models in terms of the entrance order of the explanatory variables in the decomposition. Second, it overcomes the identification problem with the reference categories of explanatory dummy variables (Yun, 2005). After estimating the labor force participation (*LFP*) model for both married (*m*) and unmarried (*u*) women, the differences in the average likelihood of labor force participation between married and unmarried may be decomposed as follows:

$$LFP = \Phi(X\beta) \quad (2)$$

$$(\overline{LFP}_u - \overline{LFP}_m) = \overline{\Phi(X_u\beta_u)} - \overline{\Phi(X_m\beta_m)} \quad (3)$$

$$(\overline{LFP}_u - \overline{LFP}_m) = \{\overline{\Phi(X_u\beta_u)} - \overline{\Phi(X_m\beta_u)}\} + \{\overline{\Phi(X_m\beta_u)} - \overline{\Phi(X_m\beta_m)}\} \quad (4)$$

where *LFP* is a  $N \times 1$  dependent variable vector, *X* is an  $N \times K$  matrix of explanatory variables,  $\beta$  is a  $K \times 1$  vector of coefficients, and  $\Phi$  is the cumulative normal distribution function. The first and second components in the right hand of Equation 4 indicate characteristics effects and coefficients effects, respectively. Basically, the characteristics effect reflects the composition in labor force participation by marital status and indicates the contributions to the marital status differences in labor force participation due to explanatory variables. The coefficient effect, however, indicates contributions associated with marital status differences who share the same characteristics in the differences in the likelihood of labor force participation by marital status. Furthermore, in order to calculate the contributions of these effects, after calculating the overall decomposition from Equation 4, the equations below need to be employed

$$W_{\Delta X}^k = \frac{(\bar{X}_u^k - \bar{X}_m^k)\beta_u^k}{(\bar{X}_u - \bar{X}_m)\beta_u} \quad (5)$$

$$W_{\Delta \beta}^k = \frac{\bar{X}_m^k(\beta_u^k - \beta_m^k)}{\bar{X}_m(\beta_u - \beta_m)} \quad (6)$$

$$(\overline{LFP}_u - \overline{LFP}_m) = \sum_{k=1}^K W_{\Delta X}^k [\overline{\Phi(X_u\beta_u)} - \overline{\Phi(X_m\beta_u)}] + \sum_{k=1}^K W_{\Delta \beta}^k [\overline{\Phi(X_m\beta_u)} - \overline{\Phi(X_m\beta_m)}] \quad (7)$$

where  $\sum_{k=1}^K W_{\Delta X}^k = \sum_{k=1}^K W_{\Delta \beta}^k = 1$  (Pritchett & Yun, 2009). According to the explanations, this study aims to analyze how much of the differences in labor force participation between married and unmarried women can be explained by differences in the determined indicators above.

## 5. Empirical Findings

The following sub-sections contain a four-fold analysis: an examination of descriptive statistics, the estimation of probit model, findings of the nonlinear decomposition method, and conduction of the sensitivity analysis.

### 5.1. Descriptive Statistics

Table 1 reveals the summary statistics of the explanatory variables explained in Equation 1 by women's labor force participation and marital status. The table also reports the Variance Inflation Factor (VIF) values to analyze the relationship between the explanatory variables and diagnose the possible multicollinearity.

Table 1. Descriptive statistics and VIF by labor force participation and marital status

	<i>LFP = 1</i>		<i>LFP = 0</i>		<i>VIF</i>	
	<i>Married</i>	<i>Unmarried</i>	<i>Married</i>	<i>Unmarried</i>	<i>Married</i>	<i>Unmarried</i>
<i>Age categories</i>						
25 - 29	0.09	0.33	0.14	0.27	2.12	3.20
30 - 34	0.17	0.17	0.18	0.13	2.47	2.12
35 - 39	0.21	0.13	0.19	0.13	2.38	1.91
40 - 44	0.22	0.13	0.17	0.13	2.04	1.81
45 - 49	0.18	0.16	0.16	0.15	1.79	1.83
50 - 54	0.13	0.08	0.16	0.19	-	-
<i>Education degrees</i>						
No schooling	0.11	0.05	0.18	0.19	-	-
Compulsory	0.54	0.28	0.59	0.42	2.15	2.64
High school	0.13	0.19	0.15	0.17	1.79	2.2
Associate	0.04	0.09	0.03	0.06	1.25	1.65
Bachelor's	0.15	0.35	0.04	0.15	1.55	2.7
Postgraduate	0.03	0.04	0.00	0.01	1.09	1.24
<i>Household economic status</i>						
Household annual wage* (in logarithmic Turkish Liras)	7.30	4.11	7.79	4.85	1.09	1.17
Household annual nonlabor income* (in logarithmic Turkish Liras)	5.54	7.35	5.51	9.06	1.17	1.15
<i>Household demographic characteristics</i>						
Children aged 0-5*	0.34	0.09	0.53	0.15	1.41	1.07
Children aged 6-14*	0.75	0.23	0.87	0.31	1.31	1.12
Reference person	0.05	0.41	0.03	0.32	1.03	1.50
Observations	2,606	720	4,544	712		
<i>Diagnostics tests</i>						
Pearson chi-square	96.16 (0.000)					
Fisher's exact test	(0.000)					
Notes: (i) * denotes continuous variable; other variables are categorical. (ii) The probability of the tests is in brackets.						

The observation numbers show that the labor force participation rates for married and unmarried women are 36% and 50%, respectively. According to the ratios, there is a 14% gap between married and unmarried women in the labor force participation. Pearson independence chi-square and Fisher's exact tests which were used to statistically examine this labor force participation gap between married and unmarried women indicate that there is a significant statistical relationship between the women's labor force participation and their marital status. As for the VIF values in Table 1, there is no significant multicollinearity among explanatory variables detected.

Table 1 reveals that younger married women are less likely to take place in the labor force may be because of having children at mid-twenties, however, as they get older, their likelihood of working increases. As for unmarried women, this situation is reversed and unmarried women in the prime working-age group are more likely to take place in the labor force at early ages. Education categories show that most of the prime working-age married women have only compulsory education in 2018 regardless of taking place in the labor force. There is a slight difference between working and not working married women in terms of compulsory education and not working married women almost 60% likely to have compulsory education. While married women have less education, unmarried working women have 35% likely to have bachelor's degrees. This ratio is 15% for unmarried women

who do not work whereas the ratio is only 4% for married women. Higher education drives women to join the labor force more. Within the context of the household’s economic situation, more income in the household causes less women’s labor force participation since the women’s labor force participation is considered as a choice rather than a necessity for both married and unmarried women. This means that if other people especially men in the household earn, women do not have to work so they can stay home and engage with the chores and childcaring. According to the household characteristics, if the number of children in the household is high, the married woman is 1.4 times more likely not participating in the labor force. This ratio is lower for unmarried and working married women. As expected, being a reference person in the household increases the likelihood of working for both married and unmarried women. However, being a reference person in the household is highly correlated with marital status; unmarried working women are referred to 41% likely to be a reference person whereas only 5% of married working women are considered as a reference person in the household. This structure only changes for unmarried women.

The questionnaire used in this study contains a question as “reason for not seeking a job” regarding the reasons for not participating in the labor force. This question has eleven different answers such as “found a job-established a business; but waiting to start”, “doing a seasonal work or waiting to be back to the old job”, “previously sought for a job but could not find”, “does not believe that she can find a suitable job in her skills”, “education or training continues”, “engage with the household chores (taking care of children, elderly, sick, etc., are included)”, “retired”, “disabled or ill”, “elderly”, “income recipient”, or “other”. Engaging with the household chores is the highest selected answer to this question with the ratios of 91.5% and 51.1% for married and unmarried women, respectively<sup>5</sup>. Although the highest selected reason is the same for married and unmarried women, there is a significant difference between the ratios. Married women are most likely not in the labor force to do chores, take care of children, or take care of other people in the household.

## 5.2. Probit Model Findings

Table 2 reveals the estimation results of the probit model for married and unmarried women.

Table 2. Probit regression findings by marital status

	Married (n = 7,150)		Unmarried (n = 1,432)	
	OR	ME	OR	ME
<b>Age categories</b>				
25 - 29	0.0932 (0.0720)	0.0328 (0.0253)	0.3215** (0.1435)	0.1086** (0.0481)
30 - 34	0.2418*** (0.0667)	0.0852*** (0.0234)	0.5138*** (0.1498)	0.1735*** (0.0499)
35 - 39	0.3207*** (0.0619)	0.1130*** (0.0217)	0.3311** (0.1475)	0.1118** (0.0495)
40 - 44	0.3431*** (0.0576)	0.1209*** (0.0201)	0.4175*** (0.1411)	0.1410*** (0.0471)
45 - 49	0.2198*** (0.0555)	0.0775*** (0.0195)	0.4683*** (0.1352)	0.1581*** (0.0450)
50 – 54 (reference)				
<b>Education degrees</b>				
No schooling (reference)				
Compulsory	0.1235*** (0.0469)	0.0435*** (0.0165)	0.4915*** (0.1252)	0.1660*** (0.0415)
High school	0.1163* (0.0597)	0.0410* (0.0210)	0.8623*** (0.1395)	0.2912*** (0.0452)
Associate	0.4822*** (0.0932)	0.1699*** (0.0326)	1.1691*** (0.1806)	0.3947*** (0.0582)
Bachelor’s	0.9649*** (0.0701)	0.3399*** (0.0237)	1.2859*** (0.1395)	0.4342*** (0.0427)
Postgraduate	1.2725*** (0.1597)	0.4483*** (0.0555)	1.8065*** (0.2942)	0.6100*** (0.0961)

<sup>5</sup> 173 women did not answer this question.

<i>Household economic status</i>				
Household annual wage	-0.0179*** (0.0036)	-0.0063*** (0.0013)	-0.018** (0.0076)	-0.0061** (0.0026)
Household annual nonlabor income	-0.00044 (0.0037)	-0.00016 (0.0013)	-0.0886*** (0.0113)	-0.0299*** (0.0036)
<i>Household demographic characteristics</i>				
Children aged 0-5	-0.2515*** (0.0298)	-0.0886*** (0.0103)	-0.0370 (0.0921)	-0.0125 (0.0311)
Children aged 6-14	-0.0605*** (0.0195)	-0.0213*** (0.0069)	-0.0687 (0.0622)	-0.0232 (0.0210)
Reference person	0.1550* (0.0830)	0.0546* (0.0292)	0.3603*** (0.0909)	0.1217*** (0.0303)
Constant	-0.4798*** (0.0699)		-0.3994** (0.2041)	
<i>Goodness of fit</i>				
Pseudo R <sup>2</sup>	0.06		0.14	
McKelvey and Zavoina's R <sup>2</sup>	0.12		0.29	
Pearson chi-square test	6,898.56 1.0488 (0.0542)		1,408.19 0.9961 (0.0657)	
Link test	0.107 (0.0681)		0.0364 (0.0773)	
	-0.0092 (0.0224)		-0.0108 (0.0421)	

Notes: (i) OR stands for odds ratio; ME stands for marginal effects. (ii) \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . (iii) Numbers in brackets are the robust standard errors for the odd ratios and are delta-method standard errors for the marginal effects. (iv) The three variables in the link test indicate the prediction, the prediction squared of the dependent variable, and the constant. The numbers in the brackets present the standard errors of the variables.

The Pearson chi-square test indicates that the data come from a specified distribution and the model fits well enough. According to the link test, since the prediction square of the dependent variable is not statistically significant, link tests do not reveal any problem with the specification for both models<sup>6</sup>.

Probit model findings for married women demonstrate an inverted U-shaped relationship between age and labor force participation and that married women are more likely to join the labor force between the ages of 40 and 44. This supports the conventional Turkish household structure; when the children start a higher education level mother may start to seek a job to support the children's education expenditures. Yet, unmarried women are most likely to participate in the labor force at an early age between the ages of 30-34. Although higher education leads to more labor force participation, education is a more significant pattern in unmarried women's labor force participation decision. Unmarried women with compulsory education, high school, associate, bachelor's, and postgraduate degrees are almost 12%, 25%, 22%, 9%, and 16% more likely to join the labor force than married women, respectively. The gap in labor force participation is the highest among the women with a high school degree and is the lowest among the women with a bachelor's degree. As expected, the annual household wage and nonlabor income decrease the likelihood of labor force participation for both married and unmarried women. While the numbers of children in the household are significant and causes a decrease in the labor force participation of married women, it is insignificant in the labor force participation decision of unmarried women. The presence of a preschooler/s in the household reduces the likelihood of working for married women maybe because of the high childcare services costs. As expected, being responsible for the household has a contribution to the likelihood of taking place in the labor force, yet this contribution is higher for unmarried women.

### 5.3. Nonlinear Decomposition Findings

Table 3 reports detailed nonlinear decomposition findings by marital status in the labor force participation based on the probit nonlinear model expressed in Equation 1.

<sup>6</sup> For more information, please see Pregibon (1980).

Table 3. Detailed nonlinear decomposition findings

	Characteristics effect	
	Estimate	Contribution
<b>Age categories</b>		
25 - 29	0.0072** (0.0029)	-5.18
30 - 34	0.0003 (0.0003)	-0.24
35 - 39	0.0027*** (0.0008)	-1.98
40 - 44	0.0027*** (0.0007)	-1.99
45 - 49	0.0001 (0.0003)	-0.1
50 - 54	-0.0011*** (0.0002)	0.84
<b>Education degrees</b>		
No schooling	-0.0073*** (0.0007)	5.28
Compulsory	-0.0288*** (0.0028)	20.85
High school	0.0052*** (0.0006)	-3.73
Associate	0.0001 (0.001)	-0.11
Bachelor's	-0.0285*** (0.0032)	20.61
Postgraduate	-0.0036*** (0.0006)	2.64
<b>Household economic status</b>		
Household annual wage	-0.0204*** (0.0041)	14.78
Household annual nonlabor income	0.0004 (0.0037)	-0.32
<b>Household demographic characteristics</b>		
Children aged 0-5	-0.032*** (0.0033)	23.17
Children aged 6-14	-0.0124*** (0.0038)	8.94
Reference person	-0.0186* (0.01)	13.46
Aggregate effect	-0.1340*** (0.0126)	96.92

Notes: (i) \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . (ii) The dummy variables in the model are normalized to overcome the identification problem.

As presented in the last row of Table 3, the aggregate characteristics effect<sup>7</sup> accounts for 96.92% of the difference in labor force participation between married and unmarried women. This means, if the married and unmarried women had the same characteristics in the determined patterns such as age, education, household economic status, and household characteristics, then 96.92% of the labor force participation gap between them would have disappeared. Table 3 also shows that most of the sub-aggregate characteristics effects are statistically different from zero at the 1%, 5%, and 10% levels of significance. Significant characteristics effects for the age categories reflect that the marital status differences among people who are in 25-29 (5.18%), 35-39 (1.98%), and 40-44 (1.99%) age intervals contribute to a reduction in marital status differences in labor force participation whereas who are in 50-54 age interval contribute promotion. The characteristics effects of education for no schooling, only

<sup>7</sup> Since the coefficient effect is around 3% and mostly insignificant, only characteristics effect findings are presented. Characteristics effect findings are available upon request.

compulsory education, bachelor's, and postgraduate degrees are statistically significant at 1% significance level and their characteristics effects are 5.28%, 20.85%, 20.61%, and 2.64% percent, respectively. The education level of unmarried women is relatively higher than married women and this increases the labor force participation gap. On the contrary, small differences in marital status in high school degrees (3.73%) contribute to a smaller labor force participation gap. The characteristics effect for household annual logarithmic wage equals to 14.78% and is statistically significant at the 1% significance level. The differences in household annual logarithmic wage between married and unmarried women contribute to the labor force participation gap between married and unmarried women and increase the size of the gap. Household characteristics indicate that having children between the ages of 0-5 and 6-14 increases the labor force participation gap by 23.17% and 8.94%, respectively. The contribution of the number of preschooler/s is 14.23% higher than the number of older children. Lastly, the difference between married and unmarried women being the reference person in the household contributes to the 13.46% increase in the labor force participation gap.

#### **5.4. Sensitivity Analysis**

This study also conducts a sensitivity analysis to investigate the robustness of the findings in Tables 2 and 3. To do so, four different probit models were estimated and decomposed. All sensitivity analysis findings are available in Appendix A. The first model changes the reference group of education categories from no schooling to postgraduate degree and there is no significant change in the coefficients other than education variables which depend on the change in the reference category. The second model adds continuous age and squared age variables instead of the dummy variables of the age categories in the model. These continuous variables confirm that there is an inverted U-shaped relationship between age and labor force participation just like the age categories. The third model combines the two variables that indicate the numbers of children aged between 0-4 and 6-14 and the findings remain robust in the case of adding this variable in the model. The last model changes the dummy variables of education categories to one continuous schooling variable. In the course of this change, the coefficients remain robust. According to the nonlinear decomposition results of these four models, although there are small changes in the percentages presented below the standard errors, the aggregate characteristics effects are robust against the change in the variables.

## **6. Conclusion**

The first task of this study was to estimate the probit model structured for the labor force participation decision for prime working-age married and unmarried women and to reveal the differences in the determinants of the labor force participation decision by marital status. As indicated in this study, although the marginal effects are higher for unmarried women, higher education degrees and being the reference person in the household increase labor force participation whereas household wage decreases the labor force participation of both married and unmarried women. However, the labor force participation age differs by marital status. Married women tend to involve in the labor force at later ages whereas unmarried women tend to involve earlier. Furthermore, neither the number of preschooler/s nor children aged between 6 and 14 have a significant influence on unmarried women's decision to labor force participation. On the other hand, the number of children is a major reducing factor on married women's decision to join the labor force, considering child care. Moreover, having younger children reduces the likelihood of labor force participation more. The second task of this study was to examine the labor force participation gap between married and unmarried women by employing a nonlinear decomposition method. The aggregate decomposition findings reveal that the differences among women by marital status lead to a labor force participation gap and if the married and unmarried women had the same characteristics, then 96.92% of the labor force participation gap between them would disappear.

The educational attainment gap between married and unmarried women and unmarried women being more educated is one of the most important determinants that increase the size of the labor force gap between them. Facilitating access to education, expanding opportunities for competence in different business sectors, and encouraging married women are the relevant ways of reducing this gap by increasing married women's education level. In recent years, the Ministry of Education has introduced an important opportunity named Public Education Centers, especially for housewives. These centers provide free courses and supply most women to gain competence in certain areas such as computer use, beauty and hair care services, skin care, massage, food and beverage services, photography, painting, crafts, foreign language, gardening, animal breeding, music, sports, etc. In line with the demands, Public Education Centers can provide 3,701 different courses in 72 areas. These centers help women not only to gain competence in certain workforces and benefit from these competencies but also help to gain self-confidence. Expanding these Public Education Centers in more areas considering the labor market demands, encouraging especially married women to participate in the labor force after they gain this specific competence,

and supporting them to become an entrepreneur by giving interest-free loans and reducing taxes, may increase labor force participation of married women. In this way, the labor force participation gap between educated unmarried women and uneducated married women can be reduced.

Turkey has a conventional household type and social values that support both men and women to marry and have children in their early life. Yet, encouraging especially women to participate in higher levels of education rather than marrying at an early age may lead to a higher labor force participation. Having children in the household is one of the most significant determinants of the size of the labor force gap and it increases the gap among married and unmarried women. Although long maternity leaves assure women, it is considered as a disadvantage for employers. If the woman is married, the employers consider that she will soon become a mother. Since they consider this as a long-time labor loss for them, they reduce the likelihood of hiring them and prefer to hire men and single women. An arrangement of maternity and paternity leaves as in mostly equal, increasing kindergarten opportunities, making them accessible and affordable are of great importance in terms of increasing married women's labor force participation. Most importantly, discrimination against especially married women in the labor market should be eliminated in order to reduce the labor force participation gap. In addition, increasing the mobility of women and increasing the labor force participation of married women should be encouraged by the authorities at every opportunity.

Understanding the driving or reducing forces behind the labor force participation gap between married and unmarried women is crucial to developing policies and campaigns to reduce the gap and increase women's involvement in the economy. This study may provide a framework in which women's marriage and labor force participation decisions can be examined.

## References

- Baslevent, C., & Onaran, O. (2003). Are Married Women in Turkey More Likely to Become Added or Discouraged Workers? *Labour*, 17(3): 439–458.
- Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Source: The Journal of Human Resources*, 8(4): 436–455.
- Chin, S. H. (1995). The Determinants and Patterns of Married Women's Labor Force Participation in Korea. *Korea Journal of Population and Development*, 24(1): 95–129.
- Contreras, D., & Plaza, G. (2010). Cultural Factors in Women's Labor Force Participation in Chile. *Feminist Economics*, 16(2): 27–46.
- Dayioglu, M., & Kirdar, M. G. (2010). *Determinants of and trends in labor force participation of women in Turkey*. State Planning Organization of the Republic of Turkey and World Bank, Welfare and Social Policy Analytical Work Program, Working Paper Number 5.
- Diwan, I., & Vartanova, I. (2017). *The Effect of Patriarchal Culture on Women's Labor Force Participation*. Working Paper Series No. 1101.
- Gedikli, C. (2014). Female Labour Supply in Turkey: Do Traditional Gender Roles Matter? In *33rd IARIW General Conference* (pp. 1–47). Netherlands.
- Goksel, I. (2013). Female Labor Force Participation in Turkey: The Role of Conservatism. *Women's Studies International Forum*, 41: 45–54.
- Ilkkaracan, I. (2012). Why so Few Women in the Labor Market in Turkey? *Feminist Economics*, 18(1): 1–37.
- Ince, M., & Demir, M. H. (2006). The Determinants of Female Labor Force: Empirical Evidence from Turkey. *Eskişehir Osmangazi Üniversitesi İİBF Dergisi*, 1(1): 71–90.
- Kasnakoglu, Z., & Dayioglu, M. (1997). Female Labor Force Participation and Earnings Differentials Between Genders in Turkey. In J. M. Rives & M. Yousefi (Eds.), *Economic Dimensions of Gender Inequality: A Global Perspective* (1st ed., pp. 95–117). United States of America: Greenwood Publishing Group.
- Komuryakan, F., & Yilgor, M. (2021). Has the gender wage gap narrowed or widened in a decade? Some recent evidence from the Turkish labor market. *Journal of Applied Microeconometrics*, 1(1): 1–17.
- Lee, B. S., Jang, S., & Sarkar, J. (2008). Women's labor force participation and marriage: The case of Korea. *Journal of Asian Economics*, 19(2): 138–154.
- Mincer, J. (1962). Labor Force Participation of Married Women: A Study of Labor Supply. In *Aspects of Labor Economics* (pp. 63–105). Princeton University Press.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3): 693–709.
- Ozar, S., & Gunluk-Senesen, G. (1998). Determinants of Female (Non-) Participation in the Urban Labour Force in Turkey. *METU Studies in Development*, 25(2): 311–328.
- Palaz, S. (2010). Women's Labour Force Participation in Turkey. *Journal of Social Policy Conferences*, 50: 591–606.
- Petrakis, I. (2021). Determinants of female labour force participation: Evidence from Greece. *Labour*, 1-30.
- Pregibon, D. (1980). Goodness of Link Tests for Generalized Linear Models. *Applied Statistics*, 29(1): 15–23.
- Pritchett, J., & Yun, M. S. (2009). The in-hospital mortality rates of slaves and freemen: Evidence from Touro Infirmary, New Orleans, Louisiana, 1855-1860. *Explorations in Economic History*, 46(2): 241–252.



- Tansel, A. (2002). *Economic Development and Female Labor Force Participation in Turkey: Time-series Evidence and Cross-province Estimates*. ERF, 1-59.
- Varol, F. (2017). The Determinants of Labor Force Participation of Women In Turkey: A Binary Logit Analysis. *Journal of History Culture and Art Research*, 6(2): 92–108.
- Verick, S. (2014). Female labor force participation in developing countries. *IZA World of Labor*, 1–10.
- Xiao, S., & Asadullah, M. N. (2020). Social Norms and Gender Differences in Labor Force Participation in China. *Feminist Economics*, 26(4): 114–148.
- Yun, M. S. (2004). Decomposing differences in the first moment. *Economics Letters*, 82(2): 275–280.
- Yun, M. S. (2005a). A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry*, 43(4): 766–772.
- Yun, M. S. (2005b). Normalized Equation and Decomposition Analysis: Computation and Inference. Bonn, Germany: The Institute for the Study and Labor (IZA).
- Yun, M. S. (2008). Identification Problem and Detailed Oaxaca Decomposition: A General Solution and Inference. *Journal of Economic and Social Measurement*, 33(1): 27-38.

Appendix A. Sensitivity analysis

	Married				Unmarried			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
<i>Age categories</i>								
25 - 29	0.0932 (0.0720)	-	-0.0032 (0.0692)	0.0168 (0.0719)	0.3215** (0.1435)	-	0.3199** (0.1433)	0.2330 (0.1451)
30 - 34	0.2418*** (0.0667)	-	0.2166*** (0.0661)	0.2044*** (0.0669)	0.5138*** (0.1498)	-	0.5124*** (0.1496)	0.4288*** (0.1514)
35 - 39	0.3207*** (0.0619)	-	0.3313*** (0.0617)	0.3019*** (0.0617)	0.3311** (0.1475)	-	0.3295** (0.1472)	0.2863* (0.1494)
40 - 44	0.3431*** (0.0576)	-	0.3694*** (0.0573)	0.3385*** (0.0572)	0.4175*** (0.1411)	-	0.4144*** (0.1402)	0.3803*** (0.1427)
45 - 49	0.2198*** (0.0555)	-	0.2340*** (0.0552)	0.2243*** (0.0554)	0.4683*** (0.1352)	-	0.4665*** (0.1350)	0.4444*** (0.1365)
50 - 54	-	-	-	-	-	-	-	-
Age	-	0.1586*** (0.0214)	-	-	-	0.1642*** (0.0399)	-	-
Age <sup>2</sup>	-	-0.002*** (0.0003)	-	-	-	-0.0022*** (0.0005)	-	-
<i>Education degrees</i>								
No schooling	-1.2725*** (0.1596)	-	-	-	-1.8065*** (0.2942)	-	-	-
Compulsory	-1.1491*** (0.1550)	0.1196** (0.0470)	0.1299*** (0.0467)	-	-1.3149*** (0.2783)	0.4892*** (0.1255)	0.4909*** (0.1254)	-
High school	-1.1562*** (0.1586)	0.1119* (0.0595)	0.1128* (0.0595)	-	-0.9441*** (0.2818)	0.8766*** (0.1402)	0.8616*** (0.1396)	-
Associate	-0.7904*** (0.1735)	0.4834*** (0.0932)	0.4693*** (0.0930)	-	-0.6374* (0.3000)	1.1893*** (0.1815)	1.1684*** (0.1808)	-
Bachelor's	-0.3077* (0.1621)	0.9605*** (0.0699)	0.9408*** (0.0696)	-	-0.5205** (0.2788)	1.3055*** (0.1411)	1.2857*** (0.1396)	-
Postgraduate	-	1.2588*** (0.1596)	1.2394*** (0.1581)	-	-	1.7992*** (0.2986)	1.8058*** (0.2942)	-
Schooling	-	-	-	0.0409*** (0.0033)	-	-	-	0.0797*** (0.0073)

Appendix A. Continued

<i>Household economic status</i>								
Household annual wage	-0.0178*** (0.0036)	-0.0179*** (0.0036)	-0.0188*** (0.0036)	-0.0167*** (0.0036)	-0.018** (0.0076)	-0.0179** (0.0076)	-0.0179** (0.0076)	-0.0181** (0.0076)
Household annual nonlabor income	-0.0004 (0.0037)	-0.0002 (0.0037)	-0.0007 (0.0037)	0.0003 (0.0037)	-0.0886*** (0.0113)	-0.0902*** (0.0113)	-0.0884*** (0.0113)	-0.0882*** (0.0113)
<i>Household demographic characteristics</i>								
Children aged 0-5	-0.2515*** (0.0298)	-0.2475*** (0.0298)	-	-0.2416*** (0.0294)	-0.0370 (0.0921)	-0.0480 (0.0927)	-	-0.0322 (0.0930)
Children aged 6-14	-0.0605*** (0.0195)	-0.0682*** (0.0193)	-	-0.069*** (0.0196)	-0.0686 (0.0622)	-0.0821 (0.0623)	-	-0.0751 (0.0621)
Total number of children	-	-	-0.1216*** (0.0169)	-	-	-	-0.0578 (0.0488)	-
Reference person	0.1550* (0.0830)	0.1513* (0.0831)	0.1551*** (0.0826)	0.1510* (0.0828)	0.3603*** (0.0909)	0.3519*** (0.0915)	0.3591*** (0.0908)	0.3725*** (0.0913)
Constant	0.7927*** (0.1649)	-3.2131*** (0.4207)	-0.4723* (0.0697)	-0.5428*** (0.0611)	1.4071*** (0.3259)	-2.9328*** (0.7722)	-0.3993*** (0.2041)	-0.3397* (0.1822)
<i>Aggregate Nonlinear Decomposition</i>								
Aggregate characteristics effect <i>Model 1</i>	-0.1339*** (0.0126) 96.92							
Aggregate characteristics effect <i>Model 2</i>	-0.1291*** (0.0127) 93.33							
Aggregate characteristics effect <i>Model 3</i>	-0.1225*** (0.0125) 88.6							
Aggregate characteristics effect <i>Model 4</i>	-0.1128*** (0.0128) 81.85							
Notes: (i) *** $p < 0.01$ , ** $p < 0.05$ , * $p < 0.1$ . (ii) Numbers in brackets are the robust standard errors for the odd ratios.								



## **Econometric Analysis of Population Increase and Population Projections in Turkey**

Cemil İSKENDER  
Retired Research Economist  
iskenderc@ttmail.com  
Orcid No: 0000-0003-2841-5964

### **Abstract**

Turkish Statistical Institute (TurkStat), which has been carrying out the official censuses in Turkey since 1927, published backward mid-year population estimates for all years from the first census year of 1927 to 2006, address-based census results and mid-year estimates for 2007-2017 and population projections for 2018-2080 period. According to the main scenario of the projections, population will reach 107.7 million in 2069, and from this point, it will decrease to 107.1 million in 2080. TurkStat population projections method covers the dimensions which is based on the institution's long term experience and data, the principles of the discipline of demography e, and international data compiled. My aim in this econometric study is to investigate the representation of the main scenario projections and the results of past censuses with five and six parameter Gompertz population growth model which was developed during this study and also sinusoidal function applications for the residuals. In addition, I also applied Verhulst five and six parameter functions and reached positive statistical and linearity test results made with all functions. The results of the econometric study confirmed that the main scenario projections of the Statistical Institute are valid and realistic.

*Keywords: Gompertz Function, Nonlinear Functions, SAS MODEL Procedure, SAS NLIN Procedure, Turkish Population Study, Verhulst Function.*

---

**Corresponding Author / Sorumlu Yazar:** 1-Cemil İSKENDER, Retired Research Economist.

**Citation / Atıf:** İSKENDER C. Econometric Analysis of Population Increase and Population Projections in Turkey. İstatistik Araştırma Dergisi, 2021, 11 (2), 30-55.

## Türkiye'deki Nüfus Artışı ve Nüfus Projeksiyonlarının Ekonometrik Analizi

### Özet

Türkiye'de 1927 yılından bu yana resmi nüfus sayımını yapan Türkiye İstatistik Kurumu (TÜİK), 1927 yılının birinci nüfus sayımı yılından başlayarak 2006 yılına kadar tüm yıllar için geriye dönük yıl ortası nüfus tahminleri, 2007-2017 yıllarına ait adrese dayalı nüfus sayımı sonuçları ve yıl ortası tahminleri ile 2018-2080 dönemine ilişkin nüfus projeksiyonlarını yayınlamıştır. Tahminlerin ana senaryosuna göre, nüfus 2069'da 107,7 milyona ulaşacak ve bu noktadan sonra 2080'de 107,1 milyona inecektir. TÜİK nüfus projeksiyonları yöntemi çalışması, kurumun uzun vadeli deneyim ve verilerine, demografi disiplini ilkelerine ve derlenen uluslararası verilere dayanan boyutları kapsamaktadır. Bu ekonometrik çalışmadaki amacım, çalışma sırasında geliştirdiğim beş ve altı parametrelili Gompertz nüfus artış modeli ve ayrıca artıklar için sinüzoidal fonksiyon uygulamaları ile ana senaryo projeksiyonlarının ve geçmiş sayımların sonuçlarının temsilini araştırmaktır. İlaveten Verhulst beş ve altı parametre fonksiyonları için de uygulama yaptım. Tüm fonksiyonlarla yapılan pozitif istatistiksel ve doğrusallık test sonuçlarına ulaştım. Ekonometrik çalışma sonuçları, İstatistik Kurumu'nun ana senaryo projeksiyonlarının geçerli ve gerçekçi olduğunu doğrulamaktadır.

*Anahtar Kelimeler: Doğrusal Olmayan Fonksiyonlar, Gompertz Fonksiyonu, SAS MODEL Yöntemi, SAS NLIN Yöntemi, Türkiye Nüfus Çalışması, Verhulst Fonksiyonu.*

### 1. Introduction and Scope of Research

The aim of this econometric study is to investigate the representation of Turkey's population projections published by TurkStat (Turkish Statistical Institute) for the 2018-2080 period and the past census results (1927-2017) together by applying growth functions of Gompertz and Verhulst. As based on economic theories, explained by mathematical models, applied and supported by data and includes advanced statistical applications, we define our research as econometric analysis.

TurkStat has prepared population projections of Turkey for 2018-2080 period and published them on its website in 2018. Projections are an important source of data when used together with the past censuses. With an advanced statistical study of this data, it is possible to obtain concrete results for the population growth trend of Turkey. However, this long-term data need to be handled within a certain economic and statistical framework. Otherwise, our statistical study remains at a simple level of regression work: If we had done a short-to-medium-term work for 10-15 years, inter-census interpolation study or a two to three-year forecast, we could have used linear, polynomial, etc. regression methods. But that is not an affair and purpose of this study and the things are quite different than that of ordinary regression analysis. At this point, the theory of population growth, advanced mathematics and statistical analysis is included into the phases of the study.

Theorists who studied population growth for this purpose also established the theory of population development. There must be a consistent approach behind the statistical study. The first and most important examples of this approach are theorists Malthus, Gompertz and Verhulst, whose theories survived to the present day. While population growth in human communities in a certain geography increases according to biological and economic principles of scarce resources, growth rate is limited and may even fall to zero with the increase of factors against population growth after a certain stage. Verhulst wrote<sup>1</sup>:

There is a limit to the growth of that population to the extent permitted by the land that meets the need for a population's housing... Therefore, all formulas that will try to represent the laws of the population must accept the requirement that *maximum* can be reached in an extremely distant period. This *maximum* will be a population that has become stable. (1838, pp. 113–115)

The mathematical population growth theories developed by Verhulst and Gompertz will be the basis for our study of population growth in Turkey. We used Verhulst logistic and Gompertz functions as they were specifically developed for the analysis of human population. In addition to the five and six-parameter versions of Verhulst

---

<sup>1</sup> For a very detailed explanation of economic fundamentals and mathematics of population theories see (İskender, 2021)

logistic function, we also used five and six parameter versions of Gompertz functions including sinusoidal ones which we developed in this study.

TurkStat population projections methods are very different from the mathematical and statistical applications used in this research and in our previous “Turkish Population Growth Study” (İskender, 2018). In that study which was based on population counting and forecasting data for the 1927-2015 period, we had made projections with mathematical growth functions for the 2016-2080 period and predicted the final levels of the population. In this study, we will use same methodology of the previous manuscript and newly developed Gompertz function for the evaluation of TurkStat population projections.

In the last section, the final evaluation of the study and its conclusions are presented.

## **2. Turkish Population Data, Projections and Methodology of Turkstat**

Population projections of 2018-2080 period published in three series, namely; *the main scenario*, *high scenario*, and *low scenario*, all are extending until 2080. According to the *main scenario*, the population of 81.9 million<sup>2</sup> people in 2018 will reach 107.7 million in 2069, and from this point, it will decrease to 107.1 million in 2080. At the *low scenario*, the population will reach a maximum of 99.7 million in 2055, then with a decrease to 94.1 million in the year 2080. As to the *high scenario*, the population will be 121 million people in 2080, and although not too fast, it will continue to increase as far as the Figure 1 indicates.

TurkStat has explained the assumptions and calculation methods on which the population projections are based, as follows:

(i) Analytical Framework, Concepts, Definitions, and Classifications:

The aim of this study is to produce national and provincial population projections by using the current data sources. Cohort-component method, which reflects the effects of various demographic indicators on age-sex structure of the population, was used. Cohort-component method is based on lifelong monitoring of cohorts of the same age group according to their fertility, mortality and migration patterns. The cohorts are annual birth cohorts (age-cohorts). The components are births, deaths and migration... (TurkStat, 2021)

(iii) Accounting Conventions:

During the design and production stages of the projections, a committee that consists of representatives from academic and official institutions/organizations and sub-working groups for fertility, mortality and migration were constituted under chairmanship of TurkStat. Within the context of activities carried out by sub-working groups, evaluations about the current and future demographic circumstances of Turkey were made; backward analysis of historical data, problematic points and limitations were discussed. Provincial age and sex distribution of demographic data and their trends were analyzed. Thus, projections and assumptions regarding future dynamics of fertility, mortality and migration have been determined...(TurkStat, 2021)

After almost a century of successful activities and achievements of important statistical publications, TurkStat has changed its mode of operation and vision to the next level by publishing the population projections for the 2018-2080 period. In addition to the population figures of the past years, the projections published on the TurkStat website are also vital reference source for researchers.

Before explaining the data we use, it is useful to look at the population curve of Turkey indicated in Figure 1.

The major event that disrupted the population trend of the 1925 -2080 period was the Second World War. Even though Turkey was not in the war, the Second World War had a negative effect on the population growth due to the high number of soldiers and the longer than normal military service period. In 1935-40, five-year population growth was 10.3%, compared to 5.4% in 1940-45. Despite the negative effects of the Second World War, the statistical representation of the curves remained intact.

TurkStat has carried out its first census in 1927 with a daily curfew and by visiting households<sup>3</sup>, starting from the establishment of the Republic of Turkey in 1923. Following the negative effects of the world-wide economic crisis in 1930, the second census was held in 1935, two years after the first five-year industrial development plan which

---

<sup>2</sup> By following the SAS Studio statistical software methodology, we use dots to separate the decimal places in our manuscript.

<sup>3</sup> For a detailed explanation of census methodologies in Turkey see Canpolat & Gök, (2020).

began in 1933 and was the first major economic initiative of the Republic. In the subsequent years, the censuses were maintained for every five-year, until 1990 (including). Last census with daily curfew carried out in 2000.

Based on the data of 12 censuses for 1927-1985 period, TurkStat has prepared and included census mid-year estimates, as well as the census years figures in its published tables (59 observations). In determining the mid-year population figures for the 1986-1999 period, estimates produced from the Turkey Demographic and Health Surveys and 2008 address-based population registration system were used (14 observations). Address-based population registration system results of 2007-2020 and population projections of 2018-2080 period have been used for the population estimates of 2000-2006 period (7 observations). Mid-year census results for the 2007-2017 period were produced from annual results of the address-based registration system for 2007-2017 period (11 observations). Thus, we have 91 mid-year-based data for each year of 1927-2017 period.

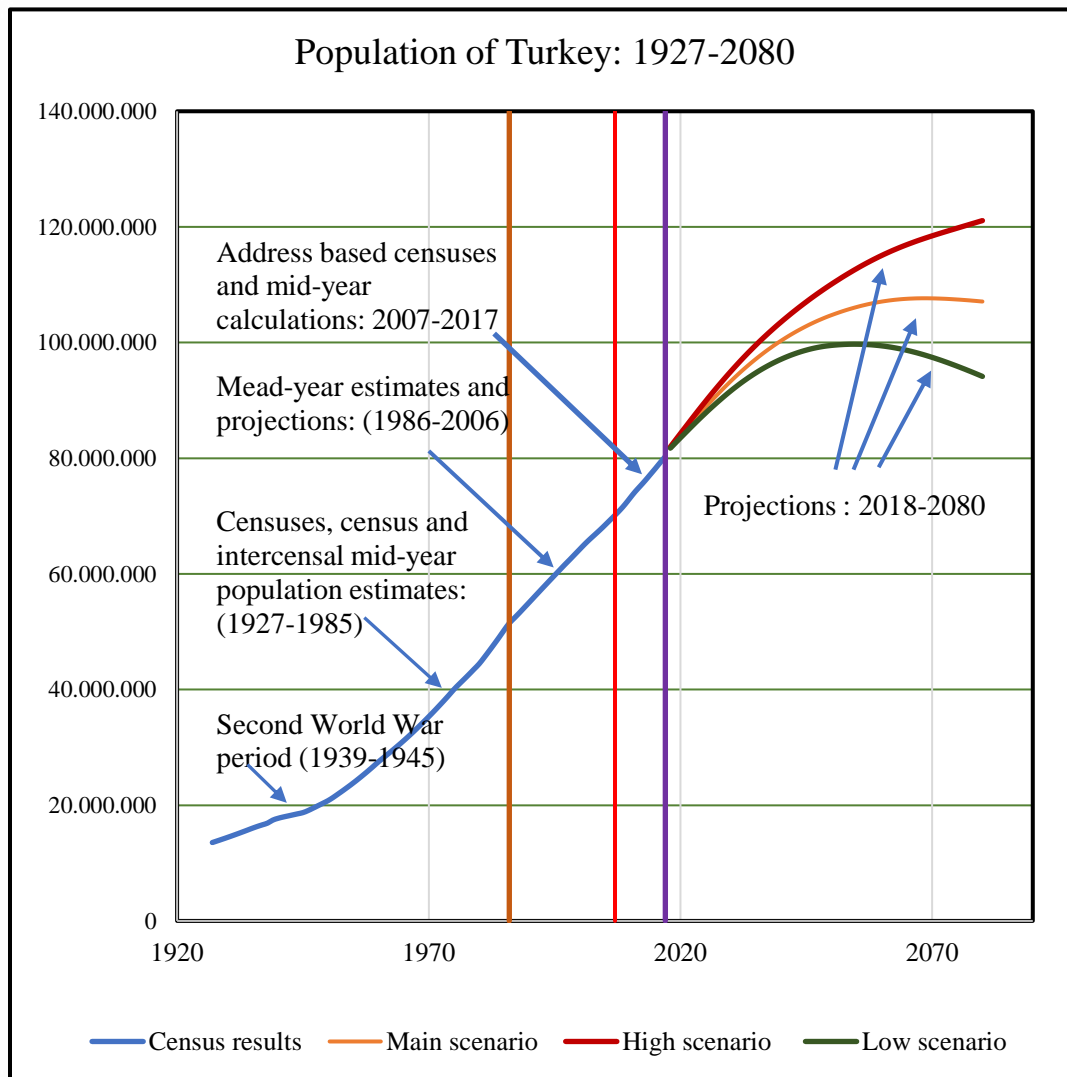


Figure 1: Turkish Population Curves of TurkStat Data

When we combine the estimates of the sixty-three years of the main scenario projections with historical data, we reach a total of 154 observations. 154 figures will represent all sections of the mathematical growth curves, including the lower and upper asymptote regions which are critical to statistical application. Our objective is to statistically analyze the consistency of *main scenario projections* with the census results.

Population statistics compiled and used are summarized in Table 1.

Table 1: Summary Table of Population Statistics and Their Sources

Index	Years	Number of observations	Census results (summary)	Source
1-59	1927-1985	59	13,538,000...50,307,000	TurkStat counts and estimates
60-80	1986-2006	21	51,480,000...69,295,000	TurkStat estimates and projections
81-91	2007-2017	11	70,158,000-80,313,000	TurkStat censuses of population registration system based on resident addresses
92-154	2018-2080	63	81,867,223...107,100,904	TurkStat <i>main scenario</i> projections

### 3. Mathematics of Particular Integral or Trend Functions

In accordance with our purpose of this study, we have selected five and six-parameter Gompertz and Verhulst growth functions. We developed and implemented five and six-parameter Gompertz functions in this study for the first time. Gompertz function will be discussed and presented with the addition of fifth and sixth parameters of a new approach and will be demonstrated with the Turkish population data. In addition, the application of the complementary sinusoidal function for residual series of growth functions will be discussed.

In our previous study, we already applied the six-parameter Verhulst growth function to Turkish population data. Here, the reason we examined the population projections of Turkey with the newly developed Gompertz growth function is to show that the successful statistical results obtained in all cases are not accidental or coincidental. Verhulst and Gompertz growth functions come from completely different differential equation definitions. Verhulst's logistic growth function is based on the dilemma of population growth and forces opposing to growth, while the Gompertz function is based on a geometric decrease in survivors and increase in deaths. The upper asymptotes obtained with each group of functions are different and in general Gompertz asymptote figures in Turkey's population data are higher than Verhulst's. On the other hand, this curve has less curvature flexibility when coming to the upper asymptotic zone. It takes starting level of population as lower asymptote. However, the logistic curve is based on a lower asymptote which is less than the initial level of the population and this gives flexibility to Verhulst curve.

#### 3.1. Gompertz Function

Although five and six-parameters forms of Gompertz function come from the roots of the main function (three-parameter), by definition they are quite different: Two more parameters; lower asymptote ( $L$ ) and allometric coefficient ( $\nu$ ), and two new definitions; separated growth base coefficients ( $Q_1$  and  $Q_2$ ) and weight's function ( $w_{(t-T)}$ ) have been added to growth formula. These changes make formulas completely different. Despite these mathematical differences, statistical test results obtained from the application of Gompertz's five and six-parameter functions are very close to each other and are more preferable relative to four-parameter forms when applied with Turkish population data.

Symbols used for Gompertz functions are:

$Y$ : Population (dependent variable),

$t$ : Time (explanatory variable),



$T$ : Base year<sup>4</sup>,

$K$ : Upper asymptote,

$L$ : Lower asymptote,

$N$ : Starting level;

$Q$ : Coefficient of exponential growth base ( $e$ ).

$Q_1$  and  $Q_2$ : Coefficients of exponential growth base of ( $e$ ),

$Y_0$ : Initial population,

$r$ : Intrinsic growth rate,

$v$ : Allometric coefficient which indicates the direction and magnitude of the structure of the curve,

$w_{(t-T)}$ : Weight's distribution function as being the key to distributing weights in the function and,

$\bar{H}_w$ : As harmonic mean of  $Q_1$  and  $Q_2$ , is the coefficient of  $e$  (*exponential base*) in the weight's function.

Based on approach of Ricketts and Head (1999), we separated  $Q$  - multiplier of the naturel logarithm ( $e$ ) base of the growth function- into two multipliers as  $Q_1$  and  $Q_2$ . Also, the weight's distribution function as weighted with the harmonic, arithmetic or geometric mean of  $Q_1$  and  $Q_2$  were included in the equation defining  $w_{(t-T)}$  for  $Q_1$  and  $(1 - w_{(t-T)})$  for  $Q_2$ . Thus, asymmetry of the Gompertz function happened to be more asymmetrical. The  $Qe^{-r(t-T)}$  term of the Gompertz equation is as follows in the five-parameter form:

$$Qe^{-r(t-T)} \equiv w_{(t-T)}Q_1e^{(-r(t-T))} + (1 - w_{(t-T)})Q_2e^{(-r(t-T))} \quad (1)$$

When  $Q_1$  becomes equal to  $Q_2$ , we obtain four-parameter form. We also applied the weight's function of (2) to the coefficients in the denominator of Gompertz function. It defines a Gompertz weighting function varying between 0 and 1 similar to main curve.

The generalized five-parameter Gompertz growth function is a mathematical function with two variables and six parameters ( $K, L, Q_1, Q_2, r$ ) and has a nonlinear structure.

The weight's function and separation of  $Q$  into two multipliers as  $Q_1$  and  $Q_2$  overcome the constraints imposed by the symmetry event and to better reflect the asymmetrical nature of the data. The flexibility of this function is as much an advantage as it is, one and most important drawback is that the number of parameters to be estimated increases by one. Difficulties are felt in estimating an increased number of parameters during the computer software application. Same argument is valid in a wider scale for six-parameter case. More parameters mean more conformity with the data on the one hand, and calculation difficulties on the other hand. To overcome this situation, a dataset which is refined, high quality, smooth (without acute changes) and having large number of observations (including sufficient data for the lower and upper asymptotic regions) needs to be implemented.

Consequently, we defined the generalized five - parameter population growth function of Gompertz (G5) based on the changes and additions of the components given below;

- (i) the addition of lower asymptote ( $L$ ),
- (ii) the separation of the base coefficient ( $Q$ ) of ( $e$ ) into two (as  $Q_1$  and  $Q_2$ ),
- (iii) the inclusion of the weight's function  $w_{(t-T)}$  in the denominator of equation.

Constituents:

Harmonic mean:

$$\bar{H}_w = \frac{2Q_1Q_2}{|Q_1 + Q_2|} \quad (2)$$

---

<sup>4</sup>  $T$  and  $t$  are different.

Weight's distribution function:

$$w_{(t-T)} = \frac{1}{e^{\bar{H}_w} e^{(-r(t-T))}} \quad (3)$$

Starting values and  $e$ -bases:

$$A = Q_1 e^{(-r(t-T))} \quad (4)$$

$$B = Q_2 e^{(-r(t-T))} \quad (5)$$

Five-parameter Gompertz function (G5):

$$Y_{(t-T)} = Y_0 + \frac{K - Y_0}{e^{[w_{(t-T)}A + (1-w_{(t-T)})B]}} \quad (6)$$

or:

$$Y_{(t-T)} = Y_0 + \frac{K - Y_0}{e^{\left[ \left( \frac{1}{e^{\left[ \frac{2Q_1Q_2}{|Q_1+Q_2|} e^{(-r(t-T))} \right]}} \right) e^{(-r(t-T))} + \left( 1 - \frac{1}{e^{\left[ \frac{2Q_1Q_2}{|Q_1+Q_2|} e^{(-r(t-T))} \right]}} \right) Q_2 e^{(-r(t-T))} \right]} \quad (7)$$

We also included allometric coefficient of Richards (1959) to denominator of equation. The variable  $\nu$  provides flexibility to the predicted dependent variable of the function and increases the compliance of the curve to the actual data.  $\nu$  is specified as the "allometric coefficient." Thus, we reach six-parameter form of Gompertz function. When  $\nu=1$ , (8) becomes (7).

$$Y_{(t-T)} = Y_0 + \frac{K - Y_0}{\left[ e^{\left[ \left( \frac{1}{e^{\left[ \frac{2Q_1Q_2}{|Q_1+Q_2|} e^{(-r(t-T))} \right]}} \right) e^{(-r(t-T))} + \left( 1 - \frac{1}{e^{\left[ \frac{2Q_1Q_2}{|Q_1+Q_2|} e^{(-r(t-T))} \right]}} \right) Q_2 e^{(-r(t-T))} \right]} \right]^{(1/\nu)}} \quad (8)$$

We defined the generalized six - parameter population growth function of Gompertz as the addition of the allometric coefficient ( $1/\nu$ ) to five-parameter form.

### 3.2. Verhulst Logistic Function

Six-parameter Verhulst logistic growth function is a mathematical function with two-variables as time and population, six-parameters ( $K, L, Q_1, Q_2, r, \nu$ ) and has a nonlinear structure.

Constituents of Function:

Harmonic mean:

$$\bar{H}_w = \frac{2Q_1Q_2}{|Q_1 + Q_2|} \quad (9)$$

Weight's distribution function:

$$w_{(t-T)} = \frac{1}{1 + \bar{H}_w e^{(-r(t-T))}} \quad (10)$$

Starting values and  $e$  bases:

$$A = Q_1 e^{(-r(t-T))} \quad (11)$$

$$B = Q_2 e^{(-r(t-T))} \quad (12)$$

Six-parameter short form of Verhulst function (V6):

$$Y_{(t-T)} = L + \frac{K - L}{(1 + w_{(t-T)}A + (1 - w_{(t-T)})B)^{(1/v)}} \quad (13)$$

Six-parameter long form of Verhulst function (V6):

$$Y_{(t-T)} = L + \frac{K - L}{\left(1 + \frac{1}{1 + \frac{2Q_1Q_2}{|Q_1 + Q_2|} e^{(-r(t-T))}} Q_1 e^{(-r(t-T))} + \left(1 - \frac{1}{1 + \frac{2Q_1Q_2}{|Q_1 + Q_2|} e^{(-r(t-T))}}\right) Q_2 e^{(-r(t-T))}\right)^{(1/v)}} \quad (14)$$

Five-parameter Verhulst logistic growth function (V5) is a growth function without the allometric coefficient ( $1/v$ ) as exponent of denominator of six-parameter form of Verhulst logistic growth function.

#### 4. Mathematical Model of Analysis Applied in Development of Population

Population developments must be addressed under two main headings: (i) The main development trend of the population according to the natural birth rate and mortality rates and (ii) increases and decreases in population caused by external events and data errors. The mathematical growth models of Gompertz and Verhulst (particular solution equations) explain the first heading. For the second title, which describes autonomous changes outside the trend in the population, it will be required to add a new constituent to the growth function. These additional functions will explain most of the residuals at statistical application phase of the study.

The general solution of the mathematical model (*complete primitive*)<sup>5</sup> is equation (18). “ $\bar{Y}(x)$  is called the *particular integral* and the remainder makes up the *complementary function*...” (Allen, 1956, p. 137). In our case particular integral -*trend, equilibrium path*- is the Gompertz population growth function. For the path to follow in statistical application Allen wrote:

...Some particular solution  $Y = \bar{Y}(x)$  is first sought. Any particular solution will serve but the most relevant, from the point of view of interpreting the result, is a solution which represents the *trend* or equilibrium variation of  $Y$ . Interest is usually concentrated on the path of  $Y$  as  $x$  increases indefinitely, e.g. the time path of  $Y$  if  $x$  is time. The particular solution  $Y = \bar{Y}(x)$  sought then describes that path which is consistent with the given equation (15) and which is in some sense a norm or equilibrium. The next step is to write  $y = Y - \bar{Y}$  as the deviation of  $Y$  from the trend or equilibrium value. It follows that  $y$  must satisfy the homogeneous form (16), as is seen by substituting  $Y$  and  $\bar{Y}$  in (15)—both being solutions—and subtracting. The last step is to obtain the general solution of the homogeneous form (16) in the way indicated in (17). The solution is now both complete and ready for interpretation. The deviations  $y$  of  $Y$  from  $\bar{Y}$  are given by (17); translating back into the original variables gives solution (18). Of the two terms in this solution, one is the trend or equilibrium path (the particular integral) and the other is deviation from trend or equilibrium (the complementary function)...(Allen, 1956, pp. 137–138)

While a little long, the quote explains the state of affairs very well, so we kept it as it was. As a complement to the population growth functions, contributions of mathematical economics we have taken from the 20th century economics and statistics lecturer and author Allen (1956, pp. 91-173; 1967, pp.342-363),, are as valuable and

<sup>5</sup> (Piaggio, 1920, p. 4)

explanatory as the population theory models that are taken from the early 19th century authors Gompertz and Verhulst<sup>6</sup>.

Differential equation with constant coefficients of non-homogenous form of order n:

$$a_0 \frac{d^n Y}{dx^n} + a_1 \frac{d^{n-1} Y}{dx^{n-1}} + \dots + a_{n-1} \frac{dY}{dx} + a_n Y = f(x) \quad (15)$$

Homogenous form of corresponding equation:

$$a_0 \frac{d^n y}{dx^n} + a_1 \frac{d^{n-1} y}{dx^{n-1}} + \dots + a_{n-1} \frac{dy}{dx} + a_n y = 0 \quad (16)$$

Solution of (16):

$$y = A_1 y_1(x) + A_2 y_2(x) + \dots + A_n y_n(x) \quad (17)$$

General solution (Complete Primitive):

$$Y = \bar{Y}(x) + A_1 y_1(x) + A_2 y_2(x) + \dots + A_n y_n(x) \quad (18)$$

Since we give trend functions in section 3 as mathematics of population growth of Gompertz and Verhulst which were based on population growth theories by authors (İskender, 2021), now we will explain complementary functions.

Second order differential equation (19) and its homogeneous form (20) can be written as follows:

$$\frac{d^2 Y}{dx^2} + a \frac{dY}{dx} + bY = f(x) \quad (19)$$

$$\frac{d^2 y}{dx^2} + a \frac{dy}{dx} + by = 0 \quad (20)$$

With two equations, values of  $a$  and  $b$  can be found. However, a particular solution  $Y = \bar{Y}(x)$  is found first. Then complementary functions are obtained.

In case of  $a^2 < 4b$ , solution equation of four-parameter sinusoidal function as complementary of growth function:

$$y_{(t-T)} = Ae^{\alpha(t-T)} \cos(\omega(t-T) - \varepsilon) \quad (21)$$

or eight-parameter form<sup>7</sup> is;

$$y_{(t-T)} = Ae^{\alpha(t-T)} \cos(\omega(t-T) - \varepsilon) + Be^{a_1(t-T)} \cos(\theta(t-T) - \varphi) \quad (22)$$

(21) and (22) are complementary sinusoidal functions with two variables (as  $y$  and  $t$ ), have four and eight parameters respectively and a non-linear structure. Since it has a nonlinear structure, data applied with this function has to provide statistical and linearity tests of the NLIN procedure<sup>8</sup>.

$y$  is the autonomous population which will be added to main trend function of  $F_{(t-T)}$ .

$\pm A$  and  $B$ : Initial amplitude of oscillation or peak and trough coefficient,

<sup>6</sup> There are also quite wide and useful explanations on sinusoidal functions in Tenenbaum and Pollard (1963, pp. 211-220, 313-392)

<sup>7</sup> See Allen (1956, pp. 123-124, 147-148) for proof from differential equation to general solution.

<sup>8</sup> All of the auxiliary sinusoidal functions we applied have had satisfactory statistical tests through this study.

$\omega, \varepsilon, \theta, \varphi$  are constants of cosines functions.

$a$  and  $aI$ : Damping or anti-damping multipliers, growth rates

If  $a$  and/or  $aI > 0$  *anti-damped* or *explosive* oscillations occur, when  $a$  and/or  $aI < 0$  amplitude *damped* and when  $a$  and/or  $aI = 0$  amplitude has *regular* oscillations.

In sinusoidal study, we found  $a$  very close to zero in conformity with theory. This means that functions of residuals we studied have a *regular* amplitude structure, neither dumped nor oscillates.

Consecutive application of sinusoidal function in the five-parameter form of Gompertz function is considered similar to Fourier series whose parameters change at each stage, not only once as in (21), since the next stage parameters are determined, calculated and included in the function through the residuals formed at previous stage. Statistics and linearity tests are performed as in the growth function in determining the coefficients of each of the sinusoidal functions. Thus, if it satisfies statistical tests of NLIN procedure, the residual series might be included in the main function via auxiliary sinusoidal functions.

Allen wrote:

“The interpretation of Fourier series is that any type of oscillation with given period  $\left(\frac{2\pi}{\omega}\right)$  can be approximated by the sum of *sinusoidal* oscillations by taking a sufficient number of terms in the series... The first of the sinusoidal oscillations has the same period or frequency as  $F(t)$ , the fundamental frequency. The latter sinusoidal oscillations in the sum have shorter periods, or higher frequencies in multiples of the fundamental frequency. Hence, in expressing  $F(t)$  as a sum of sinusoidal functions, there is a range of decreasing periods  $\left(\frac{2\pi}{n\omega}\right)$  ... Where the series is stopped is a matter of convenience or closeness of approximation. It is to be noticed that terms can be added without affecting those already written.”(1956, p. 121)

## 5. SAS® Studio Software

In our statistical studies, we used the SAS/STAT NLIN procedure, SAS/ETS ARIMA, AUTOREG and MODEL procedures of “SAS® OnDemand for Academics (SAS® Studio)” which is an advanced web-based SAS development environment for statistical and econometric analysis.

NLIN procedure which is developed for the analysis of nonlinear mathematical functions and provides very wide and flexible results in the statistical applications. Partial derivatives are taken according to the parameters in the function, the linearity of the function is checked with the results found and the relevant linearity criteria are included in the output in addition to known statistical tests: Hougaard’s skewness, Box’s bias measure and Bates and Watts’ global nonlinearity measures are among them<sup>9</sup>. Linearity tests provide close-to linear behavior of functions.(Ratkowsky, 1993)

ARIMA uses autoregressive moving-average errors to perform multiple regression analysis.

AUTOREG implements regression models that use time series data in which the errors are autocorrelated.

MODEL handles nonlinear simultaneous systems of equations, such as econometric models.

Procedure analyzes models in which the relationships among the variables form a system of one or more nonlinear equations. Primary uses of the MODEL procedure are estimation, simulation, and forecasting of nonlinear simultaneous equation models.

PROC MODEL features include the following:

- \_ SAS programming statements to define simultaneous systems of nonlinear equations
- \_ tools to analyze the structure of the simultaneous equation system
- \_ ARIMA, PDL, and other dynamic modeling capabilities
- \_ tools to specify and estimate the error covariance structure
- \_ tools to estimate and solve ordinary differential equations
- \_ the following methods of parameter estimation:

---

<sup>9</sup> For details of nonlinearity see (Gebremariam, 2014; SAS Institute Inc., 2018b).

- ordinary least squares (OLS)
- two-stage least squares (2SLS)
- seemingly unrelated regression (SUR) and iterative SUR (ITSUR)
- three-stage least squares (3SLS) and iterative 3SLS (IT3SLS)
- generalized method of moments (GMM)
- simulated method of moments (SMM)
- full information maximum likelihood (FIML)
- general log-likelihood maximization
- \_ simulation and forecasting capabilities
- \_ Monte Carlo simulation
- \_ goal-seeking solutions. (SAS Institute Inc., 2018a, pp. 1423–1424)

## **6. Statistical Application Results of Turkish Population Data**

Aiming to research the statistical representability of Turkish population growth through the entire series (1927-2080) with the mathematical curves, now we shall analyze the Turkish population in 2081, as if the population growth has already been completed in the country. Reliable long-term population projections published by TurkStat in 2018 facilitated the work of researchers. With the long-term projections which include sufficient data for the upper asymptotic region of the curves, we had the opportunity to focus on determining parameters, statistical test applications and obtaining statistical test values rather than deal with problems of census data.

### **6.1. Gompertz Growth Function (G5)**

At the first stage of statistical works of the Gompertz growth function (G5), we determined the values of the six parameters ( $K$ ,  $L$ ,  $Q_1$ ,  $Q_2$ ,  $r$ ,  $v$ ) simultaneously together with statistical test results of these parameters. Results are included in the Table 2a, 2b and 2c.

Convergence criterion/iteration phase is successful and the error sum of squares found as 1.883E14, which is quite reasonable amount for this study. Error sum of squares/total sum of squares (1.779E17) is at a level that is at a negligible level as 1058 per million. The time variable is successful in the explanation of population growth and the function/data compatibility is complete. Standard errors of the parameters, F-value, and 95% confidence limits are acceptable.

Hougaard's skewness figures of parameters are below 0.25%, which is the level of acceptance as the absolute value and Box's bias figures are below the acceptable level of 1% and almost zero. Nonlinear growth function of Gompertz is close to linear in case of Turkish population data (Ratkowsky, 1993). Global nonlinearity measures – first four row values of Table 2c are below the critical value of 0.6804 (5<sup>th</sup> row in the table 2c). Significant statistical values confirm the linearity of the nonlinear generalized function.

According to the Gompertz growth function, the lower asymptote is 15.4 million, and the upper asymptote is 112.7 million people. The figure to be focused on is the upper asymptote. Upper asymptote figure (112.7 million) is above the maximum figure of TurkStat's *main projections* which is 107.7 million (2069 max.). However, the Gompertz curve captured almost all of the TurkStat estimates for 2040-2080 period with zero deviation, except for 0.1 million deviation in 2050 (see table-11 in the "Results and Comments" below). We can say that Gompertz's five-parameter function is successful in identifying the details in the development of the population.

Table 2: SAS NLIN Procedure Statistical Test Results of Gompertz Curve (G5y1y2)

2a) Model Estimates								
	Source	DF	Sum of Squares	Mean Square	F Value	Approx. Pr.>F		
1	Model	4	1.777E17	4.442E16	35142.6	<.0001		
2	Error	149	1.883E14	1.264E12				
3	Corrected Total	153	1.779E17					
2b) Parameter Estimates								
	Parameter	Estimate	Approx. Std Error	Approximate 95% Confidence Limits		Hougaard's Skewness	Box's Bias	Percent Bias
1	K	1.1265E8	373008	1.1192E8	1.1339E8	0.1363	13330.1	0.012
2	N	15474709	335398	14811958	16137460	-0.0639	-7386.9	-0.05
3	Q1	7.9392	0.2561	7.4333	8.4452	0.00227	-0.00739	-0.09
4	Q2	2.8475	0.0770	2.6953	2.9997	0.000859	-0.00059	-0.02
5	r	0.0433	0.000644	0.0420	0.0446	-0.1261	-0.00002	-0.05
2c) Global Nonlinearity Measures								
1	Max Intrinsic Curvature				0.1068			
2	RMS Intrinsic Curvature				0.0562			
3	Max Parameter-Effects Curvature				0.2124			
4	RMS Parameter-Effects Curvature				0.1188			
5	Curvature Critical Value				0.6630			
6	Raw Residual Variance				1264E9			
7	Projected Residual Variance				249E8			

The first stage of the statistical study has provided the results we wanted until now: ESS/TSS figures are 104 and 1058 per million (0.0104% and 0.1058%) for Verhulst V6 and Gompertz G6 respectively. These figures are more than reasonable to work with. There's nothing to worry about. On the other hand, we cannot be satisfied that the analysis is sufficient and results are perfect by saying that time explains 99.9% of the population growth. Although the ESS's figures are low and relatively insignificant as a percentage, the residual figures for some years are large enough to affect the estimates as an absolute figure. Therefore, investigating possibilities of further lowering the absolute levels of ESSs' should be the next important research subject. On the other hand, we will explain the theoretical framework indicated above.

When the residuals curve of G5 is examined (blue line of figure 2), it is seen that the curve shows a structure of sinusoidal trend. The sinusoidal trend moves around the timeline, it does not disappear during the 155-year population growth period, on the contrary its period grows. Last period duration has grown significantly. The upper and lower boundaries of the residuals curve have an interval of -3 and +2 million. Figure 2 shows sinusoidal periods of residuals approximately:

- 1) 1935-1975: 40 years,
- 2) 1976-2026: 50 years,
- 3) 2027-2080: 53 years (three-quarter of period).

First and second periods run forty-fifty years successively, but projections for 2027-2080 period differ with three-quarters being 53 years and the 2060-2080 period is still exploding towards downside. By doing an additional

mathematical study, we have to investigate and explain this autonomous behavior which is second constituent of the main function.

When we include the sinusoidal functions of the residual elements which are two in our case, the main growth function becomes as follows:

$$Y_{(t-T)} = y1_{(t-T)} + y2_{(t-T)} + Y_0 + \frac{K - Y_0}{\left[ e^{[w_{(t-T)}A + (1-w_{(t-T)})B]} \right]} \quad (23)$$

Note:  $y1$  applied as eight-parameter, then  $y2$  as four-parameter respectively.

$y$ 's are sinusoidal functions that are independent of trend of growth function. They have their own growth coefficients. The last part of the equation is the Gompertz growth function driven by intrinsic rate of growth. During the application of sinusoidal function, linearity of the main function was almost not affected by changes, making it easier to work.

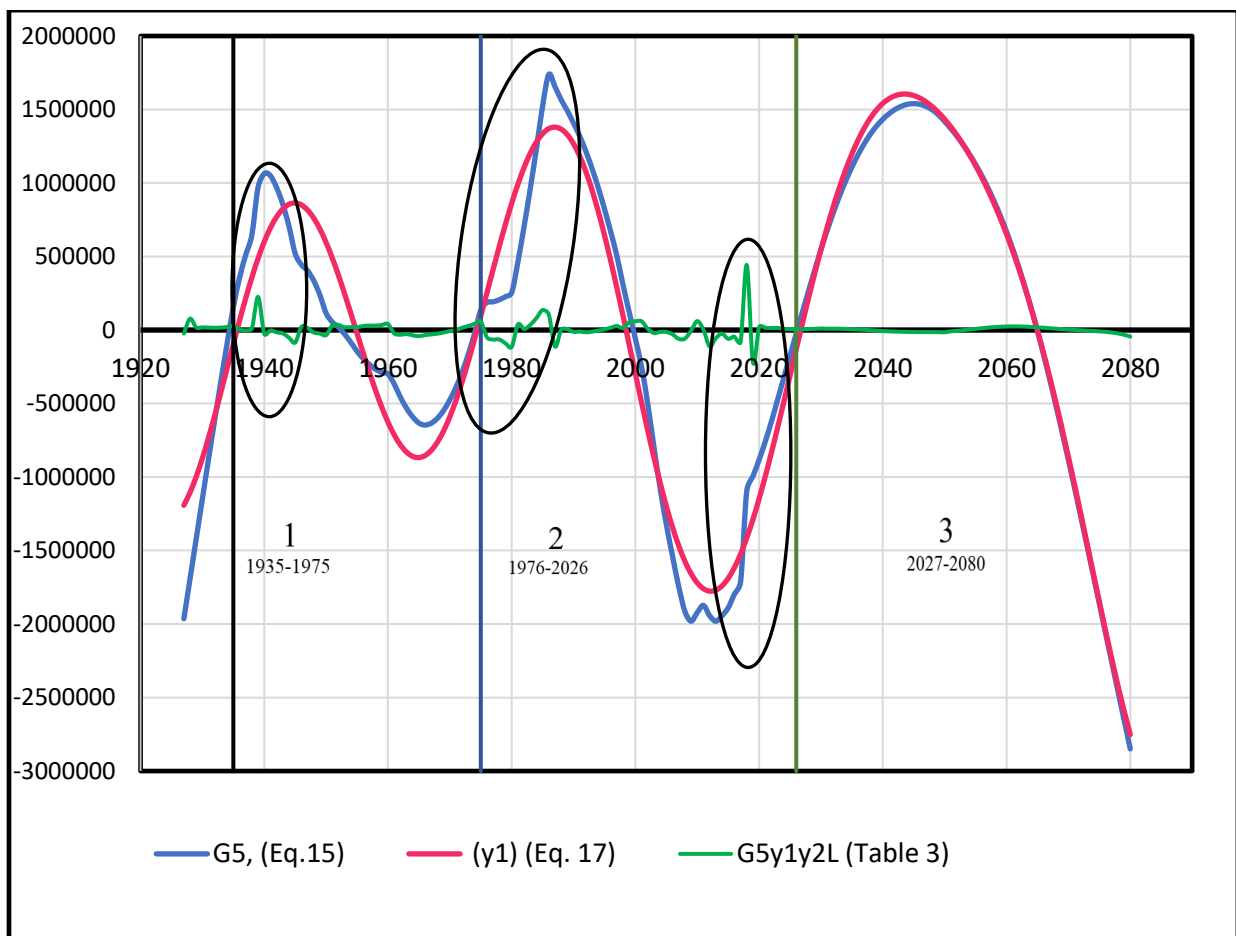


Figure 2: Residuals Curves of Five-parameter Gompertz Curve Family

Note to Figure 2: Black circles show data inconsistencies.

At second stage, curve fitting was made for the residuals series of the main equation obtained at first stage and  $y1$ -equation of the residuals- was obtained as a complementary function<sup>10</sup>.

<sup>10</sup> We found that with eight parameters (equation 7) in the first stage then with four parameters in second stage (equation 8) gave more advantageous results in statistical application.



$$y1_{(t-1951)} = -366560e^{(0.0167(t-1950))} \cos(0.0914(t - 1951) - 0.4816) + 1163610e^{(-0.00678(t-1950))} \cos(0.1443(t - 1950) + 0.6589) \quad (24)$$

At third stage, trend function (G5) plus y1 were combined to obtain the statistical results of the G5y1 function.

At fourth stage, y2 function was obtained through the residuals of the G5y1 function.

$$y2_{(t-1950)} = -414476e^{(2.6762(t-1950))} \cos(0.2225(t - 1950) - 0.3572) \quad (25)$$

At fifth stage, G5y1y2 was run together to achieve its statistical results.

In five stages, positive and satisfactory results were obtained from the statistical and linearity tests of both the main equation and complementary equations.

The application was terminated when the study on the residuals of the G5y1y2 equation did not achieve sufficient linearity results. However, an important step has already been taken to reduce residuals with y1 and y2 in these two-stage applications (as main and complementary functions).

After obtaining the main function together with complementary functions, the results of general statistical validity were investigated at the sixth stage.

At the statistical studies of G5y1y2, we applied the following methods of parameter estimation: Nonlinear Ordinary Least Squares (NOLS), Nonlinear Seemingly Unrelated Regression (NSUR) and Nonlinear Full Information Maximum Likelihood (NFIML). Since the results of the other two methods are very close to each other and to NFIML, we have only given the results of NFIML below to save some space in manuscript.

On the relationship between population development and time yielding autoregressive results, the model was redefined and the residuals of population's series -lag1 and lag2 -were added to the function as explanatory lagged variables and then desired results were reached with three explanatory variables and seven parameters. The final state of the function is as follows. The results were collected in thirteen subheadings in five tables and two figures.

$$Y_{(t-T)} = y1_{(t-T)} + y2_{(t-T)} + Y_0 + \frac{K - Y_0}{\left[ e^{[w_{(t-T)}A + (1-w_{(t-T)})B]} \right]} + X1 * lag1(\hat{Y} - Y) + X2 * lag2(\hat{Y} - Y) \quad (26)$$

Table 3: Main Statistical Test Results of SAS Software Proc MODEL Procedure

3a) Nonlinear FIML Summary of Residual Errors								
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq	Durbin Watson
Population	7	147	4.914E11	3.3425E9	57814.6	1.0000	1.0000	-
Residual Population	-	147	4.898E11	3.3317E9	57720.5	-	-	2.1747

3b) Nonlinear FIML Parameter Estimates					
Parameter	Estimate	Approx Std Err	t Value	Approx Pr >  t	Label
K	1.1263E8	231050	487.46	<.0001	Upper asymptote, at most level of population
N	15196918	108975	139.45	<.0001	Starting value
Q1	7.848847	0.0784	100.16	<.0001	First coefficient of exp growth base
Q2	2.794677	0.0314	89.01	<.0001	Second coefficient of exp growth base
r	0.043219	0.000232	186.03	<.0001	Intrinsic growth rate of population
Population_l1	1.228335	0.0441	27.84	<.0001	AR(Population) Population lag1 parameter
Population_l2	-0.46892	0.0722	-6.50	<.0001	AR(Population) Population lag2 parameter

3c) FIML Log Likelihood: -1903

Note: Population\_l1: lag 1 variable. Population\_l2: lag 2 variable.

The t-values of parameters are significantly above the t-table values and estimates are valid. The probabilities of invalidity of estimates are less than one in ten thousand.

SSE/TSS (4.898E11/1.779E+17) (2.75 per million) is at the acceptable level. The Durbin Watson coefficient indicates that there is no autocorrelation.

SAS Institute writes:

The option FIML (full information maximum likelihood) requests full information maximum likelihood estimation. If the errors are distributed normally, FIML produces efficient estimators of the parameters. If instrumental variables are not provided, the starting values for the estimation are obtained from a SUR (seemingly unrelated regression) estimation. If instrumental variables are provided, then the starting values are obtained from a 3SLS (three-stage least squares) estimation. The log-likelihood value and the l2 (lag 2) norm of the gradient of the negative log-likelihood function are shown in the estimation summary. (SAS Institute Inc., 2018a, p. 1500)

Table 4: Parameter Wald and Likelihood Ratio 95% Confidence Intervals

Parameter	Value	a) Parameter Wald 95 % Confidence Intervals		b) Parameter Likelihood Ratio 95% Confidence Intervals	
		Lower	Upper	Lower	Upper
K	1.1263E8	1.1218E8	1.1308E8	1.1257E8	1.1269E8
N	15196918	14983331	15410506	15170220	15223617
Q1	7.8488	7.6953	8.0024	7.8296	7.8680
Q2	2.7947	2.7331	2.8562	2.7870	2.8024
r	0.0432	0.0428	0.0437	0.0432	0.0433
Population_I1	1.2283	1.1419	1.3148	1.1419	1.3148
Population_I2	-0.4689	-0.6104	-0.3274	-0.4866	-0.4512

Note: Population\_I1: lag 1 variable. Population\_I2: lag 2 variable.

Table 5: Collinearity Diagnostics and Heteroscedasticity Test

5a) Collinearity Diagnostics									
Number	Eigenvalue	Condition Number	Proportion of Variation						
			K	N	Q1	Q2	r	Pop_I1	Pop_I2
1	3.103608	1.0000	0.0102	0.0116	0.0033	0.0053	0.0021	0.0000	0.0001
2	1.830841	1.3020	0.0000	0.0000	0.0000	0.0000	0.0000	0.0842	0.0841
3	1.452319	1.4618	0.0075	0.0437	0.0009	0.1158	0.0012	0.0000	0.0000
4	0.299204	3.2207	0.2499	0.0682	0.0299	0.1239	0.0007	0.0020	0.0002
5	0.168533	4.2913	0.0010	0.0000	0.0002	0.0000	0.0000	0.9129	0.9151
6	0.131736	4.8538	0.0918	0.7735	0.0272	0.7365	0.0097	0.0007	0.0000
7	0.013760	15.0184	0.6396	0.1031	0.9385	0.0185	0.9863	0.0001	0.0004

5b) Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
Pop	White's Test	49.32	29	0.0107	Cross of all vars
	Breusch-Pagan	2.28	2	0.3192	1, s, Population

Condition numbers of Table 5a are well below dangerous level of thresholds. Authors consider levels 10, 30 and 100 as thresholds. In general researchers agree that no single number can handle all situations. There is no collinearity in our study according to the thresholds 30 and 100. By threshold 10, the K, Q1, and r parameter values exceed 0.5 in 7th row (conditional number: 15.02 > 10). But, values of these parameters in 1-6 rows are very close to zero. Generally, we can say that there is no collinearity in the population study.

One assumption of regression analysis is heteroscedasticity. Accordingly, the variance of errors in observations are constant. Although the variance of errors is not distributed exactly normally in our study, our tests appear to be valid enough.

Table 6: Godfrey's Serial Correlation Test, Durbin-Watson Statistics and Normality Test Tables

6a) Godfrey's Serial Correlation Test				
Equation	Alternative		LM	Pr>LM
Pop	1		5.74	0.0165
	2		6.97	0.0306
	3		7.53	0.0567
6b) Durbin-Watson Statistics				
Equation	Order	DW	Pr<DW	Pr>DW
Pop	1	2.17	0.7430	0.2570
	2	1.81	0.0648	0.9352
	3	1.95	0.3209	0.6791
	4	1.93	0.3034	0.6966
6c) Normality Test				
Equation	Test Statistic		Value	Prob
Pop	Shapiro-Wilk W		0.71	<.0001
System	Mardia Skewness		236.5	<.0001
	Mardia Kurtosis		63.67	<.0001
	Henze-Zirkler T		12.16	<.0001

SAS Institute Inc. writes:

The GODFREY= option in the FIT statement produces the Godfrey Lagrange multiplier test for serially correlated residuals for each equation... The three variations of the test reported by the GODFREY=3 options are designed to have power against different alternative hypothesis. Thus, if the residuals in fact have only first-order autocorrelation, the lag 1 test has the most power for rejecting the null hypothesis of uncorrelated residuals. If the residuals have second- but not higher-order autocorrelation, the lag 2 test might be more likely to reject; the same is true for third-order autocorrelation and the lag 3 test. (SAS Institute Inc., MODEL Procedure, 2018a, p. 1531)

The generalized Durbin-Watson Statistics are above p-values in all four levels.

The null hypothesis of normality test of the Shapiro-Wilks assumes that residuals distributed normally. In the preparatory phase of our study, although we supported main function with complementary functions, inconsistencies from the data on the residuals have not been lost. In fact, the outliers have become even more obvious in the end. It is clear that the final series of residuals in Figure 2 is not normally distributed, and the outliers in the black circles specifically indicate this situation. The first two is due to the figures of the 1940 and 1986 censuses, and the last one is due to an outlier emerging from difference between 2017 census figure and projection figure of 2018. We chose not to do any corrective work with the data. However, arbitrariness did not disappear completely after complementary function studies. Although inconsistencies are still valid, the probability of the hypothesis that the normality test is invalid is less than one in ten thousand.

The covariance and correlation values of the parameters in Table 5a and b respectively are also at reasonable levels: only K and N highly interact with other parameters in the covariance table. There are three values in the correlation table that exceed 0.5: A high correlation of r with K and Q1 on the one hand and Q2 and N appear on the other. The figures are not sufficient high to adversely affect the validity of the model.

SAS Institute Inc. proposes: "Hausman's m-statistic can also be used, in principle, to test the null hypothesis of normality when comparing 3SLS to FIML. Because of the poor performance of this form of the test, it is not offered in the MODEL procedure. For a discussion of why Hausman's test fails for common econometric models, see Fair (1984, pp. 246-247)" (SAS Institute Inc., 2018a, p. 1556).

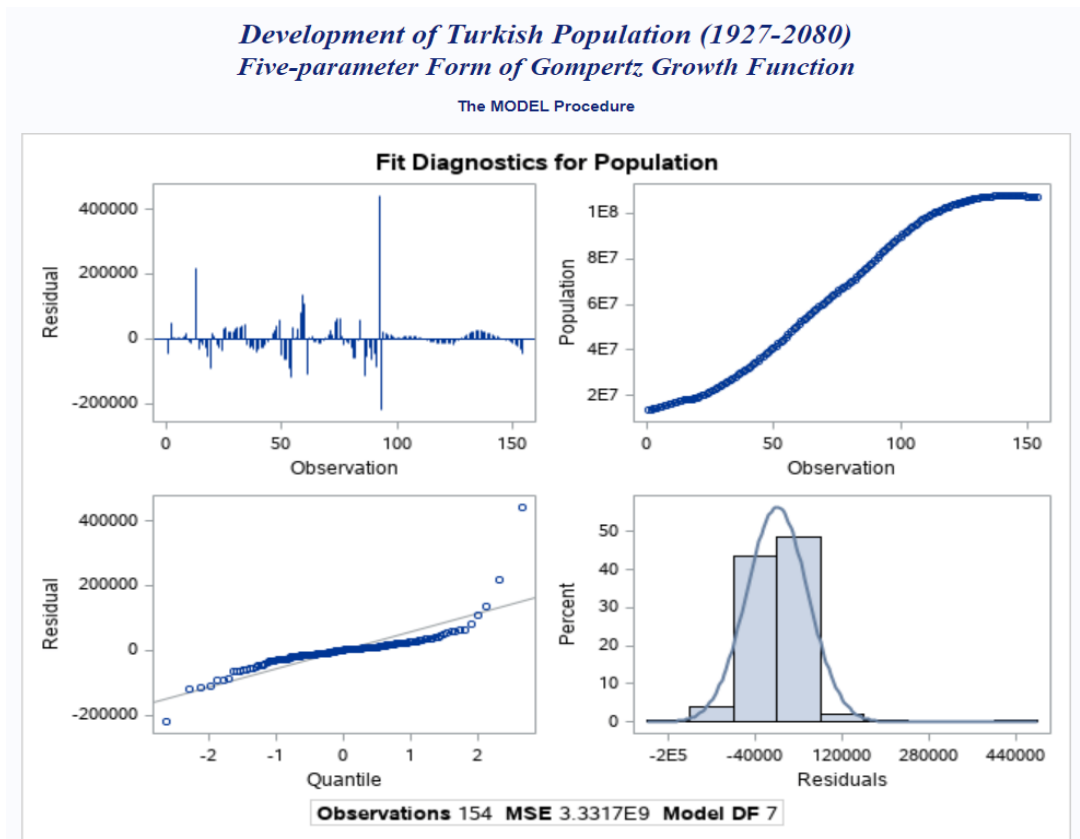


Figure 3: Fit Diagnostic Figures of Gompertz Function

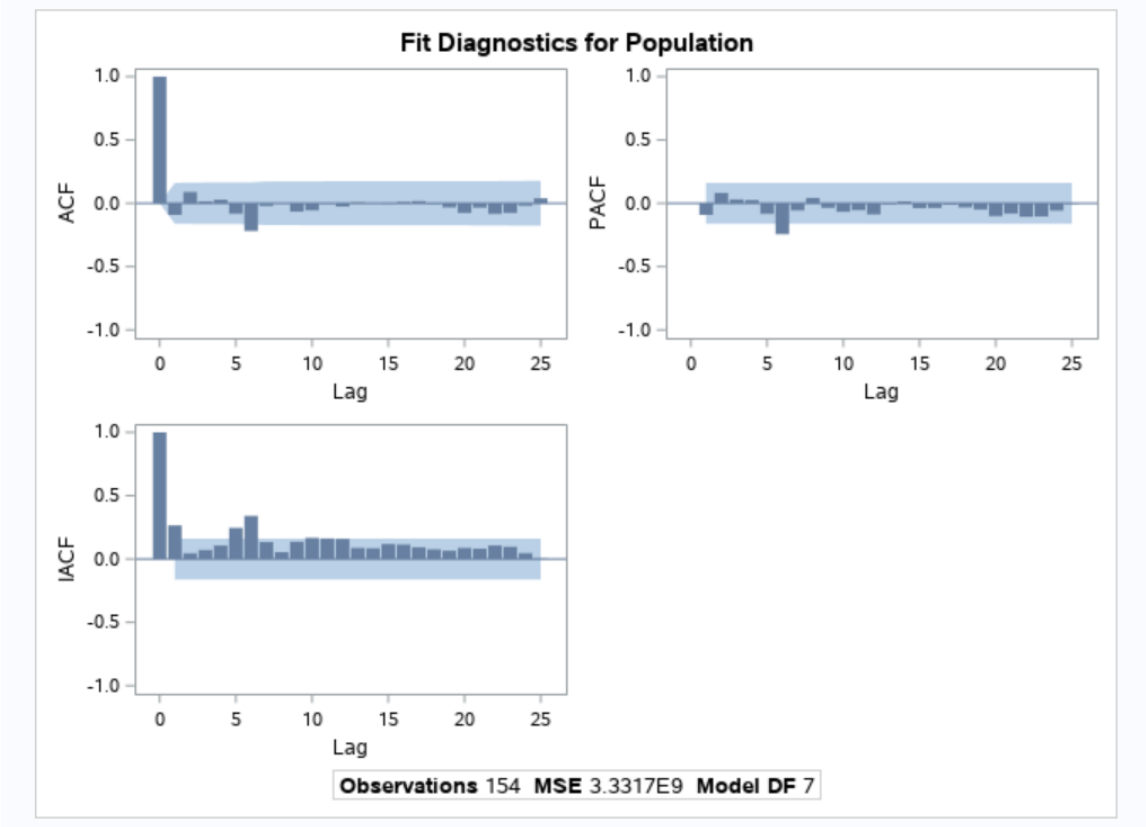


Figure 4: Fit Diagnostic Figures of Gompertz Function

Table 7: Covariances and Correlations of Parameter Estimates

7a) Covariances of Parameter Estimates							
	K	N	Q1	Q2	r	Pop_11	Pop_12
K	5.3384E10	-2.4356E9	-6737.432	421.966	-39.75519	-183.4953	-1917.902
N	-2.4356E9	1.1876E10	3666.762	2839.149	5.79392	-420.9370	710.099
Q1	-6737.432	3666.7617	0.006	0.000	0.00002	0.0002	0.002
Q2	421.96617	2839.1488	0.000	0.001	0.00000	-0.0001	0.000
r	-39.75519	5.7939243	0.000	0.000	0.00000	0.0000	0.000
Pop_11	-183.4953	-420.937	0.000	-0.000	0.00000	0.0019	-0.001
Pop_12	-1917.902	710.09938	0.002	0.000	0.00001	-0.0013	0.005

7b) Correlations of Parameter Estimates							
	K	N	Q1	Q2	r	Pop_11	Pop_12
K	1.0000	-0.0967	-0.3721	0.0582	-0.7406	-0.0180	-0.1150
N	-0.0967	1.0000	0.4294	0.8298	0.2288	-0.0876	0.0903
Q1	-0.3721	0.4294	1.0000	0.1836	0.8720	0.0455	0.4336
Q2	0.0582	0.8298	0.1836	1.0000	0.0064	-0.1010	0.0755
r	-0.7406	0.2288	0.8720	0.0064	1.0000	0.0253	0.4750
Pop_11	-0.0180	-0.0876	0.0455	-0.1010	0.0253	1.0000	-0.4001
Pop_12	-0.1150	0.0903	0.4336	0.0755	0.4750	-0.4001	1.0000

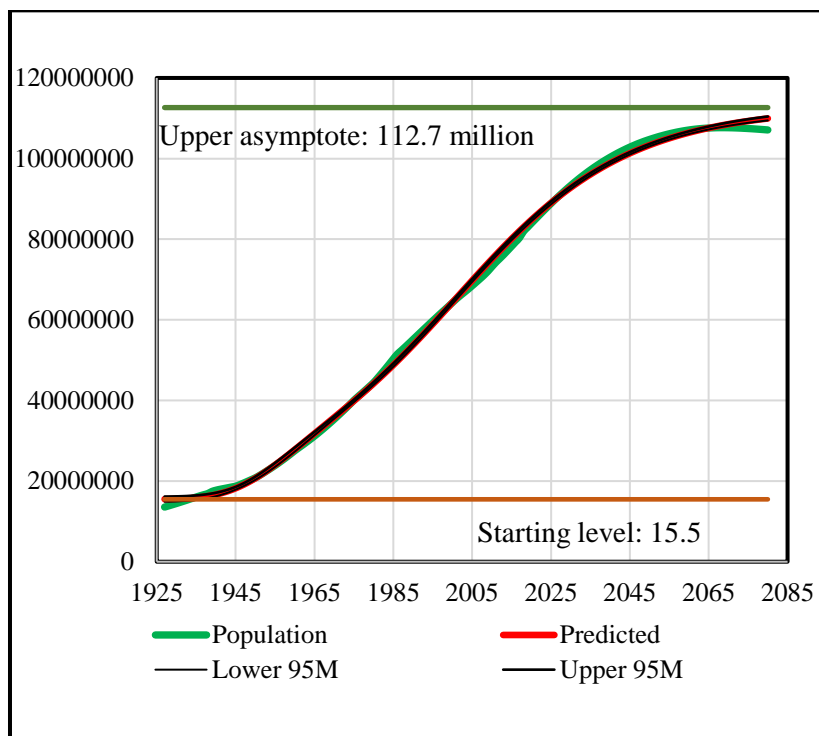


Figure 5: Population Curve of Five-parameter Gompertz Growth Function (G5)

After two complementary applications, ESS of main function (G5y1y2) which is 2.053E+12 reduced to 0.011% of the main function (G5) ESS value of 1.883E+14 (table 6). Gain is 1.862E+14. In this way, a further curve representation was realized within the narrower limits of confidence. We see that population growth curve, which has not yet reached the upper asymptote (112.6 million) from the path, is leveled off with the TurkStat projections (107 million) in the upper asymptote region. After application of sinusoidal equations, when we examine the Figure 3, we will see that the range of residuals has reduced from an interval of -3 and +2 million to almost interval of less than plus-minus two hundred thousand, and the regular trend of oscillations before 2000 have almost been disappearing after 2018 till 2080 (projections period of TurkStat) except for the tail which indicates population decrease estimate of TurkStat.

Table 8: Four-parameter (G4) Summary and Complementary Function Application Results of Five-parameter Gompertz Function (G5)

Equation name	G4	G5	G5y1	G5y1y2	G5y1y2L
Equation	-	-	-	Table 2	Table 3
Parameters	(K, L, Q, r)	(K, L, Q <sub>1</sub> , Q <sub>2</sub> , r)			
Upper asymptote (million)	119.1	112.7	112.6	112.6	112.6
Starting value (million)	15.3	15.5	15.4	15.2	15.2
Error sum of squares	4.111E+14	1.883E+14	7.671E+12	2.053E+12	4.898E11
Total sum of squares	1.779E+17	1.779E+17	1.779E+17	1.779E+17	1.779E+17
ESS/TSS (per million)	2311	1,058	43	12	2.76
ESS as % of four-parameter ESS	100	45.80	1.87	0.50	0.001191

Although we have obtained almost same results in G6 works- the 6-parameter Gompertz function (eq. 14)- we did not include results in our manuscript to save space because the ESS figure of G6 is already very close to the G5 ESS figure.

## 6.2. Verhulst Logistic Growth Function (V5 and V6)

In this section, we have decided to provide summary information about the Verhulst V6 function. Positive results were obtained in other test tables that we didn't provide. All comments on the Gompertz function also apply here. As can be seen from the pivot tables, the Verhulst function is more successful in capturing 108 million, the highest level of TurkStat estimates.

Complementary functions of Verhulst V6:

$$\hat{y}1_{(t-1960)} = 393510e^{(-0.0130(t-1960))}\cos(-0.2071(t - 1960) + 0.9723) \quad (27)$$

$$\hat{y}2_{(t-1960)} = -221926e^{(0.00515(t-1960))}\cos(-0.1529(t - 1960) - 0.0483) \quad (28)$$

$$\hat{y}3_{(t-1960)} = -0.00066(t - 1960)^5 + 0.1611(t - 1960)^4 - 11.6469(t - 1960)^3 + 141.1(t - 1960)^2 + 8153.2(t - 1960) - 93450.2 \quad (29)$$

Statistical tables of V6y1y2y3L:



Table 9: Statistical Test Results of Verhulst (V6) Growth Function

9a) Nonlinear FIML Summary of Residual Errors								
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq	Durbin Watson
Population	8	146	5.008E11	3.4303E9	58568.9	1.0000	1.0000	
Residual Population		146	5.014E11	3.4343E9	58603.1			2.1857
9b) Nonlinear FIML Parameter Estimates								
Parameter	Estimate	Approx Std Err	t Value	Approx Pr >  t	Label			
Q1e	6.417653	0.2307	27.82	<.0001	Log of first coefficient of exp growth base			
re	-2.33339	0.0300	-77.74	<.0001	Log of intrinsic growth rate of population			
Q2e	4.615833	0.1828	25.25	<.0001	Log of second coefficient of exp growth base			
v	3.048595	0.1170	26.05	<.0001	Allometric coefficient			
L	5158064	277384	18.60	<.0001	Lower asymptote, at least level of population			
K	1.0797E8	186188	579.91	<.0001	Upper asymptote, at most level of population			
Population_11	1.260139	0.0465	27.09	<.0001	AR(Population) Population lag1 parameter			
Population_12	-0.48668	0.0767	-6.35	<.0001	AR(Population) Population lag2 parameter			
a) FIML Log Likelihood: -1905								

Equivalences of log parameters:

$$Q_1 = e^{(Q1e)} = e^{6.417653} = 612.5637$$

$$Q_2 = e^{(Q2e)} = e^{4.615833} = 101.072$$

$$r = e^{(re)} = e^{-2.33339} = 0.096976$$

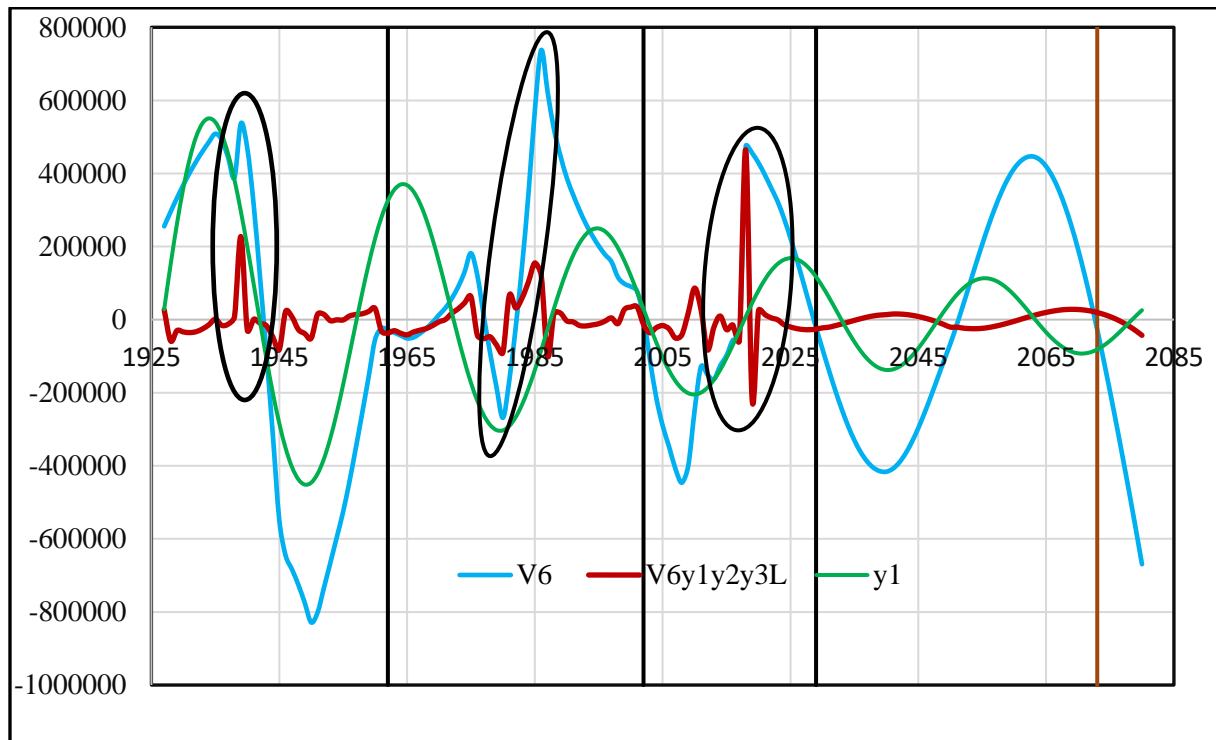


Figure 6: Residual Curves of Verhulst Function Family

Figure 6 note: Data inconsistencies of Figure 2 appear here as well.

Table 10: Summary Statistics of Verhulst Logistic Growth Functions

Equation name	5-prm V5	6-prm				
		V6	V6y1	V6y1y2	V6y1y2y3	V6y1y2y3L
Parameters	$(K, L, Q_1, Q_2, r)$	$(K, L, Q_1, Q_2, r, v)$				
Upper asymptote(million)	110.1	108.0	108.0	108.0	108.0	108
Lower asymptote (million)	9.7	5.8	5.4	5.1	5.1	5.2
Error sum of squares	1.095E+14	1.858E+13	1.109E13	3.540E+12	2.303E+12	5.014E+11
Total sum of squares	1.779E+17	1.779E+17	1.779E+17	1.779E+17	1.789E+17	1.789E+17
ESS/TSS (per million)	616	104	62	20	13	2.8
ESS as % of five-prm ESS	100	17	10	3.2	2.1	0.46
ESS as % of six-prm ESS	-	-	59.7	19.1	12.4	2.7

Prm: parameter.

## 7. Results and Comments

Even though we did not have data for the upper asymptotic region ( $K$ ) of the mathematical curves in our previous population study (İskender, 2018), we made population estimates with ten mathematical functions that we defined and statistically implemented. As a result of this, we were not able to determine the statistical test values for the parameters of the lower and upper asymptotic regions ( $L$  and  $K$ ), since we followed the minimization of the sum of squares by taking the lower and upper asymptotic parameters externally. In other words, we set our expectations for the mathematical functions we applied somewhat strongly high. However, we see that the results obtained<sup>11</sup> in this study are confirmed by the results obtained here in this study of TurkStat projections. The results of the study met our expectations. Turkey's population is expected to reach 108-114 million by the end of the twenty first century.

Both the known statistical and linearity test results obtained in this study demonstrate that it is possible to represent the *main scenario*. The projection methods used by Institute overlap with the mathematical modeling applied here. In the summary, Table 11 predictions for the years selected from the *main scenario* projections are given together with the estimates obtained from the six-parameter Verhulst and Gompertz functions. Although the TurkStat projections have foreseen the decline in the last decade of forecasts, the six-parameter logistic and Gompertz growth functions almost catch up with the annual estimates of the Institute.

Table 11: Population Figures Comparison for Selected Years (million)

Selected Years	Main Scenario Projections of TurkStat	Verhulst (V6y1y2y3L)	Difference of Verhulst	Gompertz (G5y1y2L)	Difference of Gompertz
2020	83.9	83.8	0.1	83.9	0.0
2030	93.3	93.4	-0.1	93.3	0.0
2040	100.3	100.3	0.0	100.3	0.0
2050	104.7	104.8	-0.1	104.8	-0.1
2060	107.1	107.1	0.0	107.1	0.0
2070	107.6	107.6	0.0	107.6	0.0
2080	107.1	107.1	0.0	107.1	0.0

We may summarize results of the econometric study as follows:

Our econometric analysis is fully integrated with population theory, economic modeling and statistical applications. These three aspects of the article were especially taken into account during this study. We think that it is the first study in our knowledge of mathematical economics and applied statistical methods. The development of Gompertz function as a population growth function with five and six parameters and its application in long-term population statistics data is the first study with successful results. Preferred Verhulst and Gompertz growth functions and Allen's approach of mathematical economics analysis have produced successful results in the Turkish population data. In this statistical study, reducing residuals with sinusoidal approach played an important role in achieving successful results from statistical criteria. Our analysis of TurkStat's 1927-2018 population data and 2018-2080 population estimates are a new approach to address in econometric study with the methods described.

Statistical applications and solutions that cannot be done manually, in a short time and accurately, have been obtained and achieved with the opportunities offered by SAS Studio software which is an advanced econometrics and statistics online environment. There was no way to carry out our work with ordinary softwares. On the other hand, only a small number of the obtained tables and results were included to the article for the sake of brevity. The software we use has much more possibilities. New statistical tests developed in the last fifty years namely;

<sup>11</sup> Upper asymptote values of 2018 study are: Four-parameter logistic 94000000, six-parameter logistic 115000000, four-parameter Gompertz 113800000.

Hougaard, Box and Bates, Ratkovsky etc. linearity tests, Godfrey serial correlation test, normality tests, White and Breusch-Pagan and others were extensively used in our study. Our software allows this. The linearity tests we mentioned are not available in any software other than SAS Studio. And it is necessary for a healthy application of mathematically nonlinear growth functions. Advanced statistical applications have been made possible with this software. All contributions of this software have developed this manuscript in the direction of both the purpose and results of our article.

TurkStat's 2018-2080 population projections are a valuable and important source to be considered by policy makers involved in population policies of Turkey. With its approach, the institution has achieved a significant success not only in preparation and publishing the statistics of the past years, but also with its future works: we can perceive this approach as a first in the *back to the future* way of understanding.

Five and six-parameter forms (inc. sinusoidal) of all growth curves applied in this study can be taken as a reference for population analysis of Turkey.

Econometric study has confirmed that Turkstat's *main scenario* projection is *consistent* in terms of statistical evaluation. All statistical test results of functions and curves successfully represent the Turkish population growth and its projections.

At the statistical analysis of economic categories such as income, demand, population, etc., the explanation of the residuals should be included in the study as well as the explanation of the main trend. This approach makes the results more valuable. Application of sinusoidal functions has led to significant decreases in ESS values and increased representation of curves, while it has highlighted features that were not appear in the trend of the population. My first excerpt from Allen (subdivision 7.1) explains the theoretical basis of state of affairs from mathematical economics to statistics. This is the method to be followed in econometrics.

Although sinusoidal approach reduced absolute level of residuals at a great extent, linearity measures are left unchanged which is an advantage of applied methodology.

It is possible to represent Turkish population development not with simple regression analyses, but with advanced econometric models, statistical applications and software in accordance with the growth theory taken from Allen, Verhulst and Gompertz.

From 2017 to 2080 population growth figure of 26.3 million is very close to Turkey's population level of 27.7 million people in 1960. In other words, in the next half century Turkey's population figure close to that of 1960 will be added to the 2017 population.

Considering Turkey's current political, social, economic and geographical problems, it is necessary to consider the planning for the next fifty years from today onwards.

It is clear that Turkey will need significant capital accumulation and investment when the economic demands of the future population to live by at least today's standards, not to mention of standards of nineteen sixties.

It is difficult to plan what kind of geography the future population of 26.3 million will live in. It is quite obvious that Turkey will face significant settlement problems when considering the earthquake structure, water resources, the settlement difficulties brought by the mountainous structure etc. It will be an important task of the public due to the fact that horizontal or vertical urbanization approaches are widely discussed.

It is inevitable that the population's settlement area demands will conflict with the needs for agricultural land.

Turkey is a state that has successfully protected its borders and geography throughout the history of the Republic. It is realistic to continue this assumption for future. Inevitably, population growth, problems and solutions will be shaped and located within these borders and framework.

## References

- Allen, R. G. D. (1956). *Mathematical Economics*. London: Macmillan & Co Ltd.
- Allen, R. G. D. (1969). *Mathematical Analysis for Economists*. London, UK: Macmillan and Co. Ltd. (Original work published 1938).
- Canpolat, S. B. & Gök, M. G. (2020). Census Methodology in Turkey: Transformation from Traditional to Register-based Administrative Data. *Statistical Journal of the IAOS*, 36(1), 177–185. <https://doi.org/10.3233/SJI-190561>
- Gebremariam, B. (2014). *Is Nonlinear Regression Throwing You a Curve? New Diagnostic and Inference Tools in the NLIN Procedure* (No. Paper SAS384-2014). Retrieved from <https://support.sas.com/resources/papers/proceedings14/SAS384-2014.pdf>
- İskender, C. (2018). Türkiye Nüfus Büyümesi ve Tahminleri: Matematiksel Büyüme Modelleri ve İstatistiksel Analiz ile Kuramsal ve Uygulamalı bir Yaklaşım. *Istanbul University Econometrics and Statistics E-Journal*, 14(28), 75–141. <https://doi.org/10.26650/ekoist.2018.14.28.0004>
- İskender, C. (2021). Mathematical Study of the Verhulst and Gompertz Growth Functions and Their Contemporary Applications. *Ekoist: Journal of Econometrics and Statistics*, (34). Retrieved from <https://dergipark.org.tr/pub/ekoist/issue/64260/876749>
- Piaggio, H. T. (1920). *An Elementary Treatise on Differential Equations and Their Applications* (W. P. Milne, ed.). London: G. Bell and Sons, Ltd.
- Ratkowsky, D. A. (1993). Principles of Nonlinear Regression Modeling. *Journal of Industrial Microbiology*, 12(3–5), 195–199. <https://doi.org/10.1007/BF01584190>
- Richards F. J. (1959). A Flexible Growth Function for Empirical Use. *Journal of Experimental Botany*, 10(2), 290–300.
- Ricketts, J. H. & Head, G. A. (1999). A Five-Parameter Logistic Equation for Investigating Asymmetry of Curvature in Baroreflex Studies. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 277(2), R441–R454. <https://doi.org/https://doi.org/10.1152/ajpregu.1999.277.2.R441>
- SAS Institute Inc. (2011). SAS/STAT® 9.3 User’s Guide The NLIN Procedure (Chapter). In *SAS Institute Inc. 2011. SAS/STAT® 9.3 User’s Guide*. Cary, NC: SAS Institute Inc. Retrieved from <https://support.sas.com/documentation/onlinedoc/stat/930/whatsnew.pdf%0A>
- SAS Institute Inc. (2013). SAS/STAT® 13.1 User’s Guide The NLIN Procedure. In *SAS Institute Inc. 2013. SAS/STAT® 13.1 User’s Guide*. Cary, NC: SAS Institute Inc. Retrieved from <https://support.sas.com/documentation/onlinedoc/stat/131/nlin.pdf>
- SAS Institute Inc. (2018a). SAS/ETS® 15.1 User’s Guide The MODEL Procedure. In *SAS Institute Inc. 2018. SAS/ETS® 15.1 User’s Guide*. Cary, NC: SAS Institute Inc. *SAS/ETS®* (2018th ed.). Retrieved from <https://support.sas.com/documentation/onlinedoc/ets/151/model.pdf>
- SAS Institute Inc. (2018b). SAS/STAT® 15.1 User’s Guide The NLIN Procedure. In *SAS Institute Inc. 2018. SAS/STAT® 15.1 User’s Guide*. Cary, NC: SAS Institute Inc. Retrieved from <https://support.sas.com/documentation/onlinedoc/stat/143/nlin.pdf>
- Tenenbaum, M., & Pollard, H. (1963). *Ordinary Differential Equations*. New York: Dover Publications.
- TurkStat. (2021). Population and Demography Statistics. Retrieved from <https://data.tuik.gov.tr/Kategori/GetKategori?p=Population-and-Demography-109>
- Verhulst, P.-F. (1838). Notice Sur La Loi Que La Population Suit Dans Son Accroissement. In A. Quetelet & J. G. Garnier (Eds.), *Correspondance Mathématique et Physique* (Vol. 4, pp. 113–121). Retrieved from <https://archive.org/details/correspondancem02belgoog/page/n122/mode/2up>



## Türkiye'de Hanehalkı Yoksulluğunu Etkileyen Faktörlerin Belirlenmesi

Onur ŞENTÜRK

Türkiye İstatistik Kurumu / TÜİK Uzmanı

[onur.senturk@tuik.gov.tr](mailto:onur.senturk@tuik.gov.tr)

Orcid No: 0000-0002-6752-4963

Nuri ÇELİK

Gebze Teknik Üniversitesi / Doç. Dr.

[nuricelik@gtu.edu.tr](mailto:nuricelik@gtu.edu.tr)

Orcid No: 0000-0002-4234-2389

### Özet

21. yüzyılda günümüzün en önemli sorunlarından biri olarak hayatımızda olan yoksulluk olgusu, ferdi hayatı üzerinde farklı derecelerde de olsa etkisini sürdürmektedir. Bu nedenle, yoksullukla mücadele için yoksulluğa neden olan etmenlerin bilinmesi önem teşkil etmektedir. Bu çalışmada amaç hanehalkı yoksulluğunu etkileyen faktörleri tespit etmek ve ANOVA aracılığıyla bunların söz konusu yoksul hanelerin kullanılabilir gelirleri üzerindeki etkilerinin büyüklüğünü tahmin etmek ve belirlemektir. Bu amaçla çalışmada Türkiye İstatistik Kurumu tarafından üretilmiş olan Gelir ve Yaşam Koşulları Araştırması'na ait olan 2016 yılı verileri kullanılmıştır. Hanehalkı yoksulluğuna etki eden faktörlerin belirlenmesinde ANOVA kullanılmıştır ve veriler SPSS programıyla analiz edilmiştir.

*Anahtar Sözcükler: ANOVA, Gelir ve Yaşam Koşulları Araştırması, Yoksulluk.*

**Sorumlu Yazar / Corresponding Author:** 1-Onur ŞENTÜRK, Türkiye İstatistik Kurumu, İşgücü ve Yaşam Koşulları Daire Başkanlığı, Hanehalkı Gelir ve Tüketim İstatistikleri Grup Başkanlığı.

2-Nuri ÇELİK, Gebze Teknik Üniversitesi, Temel Bilimler Fakültesi, Matematik Bölümü.

**Atf / Citation:** ŞENTÜRK O., ÇELİK N. Türkiye'de Hanehalkı Yoksulluğunu Etkileyen Faktörlerin Belirlenmesi. İstatistik Araştırma Dergisi, 2021, 11 (2), 56-70.

## Determination of Factors Affecting Household Poverty in Turkey

### Abstract

In the 21<sup>st</sup> century, the phenomenon of poverty, which is one of the most important problems of today, continues to affect the life of the individual, even if in different levels. Therefore, in order to struggle against poverty, it is important to know the factors that cause poverty. The aim of this study is to determine the factors affecting household poverty and to estimate and determine the magnitude of the effects of these factors on the disposable income of the poor households through ANOVA. In this context, 2016 Income and Living Conditions Survey data serviced by Turkish Statistical Institute (TurkStat) were used in this study. In order to determine the factors effecting household poverty, ANOVA was applied, and analyzed with the SPSS program.

*Keywords: ANOVA, Poverty, Study of Income and Living Conditions.*

### 1. Giriş

Yoksulluk, insanlık tarihinin en eski konularından biridir. Buna rağmen yoksulluğun tanımının her kesim tarafından kabul edilen bir şekli yoktur. Yoksulluk kavramı birden fazla boyuta sahip bir kavramdır. Bu kavramın, yeterli besin maddesine ulaşılamaması biçiminde ifadesi ya da yalın şekilde aç kalmak olarak ifadesi doğru değildir. Yoksulluğun genel bir tanımı, fertlerin belirli bir ekonomik refahın altında yaşamlarını sürdürmesidir (Yaşar ve Taşar, 2019).

Yoksulluk, toplumların gelişmişlik düzeylerine bakılmaksızın meydana gelen genel bir sorundur. Yani dünyanın birçok ülkesinde bu soruna rastlanmaktadır. Nedeni toplumdaki topluma göre farklılık gösterdiğinden çözümünde de öncelikle yoksulluğun nedenlerin belirlenmesi gerekmektedir. Gelişmekte olan diğer ülkelerde olduğu şekilde ülkemizde de özellikle son yirmi yıldır yoksulluk önemli bir problem haline gelmiştir. Yoksulluk olgusunun incelenmesinde Türkiye İstatistik Kurumu (TÜİK) başta olmak üzere bazı kamu kurum ve kuruluşları ile birçok araştırmacı çalışmaktadır. Ancak buna rağmen sorun hala geçerliliğini korumaktadır. Yapılan çalışmalar sayesinde yoksulluğu etkileyen faktörler tespit edilip politika yapımcılarının da bunlara göre mücadele yöntemleri geliştirmeleri yoksulluğun önlenmesi açısından büyük önem arz etmektedir.

Yoksulluk kavramı çok yönlü bir kavramdır ve oldukça tartışmalıdır. Aksan (2012) tarafından da ifade edildiği gibi Tanımlar, bir toplumsal yapıdan diğerine, zamanın etkisi ile de farklılık göstermektedir. Bu farklılığın sebebi yoksullara bakış açılarında göre farklı değerler sistemini içinde barındırmasından kaynaklanmaktadır.

Kelime olarak yoksulluk, yaşamın gerektirdiği olanaklara sahip olmama durumunu ifade eder. Yaşamı devam ettirmek için gerekli olan şeyler fertlere, içinde bulunulan topluma, çevreye ve koşullara bağlı olarak değişiklik gösterir. Bu yönüyle yoksulluk kavramı her zaman gözlemlenen durumla standart durumun karşılaştırılmasını sağlar (Aktan ve Vural, 2002).

TÜİK'e göre yoksulluk ise insanların temel ihtiyaçlarını karşılayamaması durumudur. Yoksulluk kavramı iki farklı şekilde ele alınmaktadır. Dar anlamda yoksulluk, açlıktan ölme ve barınmadan yoksunluk iken geniş anlamda yoksulluk ise insan hayatının sürdürülebilmesini sağlayan gıda, giyim ve barınma gibi temel ihtiyaçlara sahip olmasına rağmen yaşadığı çevrenin genel seviyesinin gerisinde kalmayı ifade etmektedir (TÜİK, 2008).

Literatürde daha başka yoksulluk tanımlarını da görmek mümkündür. Yapılan bütün tanımlar gözden geçirildiğinde yoksulluğa ilişkin genel bir tanım; minimum yaşam standardına erişememe şeklindeki temel harcama ihtiyaçları ya da onları tatmin için gerekli olan gelir düzeyine erişememe şeklinde yapılabilir.

Ayrıca yoksulluk literatürü incelendiğinde mutlak, göreceli ve öznel (sübjektif) yoksulluk gibi farklı yoksulluk çeşitlerinin de olduğu görülmektedir. Mutlak yoksulluk, temel fiziksel kriterleri karşılamak için minimum tüketim standardına erişebilme becerisinden yoksun olma şeklinde ifade edilebilir. Başka bir deyişle hanehalkının ya da ferdin asgari yaşam düzeyini sürdürebilmesi için gerekli, yalnızca en temel ihtiyaçlarını bile karşılayamaması durumudur (Taş ve Özcan, 2012). Ülkelerdeki yoksulluk düzeyleri söz konusu ülkelerin gelişmişlik düzeyi ile alakalıdır.

Gelişmiş ülkeler için, mutlak yoksulluk kavramında kullanılan eşik değerler minimum yaşam standartları üzerinde kurulmaktadır. Bu sebeple, mutlak yoksulluk yaklaşımı gelişmiş ülkeler için kullanılmamaktadır. Ancak, gelişmekte veya az gelişmiş ülkeler için nüfuslarının çok büyük bir kısmı geçim düzeyinin altında olduğu için mutlak yoksulluk analizlerinin yapılması gerekmektedir. (Foster ve Shorrocks, 1991). Göreceli yoksulluk, ihtiyaçlarını mutlak olarak karşılayabilen fakat içinde yaşadıkları toplumun ortalama refah düzeyini yakalayamayanların yoksulluk durumu

olarak görülmektedir (Şenses, 2002). Gelişmekte olan ülkeler için mutlak yoksulluk yaklaşımını kullanmak daha uygunken, gelişmiş ülkeler için göreceli yoksulluk yaklaşımı daha uygundur (Arabacı, 2016). Göreceli ve mutlak yoksulluk tanımları arasındaki en önemli fark, mutlak yoksulluğun sabit bir çizgide ilerlemesidir. Göreceli yoksulluk sınırı, ülke içindeki ortalama refah (gelir ve harcama) düzeylerine göre değişmektedir. Ortalama harcama seviyesi ne kadar yüksekse, nispi yoksulluk oranı da o kadar yüksektir. Göreceli bir yoksulluk sınırının benimsenmesiyle yaşam standartlarında genel bir artış varsa, yoksul tabakaların payı mutlaka azaltılmamalıdır (Coşkun ve Tireli, 2008). Gelir ve tüketim esasına dayalı olan mutlak ve göreceli yoksulluktan oldukça farklı bir yaklaşım olan öznel yoksulluğun temeli Adam Smith'in yoksullukla ilgili düşüncelerine dayanmaktadır. Adam Smith'e göre bütün insanlar eşit fırsatlara ve seçimlere sahiptir ve fertlerin bu olasılıkları ve seçeneklerini kullanmaları gerektiğini savunur. Fertler içinde buldukları toplumda var olmaktan utanmamalı, diğer insanlarla bütünleşebilmeli ve topluma dahil olmalıdır. Adam Smith tarafından geliştirilen bu tanım günümüzde "öznel yoksulluk" yaklaşımı olarak yeniden tanımlanmıştır (Yüce ve Özbek, 2006).

Yoksulluk kavramı ve türleri her ülkeyi ilgilendiren bir kavram olduğundan ülkelerde birbirlerinden farklı yapılarda olduğundan yoksulluğu ölçmekte kolay değildir. Sen (1976), yoksulluğun ölçümü ile ilgili iki farklı problemin varlığına işaret etmektedir. Bu sorunlar sırasıyla, tüm nüfus içindeki yoksulluğu tanımlamak ve yoksulluk endekslerini oluşturmak için yoksullar hakkında önemli bilgiler toplamaktır. Literatürde bazı çalışmalarda yoksulluk sadece yoksulluk sınırının altında kalan fertlerin değerlendirilmesiyle belirlenmektedir. Öte yandan yoksulluk sınırı, yoksul fertler arasında yoksulluğun yoğunluk derecesi, yoksulluğun şiddeti, derinliği ya da yoğunluğu hakkında bilgi vermektense uzaktır. Yoksulluk sınırının altındaki iki grup incelendiğinde, gruplardan biri diğerine göre yoksulluk sınırına daha yakın veya daha uzak olabilir. Bu durumda kimin daha yoksul olduğunu değerlendirmek için ulusal ve uluslararası alanlarda çeşitli yoksulluk endeksleri geliştirilmiştir. Bu çalışmada yoksulluğu ölçme yöntemleri; Kafa Sayısı Yoksulluk Endeksi (KSE), Yoksulluk Açığı Endeksi, Sen Endeksi, Foster-Greer-Thornbecke Endeksi (FGT Endeksi), İnsani Gelişmişlik Endeksi, İnsani Yoksulluk Endeksi ve Çok Boyutlu Yoksulluk Endeksi incelenmiştir.

Kafa sayısı yoksulluk endeksi, yoksulluğun ölçülmesinde hesaplanması, yorumu kolay, veri gereksinimi az, basit ve net olması nedeniyle en yaygın olarak kullanılan yöntemdir. KSE ölçüm yöntemine göre; fert başına düşen geliri, belirlenen yoksulluk sınırının altında olan fertlerin, toplam nüfus içindeki yüzdesi olarak ifade edilmektedir (Ravallion ve Datt, 1994). KSE'ye eleştirilerden yola çıkılarak geliştirilmiş olan ikinci bir ölçüm yöntemi yoksulluk açığı endeksidir. Yoksulluk açığı, yoksulluğun derecesine ilişkin bilgi verir. Yani, ferdin yoksulluk sınırı ile şahsi geliri arasındaki fark yoksulluk açığını vermektedir. Sonucun 100'e yaklaşması toplumun yoksulluk riskinin fazla olduğunu, sonucun 0'a yaklaşması ise yoksulluk riskinin azaldığını gösterir (TÜİK, 2016). Amartya Sen tarafından geliştirilen "Sen Endeksi" yoksullar arasındaki gelir dağılımını göz önünde bulundurarak yoksulların sayısını ve yoksulluğun boyutunu da kapsayan bir endekstir. Sen'in bu endeksi oluşturmasındaki amaç, yoksullar arasında en iyi ve en kötü, en yoksul şeklinde karşılaştırmalar yapılabilmesini sağlamaktır (Şenses, 2002). FGT Endeksi, diğer endekslere göre yoksul kitleyi de birbirinden ayırarak, toplam yoksulluğu farklı alt gruplara ayırarak, bu endeks yoluyla, yoksul kitle birbirinden kesin çizgilerle ayrılmış değişik sosyo-ekonomik veya bölgesel gruplara bölünebilmekte ve bunlara ilişkin hem kafa sayısına hem de yoksulluk sınırının altında kalan fert sayısının hesaplanabilmesine olanak verir ve yoksulluğu giderecek imkanların belirlenmesini sağlar. İnsani Gelişmişlik Endeksi (İGE), 1990 yılından itibaren UNDP tarafından geliştirilen İnsani Gelişme Raporlarında yer alan ve yoksulluğu ölçen bir araç olarak tanımlanmaktadır. Bu yöntem, Amartya Sen tarafından yönetilen Kapasite Yaklaşımına ilkesine dayanmaktadır. Söz konusu yaklaşımda yoksulluk gelir-harcama eksikliği değil, kapasite eksikliği olarak tanımlanmaktadır. İnsani Yoksulluk Endeksleri, insani gelişmişliğin dağılımını ve hâlâ mevcut olan yoksunluğun (deprivations) ana kaynaklarını ölçmektedir (Coşkun ve Tireli, 2008). UNDP desteğiyle Oxford Üniversitesi Yoksulluk ve İnsani Gelişme Girişimi ve UNDP'nin İnsani Gelişme Rapor Ofisi tarafından 2010 yılında İnsani Gelişme ve İnsani Yoksulluk Endekslerine ek olarak, yaratıcılarının gelişme kaynaklarının daha etkin olmasına yardımcı olacağını belirttikleri ve yoksulluk içinde yaşayan insanların görüntüsünü yansıtmakta olan yeni bir "çok boyutlu" yoksulluk ölçütü olarak geliştirilmiştir.

Literatürde gelir ve yaşam koşulları araştırması (GYKA) kullanılarak yapılan çalışmaların sayısı oldukça fazladır. Bunlardan bazılarına bakılacak olursa: Castellona ve Punzo (2012), yaptıkları çalışmalarında, hane düzeyinde toplam cevapsızlığın çeşitli bileşenleriyle ilgili örnekleme dışı hatalara odaklanarak, EU-SILC verilerinin İtalya bölümünün doğruluğunun değerlendirilmesi için bir çerçeveyi tartışmaktadır. Bazı geçici temel kalite oranlarının bir araya getirilmesiyle kalite göstergelerinin sınıfları elde edilir. Ardından, bölgesel perspektifi ve bunun toplam cevapsızlık üzerindeki etkilerini araştırmak için tek yönlü ANOVA modeli ve çok terimli lojistik regresyon modelini kullanmışlardır. Arbia ve Pace (2018), eşitsizlik endeksinin gereksinimlerine daha iyi uyması için değişim katsayısını araştırmak üzere uyarlanmış ANOVA çerçevesini kullanarak iki eşitsizlik kaynağı arasındaki farkları test etmeyi önermişlerdir. Önerilen prosedür, Gelir ve Yaşam Koşullarına İlişkin İtalya Anketine uygulanmıştır. Sonuçlar, İtalya'da fırsat eşitsizliğinin artmasını teşvik eden koşulların belirlenmesine yardımcı olmuştur. Karcı ve Arlı (2018), yaptıkları çalışmalarında amaç olarak maddi yoksunluğu etkilediklerini düşündükleri değişkenlerin istatistiki etkinliklerini belirlemektir. Kurdukları modelde hanehalkı sorumlusunun eğitim düzeyinin, istihdam durumunun,



yaşının ve sağlık durumunun yanı sıra konutun oda sayısı, ısıtma sisteminin ve Türkiye'nin bölgeleri de maddi yoksunluğu anlamlı bir şekilde etkilediğini görmüşlerdir. Evcim ve Ark. (2019), Türkiye'de toplumsal refah artışını engelleyen hanehalkı göreceli yoksulluğunun, boyutları ve temel belirleyici faktörlerini incelemişlerdir. Bu amaçla analizde TÜİK 2017 yılı GYKA mikro veri seti kullanılmışlardır. Bağımlı değişkenin yoksulluk sınırına göre yoksul ve yoksul olmayanlar olarak ele alındığı İkili Lojistik Regresyon analizinde, hanehalkı demografik özellikleri, işgücü piyasası ve coğrafi özellikleri içeren açıklayıcı değişkenlerini kullanmışlardır. Ünver ve Alkan (2020), çalışmalarında TÜİK tarafından 2017 ve 2018 yıllarında gerçekleştirilen GYKA'nın mikro veri setlerini kullanarak bireylerin maddi yoksunluk durumunu etkileyen faktörlerin araştırılmasını amaçlamışlardır. Bu amaç için probit regresyon yöntemini uygulamışlardır.

Yapılan bu çalışmada, Türkiye'de yoksulluğu etkilediği düşünülen hane bazında seçilen göstergelerin yoksulluğu ne kadar etkilediği incelenmiştir ve yoksul olanları en iyi şekilde tahmin edebilen modelleri belirlemek amaçlanmıştır. Bu amaçla, TÜİK tarafından yapılan GYKA 2016 yılı anket verisi kullanılmıştır. Bu kapsamda ilk olarak yoksulluk kavramı hakkında genel bir bilgi verilmiş, kavramsal ve ölçümsel düzeyleri gösterilmiştir. Ardından deney tasarımı hakkında bilgi verilmiş ve kullanılan deney tasarımı modelleri tanıtılmıştır. Çalışmanın uygulama bölümünde ise Türkiye'deki hanehalklarının yoksulluğunu etkileyen faktörleri ortaya koymak amacıyla medyan gelirin %40'ına göre yoksul olan haneler incelenmiş bunlara ilişkin betimsel istatistikler verilmiştir. Yoksulluk oranı EUROSTAT metodolojisi gereği medyan gelirin %40, %50, %60 ve %70 i kullanılarak açıklanmaktadır (TÜİK, 2016). Bu çalışmada %40 seçilme sebebi en yoksul haneleri incelemektir. Analizler IBM SPSS 22 programı kullanılarak yapılmıştır.

## 2. Deney Tasarımı

Deney tasarımı, Fisher (1920) tarafından yürütülen tarım alanındaki deneyler uygulanırken keşfedilmiş ve geliştirilmiştir. Fisher, ayrıca bu deneyleri günümüzde halen kullanılmakta olan “varyans analizi” yöntemini de uygulamıştır. Deney tasarımı (DT), ele alınan süreci etkileyen etkenler üzerinde ayarlamalar yaparak sürecin performansının geliştirilmesini ve çıktılar üzerindeki değişikliklerin gözden geçirilmesidir. Deney tasarımının amacı, gözlenen veriler yarımıyla, süreci etkileyen faktörlerin gözlenmesi ve hangi faktörlerin ne şekilde sürece dahil edileceğini tespit etmektir. Böylece deney süreci için optimal koşullara ulaşılmış olunur. Dolayısıyla, deney tasarımı süreç geliştirmekte ve iyileştirmekte önemli bir rol oynar. Ayrıca, bir ürün veya sürecin kontrol dışı dış etkenler ışığında istenen performansı göstermesi biçiminde ifade edilen robust (sağlam) tasarımların ortaya konulmasında DT yaygın bir şekilde kullanılmaktadır (Montgomery, 1991).

Literatürde kullanılan farklı deney tasarım yöntemleri mevcuttur. Bu çalışmada tek yönlü ANOVA, iki yönlü ANOVA ve faktöriyel tasarım kullanılmıştır.

### 2.1. Tek Yönlü Varyans Analizi

Varyans analizinde en çok kullanılan modeldir. Buradaki ‘tek’ in anlamı, farklı düzeyleri olan bir faktörün olmasıdır. Tek yönlü ANOVA modelinde, deney birimlerinin homojen olduğu ve denemelerin deney birimlerine rasgele olarak atıldığı varsayılır. Böylelikle, deney birimleri denemelere eşit olasılıklarla atanmış olur. Rasgelelik sağlanmazsa, denemeler arasındaki farklar ile hataların varyansının tahmin değerleri yanlış olacaktır (Rangaswamy 1995). Tek yönlü ANOVA için matematiksel model,

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, i = 1, 2, \dots, a; j = 1, 2, \dots, n \quad (1)$$

Burada,  $y_{ij}$ : inci denemede jnci gözlem değerini,  $\mu$ : genel ortalamayı,  $\tau_i$ -inci deneme etkisini,  $\varepsilon_{ij}$  rasgele hata terimlerini göstermektedir. Benzer şekilde a, deneme sayısını, n ise her bir denemede gözlem sayısını göstermektedir. Böyle toplam gözlem sayısı  $N=an$  şeklinde ifade edilmektedir.

Tek yönlü ANOVA'da, denemeler arası farkın önemli olup olmadığının araştırılmaktadır. Varyans analizi tablosu:

Tablo 1. Tek yönlü ANOVA Tablosu

Değişim Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	F
Deneme	a-1	$KT_{Deneme}$	$KO_{Deneme}$	$KO_{Deneme} / KO_{Hata}$
Hata	N-a	$KT_{Hata}$	$KO_{Hata}$	
Toplam	N-1	$KT_{Genel}$		

Burada,  $KT_{Genel} = \sum_{i=1}^a \sum_{j=1}^n (Y_{ij} - \bar{Y}_{..})^2$  genel kareler toplamıdır. Genel Kareler toplamı ise deneme kareler toplamı ve hata kareler toplamına eşittir. Deney için kullanılan toplam gözlem sayısı N olduğu için, GKT için N-1 serbestlik derecesine sahiptir. Her bir denemenin n tane gözlemi olduğundan, her bir deneme için serbestlik derecesi n-1 olup, a adet deneme bulunduğu için, hata için serbestlik derecesi a(n-1)=N-a şeklinde olacaktır. Hata kareler toplamı da,  $KT_{Hata} = \sum_{i=1}^a \sum_{j=1}^n (Y_{ij} - \bar{Y}_{.j})^2$  ile elde edilir. Sonuç olarak hesaplanan F istatistiği tablo değerinden büyükse denemeler arasında anlamlı bir farklılık olduğu sonucuna varılır.

ANOVA sonucunda süreç, gruplar arası farklılığın bulunmaması durumunda sonlanır. Ancak gruplararası farkın olduğu durumda, bu farklılığa hangi grubun yol açtığını belirleyen istatistik post-hoc olarak bilinmektedir (Roscoe, 1975). Yapılan çalışmalarda, grup içi farklılığa hangi grubun neden olduğunu belirlemek için kullanılan bir çok farklı post-hoc tekniği kullanılabilir. Gruplararası varyansların eşit olup olmaması özelliği hangi post-hoc testinin uygulanacağı konusunda önem taşımaktadır (Ramig, 1983). Özetle, post-hoc istatistikleri, gruplar arası varyansın aynı olması veya olmaması durumunda kullanılanlar olmak üzere iki farklı şekilde ele alınmaktadır (Nelson, 1983).

Gruplararası varyansların eşit olduğu durumlarda, LSD (Least Significant Difference), Bonferroni, Tukey, Hochberg's GT2, Sidak, Gabriel ve Scheffe testleri tercih edilirken, gruplararası varyansın eşit olmadığı durumlarda ise Tamhane's T2, Tamhane's T3, Dunnet's Cü, Games-Howell, ve Dunnet's T3 yöntemleri tercih edilmekte ve sadece "çoklu aralık testi" genel kullanım alanı bulmaktadır.

## 2.2. İki Yönlü Varyans Analizi

Tek yönlü ANOVA modelinde olduğu gibi iki yönlü ANOVA modelinde de faktör sayısı birdir. Ancak bir-yönlü ANOVA modelinden farklı olarak, deney birimleri arasında sistematik farklılıklar vardır. Bu sistematik farklılıkların etkisi kendi içlerinde homojen, kendi aralarında ise heterojen bloklar kullanılarak giderilmeye çalışılır. Bloklama ile deneysel hatanın azaltılması sağlanarak, deneyin hassaslığı artırılmış olur (Montgomery, 1991). Birden fazla iki-yönlü ANOVA modeli vardır. İki yönlü ANOVA modeli için matematiksel model,

$$y_{ijk} = \mu + \tau_i + \beta_j + \tau\beta_{ij} + \varepsilon_{ijk}, \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, b, \quad k = 1, 2, \dots, n \quad (2)$$

Burada,  $y_{ijk}$ : inci deneme jnci blok k.ıncı gözlem değerini,  $\mu$ : genel ortalamayı,  $\tau_i$ -inci deneme etkisini,  $\beta_j$ , jinci blok etkisini,  $\tau\beta_{ij}$  deneme ve blok etkileşim etkisini ve  $\varepsilon_{ijk}$  rasgele hata terimlerini göstermektedir. Benzer şekilde a, deneme sayısını, b, blok sayısını ve n ise her bir deneme ve blok etkileşimdeki gözlem sayısını göstermektedir. Böyle toplam gözlem sayısı  $N=abn$  şeklinde ifade edilmektedir.

İki yönlü ANOVA'da, denemeler arası farkın önemli olup olmadığı, bloklar arası farkın önemli olup olmadığı ve etkileim etkisinin önemli olup olmadığı araştırılmaktadır. Varyans analizi tablosu:

Tablo 2. İki yönlü ANOVA Tablosu

Değ. Kayn.	Serb. Derc.	Kar. Top.	Kar. Ortl.	F
Deneme	a-1	$KT_{Deneme}$	$KO_{Deneme}$	$KO_{Deneme} / KO_{Hata}$
Blok	b-1	$KT_{Blok}$	$KO_{Blok}$	$KO_{Blok} / KO_{Hata}$
Etkileşim	(a-1)(b-1)	$KT_{Etkileşim}$	$KO_{Etkileşim}$	$KO_{Etkileşim} / KO_{Hata}$
Hata	N-ab	$KT_{Hata}$	$KO_{Hata}$	
Toplam	N-1	$KT_{Genel}$		

Tek yönlü ANOVA’da olduğu gibi burada genel KT, hata KT, blok KT, deneme kareler toplamı ve etkileşim KT’na eşittir. Serbestlik dereceleri ise Tablo 2’deki gibi bulunmaktadır. Burada, deneme etkisi için bulunan F istatistiği tablo değerinden büyükse denemeler arası anlamlı bir farklılık vardır denilmektedir, benzer şekilde blok etkisini ölçmek için kullanılan F test istatistiği tablo değerinden büyükse bloklar arası anlamlı bir farklılık vardır denilmektedir. Ayrıca, etkileşim etkisi için bulunan F istatistiği tablo değerinden büyükse etkileşim etkisi anlamlıdır denilmektedir.

### 2.3. Faktöriyel ANOVA

Çoğu deneyde, iki veya daha fazla faktörün etkilerinin incelendiği görülmüş olup kayda değer sonuçlar elde edilmiştir. Genel olarak, faktöriyel tasarımlar bu tür deneyler için en verimli yöntemlerdir. Bir faktöriyel tasarım ile faktörlerin seviyelerinin tüm olası kombinasyonları araştırılmaktadır. Bu kombinasyonlardan yola çıkarak cevaba yönelik en iyi kombinasyonları tespit etmek mümkündür (Myers vd, 2016). Literatürde birçok faktöriyel deneme türü gözlemlenmektedir.

### 3. Uygulama

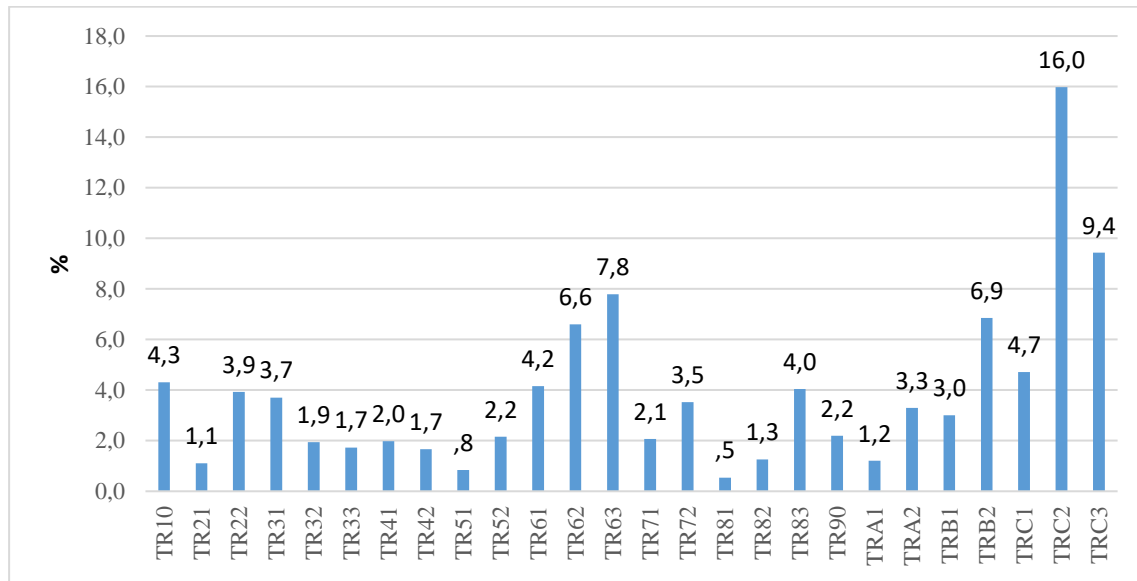
Ülke içinde yaşayan insanların refah seviyelerini belirleyebilmek ve sosyal devlet sisteminin sağlıklı bir şekilde değerlendirmesini yapabilmek için o ülkede yaşayan insanların gelir dağılımı verilerine ve göstergelerine ihtiyaç duyulmaktadır. O ülkede yaşayan insanlar arasındaki gelir adaletsizliği ve yoksulluk artık günümüz dünyasında ekonomik bir sorun olmaktan çıkmış, siyasi ve sosyal bir problem olarak ele alınmaya başlanmıştır. Bu sebeple, “gelir yoksulluğu”, “sosyal imkân yoksulluğu”, “sosyal dışlanma” gibi yeni kavramlar için verilerin oluşturulması istenmiştir. GYKA, TÜİK tarafından 2006 yılından itibaren Avrupa Birliğine uyum yasaları çerçevesinde yapılmış olduğu bir panel ankettir. Anketin amacı, gelir dağılımının yanı sıra, yaşam koşulları, sosyal dışlanma ve görelî gelir yoksulluğunun saptanmasıdır. Bu sebeple, GYKA, Türkiye’deki gelir dağılımına, yoksulluğun saptanmasına ve şekline, yaşam koşullarının ve sosyal dışlanmaya ışık tutan önemli bir bilgi kaynağı niteliği taşımaktadır.

Tablo 3. Hanehalkı Kullanılabilir Geliri Normallik Testi

	Kolmogorov-Smirnov Testi		
	Test İstatistiği	Serbestlik Derecesi	Sig.
Hanehalkı Kullanılabilir Geliri	,048	1413	,000

Bu makalede ilk olarak Kolmogorov-Smirnov testi ile verilerin normal dağılıma uygunluğu incelenmiştir. Sonuçlar Tablo 3 de verilmiştir. Kullanılan teste ilişkin hipotezler:  $H_0$ : Hanehalkı Kullanılabilir Geliri normal dağılmaktadır.  $H_1$ : Hanehalkı Kullanılabilir Geliri normal dağılıma uygun değildir şeklinde kurulur. Anlamlılık değeri  $0,000 < 0,05$

olduğundan  $H_0$  reddedilmiştir. Ölçekte yer alan bağımlı değişken olan hanenin yıllık kullanılabilir geliri normal dağılıma sahip olmadığı tespit edildiğinden ötürü Box-Cox transformation (dönüşümü) metodu kullanılarak veriler normal dağılıma sağlanmıştır. Bu tür sorunlar karşısında veri setine logaritmik, ters, karekök, arcsin vb. dönüşümler ile normal dağılıma benzer seriler (hatta normal dağılan) elde edilir. Box-Cox dönüşümü George E.P. Box ve David Cox (1964) tarafından ilk kez açıklanan  $\lambda$  parametrelili üstel bir dönüşümdür, normal dağılmayan nicel verilere, optimum  $\lambda$  değeri belirlenip Box-Cox transformasyonu uygulandığında verilerin normal dağılıma benzediğinin aktarılması hedeflenmiştir. İki'den fazla grubun olduğu karşılaştırmalarda tek yönlü ANOVA kullanılmıştır. Eğer gruplar arası farklılıklar istatistiksel olarak anlamlı bulunmuş ise hangi grupların birbirinden farklı olduğunun tespit edilmesi için varyanslar eşit tahmin edilmiş ise çoklu karşılaştırma (multiple comparisons) için Tukey testi, varyanslar eşit tahmin edilmemiş ise Tamhane's T2 testi kullanılarak sonuçlar yorumlanmıştır. Çalışmada analizler için elde edilen anlamlılık düzeyleri 0,05 hata düzeylerinde değerlendirilmiştir. Dolayısıyla çalışmadaki analiz sonuçları %95 güven seviyesinden yorumlanmıştır. Çalışmada yer alan incelemeler SPSS 22.0 paket programı ile yürütülmüştür.



Şekil 1: Yerleşim yerine göre yoksul hanelerin dağılımı (%)

Şekil 1'de görüldüğü üzere, en fazla yoksul hane %16 ile TRC2- Şanlıurfa, Diyarbakır bölgesinde en az yoksul hane ise %0,5 ile TR81- Zonguldak, Karabük, Bartın bölgesindedir. Çalışma için kullanılan verilerin betimsel istatistiklerinden elde edilen bazı önemli sonuçlar ise şunlardır; yoksul hanelerin %70.8'i müstakil konutta, %15.1'inin Apartmanda (10 daireden az), %9.4'ünün Apartman (10 daireden fazla) ve %4.6'sının ise ikiz ya da sıralı evde oturdukları görülmektedir. Yoksulların %48.7'sinin oturduğu konutun yaşı 20-49 yıllık iken, %21.6'sı 10-19 yıllık binalarda oturmaktadır. Yoksulların %15.5'i 50 yıl ve daha fazla yıllık binalarda otururken, %14.2'si 0 – 9 yıllık yeni binalarda oturmaktadır. Yoksul hanelerin %74.7'sinin yaşadığı çevrede; hava kirliliği, çevre kirliliği, toz, hoş olmayan koku, pis su veya diğer çevresel sorunlar görülürken, %25.3'ünde ise böyle bir problem yaşanmamaktadır. Yoksulların %65.5'inin hanelerinin genellikle gerekli harcamalarını zor yapabildiklerini, %33.7'sinin ne zor ne kolay ve %0.8'inin ise gerekli harcamaları kolay yapabildiklerini beyan ettikleri görülmektedir. Yoksul hanelerin %60.3'ü dört ve daha fazla kişiden oluşurken, %18.6'sı üç ve dört kişilik hanelerden, %12.3'ü tek kişilik hanelerden ve %8.8'i ise iki kişilik hanelerden oluşmaktadır.

Aşağıda verilen tablolarda kullanılan bazı değişkenlere ilişkin ANOVA sonuçları verilmiştir.

Tablo 4. Düzey 2 Bazında Bölgelerin Yoksulluğa Etkisine İlişkin ANOVA Sonucu

Varyans Homojenlik Testi					
Levene İstatistiği	Serbestlik derecesi1	Serbestlik derecesi2	P		
2.045	25	1386	0.002		
Levene testi grup varyanslarının [ $p < 0.05$ ] eşit olmadığını göstermektedir.					
ANOVA					
Değişimin kaynağı	Kareler toplamı	Serbestlik derecesi	Kareler ortalaması	F	p
Gruplararası	5196515979	25	207860639.2	8.529	0
Grup içi (hata)	33777260586	1386	24370317.88		
Genel	38973776565	1411			

H<sub>0</sub>: Hanenin kullanılabilir gelirin bölgelerin etkisi önemsizdir.

H<sub>1</sub>: Hanenin kullanılabilir gelirin bölgelerin etkisi önemlidir.

Tablo 5. Düzey 2 Bazında Bölgelere Göre Tamhane Testi

Bölge (Düzey 2)	Tamhane Testi Anlamlı Fark
TR10- İstanbul	-
TR21- Tekirdağ, Edirne, Kırklareli	TRA2-TR21, TRB2-TR21, TRC1-TR21, TRC3-TR21
TR22- Balıkesir, Çanakkale	TR62-TR22, TRA2-TR22, TRB1-TR22, TRB2-TR22, TRC1-TR22, TRC2-TR22, TRC3-TR22
TR31- İzmir	TRA2-TR31, TRB2-TR31, TRC1-TR31, TRC2-TR31, TRC3-TR31
TR32- Aydın, Denizli, Muğla	TRA2-TR32, TRB2-TR32, TRC1-TR32, TRC2-TR32, TRC3-TR32
TR33- Manisa, Afyon, Kütahya, Uşak	-
TR41- Bursa, Eskişehir, Bilecik	-
TR42- Kocaeli, Sakarya, Düzce, Bolu, Yalova	-
TR51- Ankara	-
TR52- Konya, Karaman	TRA2-TR52, TRB2-TR52, TRC1-TR52, TRC2-TR52, TRC3-TR52
TR61- Antalya, Isparta, Burdur	TRA2-TR61, TRB2-TR61, TRC1-TR61, TRC2-TR61, TRC3-TR61
TR62- Adana, Mersin	TR62 – TR22
TR63- Hatay, Kahramanmaraş, Osmaniye	TRA2-TR63, TRB2-TR63, TRC3-TR63
TR71- Kırıkkale, Aksaray, Niğde, Nevşehir, Kırşehir	-
TR72- Kayseri, Sivas, Yozgat	TRA2-TR72, TRB2-TR72, TRC1-TR72, TRC2-TR72, TRC3-TR72
TR81- Zonguldak, Karabük, Bartın	-
TR82- Kastamonu, Çankırı, Sinop	TRA2-TR82, TRB2-TR82, TRC1-TR82, TRC3-TR82
TR83- Samsun, Tokat, Çorum, Amasya	TRB2-TR83, TRC3-TR83
TR90- Trabzon, Ordu, Giresun, Rize, Artvin, Gümüşhane	-
TRA1- Erzurum, Erzincan, Bayburt	-
TRA2- Ağrı, Kars, Iğdır, Ardahan	TRA2-TR21, TRA2-TR22, TRA2-TR31, TRA2-TR32, TRA2-TR52, TRA2-TR61, TRA2-TR63, TRA2-TR72, TRA2-TR82
TRB1- Malatya, Elazığ, Bingöl, Tunceli	TRB1-TR22
TRB2- Van, Muş, Bitlis, Hakkari	TRB2-TR21, TRB2-TR22, TRB2-TR31, TRB2-TR32, TRB2-TR52, TRB2-TR61, TRB2-TR63, TRB2-TR72, TRB2-TR82, TRB2-TR83
TRC1- Gaziantep, Adıyaman, Kilis	TRC1-TR21, TRC1-TR22, TRC1-TR31, TRC1-TR32, TRC1-TR52, TRC1-TR61, TRC1-TR72, TRC1-TR82
TRC2- Şanlıurfa, Diyarbakır	TRC2-TR22, TRC2-TR31, TRC2-TR32, TRC2-TR52, TRC2-TR61, TRC2-TR72
TRC3- Mardin, Batman, Şırnak, Siirt	TRC3-TR21, TRC3-TR22, TRC3-TR31, TRC3-TR32, TRC3-TR52, TRC3-TR61, TRC3-TR63, TRC3-TR72, TRC3-TR82, TRC3-TR83

Tablo 6. Hanehalkı Büyüklüğünün Hanelerin Kullanılabilir Gelirine Etkisine İlişkin Tek Yönlü ANOVA

<b>Varyans Homojenlik Testi</b>					
<b>Levene İstatistiği</b>	<b>Serbestlik derecesi 1</b>	<b>Serbestlik derecesi 2</b>	<b>p</b>		
38.36	3	1408	0.00		
Levene testi grup varyanslarının [ $p < 0.05$ ] eşit olmadığını göstermektedir.					
<b>ANOVA</b>					
<b>Değişimin kaynağı</b>	<b>Kareler toplamı</b>	<b>Serbestlik derecesi</b>	<b>Kareler ortalaması</b>	<b>F</b>	<b>p</b>
Gruplararası	18734223288	3	6244741096	434.43	0.00
Grup içi (hata)	20239553276	1408	14374682		
Genel	38973776565	1411			

H<sub>0</sub>: Hanenin kullanılabilir gelirine hanehalkı büyüklüğünün etkisi önemsizdir.

H<sub>1</sub>: Hanenin kullanılabilir gelirine hanehalkı büyüklüğünün etkisi önemlidir.

Tablo 7. Hanelerin Kullanılabilir Gelirleri Arasındaki Farkın Kaynağını Belirlemek Üzere Hanehalkı Büyüklüğüne Göre Tamhane Testi

<b>(I) HHB</b>	<b>(J) HHB</b>	<b>Ortalama Farkı (I-J)</b>	<b>Standart Hata</b>	<b>p</b>
1	2	-2865.37626*	243.93	0.00
	3-4	-5439.24187*	254.53	0.00
	5+	-10388.47604*	232.25	0.00
2	1	2865.37626*	243.93	0.00
	3-4	-2573.86561*	255.24	0.00
	5+	-7523.09979*	233.03	0.00
3-4	1	5439.24187*	254.53	0.00
	2	2573.86561*	255.24	0.00
	5+	-4949.23418*	244.11	0.00
5+	1	10388.47604*	232.25	0.00
	2	7523.09979*	233.03	0.00
	3-4	4949.23418*	244.11	0.00

Tablo 8. Gerekli Harcamaları Yapabilme Durumunun Hanelerin Kullanılabilir Gelirine Etkisine İlişkin Tek Yönlü ANOVA

Varyans Homojenlik Testi			
Levene İstatistiği	Serbestlik derecesi 1	Serbestlik derecesi 2	p
38.36	3	1408	0.00

Levene testi grup varyanslarının [ $p < 0.05$ ] eşit olmadığını göstermektedir.

ANOVA					
Değişimin kaynağı	Kareler toplamı	Serbestlik derecesi	Kareler ortalaması	F	p
Gruplararası	18734223288	3	6244741096	434.43	0.00
Grup içi (hata)	20239553276	1408	14374682		
Genel	38973776565	1411			

H<sub>0</sub>: Hanenin kullanılabilir gelirine geçinme durumunun etkisi önemsizdir.

H<sub>1</sub>: Hanenin kullanılabilir gelirine geçinme durumunun etkisi önemlidir.

Tablo 9. Hanelerin Kullanılabilir Gelirleri Arasındaki Farkın Kaynağını Belirlemek Üzere Geçinme Durumuna Göre Tamhane Testi

(I) GECINME_DURUM	(J) GECINME_DURUM	Ortalama Farkı (I-J)	Standart Hata	p
Zor	Ne zor ne kolay	-1234.76340*	295.86	0.00
	Kolay	4509.92	2965.89	0.43
Ne zor ne kolay	Zor	1234.76340*	295.86	0.00
	Kolay	5744.68	2971.04	0.26
Kolay	Zor	-4509.92	2965.89	0.43
	Ne zor ne kolay	-5744.68	2971.04	0.26

Yoksulluğa etki eden faktörler için yapılan ANOVA testi sonuçlarının ise sonuçları aşağıda verilmiştir.

- 1) TR10 (İstanbul), TR33 (Manisa, Afyon, Kütahya, Uşak), TR41(Bursa, Eskişehir, Bilecik), TR42 (Kocaeli, Sakarya, Düzce, Bolu, Yalova), TR51 (Ankara), TR71 (Kırıkkale, Aksaray, Niğde, Nevşehir, Kırşehir), TR81 (Zonguldak, Karabük, Bartın), TR90 (Trabzon, Ordu, Giresun, Rize, Artvin, Gümüşhane) ve TRA1(Erzurum, Erzincan, Bayburt) bölgelerinin diğer bölgelerle arasında anlamlı bir fark yoktur.
- 2) Bina yaşı 0-9 yıl olanlar, diğer üç bina yaşı ile kıyaslandığında, bu binanın (0-9 yıl) sahip olunan grup ortalama gelirleri ile bina yaşı 20-49 yıl ve bina yaşı 50 yıl ve üstü yaşa sahip binaların grup ortalama gelirleri arasında anlamlı fark olduğu belirlenmiştir.
- 3) Yoksul hanelerin ölçek alışveriş değişkenine göre gelirlerine etkisi Tukey Testi sonucunda; yaşadığı çevrede alışveriş merkezine kolay erişebilen hanelerin çok zor olan hanelere göre gelirleri daha yüksek çıkmıştır. Alışveriş merkezine zor erişebilen hanelerin çok zor erişebilen hanelere göre gelirleri daha olumlu görülmüştür.



- 4) Yoksul hanelerin ölçek toplu taşıma hizmeti değişkenine göre gelirlerine etkisi Tukey Testi sonucunda; yaşadığı çevrede toplu taşıma hizmetine kolay ve zor erişebilen hanelerin çok zor erişebilen hanelere göre gelirlerinin etkisi daha yüksek çıkmıştır.
- 5) Yoksul hanelerin ölçek sağlık merkezi değişkenine göre gelirlerine etkisi Tukey Testi sonucunda; yaşadığı çevrede sağlık merkezine kolay erişebilen hanelerin çok zor erişebilen hanelere göre gelirlerinin etkisi daha fazla olmuştur.
- 6) Yoksul hanelerin ölçek banka değişkenine göre gelirlerine etkisi Tukey Testi sonucunda; yaşadığı çevrede bankacılık hizmetine kolay erişebilen hanelerin çok zor erişebilen hanelere göre gelirlerinin etkisi daha fazla olmuştur.
- 7) Hanehalkı büyüklüğü değişkenine göre gelirlerine etkisi Tamhane testi sonucunda beş ve daha fazla kişilik hanelerden oluşan hanehalklarının ortalama gelirleri tek kişilik, iki kişilik ve üç veya dört kişilik hanehalklarına göre daha fazladır. Üç veya dört kişiden oluşan hanehalklarının ortalama gelirleri tek kişilik hanelerden ve iki kişilik hanelerden daha yüksektir. İki kişilik hanelerin gelirleri de tek kişilik hanelerden daha fazladır. Sonuç olarak tek kişilik haneler daha az ortalama gelire sahip olduğundan bu haneler daha çok yoksullardır.
- 8) Bina yaşı \* sıcak su etkileşim etkisinin yoksul hanelerin gelirini etkilediği söylenebilir. Yani 0-9 yıllık binalarda oturanlarda sıcak su sistemi mevcut olan hanelerin gelir düzeyleri en fazla iken, 50 yıl ve üstü binalarda oturanlarda ise sıcak su sistemi olmayanların gelir düzeylerinin en düşük olduğu gözlenmektedir.
- 9) Bina yaşı \* sahip olunan bulaşık makinesi etkileşim etkisinin yoksul hanelerin gelirini etkilediği söylenebilir. Genel olarak oturlan konutta bina yaşı arttıkça, bulaşık makinesi olmayan hanelerin gelir düzeyleri azalma eğiliminde olduğu görülmektedir.
- 10) Yoksul hanelerin hanehalkı kullanılabilir gelirlerine taksit borç ve alan etkileşim etkisi önemli olduğu görülmektedir. Yani hanedeki herhangi bir fert tarafından taksit borç ödemesi yapan hanelerde, oturduğu konutun kullanım alanı yeterli olmayanların gelir düzeyleri daha fazla iken, taksit borç ödemesi olmayan hanelerde ise oturduğu konutun kullanım alanı yeterli olan hanelerin gelir düzeyleri en düşük olduğu görülmektedir.

#### 4. Tartışma ve Sonuç

Türkiye’de yoksulluğu etkileyen faktörleri belirlemek ve bunların söz konusu yoksul hanelerin gelirleri üzerindeki etkilerinin büyüklüğünü belirlemeyi amaçlayan çalışmada, TÜİK tarafından yürütülen 2016 yılında uyguladığı GYKA verileri kullanılmıştır.

Yoksulluk incelenirken bağımsız değişkenler olarak bölgesel düzey, bina yaşı, alışveriş merkezine, toplu taşıma hizmetine, sağlık merkezine, zorunlu eğitim hizmetine ve bankacılık hizmetine erişebilme durumu, yaşadığı çevrenin gelişmişlik düzeyi, konutun mülkiyet durumu, geçinme durumu, hanehalkı büyüklüğü, bina yaşı ve sıcak su sistemi kolaylığının birlikte etkisi, bulaşık makinesine sahiplik durumu ile bina yaşının etkisi, geçinme durumu ve konutun ısınma durumunun etkisi, beklenmeyen harcamalar ile ısınma masrafını karşılama durumunun etkisi ele alınmıştır. Bağımlı değişken olarak ise hanehalkı kullanılabilir geliri temel alınmıştır. Bunun nedeni hanehalkı toplam yıllık geliri hanehalkının eşdeğer hanehalkı değerine oranlanarak ilgili eşdeğer fert başı gelirin hesaplanmasıdır. Eşdeğer fert başına düşen gelir elde edildikten sonra bu değer medyan değerinin %40’ı yoksulluk sınırını vermektedir. Eşdeğer fert başına gelirden çalışmada ele alınan yoksulluk sınırı çeşidi olan göreceli yoksulluk sınırının oluşturulmasında temeldir.

Doğu bölgelerinde yaşayan yoksulların gelirlerinin ortalaması batı bölgelerindeki yoksul hanelere göre daha yüksek görülmektedir. Bunun sebebi ise doğu bölgelerine sosyal transferlerin önemli seviyede aktarılmasından dolayı olabilir. Ayrıca TR10 (İstanbul), TR33 (Manisa, Afyon, Kütahya, Uşak), TR41 (Bursa, Eskişehir, Bilecik), TR42 (Kocaeli, Sakarya, Düzce, Bolu, Yalova), TR51 (Ankara), TR71 (Kırıkkale, Aksaray, Niğde, Nevşehir, Kırşehir), TR81 (Zonguldak, Karabük, Bartın), TR90 (Trabzon, Ordu, Giresun, Rize, Artvin, Gümüşhane) ve TRA1 (Erzurum, Erzincan, Bayburt) da yaşayan yoksul haneler Türkiye’deki diğer yoksul hanelerle benzer profildedir. Bu bölgelerde yapılacak analizler Türkiye’deki tüm bölgelerin ortalamasını kapsamaktadır. Genel olarak bölgesel düzeyde yoksulluğun farklı çıktığı görülmüştür.

Hanehalkı büyüklüğü de yoksulluğu etkileyen bir diğer önemli özelliktir. En düşük ortalama hanehalkı geliri tek kişilik hanelerde görülmektedir. Tek kişilik hanelerde, kişi işgücü piyasasındaki faaliyeti ile evdeki faaliyetlerini eş zamanlı yürütmek ihtiyacı duyduğundan bu hanenin geliri, iki yetişkinli hanenin gelirine oranla daha azdır. Üstelik tek kişinin kadın olduğu senaryoda, kadınların erkeklere göre daha az gelir getirici işlerde çalışması genellemesine göre, hanenin geliri daha da azalacaktır. Sonuç olarak, tek yetişkinli hanelerde yoksulluk oranı daha yüksek olmaktadır (Schiller,

2004). Kullanılan veri setinde tek kişilik hanehalklarının %56,4'ü kadınlardan oluşmaktadır. Dolayısıyla Schiller'in bu çalışmasını desteklemektedir.

Sonuç olarak, bölgesel düzeyde yoksulluğun farklı çıktığı görülmüştür. Kurulan modele göre hanehalkı yoksulluğunu etkileyen faktörlerin; bina yaşı, hanenin bulunduğu yer itibariyle alışveriş merkezine, toplu taşıma hizmetine, sağlık merkezine, zorunlu eğitim hizmetine erişebilme durumu ve bankaya ulaşma durumu, mülkiyet şekli, hanehalkı büyüklüğü, geçinme durumu, hanenin bulaşık makinesine sahip olup olmaması ve bina yaşı etkileşim etkisi, bina yaşı ve sıcak su sisteminin olup olmamasının etkileşim etkisi, taksit veya borç ödemesi ile kullanım alanı yeterliliğinin etkisi ve beklenmeyen mecburi bir harcama ile ısınma masrafını karşılama durumu olduğu görülmüştür.

## Kaynakça

- Arabacı, R. (2016). Görelî Yoksulluk Yaklaşımı Ve Türkiye İçin Uygunluğu. *Ataturk University Journal of Economics & Administrative Sciences*, 30(5).
- Aksan, G. (2012). Yoksulluk ve Yoksulluk Kültürünün Toplumsal Görünümleri. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (27), 9.
- Aktan, C., C., ve Vural, İ., Y. (2002). "Yoksulluk: Terminoloji, Temel Kavramlar ve Ölçüm Yöntemleri", Yoksullukla Mücadele Stratejileri, Aktan, Coşkun Can (ed), Hak-İş Konfederasyonu Yayınları, Ankara.
- Arbia, G., & Pace, M. L. (2018). Testing inequality of opportunities in Italy using the ANOVA framework. *Empirical Economics*, 55(1), 297-322.
- Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211-243.
- Castellano, R., & Punzo, G. (2012). The impact of territorial factors on the total non-response error in the European Union-Survey on Income and Living Condition (EU-SILC). *Statistica*, 72(2), 211-233.
- Coşkun, S., ve Tireli, M. (2008). "Avrupa Birliğinde Yoksullukla Mücadele Stratejileri ve Türkiye".
- Evcim, N., Güneş, S., & Karaalp-Orhan, H. S. (2020). Türkiye'de Hanehalkı Görelî Yoksulluğunu Etkileyen Faktörler: Lojistik Regresyon Analizi. *Sosyoekonomi*, 28(43), 11-32.
- Foster, J.E. ve Shorrocks A.F. (1991). Subgroup-Consistent Poverty Indices. *Econometrica*, 59(3).
- Karcı, Z., ve Bayram, N. A. (2018). Maddî Yoksunluğu Etkileyen Değişkenlerin Lojistik Regresyon Analizi ile Belirlenmesi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 23(3), 1039-1048.
- Montgomery, D. C., (1991). "Design and Analysis of Experiments", (3rd Ed.), *John Wiley Sons*, 649 s.
- Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M., (2016), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley & Sons., Hoboken, *New Jersey*, 856 s.
- Nelson, P.R., (1983), A comparison of sample sizes for the analysis of means and the analysis of variance. *Journal of Quality Technology*, 15, 33-39.
- Ramig, P.R., (1983), Applications of the analysis of means. *Journal of Quality Technology*, 15, 19-25.
- Rangaswamy, R., (1995), A textbook of agricultural statistics. *New Age International*, 552, *New Delhi*.
- Ravallion, M., & Datt, G., (1994), How Important to India's Poor is the Urban-Rural Composition of Growth? *Policy Research Working Paper 1399*, 1-25.
- Roscoe, J. T., (1975), *Fundamental research statistics for the behavioral sciences*. New York: *Holt, Rinehart and Winston, Inc.*
- Schiller, B. R., (2004), *The Economics of Poverty and Discrimination*, *Pearson Prentice Hall*, New Jersey.
- Sen, A. (1976). Poverty: An Ordinal Approach to Measurement. *Econometrica*, 44(2), 219-231.
- Şenses, F. (2002); *Küreselleşmenin Öteki Yüzü: Yoksulluk*, 2. Baskı, İletişim Yayınları, İstanbul.
- Taş, H. Y., ve Özcan, S. (2012). Türkiye'de ve Dünya'da yoksulluk üzerine bir araştırma. In *International Conference on Eurasian Economies* (pp. 423-430).
- TÜİK., (2008), *Tüketim Harcamaları, Yoksulluk ve Gelir Dağılımı*. Sorularla Resmi İstatistikler Dizisi 6. Ankara, Türkiye İstatistik Kurumu.
- TÜİK., (2016), *Gelir ve Yaşam Koşulları İstatistikleri El Kitabı*. Ankara, *Türkiye İstatistik Kurumu*.
- TÜİK., (2016), *Gelir ve Yaşam Koşulları Araştırması*.
- Ünver, Ş., & Alkan, Ö. (2020). Türkiye'de Bireylerin Maddî Yoksunluk Durumlarını Etkileyen Faktörlerin Modellenmesi. *Business & Management Studies: An International Journal*, 8(2), 1334-1370.
- Yaşar, S., & Taşar, M. O. (2019). Kavramsal Olarak Yoksulluk Ve Türkiye'de Yoksullukla Mücadele Politikalarının Etkileri. *Sosyal Ekonomik Araştırmalar Dergisi*, 19(38), 118-144.

**Atf / Citation:** ŐENTÜRK O., ÇELİK N. Türkiye'de Hanehalkı Yoksulluğunu Etkileyen Faktörlerin Belirlenmesi. İstatistik Araştırma Dergisi, 2021, 11 (2), 56-70.

Yüce, M. ve Özbek,M.F. (2006). “Orta Asya Türk Cumhuriyetlerinde Sosyo-Ekonomik Boyutlarıyla Yoksulluk Olgusu Üzerine Bir Değerlendirme” , *İktisat ve Girişimcilik Üniversitesi, Türk Dünyası Kırgız, Türk Sosyal Bilimler Enstitüsü Celalabat, Kırgızistan, Uluslararası Hakemli Sosyal Bilimler E-Dergisi*, (100).



## Türkiye’ de Toplumsal Cinsiyetin Çok Boyutlu Ölçekleme ile İncelenmesi

Asila KOÇAK  
Türkiye İstatistik Kurumu / İstatistikçi / Dr.  
asila.kocak@tuik.gov.tr  
Orcid No: 0000-0001-5048-3424

### Özet

Kadın ile ilgili istatistikler, toplumsal cinsiyet eşitliğinin sağlanması, kadının toplum içindeki yerinin belirlenmesi ve sürdürülebilir kalkınmada kadınların katkısının güçlendirilmesi açısından oldukça önemlidir.

Bu çalışmada nüfus, doğurganlık, sağlık, evlenme, aile yaşamı, eğitim, siyasal yaşam, iş ve kazanç konuları kapsamında toplumsal cinsiyet bakış açısıyla seçilmiş 17 değişken yer almaktadır. İstatistiki Bölge Birimleri Sınıflaması İBBS 3. Düzeyinde (81 il) oluşturulan veri seti, 2019 yılına ait güncel verileri içermektedir. Bu çalışma Çok Boyutlu Ölçekleme (ÇBÖ) yöntemi kullanılarak illerin birbirine benzerlik ve farklılıklarından hareketle toplumsal cinsiyet ve toplumsal yapı ile ilgili bir karşılaştırma elde etmeyi amaçlamaktadır. Aynı zamanda, veri setinde yer alan değişkenler birbirlerine yakınlık, uzaklık ve ayrıştırıcılık yönünden incelenmiştir. Veriler SPSS 15.0 paket programı ile SPSS ALSCAL algoritmasında, gözlemler arası uzaklık matrisi elde edilerek analiz edilmiş ve yorumlanmıştır. Bu çalışmanın sonuçlarının, toplumsal cinsiyet eşitliğine yönelik bölgesel düzeyde farkındalığın artmasına ve politikaların gelişmesine katkı sunması amaçlanmaktadır.

*Anahtar Sözcükler: Çok Boyutlu Ölçekleme, Kadın, Metrik Çok Boyutlu Ölçekleme, Toplumsal Cinsiyet.*

**Sorumlu Yazar / Corresponding Author:** 1-Asila KOÇAK, Türkiye İstatistik Kurumu, Yöntem Araştırmaları Daire Başkanlığı, Kalite Denetim Grup Başkanlığı.

**Atıf / Citation:** KOÇAK A. Türkiye’de Toplumsal Cinsiyetin Çok Boyutlu Ölçekleme ile İncelenmesi. İstatistik Araştırma Dergisi, 2021, 11 (2), 71-82.

## Examining Gender in Turkey with Multidimensional Scaling

### Abstract

Statistics on women are very important in terms of ensuring gender equality, determining the place of women in the society and strengthening the contribution of women in sustainable development.

In this study, there are 17 variables selected from a gender perspective, including population, fertility, health, marriage, family life, education, political life, work and earnings. The data set created in the NUTS 3rd Level (81 provinces) of the Classification of Statistical Regional Units includes current data for 2019. This study aims to obtain a comparison of gender and social structure based on the similarities and differences of the provinces by using Multidimensional Scaling (MDS) method. In addition, the variables in the data set were also examined in terms of proximity, distance and discrimination to each other. The data were analyzed and interpreted in the SPSS 15.0 package program, and the distance matrix between observations was obtained by using SPSS ALSCAL algorithm. It is aimed that the results of this study may contribute to the increase of awareness on gender equality at the regional level and to the development of policies in this regard.

*Keywords: Gender, Metric Multidimensional Scaling, Multidimensional Scaling. Women.*

### 1. Giriş

Toplumsal cinsiyet biyolojik cinsiyetten farklı olarak toplumsal ve kültürel olarak belirlenen ve dolayısıyla içeriği yere ve zamana göre değişebilen “cinsiyet konumu” ya da “cins kimliği” dir. Yalnızca cinsiyet farklılığını belirtmekle kalmaz aynı zamanda cinsler arasındaki eşitsiz güç ilişkilerini de belirler.(Berktaş vd., 2011)

Cinsiyet biyolojik bir anlamı ifade ederken, toplumsal cinsiyet erkeklik ile kadınlık arasında toplumsal bakımdan eşitsiz bölünmeye işaret etmektedir.

Biyolojik özelliklerin toplumsal, kültürel, politik, ideolojik ve ekonomik yapılarla etkileşimi sonucu farklılaşarak, biyolojik cinsiyetten başka bir hal alışı ve iki farklı yaşam alanının oluşması, eşitsizliklere neden olmaktadır. (Sarp, 2008)

Toplumsal cinsiyette eşitlik kavramı kaynak ve fırsatlara sahipliği ve kullanımındaki eşitliği ifade eder. Bu kaynak ve fırsatlar hem evin içinde, hem ev dışında yaşamın her alanında söz konusu olabilir. Buna göre toplumsal cinsiyet ayrımcılığı sosyal, ekonomik, siyasi ve hukuki yaşamın her alanında kaynaklara ve fırsatlara sahiplikte ve bunların kullanımında eşitsizliklerle karşı karşıya kalmaktır. Toplumsal cinsiyet eşitliği, cinsiyete dayalı ayrımcılık olmadan bireyin, yaşamın her alanında eşit olarak yer alması olarak ifade edilmektedir. Kısacası toplumsal cinsiyet hem kadınlar, hem de erkekler için toplumun bireyin cinsiyeti nedeni ile ona biçtiği rol ve beklentileri şekillendirerek biyolojik farklılıktan daha fazla anlam taşır. Gündelik kullanımda cinsiyet ve toplumsal cinsiyetin birbirleriyle ilgisi devam ediyor olsa da, kavramsal boyutlarını farklı bakış açısıyla ele alan geniş bir literatür bulunmaktadır.

Zaman içinde kadınlarda doğuştan beklenen yaşam süresinin uzaması, anne ve çocuk ölümlerinin azalması, kadının eğitime ve işgücüne katılımı başta üzere birçok alanda önemli değişimlere, sanayi ve bilgi toplumunun olumlu etkilerine rağmen, sadece Türkiye’ de değil tüm dünyada cinsiyete dayalı kalıp yargılar devam etmektedir ve ne yazık ki kadınlar yaşamın her alanında hala yeterince iyi durumda değildir.

Küresel Toplumsal Cinsiyet Uçurumu 2019 Raporu’na (WEF) göre Türkiye 153 ülke arasında 130. sırada yer almaktadır. (WEF, 2019). Birleşmiş Milletler’in 2019 yılı Toplumsal Cinsiyet Eşitsizliği Endeksi’ ne göre Türkiye 162 ülke arasında 68. sırada yer almaktadır. Benzer durum, Türkiye için Türkiye Ekonomi Politikaları Araştırma Vakfı (TEPAV) tarafından 81 il düzeyinde yapılan sonuçlarda da görülmektedir. Buna göre, kadınların birçok yönden iyi durumda olmasının gelişmişlik düzeylerinde doğrudan etkisi olduğu söylenebilmektedir.

Toplumsal cinsiyet eşitsizliğinin bölgesel farklılıklarla ele alınması, kadınların toplumsal alanda karşılaştıkları engellerinin aşılması ve buna yönelik önlemlerin alınması açısından oldukça önemlidir.

## 2. Uygulama

### 2.1 Veri Seti

Tablo 1. 81 İl Toplumsal Cinsiyet Veri Setinde Yer Alan Değişkenler (Gökçin ve Koçak, 2021)

Değişken	Kapsamı	Hesaplama	Kaynak
<b>Kadın intihar oranı %</b>	Sağlık	Kadın intihar sayısı (İl) / toplam intihar sayısı(TR) *100	TÜİK
<b>Kadın milletvekili oranı %</b>	Siyasi Güç	Kadın milletvekili sayısı (İl) / toplam milletvekili sayısı (İl) *100	TBMM
<b>15+okuryazar olmayan kadın oranı %</b>	Eğitim	15+okuryazar olmayan kadın (İl) / eğitim alan genel toplam kadın sayısı(İl) *100	TÜİK
<b>Yüksekokul mezunu kadın oranı %</b>	Eğitim	Yüksekokul mezunu kadın (İl) / eğitim alan genel toplam kadın sayısı (İl) *100	TÜİK
<b>Doğurganlık hızı</b>	Doğurganlık	-	TÜİK
<b>Evlilik süresi 1 yıldan az, doğum oranı %</b>	Doğurganlık	Evlilik süresi 1 yıldan az doğum sayısı (İl) /toplam doğum sayısı (İl) *100	TÜİK
<b>Evlilik süresi 1 yıl</b>	Doğurganlık	Evlilik süresi 1 yıl doğum sayısı (İl) /toplam doğum sayısı (İl) *100	TÜİK
<b>Evlilik süresi 20 yıl, doğum oranı %</b>	Doğurganlık	Evlilik süresi 20 + doğum sayısı (İl) /toplam doğum sayısı (İl) *100	TÜİK
<b>Hiç evlenmeyen anne oranı%</b>	Evlenme	Medeni durumu "hiç evlenmeyen" anne (İl) / toplam doğum sayısı (İl) *100	TÜİK
<b>16-19 yaş kadın evlenme oranı %</b>	Evlenme	16-19 yaş evlenme sayısı (il) / toplam evlenme sayısı (İl) *100	TÜİK

<b>Erkek, kadından 11 yaş büyük oranı%</b>	Evlenme, Aile yaşamı	Erkek 11 yaş büyük evlenme sayısı (İl) / toplam evlenme sayısı (İl)	TÜİK
<b>Tek Çekirdek aile oranı %</b>	Aile yaşamı	Tek çekirdekli aile sayısı (İl) / toplam hanehalkı sayısı (İl) *100	TÜİK
<b>En az bir Çekirdek aile oranı %</b>	Aile yaşamı	En az bir çekirdekli aile sayısı (İl) / toplam hanehalkı sayısı (İl) *100	TÜİK
<b>Yabancı nüfuslu kadın oranı %</b>	Nüfus	Yabancı nüfuslu kadın (İl) / toplam kadın nüfusu (İl) *100	TÜİK
<b>Ortalama kadın evlilik yaşı</b>	Evlenme	-	TÜİK
<b>Akraba evliliği oranı %</b>	Evlenme, Aile yaşamı	Akraba evliliği sayısı (İl) / toplam evlenme sayısı (İl) *100	TÜİK
<b>4A/b/c zorunlu sigortalı kadın oranı %</b>	İş Hayatı	4a/b/c zorunlu sigortalı kadın sayısı (İl) / toplam kadın nüfusu (İl) *100	SGK

## 2.2 Yöntem

Çok değişkenli istatistiksel yöntemler, çok değişkenli veri setlerinin analizinde yaygın kullanım alanına sahip yöntemlerdir. Bu yöntemler karmaşık yapıları veri setlerinde değişkenler arası bağımlılık yapısını ortadan kaldırarak veriyi daha az boyuta indirgeyerek uygulama ve yorumlama kolaylığı sağlamaktadır. Aynı zamanda verilerin ilişki yapısını grafiksel (görsel) olarak sunarak verinin benzerlik veya farklılıklarına göre analizine olanak tanır.

Çok değişkenli istatistiksel yöntemlerden biri olan çok boyutlu ölçekleme analizi, n nesne ya da birim arasındaki p değişkene göre belirlenen uzaklıklara dayalı olarak nesnelerin k boyutlu ( $k < p$ ) bir uzayda gösterimini elde etmeyi amaçlayan, böylece nesneler arasındaki ilişkileri belirlemeye yarayan bir yöntemdir. (Özdamar, 2004)

Bu analiz ile çok boyutlu veri matrisindeki birimler arasındaki karmaşık ilişkilerin, daha kolay anlaşılabilir ve açıklanabilir boyutlara indirgenmesi sağlanabilmektedir. (Kalaycı vd. , 2005)

ÇBÖ analizi, kümeleme ve faktör analizi ile benzerlik göstermekle birlikte, normallik ve doğrusallık varsayımlarının sağlanması koşulunun aranmaması sebebiyle bu yöntemlerden ayrılır. (Tinsley, 2000)

ÇBÖ analizinde verinin hangi ölçek düzeyinde olduğu uzaklıklar matrisi kullanılması ve veri tipine göre uygun uzaklık matrislerinin oluşturulması sebebiyle önemli bir yer tutar. Veri tipine göre uzaklık elde etme yöntemleri Tablo 2’de görülmektedir.



Tablo 2. Veri Tipine Göre Uzaklık Elde Etme Yöntemleri

Veri Ölçeği	Tipi	Uzaklık Elde Etme Yöntemleri
Aralık ve orantılı Ölçek		Öklid uzaklığı, Karesel öklid uzaklığı, Minkowsky, Chebychef
İkili Ölçek		Öklid uzaklığı, Karesel öklid uzaklığı, Size difference, Pattern difference, Varyans uzaklığı
Sayım Değerleri		Ki-kare, Fi-kare Ölçüsü

Eğer veri setinde yer alan değişkenler, farklı ölçeklerde elde edilmiş ise bu değerlerin standardize edilmesi mümkündür. Uygulamada çok sayıda standartlaştırma tekniği yer alırken genelde çok boyutlu ölçeklemede kullanılan standartlaştırma teknikleri, Z skor, 0 ile 1 aralığında dönüştürme, maksimum değer 1 olacak şekilde dönüştürme, ortalama 1 olacak şekilde dönüştürme ve standart sapması 1 olan değerlere dönüştürme teknikleridir. (Özdamar 2004).

Bu çalışmada Z skor standartlaştırma tekniğinden yararlanılarak, veri tipi ölçeğine uygunluğu sebebiyle metrik ÇBÖ kullanılmıştır.

## 2.3 Çok Boyutlu Ölçekleme Analizi Türleri

### 2.3.1 Metrik Çok Boyutlu Ölçekleme

Metrik çok boyutlu ölçekleme ile n tane noktanın uzaklıkları ( $d_{ij}$ ) ve tahmini uzaklıkları ( $\delta_{ij}$ ,  $f(\delta_{ij})$ ) olmak üzere aşağıdaki formülasyonun minimize edilmesi sağlanarak en uygun boyutta ilişki belirlenmeye çalışılır. (Nemlioğlu, 2005).

$$L = \sum (d_{ij} - \delta_{ij})^2 \quad (1)$$

### 2.3.2. Metrik Olmayan Çok Boyutlu Ölçekleme

Metrik olmayan çok boyutlu ölçeklemede ise birimler arası uzaklıkların sıralaması dikkate alınır. Metrik ölçekleme ile arasındaki temel fark, metrik olmayan çok boyutlu ölçeklemede verilerin sınıflama ya da sıralama ölçek tipinde olmasıdır.

### 2.3.3. Çok Boyutlu Ölçekleme Analizi Uygulama Aşamaları

Öncelikle veri setinde yer alan değişkenlerin veri matrisinde ( $X_{ij}$ ) aynı ölçek türüne göre oluşturulması gerekmektedir. Veriler arasında farklı ölçek türüne ait değerler var ise verilerin uygun bir ölçek türüne dönüştürme veya standartlaştırma işleminin yapılmalıdır.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{bmatrix} \quad (2)$$

Uzaklık matrisi (D<sub>ij</sub>) olmak üzere birimler arasındaki uzaklıklar aşağıdaki formül yardımıyla oluşturulur. (Hair,1998).

$$d_{ij} = \sqrt{\sum (x_{ik} - x_{jk})^2} \quad (3)$$

$$D = \begin{bmatrix} 0 & & & \\ d_{21} & 0 & & \\ \cdot & \cdot & \cdot & \cdot \\ d_{p1} & d_{p2} & \dots & 0 \end{bmatrix} \quad (4)$$

En önemli kararlardan biri olan uygun boyutun belirlenmesi için, her boyutun stress istatistiği değerlerine bakılır. Stress istatistiklerinden en küçük olana ait boyut, çözümde kullanılacak en uygun boyut olarak belirlenir.

Tahmini uzaklık matrisinin elde edilmesi tahmini uzaklık değerleri (konfigürasyon uzaklık değerleri), veri tipine bağlı olarak doğrusal, polinomial veya monotonik regresyon yöntemlerinden birisi ile elde edilir. Bu aşamada regresyon yöntemlerinden hangisinin seçileceğine verilerin serpilme diyagramına bakılarak karar verilir. (Kalaycı vd, 2005)

Sonraki aşamada stress istatistiği ve RSQ uyum indeksi değerlerine göre modelin veriye uyumu hakkında sonuçlar elde edilir.

RSQ uzaklık değerlerinden hareketle hesaplanan dönüştürülmüş yakınlık değerlerindeki varyansın yüzdesidir. Stress istatistiğinin sıfıra, RSQ değerinin ise 1’e yaklaşması istenir. (Tinsley, 2000)

$$\text{Stress} = \sqrt{\frac{\sum \sum (\hat{d}_{ij} - d_{ij})^2}{\sum \sum d_{ij}^2}} \quad (5)$$

Analizin son aşamasında ise elde edilen grafiksel gösterim kolaylığı ile birimler birbirlerine ya da genel eğilime olan uzaklık ya da yakınlıklarına göre yorumlanır.

### 3. Bulgular

Türkiye’de 81 il düzeyinde toplumsal cinsiyet bakış açısıyla seçilmiş Tablo.1’deki değişkenler, 2019 yılına ait veri kaynakları kullanılarak oluşturulan veri seti ile birbirine benzerlik ve farklılıklarından yola çıkarak, toplumsal yapı ile ilgili bir karşılaştırma elde etmek üzere ÇBÖ analizi ile kümelendi. Aynı zamanda yine ÇBÖ analizi kullanılarak, veri setinde yer alan değişkenler birbirlerine yakınlık uzaklık yönünden incelenmiştir.

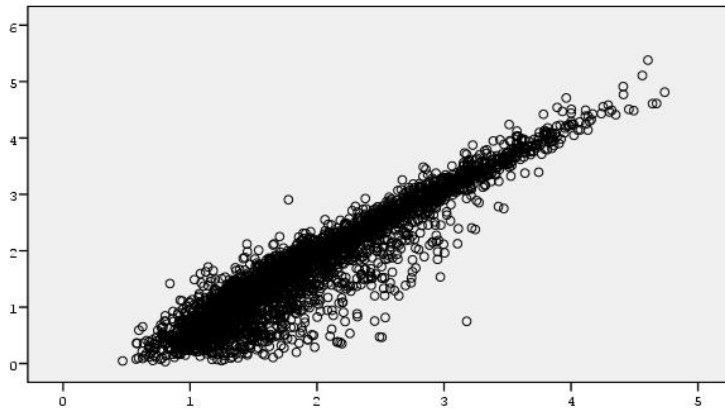
81 il için yapılan ÇBÖ analizinde elde edilen sonuçların, mevcut veri setini yeterli ölçüde yansıtmadığını gösteren stress değerleri incelenmiştir. k =2 boyut için stress istatistiğinin 0.001 değerinden küçük olduğu değere kadar iterasyon devam ettirilmiştir. 4. İterasyonda 0.00057 değerine ulaşıldığında iterasyon durdurulmuştur. 81 il için yapılan ÇBÖ analizinde RSQ değeri 0.87528 olarak hesaplanmıştır. Buna göre k=2 için stress değeri, verileri 0.87528 oranında açıklamaktadır.

81 il için yapılan ÇBÖ analizinde birbirine yakınlık gösteren illerin Türkiye’ nin coğrafi bölge ayırımına yakın bir kümelenme yapısı oluşturduğu görülmektedir. Öte yandan elde edilen sonuç illerin gelişmişlik düzeyleriyle toplumsal cinsiyet karnelerinin benzerlik gösterdiğine dikkat çekmektedir.

17 değişken için yapılan ÇBÖ analizinde elde edilen sonuçların, mevcut veri setini yeterli ölçüde yansıtır yansıtmadığını gösteren stress değerleri incelenmiştir.  $k=2$  boyut için stress istatistiğinin 0.001 değerinden küçük olduğu değere kadar iterasyon devam ettirilmiştir. 4. İterasyonda 0.00038 değerine ulaşıldığında iterasyon durdurulmuştur. 17 değişken için yapılan ÇBÖ analizinde RSQ değeri 0.86733 olarak hesaplanmıştır. Buna göre  $k=2$  için stress değeri, verileri 0.86733 oranında açıklamaktadır.

17 değişken için yapılan ÇBÖ analizinde ise yüksekokul mezunu kadın oranı, 4abc zorunlu sigortalı kadın oranı, kadın ortalama evlilik yaşı, kadın yabancı nüfus oranı birbirine benzer algılanan (yakınlık gösteren) değişkenlerdir. Pozitif en büyük değere sahip olan evlilik süresi 1 yıl doğum oranı ve 4abc zorunlu sigortalı kadın oranı 1. boyut için en ayrıştırıcı değişkendir. 2. boyut için en ayrıştırıcı değişken ise tek çekirdek aileden oluşan aile oranıdır.

Uzaklıklar

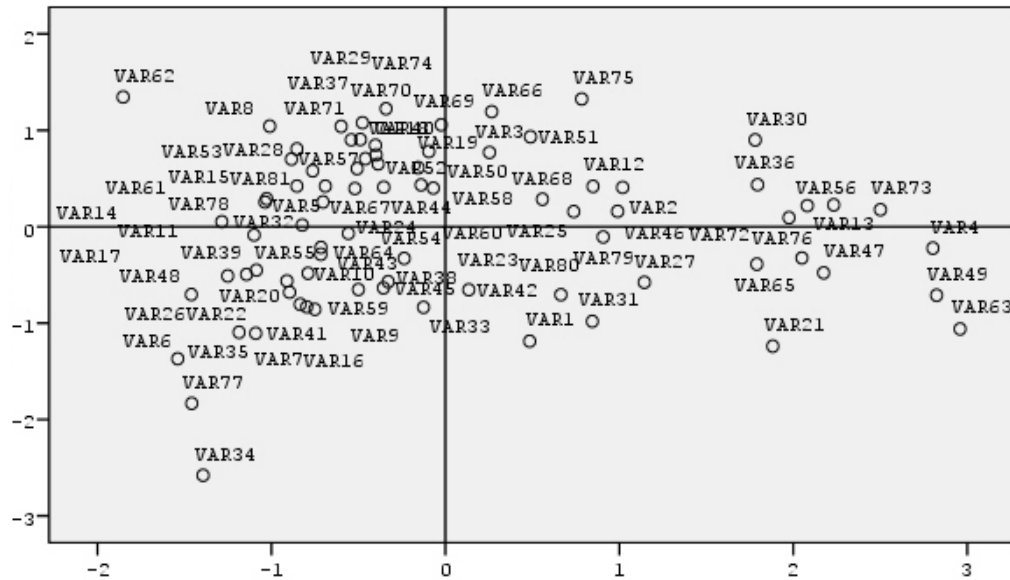


Farklar

Şekil 1. 81 İl İçin Öklid Mesafesi Modeli Serpilme Diyagramı

Şekil 1.’de birimler arası farklılıklar uzaklıklara göre doğrusallık göstermektedir. Buna göre ÇBÖ analizinde uzaklık matrisinden elde edilen tahmini uzaklıklar, veriden elde edilen uzaklıklarla uyumludur. Buna göre model veriyle uyumludur şeklinde yorumlanır.

Boyut2



Boyut1

Şekil 2. 81 İl İçin Öklid Mesafesi Modeli

Şekil 2.’ de illerin birbirine yakınlık ve uzaklıkları ve genel eğilimden farklılıkları görülmektedir. İstanbul, Tunceli ve Şanlıurfa’nın genel eğilimden farklılık gösterdiği söylenebilir.

Tablo 3. 81 İl için Uyarıcı Koordinatlar

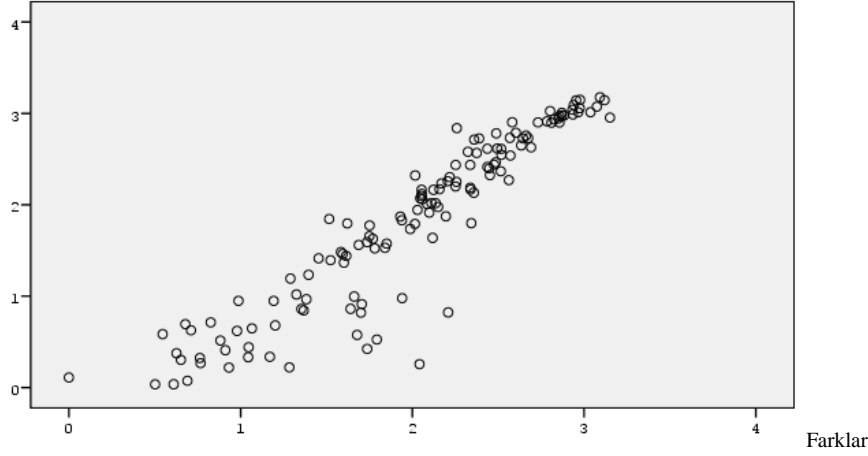
No	Değişken Adı	1. Boyut	2. Boyut
1	Adana (VAR1)	0,4846	-1,1876
2	Adıyaman (VAR2)	1,0191	0,4071
3	Afyonkarahisar (VAR3)	0,2537	0,7706
4	Ağrı (VAR4)	2,8024	-0,2228
5	Amasya (VAR5)	-0,6902	0,4212
6	Ankara (VAR6)	-1,5389	-1,3694
7	Antalya (VAR7)	-1,1859	-1,0950
8	Artvin (VAR8)	-1,0102	1,0427
9	Aydın (VAR9)	-0,5001	-0,6517
10	Balıkesir (VAR10)	-0,7898	-0,4829
11	Bilecik (VAR11)	-1,0997	-0,0859
12	Bingöl (VAR12)	0,8488	0,4185
13	Bitlis (VAR13)	2,0804	0,2185
14	Bolu (VAR14)	-1,2853	0,0524
15	Burdur (VAR15)	-0,8533	0,4213
16	Bursa (VAR16)	-0,7969	-0,8311
17	Çanakkale (VAR17)	-1,1434	-0,4951
18	Çankırı (VAR18)	-0,3984	0,7443
19	Çorum (VAR19)	-0,1563	0,6127
20	Denizli (VAR20)	-0,9109	-0,5624
21	Diyarbakır (VAR21)	1,8819	-1,2398
22	Edirne (VAR22)	0,8958	-0,6800
23	Elazığ (VAR23)	-0,2370	-0,3287
24	Erzincan (VAR24)	-0,5077	0,6032
25	Erzurum (VAR25)	0,7378	0,1568
26	Eskişehir (VAR26)	-1,4603	-0,7039
27	Gaziantep (VAR27)	1,1438	-0,5772
28	Giresun (VAR28)	-0,8536	0,8064
29	Gümüşhane (VAR29)	-0,4765	1,0784
30	Hakkari (VAR30)	1,7808	0,9005
31	Hatay (VAR31)	0,8446	-0,9816
32	Isparta (VAR32)	-0,8227	0,0166
33	Mersin (VAR33)	-0,1257	-0,8361
34	İstanbul (VAR34)	-1,3930	-2,5792
35	İzmir (VAR35)	-1,0901	-1,1057
36	Kars (VAR36)	1,7946	0,4356
37	Kastamonu (VAR37)	-0,5996	1,0406
38	Kayseri (VAR38)	-0,3291	-0,5710
39	Kırklareli (VAR39)	-1,0862	-0,4508
40	Kırşehir (VAR40)	-0,4889	0,9052

Tablo 3. 81 İl için Uyarıcı Koordinatlar (devam)

No	Değişken Adı	1. Boyut	2. Boyut
41	Kocaeli (VAR41)	-0,8344	-0,8050
42	Konya (VAR42)	0,1345	-0,6547
43	Kütahya (VAR43)	-0,7036	0,2569
44	Malatya (VAR44)	-0,3546	0,4097
45	Manisa (VAR45)	-0,3567	-0,6351
46	Kahramanmaraş (VAR46)	0,9902	0,1595
47	Mardin (VAR47)	2,1746	-0,4782
48	Muğla (VAR48)	-1,2514	-0,5103
49	Muş (VAR49)	2,8235	-0,7113
50	Nevşehir (VAR50)	-0,0950	0,7806
51	Niğde (VAR51)	0,4890	0,9330
52	Ordu (VAR52)	-0,3870	0,6506
53	Rize (VAR53)	-0,8848	0,7013
54	Sakarya (VAR54)	-0,5575	-0,0698
55	Samsun (VAR55)	-0,7155	-0,2146
56	Siirt (VAR56)	2,2322	0,2266
57	Sinop (VAR57)	-0,4597	0,7061
58	Sivas (VAR58)	-0,0695	0,4003
59	Tekirdağ (VAR59)	-0,7504	-0,8582
60	Tokat (VAR60)	-0,1401	0,4336
61	Trabzon (VAR61)	-1,0278	0,2897
62	Tunceli (VAR62)	-1,8527	1,3452
63	Şanlıurfa (VAR63)	2,9588	-1,0609
64	Uşak (VAR64)	-0,7158	-0,2856
65	Van (VAR65)	2,0492	-0,3235
66	Yozgat (VAR66)	0,2661	1,1937
67	Zonguldak (VAR67)	-0,5204	0,3958
68	Aksaray (VAR68)	0,5579	0,2850
69	Bayburt (VAR69)	-0,0241	1,0572
70	Karaman (VAR70)	-0,4023	0,8446
71	Kırıkkale (VAR71)	-0,5411	0,9012
72	Batman (VAR72)	1,7905	-0,3882
73	Şırnak (VAR73)	2,5010	0,1768
74	Bartın (VAR74)	-0,3417	1,2246
75	Ardahan (VAR75)	0,7838	1,3252
76	Iğdır (VAR76)	1,9749	0,0932
77	Yalova (VAR77)	-1,4584	-1,8336
78	Karabük (VAR78)	-1,0368	0,2555
79	Kilis (VAR79)	0,9070	-0,1061
80	Osmaniye (VAR80)	0,6639	-0,7053
81	Düzce (VAR81)	-0,7628	0,5794

Tablo 3’ e göre 1. boyutta hem pozitif hem en büyük değer alan iller en önemli ayrıştırıcı olarak yorumlanabilir. Bu durumda Ağrı, Bitlis, Mardin, Muş, Siirt, Şanlıurfa, Van, Şırnak 1. boyutta en önemli ayrıştırıcı illerdir. 1. boyutta negatif değer alan illerin sayısı oldukça fazladır. Burada dikkat çekici olan Ankara, İzmir, Antalya, Eskişehir, Denizli, Muğla başta olmak üzere bu illerin birbirine yakınlık gösterdiği ancak ayrıştırıcı özellikte olmadıklarıdır. 0’ a yakınlık gösteren illerin sayısı da pozitif ve negatif değer gösteren illere göre daha azdır. 2. boyutta ise en yüksek pozitif değer Tunceli ilinde görülmektedir. Tunceli 2. boyut için en önemli ayrıştırıcı ildir. 2. boyutta ayrıştırıcılığı önemsiz olarak belirlenen il ise İstanbul’ dur.

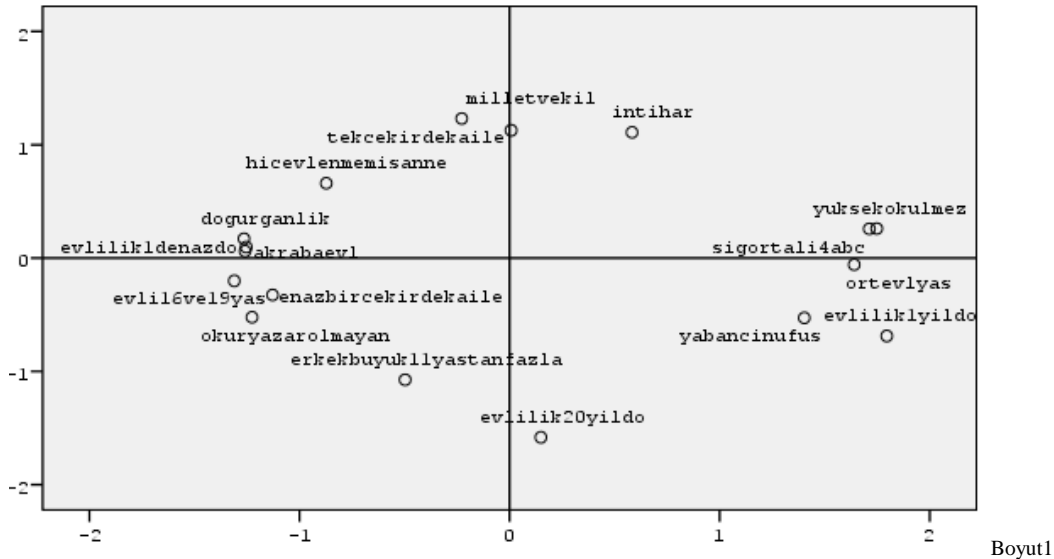
Uzaklıklar



Şekil 3. 17 Değişken İçin Öklid Mesafesi Modeli Serpilme Diyagramı

Şekil 3.’ de birimler arası farklılıkların uzaklıklara göre doğrusallık göstermesi sebebiyle ÇBÖ analizinde uzaklık matrisinden elde edilen tahmini uzaklıklar, veriden elde edilen uzaklıklarla uyumluluk göstermektedir. Buna göre model veriyile uyumludur şeklinde yorumlanır.

Boyut2



Şekil 4. 17 Değişken İçin Öklid Mesafesi Modeli

Şekil 4’ te illerin birbirine yakınlık ve uzaklıkları ve genel eğilimden farklılıkları görülmektedir. Buna göre yüksekokul mezunu kadın oranı, 4abc zorunlu sigortalı kadın oranı, kadın ortalama evlilik yaşı, kadın yabancı nüfus

oranı birbirine benzer algılanan (yakınlık gösteren) değişkenlerdir. Tablo 4’ e göre pozitif en büyük değere sahip olan evlilik süresi 1 yıl doğum oranı ve 4abc zorunlu sigortalı kadın oranı 1. boyut için en ayırıştırıcı değişkendir. Kadın milletvekili oranı ve kadın intihar oranı 0’ yakın pozitif değer alan değişkenler olup, geriye kalan tüm değişkenler de negatif değer olarak 1. boyut için birbirine yakın ve ayırıştırıcı değildir şeklinde yorumlanır. 2. boyut için en büyük pozitif değer ile en ayırıştırıcı değişken ise tek çekirdek aileden oluşan aile oranıdır.

Tablo.4. 17 Değişken İçin Uyarıcı Koordinatlar

No	Değişken Adı	1. Boyut	2. Boyut
1	kadın intihar oranı	0,5832	1,1095
2	kadın milletvekili oranı	0,0073	1,1277
3	15+okuryazar kadın oranı	-1,2257	-0,5224
4	yüksekokul mezunu kadın oranı	1,7124	0,2581
5	dogurganlık hızı	-1,2641	0,1709
6	evlilik 1 yıldanaz doğum oranı	-1,2536	0,0973
7	evlilik 1yıl doğum oranı	1,7954	-0,6871
8	evlilik 20yılıfazla dogum oranı	0,1491	-1,5819
9	hic evlenmeyen anne oranı	-0,8725	0,6603
10	kadın16-19yas evlenme oranı	-1,3104	-0,2011
11	erkek kadımdan11yasbuyuk	-0,4973	-1,0735
12	tekeçirdekaile oranı	-0,2276	1,2317
13	enazbirçekirdek aile oranı	-1,1281	-0,3259
14	yabancımüfuslu kadın oranı	1,4030	-0,5278
15	kadın ortevlenmeyaşı	1,6402	-0,0579
16	akrabaevliliğioranı	-1,2591	0,0613
17	4abc zorunlu sigortalı kadın oranı	1,7477	0,2609

#### 4. Sonuç

Bu çalışmada nüfus, doğurganlık, sağlık, evlenme, aile yaşamı, eğitim, siyasal yaşam, iş ve kazanç konuları kapsamında 81 il düzeyinde toplumsal cinsiyet bakış açısıyla seçilmiş 17 değişken yer almaktadır. 2019 yılına ait veri kaynakları kullanılarak toplumsal cinsiyet bakış açısıyla seçilmiş değişkenlerden oluşturulan veri seti 81 il ve veri setinde yer alan değişkenler birbirine benzerlik ve farklılıklarından yola çıkarak, toplumsal yapı ve toplumsal cinsiyet ile ilgili bir karşılaştırma elde etmek üzere ÇBÖ analizi ile kümelendi.

Bu çalışmada toplumsal cinsiyetin, toplumsal yapıyı oluşturan sosyal, kültürel, ekonomik etkenlerle birlikte ele alınması çok boyutlu değerlendirme imkanı sağlamaktadır. Çalışmanın sonuçları, kadının toplum ve ailede algılanma biçimi hakkında önemli ipuçları içermektedir. 81 ilin toplumsal cinsiyet açısından bölgesel farklılık gösterdiği, yanı sıra, birbirine yakınlık gösteren illerin Türkiye’nin coğrafi bölge ayrımını temsil eden bir yapıda kümelendiği görülmektedir. Bu kümeleneceğe göre illerin gelişmişlik düzeylerinin toplumsal cinsiyet bakış açısıyla seçilmiş değişkenler yönünden de benzerlik gösterdiği dikkat çekmektedir.

Çalışmanın sonuçlarının, bölgesel düzeyde toplumsal yapıya ve toplumsal cinsiyete dikkat çekmesi beklenmektedir. Bölgesel düzeyde farkındalığın artırılmasının sürdürülebilir kalkınmada kadınların katkısının güçlendirilmesi hedefinin gerçekleştirilmesi için önemli bir adım olduğu düşünülmektedir.

Toplumsal yapı ve toplumsal cinsiyete konu olabilecek değişkenlerin 81 il düzeyinde mevcut olup olmaması çalışmanın başlangıç aşamasında karşılaşılan kısıtlayıcıların başında gelmektedir. Bu sebeple mevcut veri setine eklenecek değişkenlerin il düzeyinde elde edilebilir olması, ileride yapılacak çalışmalar için önemlidir. Ayrıca bu çalışmada yer alan değişkenlerin yıllar içindeki değişimlerinin izlenmesine yönelik yapılacak çalışmalar da toplumsal yapı ve toplumsal cinsiyet için detaylı ve alternatif analizler sağlayacaktır.

## Kaynakça

- Aytaç M. Bayram N. (2001). Öğretim Elemanlarının Kariyer Tutumlarının Gruplandırılması, Çukurova Üniversitesi V. Ulusal Ekonometri ve İstatistik Sempozyumu.
- Berktaş, F. (2011). Feminist teoride açılımlar, toplumsal cinsiyet çalışmaları. Eskişehir: Anadolu Üniversitesi Yayını, 2-23
- CEİD. (2019). Toplumsal Cinsiyet Eşitsizliği Endeksi Erişim adresi: <https://ceidizler.ceid.org.tr/UNDP-Toplumsal-Cinsiyet-Esitsizligi-Endeksi-2019-Yayimlandi-d122>
- Cox T.F. Cox M. A. A. (2001). Multi dimensional Scaling. USA: Chapman & Hall/CRC (Second Edition)
- Gökçin G. M, Koçak A., (2021) Türkiye’ de Toplumsal Cinsiyetin Veri Madenciliği Sınıflama Modelleri İle Analizi, International Idea Studies Journal, 7(31), 480-490
- Hair J.F. Anderson R.E.Tatham R. L. Black W. C. (1998) Multivariate Data Analysis. USA: Prentice-Hall Inc., (Fifth Edition)
- Kalaycı Ş, vd. (2005) SPSS Uygulamalı Çok Değişkenli İstatistiksel Teknikleri. İstanbul: Asil Yayın Dağıtım Ltd.
- Nemlioğlu A.K (2005). Benzerlik Analizleri, Varsayımlara Uymayan Kategorik Verilerde Çok Değişkenli Analiz, Correspondence & Homogeneity. İstanbul: Beşir Kitabevi.
- Özdamar K. (2004). Paket Programlar ile İstatistiksel Veri Analizi -2. Yenilenmiş İstanbul: Kaan Kitabevi (5. Baskı)
- Sarp, Ü. (2008). Toplumsal cinsiyet eşitliği, T.C. Başbakanlık Kadının Statüsü Genel Müdürlüğü Rapor, Ankara, SGK. (2019). İşyeri, Sigortalı, Aylık ve Gelir Alanlara Ait İstatistikler [Veri seti]. Erişim adresi: <http://www.sgk.gov.tr/wps/portal/sgk/tr/kurumsal/istatistik>
- Tatlıdil H. (1999). Uygulamalı Çok Değişkenli İstatistiksel Analiz. Ankara: Engin Yayınları.
- TEPAV (2019) Türkiyenin Toplumsal Cinsiyet Eşitliği Karnesi Erşim adresi : [https://www.tepav.org.tr/upload/mce/2020/haberler/tobb\\_tskb\\_tepav\\_81\\_ilde\\_turkiyenin\\_toplumsal\\_cinsiyet\\_esitligi\\_karnesi.pdf](https://www.tepav.org.tr/upload/mce/2020/haberler/tobb_tskb_tepav_81_ilde_turkiyenin_toplumsal_cinsiyet_esitligi_karnesi.pdf)
- Tinsley H.E. A. (2000). Brown S.D. Handbook of Applied Multivariate Statistics and Mathematical Modeling. USA: Academic Press.
- TBMM. (2019). Türkiye Büyük Millet Meclisi Milletvekilleri Dağılımı [Veri seti]. Erişim adresi: [https://www.tbmm.gov.tr/develop/owa/milletvekillerimiz\\_sd.dagilim](https://www.tbmm.gov.tr/develop/owa/milletvekillerimiz_sd.dagilim)
- TÜİK. (2019). Toplumsal Yapı Ve Cinsiyet İstatistikleri [Veri seti]. Erişim adresi: <https://data.tuik.gov.tr/Kategori/GetKategori?p=nufus-ve-demografi-109&dil=1>
- WEF, (2019). 2019 Küresel Toplumsal Cinsiyet Uçurumu Raporu. Erişim adresi: <https://www.weforum.org/reports/gender-gap-2020-report-100-years-pay-equality>

\*Çalışmadaki ifade edilen görüşler tamamen yazara aittir ve hiçbir şekilde Türkiye İstatistik Kurumu’nu bağlamaz.

\*Bu çalışmanın veri seti International Idea Studies Journal’ da yayınlanan Türkiye’ de Toplumsal Cinsiyetin Veri Madenciliği Sınıflama Modelleri ile Analizi (Gülsüm Merve Gökçin ortak) adlı makalemle aynıdır.