



T.C.  
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ



## İKİ AŞAMALI EN KÜÇÜK KARELER YÖNTEMİ VE UYGULAMASI

Fırat DEMİR  
BİYOİSTATİSTİK ANABİLİM DALI  
YÜKSEK LİSANS TEZİ

DANIŞMAN  
Prof. Dr. Sıddık KESKİN

VAN-2019

T.C.  
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

**İKİ AŞAMALI EN KÜÇÜK KARELER YÖNTEMİ VE  
UYGULAMASI**

Fırat DEMİR

BİYOİSTATİSTİK ANABİLİM DALI  
YÜKSEK LİSANS TEZİ

DANIŞMAN

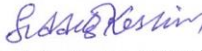
Prof. Dr. Sıddık KESKİN

VAN-2019

## KABUL VE ONAY

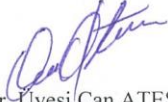
Van Yüzüncü Yıl Üniversitesi Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalında "Fırat DEMİR" tarafından hazırlanan "*İki Aşamalı En Küçük Kareler Yöntemi*" adlı tez çalışması aşağıdaki jüri tarafından YÜKSEK LİSANS TEZİ olarak OY BİRLİĞİ/ÇOKLUĞU ile kabul edilmiştir.

Tez Savunma Tarihi: 12.07.2019



Prof. Dr. Siddık KESKİN

Jüri Başkanı



Dr. Öğr. Üyesi Can ATEŞ

Jüri Üyesi



Dr. Öğr. Üyesi Fahrettin ÖZBEY

Jüri Üyesi

Tez hakkında alınan jüri kararı, Van Yüzüncü Yıl Üniversitesi Sağlık Bilimleri Enstitüsü Yönetim Kurulu tarafından onaylanmıştır.

  
Prof. Dr. Semiha DEDE  
Sağlık Bilimleri Enstitüsü Müdürü

II

## ETİK BEYAN

T.C.  
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE

Yüksek Lisans tezi olarak hazırlayıp sunduğum “İki Aşamalı En Küçük Kareler Yöntemi” başlıklı tezim; bilimsel ahlak ve değerlere uygun olarak tarafımdan yazılmıştır. Tezimin fikir/hipotezi tümüyle tez danışmanım ve bana aittir. Tezde yer alan deneysel çalışma/araştırma tarafımdan yapılmış olup, tüm cümleler, yorumlar bana aittir. Bu tezdeki bütün bilgiler akademik kurallara ve etik ilkelere uygun olarak hazırlanıp, bu kural ve ilkeler gereği, çalışmada bana ait olmayan tüm veri, düşünce ve sonuçlara atıf yapılmış ve kaynak gösterilmiştir.

Yukarıda belirtilen hususların doğruluğunu beyan ederim.

Fırat DEMİR

12.07.2019

## TEŐEKKÜR

Yüksek lisans tezimin hazırlanmasında, tecrübeleri ve bilgi birikimleriyle katkısı olan ve çalışmalarımnda desteklerini benden esirgemeyen danışman hocam Prof. Dr. Sıddık KESKİN'e sonsuz teşekkürlerimi sunarım. Ayrıca desteğini benden esirgemeyen ve bilimsel anlamda tecrübeleriyle bana ışık tutan değerli hocam Dr. Öğr. Üyesi Bahattin BULDUK ve Öğr. Gör. Dr. Sadi ELASAN'a teşekkürlerimi borç bilirim. Aynı zamanda beni her konuda destekleyen ve yanımda olan değerli eşim Tahmine'ye sonsuz sevgilerimi sunarım.



## ÖZET

**Demir F, İki Aşamalı En Küçük Kareler Yönteminin incelenmesi, Van Yüzüncü Yıl Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, Yüksek Lisans Tezi, Van 2019.**İki aşamalı en küçük kareler yöntemi aracı değişken kullanarak, açıklayıcı içsel değişkenler ile yapılan özel bir regresyon analizi yöntemi olup, analiz sonucu elde edilen tahminlerin, sapmasız ve tutarlı olabilmesini sağlamaya yardımcı olmaktadır. Bu çalışmada iki aşamalı en küçük kareler yöntemi incelenerek bir veri seti ile uygulama yapılmıştır. Uygulamada, örneklem genişliği 300 alınarak, 18 değişken içeren serbest erişimli veri seti kullanılmıştır. Verilere ilk önce adımsal regresyon analizi uygulanarak, modele girecek değişkenler belirlenmiştir.Sonraki aşamada bu değişkenler ile standart regresyon analizi ve İki aşamalı en küçük kareler regresyon analizi yapılmıştır.Sonuç olarak, veri setinde içsellik problemi, bağımsız değişkenlerin hatalı olarak ölçülmüş olması ve modelde feed-back halkasının olması durumunda İki aşamalı en küçük kareler regresyon yönteminin, standart regresyon analizi yöntemine alternatif olarak kullanılabileceği vurgulanmıştır.

**Anahtar kelimeler:**İçsellik problemi, sapmasız tahmin, ölçüm hatası



## ABSTRACT

**Demir F, Two-Stage Least Squares Methods, Van Yuzuncu Yil University, Institute of Health Sciences, Department of Biostatistics, M.Sc. Thesis, Van 2019.** The two-stage least squares method is a special regression analysis method using instrumental variables, which helps to ensure that the predictions obtained from the analysis, are accurate and consistent. In this study, two-stage least squares method was examined and an application was performed. In the application, free access data set with 18 variables was used and sample size was 300. First, stepwise regression analysis was applied to the data and the variables to be entered into the model were determined. In the next step, standard regression analysis and two-stage least squares regression analyses were performed. As a result, it is emphasized that two-stage least squares regression method can be used as an alternative to standard regression analysis method in case of endogeneity problem, errors in the X variables and feed-back loop in the model.

**Keywords:** Endogeneity problem unbiased estimation, measurement error



## İÇİNDEKİLER

|   |      |
|---|------|
| KABUL VE ONAY .....   | II   |
| ETİK BEYAN.....   | III  |
| TEŞEKKÜR.....   | IV   |
| ÖZET .....  | V    |
| ABSTRACT.....   | VI   |
| İÇİNDEKİLER .....   | VII  |
| SİMGELER VE KISALTMALAR .....   | VIII |
| ŞEKİLLER LİSTESİ .....  | IX   |
| TABLolar LİSTESİ.....   | X    |
| 1. GİRİŞ .....  | 1    |
| 2. GENEL BİLGİLER .....   | 3    |
| 3. GEREÇ YÖNTEM .....   | 5    |
| 3.1. Gereç.....   | 5    |
| 3.2. Yöntem.....  | 5    |
| 3.2.1. Regresyon analizi.....   | 6    |
| 3.2.2. Regresyon analizi varsayımları.....                                  | 6    |
| 3.2.3. En küçük kareler yöntemi .....                                       | 8    |
| 3.2.4. En küçük kareler yönteminin varsayımları .....                       | 10   |
| 3.2.5. İki aşamalı en küçük kareler yöntemi .....                           | 11   |
| 3.2.6. İki aşamalı en küçük kareler yönteminin varsayımları .....           | 13   |
| 3.2.7. İki aşamalı en küçük kareler yönteminin parametre tahminleri .....   | 13   |
| 3.2.8. İki aşamalı en küçük kareler yöntemi tahmincilerin özellikleri ..... | 14   |
| 3.2.9. Hausman testi.....   | 14   |
| 4.BULGULAR.....   | 16   |
| 5.TARTIŞMA VE SONUÇ .....   | 19   |
| KAYNAKLAR .....   | 22   |
| ÖZGEÇMİŞ .....  | 24   |
| EKLER.....  | 25   |
| Ek 1. Etik Kurul Raporu .....   | 25   |
| Ek 2. Tez Orijinallik Raporu.....   | 26   |



## SİMGELER VE KISALTMALAR

|                              |  |
|------------------------------|--|
| <b><math>\Sigma</math></b>   | : Toplam sembolü                                 |
| <b>a</b>                     | : Sabit sayı                                     |
| <b>e</b>                     | : Hata terimi                                    |
| <b>F</b>                     | : Frekans  |
| <b>HKO</b>                   | : Hata Kareler Ortalaması                        |
| <b>HKT</b>                   | : Hata Kareler Toplamı                           |
| <b>k</b>                     | : Regresyon modelindeki bağımsız değişken sayısı |
| <b>N</b>                     | : Sütun sayısı                                   |
| <b>R<sup>2</sup></b>         | : Belirleme katsayısı                            |
| <b>X</b>                     | : Bağımsız değişken                              |
| <b>Y</b>                     | : Bağımlı değişken                               |
| <b><math>\sigma^2</math></b> | : Varyans  |
| <b>EKKY</b>                  | : En küçük kareler yöntemi                       |
| <b>2AEKKY</b>                | : İki aşamalı en küçük kareler yöntemi           |
| <b>IV</b>                    | : Aracı değişken                                 |

## ŞEKİLLER LİSTESİ

Şekil 1.Hata terimi ve hata terimi dağılımının grafiği ..... 10



## TABLÖLAR LİSTESİ

|   |    |
|---|----|
| <b>Tablo 1.</b> Değişkenlere ait genel tanımlayıcı istatistikler.....                       | 5  |
| <b>Tablo 2.</b> Standart regresyon analizi ile 2AEKK analizi istatistikleri (9 değişken)... | 16 |
| <b>Tablo 3.</b> Standart regresyon analizi ile 2AEKK analizi istatistikleri (7 değişken)... | 17 |
| <b>Tablo 4.</b> Standart regresyon analizi ile 2AEKK analizi istatistikleri (5 değişken)... | 17 |
| <b>Tablo 5.</b> Standart regresyon analizi ile 2AEKK analizi istatistikleri (3 değişken)... | 18 |
| <b>Tablo 6.</b> Standart regresyon analizi ile 2AEKK analizi istatistikleri (3 değişken)... | 18 |



## 1. GİRİŞ

Bilimsel çalışmalarda, genellikle ya özellikler arası ilişki, yada ilgilenilen özellik bakımından faktör olarak nitelendirilen gruplar arası fark olup olmadığı araştırılır. İlgilenilen değişkenlerden birisinin cevap değişkeni (bağımlı değişken), diğerinin veya diğerlerinin de açıklayıcı (bağımsız) değişken olması durumunda, cevap değişkenini tahmin etmek üzere kurulan eşitlik, regresyon veya tahmin denklemi olarak adlandırılır. Regresyon analizi ise bu denklemdeki katsayıların bulunması ve sonuçların yorumlanması işlemlerini kapsar. Yaygın olarak kullanılan standart regresyon analizinin uygulanabilmesi için, cevap değişkeni ve açıklayıcı değişkenlere ait bazı varsayımların veya ön şartların sağlanması gerekir.

Standart regresyon analizindeki varsayımlardan birisi de hata terimi ( $\epsilon$ ) ile açıklayıcı değişkenlerin (X) ilişkisiz olması varsayımdır. Ancak, uygulamalarda bu varsayım çoğu zaman sağlanmaz ve bazı açıklayıcı değişkenler (X), hata terimi ( $\epsilon$ ) ile ilişkili ya da korelasyonlu olabilir. Diğer bir ifade ile “X değişkenleri hatasız ölçülmüştür” varsayımı sağlanmayabilir. Zira, günümüzde yaygın olarak kullanılan araştırmaların birçoğunda açıklayıcı değişkenlerde ölçüm hataları söz konusudur. Bu durumda, standart regresyon analizinin kullanılması, tahminlerin yanlış ve tutarsız olmasına neden olabilir. Bu sorunu gidermek üzere, Standart regresyon analizi yöntemine alternatif olarak geliştirilmiş yöntemlerden birisi de İki aşamalı en küçük kareler yöntemidir. Bu yöntem, aracı veya yardımcı değişken kullanarak, açıklayıcı içsel değişkenler ile yapılan özel bir regresyon analizi yöntemi olup, analiz sonucunda elde edilen tahminlerin, sapmasız ve tutarlı olabilmesini sağlamaya yardımcı olmaktadır.

İki aşamalı en küçük kareler yöntemi, tek bir eşitlik yöntemi olup, aracı değişkenler modeli tahmininde kullanıldığı gibi bir eşitlik sistemindeki tüm eşitliklere de ayrı ayrı uygulanabilir. Bu durum, değişkenler arası iki yönlü ilişkinin tahmini için kolaylık sağlar. Yöntemin hesaplaması kolaydır ve sayısal optimizasyon algoritmalarının kullanılmasını gerektirmez. Yöntem genel olarak iki aşamayı içerir. Birinci aşamada, aracı değişken kullanılarak yeni değişken elde edilir. İkinci aşamada ise problemleri açıklayıcı değişkenin gerçek değerleri yerine kullanılan değişkendeki yararlanılarak cevap değişkeni için standart regresyon analizi yöntemi uygulanır.

İki aşamalı en küçük kareler yöntemi, Ekonometri başta olmak üzere; Sosyal Bilimler, Sağlık Bilimleri, Ziraat ve Mühendislik alanlarda yaygın olarak kullanılmaktadır. Ancak yapılan literatür incelemesinde, yöntem ve uygulamalarda kullanılabilirliğine ilişkin yapılan çalışmaların az olduğu gözlenmiş ve bu nedenle çalışmada, İki aşamalı en küçük kareler yönteminin tanıtılması ve uygulama yapılması amaçlanmıştır. Ayrıca bu çalışma ile; İki aşamalı en küçük kareler yönteminin kullanılmasının, değişkenler arası ilişkileri nasıl etkilediğinin belirlenmesi ve değişkenlerin tipine ve aralarındaki ilişki durumuna göre uygun yöntem seçimi konusunda fikir edinilmesi amaçlanmaktadır.



## 2. GENEL BİLGİLER

İki aşamalı en küçük kareler yöntemi, doğrusal eş zamanlı eşitlikler sisteminde tek bir yapısal eşitliğin parametrelerinin tahmin edilmesinde yaygın olarak kullanılan bir yöntem olmuştur.

Murphy (1973) yaptığı çalışmada, İki aşamalı en küçük kareler yöntemini, çok denklemlili modelin yapısal denklemindeki parametrelerin tahmininde kullanmıştır. Bu yöntemde birden fazla tahmin oluşturulduğunu belirtmiş ve ilk aşamada elde edilen sonuçların, ikinci aşamada girdi olarak kullanıldığını vurgulamıştır.

Koutsoyiannis (1989), İki aşamalı en küçük kareler yönteminin bir tek denklem yöntemi olduğunu ve sistemdeki denklemlere tek tek uygulanabildiğini belirtmiştir. Yöntemin varsayımlarına değinerek, aşırı belirlenmiş modellerle birlikte, tam belirlenmiş modellere de uygulanabileceğini vurgulamıştır.

Üçdoğruk (1990) çalışmasında, İki aşamalı ve Üç aşamalı en küçük kareler yöntemlerinin karşılaştırmasını yapmıştır. İki aşamalı en küçük kareler yöntemi ile ilgili bilgiler vererek, Türkiye çimento sanayinde 1971-1986 yılları arasında çimento talep-üretimi ile ilgili çok denklemlili makro ekonometrik model denemelerinde iki aşamalı en küçük kareler yönteminin gerekliliğini vurgulamıştır.

Tatlidil (1996), çalışmasında regresyon analizinde yaygın olarak kullanılan en küçük kareler yöntemi ve varsayımlarını ele alarak, İki aşamalı en küçük kareler yöntemine değinmiştir.

Akkaya (1998), yaptığı çalışmada İki aşamalı en küçük kareler yönteminin tutarlı ve asimptotik olarak etkin olduğunu ve tek tahminler elde etmek için aşırı belirlenmiş denklemlere de uygulanabileceğini belirtmiştir. Ayrıca, yöntemin tam belirlenmiş denklemlere uygulanması durumunda, dolaylı en küçük kareler yöntemi ile aynı sonuçlar verdiğini vurgulamıştır.

Yalta (2007), çalışmasında İki aşamalı en küçük kareler yöntemi ile bulunan tahminlerin, her zaman yansızlık ve en az varyanslılık özelliklerini sağlamayabilseler de tutarlı olduğunu belirtmiştir. Örneklem genişliği büyüdükçe, yanlılığın azaldığını ve

tahminlerin giderek popülasyongerçek deęerine yaklařtıęını belirtmiřtir. alıřmada ayrıca, kk rneklerde dikkatli olunması gerektięi ve İki ařamalı en kk kareler yntemi kullanılmadan nce Hausman sınaması yapılıp aıklayıcı deęiřkenlerin hata terimi ile iliřkili olduęunun doęrulanması gerektięi vurgulanmıřtır.

akıroęlu (2007) yaptıęı alıřmada; tketim, yatırım ve gayrisafi milli hasıla deęiřkenleri arasında iliřkiyi tahmin etmede İki ařamalı en kk kareler yntemini kullanmıřtır. Denklem modeli ile var modelinin ngr bařarısı aısından karřılařtırmasında, İki ařamalı en kk kareler ynteminin bařarisına vurgu yapmıřtır.



### 3. GEREÇ VE YÖNTEM

#### 3.1. Gereç

Çalışmada, uygulama materyali olarak serbest erişimli veri sitesi olan Kaggle dan sağlanan 300 bireye ait 10 değişkenli yaşam beklentisi veri setinin bir kısmı (Dataset Kaggle, 2019) kullanılmış olup, ele alınan değişkenler ve özellikleri Tablo 1’de verilmiştir.

**Tablo 1.** Değişkenlere ait genel tanımlayıcı istatistikler (n=300)

|                                | <b>Ortalama ± Standart Sapma</b> | <b>Minimum</b> | <b>Maksimum</b> |
|--------------------------------|----------------------------------|----------------|-----------------|
| Yıl                            | 2006.903 ±3.904                  | 2000.00        | 2014.00         |
| Yaşam beklentisi               | 61.668 ±7.754                    | 44.80          | 75.80           |
| Yetişkin ölümü                 | 242.993 ±155.817                 | 2.00           | 699.00          |
| Bebek ölümü                    | 101.436 ±250.530                 | 1.00           | 1600.00         |
| Alkol                          | 2.250 ±2.604                     | 0.01           | 10.42           |
| Yüzde harcama                  | 98.858 ±143.730                  | 0.10           | 796.87          |
| Hepatit B                      | 69.546 ±29.707                   | 2.00           | 99.00           |
| Kızamık                        | 7082.483 ±18996.263              | 0.00           | 131441.00       |
| BKI                            | 23.241 ±9.404                    | 2.10           | 45.70           |
| Beş yaş altı ölüm              | 139.713 ±336.343                 | 1.00           | 2100.00         |
| Çocuk felci                    | 74.413 ±27.2091                  | 3.00           | 99.00           |
| Genel harcama                  | 5.161 ±2.162                     | 0.76           | 12.60           |
| Difteri                        | 76.553 ±24.662                   | 5.00           | 99.00           |
| HIV/AIDS                       | 4.467 ±8.733                     | 0.10           | 49.90           |
| GSYİH                          | 1159.650 ±1505.895               | 1.68           | 8513.63         |
| Popülasyon                     | 31985587.85±152693197.49         | 795.00         | 1293859294.00   |
| İncelik 1yaş ile 9 yaş arası   | 8.853 ±5.939                     | 1.00           | 27.20           |
| İncelik 5yaş 9 yaş arası       | 9.139 ±5.994                     | 1.00           | 28.20           |
| Kaynakların Gelir kompozisyonu | 0.503 ±0.148                     | 0.00           | 0.76            |
| Okullaşma oranı                | 9.8907 ±2.27496                  | 4.20           | 14.70           |



## 3.2. Yöntem

### 3.2.1. Regresyon analizi

Standart regresyon analizi, arařtırıcının bağımlı deęişken olarak varsaydığı cevap deęişkeni (Y) ile bağımsız deęişken olarak varsaydığı açıklayıcı deęişkenler (X) arasındaki doğrusal ilişkiyi belirlemek üzere, bir tahmin denklemi elde etme ve elde edilen katsayıları yorumlama sürecini kapsayan bir analiz yöntemidir. Ancak yöntemin uygulanabilmesi için bazı varsayımların yerine gelmiş olması gerekmektedir ve bu varsayımlar ařağıdaki gibi özetlenebilir.

### 3.2.2. Regresyon analizi varsayımları

**Hatasızlık:** Açıklayıcı deęişkenlerin (X) hatasız ölçüldüğü varsayılır.

**Normallik:** Her bir X (açıklayıcı deęişken) deęeri için birden çok Y (cevap deęişkeni) deęeri vardır ve X (açıklayıcı deęişken) deęerine karşılık gelen Y (cevap deęişkeni) deęerlerinin normal dağılım gösterdiği varsayılır.

**Homojen varyanslılık:** Her bir X (açıklayıcı deęişken) deęişkenine karşılık gelen Y (cevap deęişkeni) deęerlerine ait varyanslar homojendir. Yani her X deęeri için karşılık gelen Y deęerlerine ait varyanslar arasındaki fark tesadüften ileri gelmektedir.

**Doğrusallık:** Açıklayıcı deęişkenin her deęeri için elde edilen Y deęerleri arasında doğrusal ilişki vardır. Dięer bir ifade ile her bir X deęeri için elde edilen Y deęerlerinin ortalaması bir doğru üzerindedir.

**Bağımsızlık:** Açıklayıcı deęişkenin herhangi bir deęeri için elde edilen Y deęeri, dięer bir deęeri için elde edilen Y deęerini etkilenmez.

Standart regresyon analizi için yazılan;

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon \text{ eşitlięi, matris formunda}$$

$$\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1)$$

olarak yazılır. (1) no'lueşitlikte:

$y$ ;  $n \times 1$  boyutlu cevap deęişkenine ait gözlemler vektörü,

$\beta_0$ ; regresyon sabiti,

$\beta$ ;  $p \times 1$  boyutlu regresyon katsayıları vektörü,

$X$ ;  $n \times p$  boyutlu açıklayıcı deęişkenlere ilişkin veri matrisi,

$\epsilon$ ;  $n \times 1$  boyutlu, sıfır ortalamalı ve  $\sigma^2$  varyanslı hata vektörüdür.

Standart regresyon analizinde;  $X$  ve  $Y$  deęişkenleri standardize edilmiş gözlemleri içeriyorsa, eşitlik;

$$y = X\beta + \epsilon \quad (2)$$

olarak kısaltılır.

Bağımlı deęişken veya cevap deęişkeni sayısının 1'den fazla olması durumunda ise eşitlik

$$Y = XB + E \quad (3)$$

olarak ifade edilir. (3) no'lu eşitlikte:

$Y$ ;  $n \times q$  boyutlu cevap deęişkenleri matrisi,

$B$ ;  $p \times q$  boyutlu regresyon katsayıları matrisi,

$X$ ;  $n \times p$  boyutlu açıklayıcı deęişkenler matrisi,

$E$ ;  $n \times q$  boyutlu, satırları birbirinden bağımsız ve özdeş dağılımlı hatalar matrisidir.

Standart regresyon analizinde,  $X$  tam rank'lı matris ise regresyon katsayıları;

$$\hat{\beta} = (X'X)^{-1}(X'Y) \quad (4)$$

eşitliği ile bulunur. Bağımlı deęişken veya cevap deęişkeni sayısı 1'den fazla ise eşitlik;

$$\hat{B} = (X'X)^{-1}(X'Y) \quad (5)$$

olarak yazılır. Doğrusal regresyon analizinde katsayıları tahmin etmede yaygın olarak kullanılan yöntem en küçük kareler yöntemidir.

### 3.2.3. En küçük kareler yöntemi

Doğrusal regresyon analizinde parametre tahmininde yaygın olarak kullanılan en küçük kareler yöntemi, Alman Carl Frederick (1795) ve Gauss (1805) tarafından birbirinden habersiz olarak geliştirilmiştir. Yöntem ilk olarak Astronomi ve Yerbilim verilerinin analizinde kullanılmıştır (Sanford Weisberg, 1985).

$Y = \beta_0 + \beta_1 X + \varepsilon$  olarak ifade edilen eşitlik; X ve Y değişkenlerinin ait olduğu popülasyonları için geçerlidir. İstatistik çalışmaların çoğunda olduğu gibi regresyon analizinde de popülasyonun tamamına ait verilere ulaşılamadığından, bu popülasyondan alınan örneklem verileriyle analiz yapılır. Örneklem verileri kullanılarak popülasyon parametreleri olan  $\beta_0$  ve  $\beta_1$ 'in kestirimlerini elde edebilmek için en küçük kareler yönteminden yararlanır. X ve Y değerleri için  $(X_1, Y_1), (X_2, Y_2), \dots, (X_k, Y_k)$  ikili gözlemleri bir örneklem oluşturur. Bunlar yardımıyla  $Y = \beta_0 + \beta_1 X$  eşitliğinin katsayılarını bulma olanağı yoktur. Örnekten hareketle bunlar tahmin edilebilir. (Montgomery ve Peck, 1992)

Regresyon modelinde yer alan  $\beta_0$  ve  $\beta_1$  parametrelerinin tahmini için sırası ile  $\hat{\beta}_0$  ve  $\hat{\beta}_1$  terimleri olarak ele alındığında, regresyon eşitliği;

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$  olarak yazılır. Eşitlikte yer alan  $\hat{\beta}_0$  ve  $\hat{\beta}_1$  terimlerini bulmak için en küçük kareler yöntemi kullanılır. Bu yöntemin temelinde, sapmaların karelerinin toplamını en küçük yapacak değerlerin bulunması yatmaktadır.  $\hat{Y}$  kestirim değeriyle Y gözlem değerleri arasındaki farklar sapma veya hata terimlerini oluşturur. (Marquardt, 1963)

$e = Y - \hat{Y}$  olarak hesaplanan hata terimleri pozitif, negatif veya sıfır değerine sahip olurken; bu farkların toplamı ( $\sum_{i=1}^k e_i = \sum_{i=1}^k (Y_i - \hat{Y}_i) = 0$ ) sıfıra eşittir. Hatalar, pozitif ve negatif değerli olduğundan, bunların toplamının minimum yapılması yerine, karelerinin toplamı minimum yapılır. En küçük kareler yönteminin esası;  $\beta_0$  ve  $\beta_1$  parametrelerinin kestirimleri olan  $\hat{\beta}_0$  ve  $\hat{\beta}_1$  'i, hataların karelerini [ $\sum_{i=1}^k e_i^2 = \sum_{i=1}^k (Y_i -$

$\hat{Y}_i)^2]$  minimum (en küçük) yapacak şekilde belirlemektir. Burada regresyon katsayılarının en küçük kareler tahminlerini elde edebilmek için,

$\sum_{i=1}^k (Y - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2$  eşitliğinde  $\hat{\beta}_0$  ve  $\hat{\beta}_1$ 'e göre kısmi türevler alınıp sıfıra eşitlenir. Buradan

$$\frac{\partial q}{\partial \hat{\beta}_0} = \sum_{i=1}^k (-2) [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)] = 0$$

ve

$$\frac{\partial q}{\partial \hat{\beta}_1} = \sum_{i=1}^k (-2) X_i [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)] = 0$$

Eşitlikleri ve

$$\sum_{i=1}^k Y_i = \hat{\beta}_0 n + \hat{\beta}_1 \sum_{i=1}^k X_i$$

$$\sum_{i=1}^k X_i Y_i = \hat{\beta}_0 \sum_{i=1}^k X_i + \hat{\beta}_1 \sum_{i=1}^k X_i^2$$

eşitliklerine ulaşılır (Marquardt, 1963)

$\hat{\beta}_1$  için eşitlik;

$$\hat{\beta}_1 = \frac{n \left[ \sum_{i=1}^k X_i Y_i \right] - \left( \sum_{i=1}^k X_i \right) \left( \sum_{i=1}^k Y_i \right)}{n \left( \sum_{i=1}^k X_i^2 \right) - \left( \sum_{i=1}^k X_i \right)^2} = \frac{\sum_{i=1}^k (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^k (X_i - \bar{X})^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^k Y_i + \hat{\beta}_1 \sum_{i=1}^k X_i}{n} = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (6)$$

olarak bulunur. (6) no'lu eşitlikte  $\bar{X}$  ve  $\bar{Y}$  sırasıyla  $X_i$  ve  $Y_i$  değerlerinin ortalamalarını göstermektedir.

Genelleştirilmiş doğrusal regresyon modeli için en küçük kareler normal denklemi;

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

ve

en küçük kareler tahmin edicisi denklemi

$$\underset{px1}{\mathbf{b}} = \underset{pxp}{(\mathbf{X}'\mathbf{X})^{-1}} \underset{px1}{(\mathbf{X}'\mathbf{Y})}$$

Olarak ifade edilir (Marquardt, 1963)

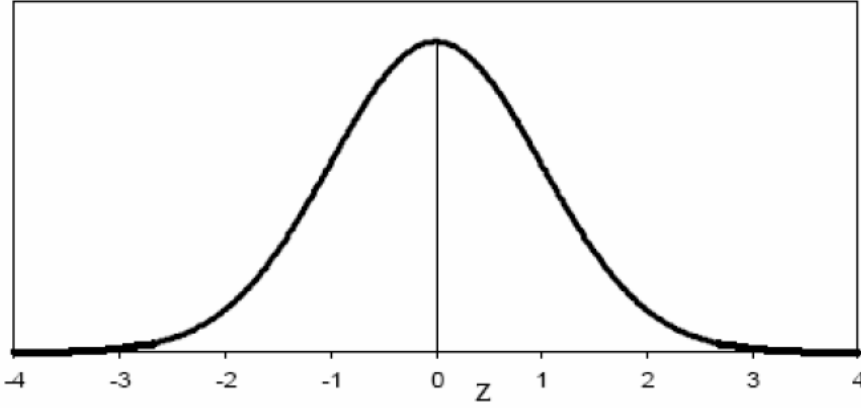
### 3.2.4. En küçük kareler yönteminin varsayımları

En küçük kareler yönteminin değişkenlere ve hata terimine ait bazı varsayımları bulunmaktadır. İyi bir tahmin için en küçük kareler yöntemi, bu varsayımları yerine getirmelidir. Bu varsayımlar aşağıda özetlenmiştir.

- Hata terimi, stokastik (tesadüfi) bir değişkendir. Diğer bir ifade ile hangi değer veya sonucu alacağı önceden bilinmeyen ve aldığı değerler rasgele (rassal)olan bir değişkendir.  $\varepsilon_i$ 'ler sürekli aynı yönde sapma göstermeyip, düzensiz bir dağılım gösterir(Myers, 1990)
- Hata terimi ortalaması sıfırdır.X'in aldığı farklı değerler için  $\varepsilon_i$ ' de artı, eksi ve sıfır gibi çeşitli değerler almaktadır. Bu değerlerin ortalaması sıfıra eşittir (Tatlıdil, 1996).

$$E(\varepsilon_i) = 0, i = 1, 2, \dots, n$$

- Hata terimi normal dağılım gösterir. Bağımsız değişkene ait her bir distürbans değeri, kendi ortalamaları etrafında normal dağılım eğrisi (Şekil 1) şeklinde simetrik bir dağılım gösterir (Raymond, 1990).



Şekil 1 Hata terimi dağılımının grafiği

- Hata terimi sabit varyanslıdır. Bu varsayıma göre, distürens varyansları, bağımsız deęişkendeki deęişmelere göre deęişmeyip sabit kalmaktadır (Tatlídil, 1996).
- Hata terimi arasında otokorelasyon yoktur (Sanford Weisberg, 1985).
- Bağımsız deęişken ile hata terimi arasında bir ilişki yoktur (Raymond, 1990).
- Bağımsız deęişkenler arasında çoklu doğrusal bağlantı yoktur.
- Bağımlı deęişken (Y) ile distürens ( $\epsilon_i$ ) arasında korelasyon yoktur (Raymond, 1990).

Yukarıda belirtilen varsayımlardan herhangi birisinin yerine gelmemesi durumunda En küçük kareler tahmincileri yanlı veya tutarsız olabilmektedir.

### 3.2.5. İki aşamalı en küçük kareler yöntemi

İki aşamalı en küçük kareler yöntemi Theil ile Basman tarafından bağımsız olarak geliştirilmiş olup, sistemin sadece bir denklemine uygulanabilen tek denklem yöntemidir. İki aşamalı en küçük kareler yöntemi aynı zamanda, Genelleştirilmiş en küçük kareler yönteminin özel bir durumu olarak geliştirilmiştir (Koutsoyiannis, 1977) Yöntem, aracı veya yardımcı (instrumental) deęişkenleri kullanır ve açıklayıcı içsel (endogenous) deęişkenler ile yapılan özel bir regresyon analizi olarak nitelendirilir.

Yöntemdeki araç deęişkenler aynı zamanda z deęişkeni olarak adlandırılır ve bu deęişkenler; hata deęişkeni ile ilişkisiz, x deęişkeni ile ilişkili olacaktır.

İki aşamalı en küçük kareler yönteminde 4 tip değişkenden bahsedilir. Bunlar; bağımlı değişken, içsel değişken, dışsal değişken ve aracı değişkendir.

**Bağımlı değişken:** Genellikle  $y$  değişkeni olarak ifade edilen cevap değişkenidir.

**Dışsal değişken:** Genelde bağımsız değişkenler olarak bilinen bu değişkenler,

İki aşamalı en küçük kareler yönteminin hem birinci aşamasında hem de ikinci aşamasında regresyon modellerine dahil edilen değişkenlerdir. Bu değişkenler, regresyonun ikinci aşamasındaki rastgele hata terimleri ile korelasyonlu değildir.

**İçsel değişken:** İçsel değişken, İki aşamalı en küçük kareler regresyonunun birinci aşamasındaki regresyon eşitliğinde bağımlı değişken gibi işlem görür. Bu değişkenler aracı değişkenler ile dışsal değişkenlerden yararlanılarak tahmin edilir. Bu regresyon eşitlikleri yardımı ile tahmin edilen değerler, ikinci aşamadaki regresyon modelinde içsel değişkenlerin orijinal değerleri yerine kullanılır yani orijinal değerler ile yer değiştirir.

**Aracı değişken:** Bir regresyon eşitliğinde, bağımlı değişken olarak alınan her bir içsel değişken için aracı değişkenler ile dışsal değişkenler kullanılarak regresyon eşitliği elde edilir. Diğer bir ifadeyle aracı değişkenler ve dışsal değişkenler alınarak her bir içsel değişkenin değeri, regresyon eşitliğinin birinci aşamasında tahmin edilir tahmin edilen değerler orijinal değerlerle yer değiştirerek ikinci aşamada kullanılır.

### 3.2.6. İki aşamalı en küçük kareler yönteminin varsayımları

- Tahmin edilecek yapısal denklemin hata teriminin bilinen varsayımları (Sıfır ortalamalı, eşit varyanslı ve otokorelasyonsuz) sağlaması gerekir.
- İndirgenmiş form hata terimleri de, eşit varyanslı, sıfır ortalamalı ve otokorelasyonsuz olmalıdır. Aynı zamanda dışsal değişkenlerle de aralarında korelasyon olmamalıdır.
- Dışsal değişkenler arasında çoklu doğrusal bağlantı olmamalıdır.
- Model dışsal değişkenler bakımından doğru kurulmuş olmalıdır.
- Örnek büyüklüğü yapısal modeldeki dışsal değişken sayısından büyük olmalıdır.

### 3.2.7. İki aşamalı en küçük kareler yönteminin parametre tahminleri

İki aşamalı en küçük kareler yönteminin tahmin edicisini geliştirmek için aşağıdaki eşitlik yazılır ( Klein, 1974).

$$Y = Y_1\beta_1 + X_1\gamma_1 + U_1 \quad (7)$$

(7) no'lu eşitlikte yer alan parametreleri doğrudan tahmin etmek üzere, En küçük kareler yönteminin uygulanabilmesi için hata terimi ( $U_1$ ) ile ilişkili olan içsel değişkenin ( $Y_1$ ) olmaması gerekir. Bu durumda En küçük kareler yönteminin uygulanması sonucu tahminler tutarsız olacaktır. Bunun için iki aşamalı en küçük kareler yönteminin uygulanması gerekir.

$Y = Y_1\beta_1 + X_1\gamma_1 + U_1$  eşitliğindeki içsel değişkenler, tüm dışsal değişkenlerin regresyonu şeklinde ifade edebilir.

$Y_1 = X\pi_1 + v_1(\pi_1)$  eşitliğinde,  $\pi_1$  daraltılmış biçim parametresi ise;

$\hat{\pi}_1 = (X'X)^{-1}X'Y_1$  şeklinde bir sonuç elde edilir. Böylece;

$$\hat{Y}_1 = X\hat{\pi}_1$$



ve

$$Y_1 = X(X'X)^{-1}X'Y_1\hat{v}_1$$

$Y_1 = \hat{Y}_1\hat{v}_1$  şeklinde deęer elde edilir.

### 3.2.8. İki aşamalı en küçük kareler tahmincilerinin özellikleri

- 2AEKK büyük örnekler için daha uygun olup, küçük örneklerde sapmalı tahminler verir.
- 2AEKK tahminleri tutarlıdır.
- 2AEKK tahminleri asimptotik etkindir.
- Tam belirlenmiş denklemlerde dolaylı en küçük kareler yöntemi ile aynı sonuçları verir.
- Aşırı belirlenmiş denklemler için idealdir.
- Hesaplanması kolay ve iyi sonuçlar verir.
- Dışsal deęişkenin çok olduđu durumlarda örnek hacminin fazla olması gereklidir.

İki aşamalı en küçük kareler yöntemi ile bulunan tahminler her zaman yansızlık ve en az varyanslılık özelliklerini sağlamayabilseler de tutarlıdır. Dięer bir ifade ile örneklem büyüdükçe yanlışlık azalır ve tahminler giderek anakütledeki gerçek deęere yaklaşır. Bu nedenle küçük örneklerde dikkatli olunmalı, iki aşamalı en küçük kareler yöntemini kullanmadan önce Hausman testi yapıp açıklayıcı deęişkenlerin hata terimi ile ilintili olduđu doğrulanmalıdır.

### 3.2.9. Hausman testi

Regresyon modelinde  $y = x\beta + \varepsilon$  önemli varsayımlardan biri verilen bir  $x$  deęerine karşılık hata teriminin  $\varepsilon$  beklenen deęerinin 0 olmasıdır.  $E(\varepsilon/x) = 0$  Bu varsayımın geçerli olmaması durumunda parametre tahminleri yanlış olacaktır. Hausman'ın geliştirdiđi tanımlama hatası testi ile bu varsayımın geçerliliđi incelenmektedir. Test tanımlama hatası olmadığında, yani sıfır hipotezinin reddedilemediđi durumlarda tutarlı, asimptotik olarak normal ve asimptotik olarak etkin

bir tahmincinin varlığına dayanır. Alternatif hipotez altında diğer bir ifade ile tanımlama hatasının varlığında bu tahminci yanlı ve tutarsız olacaktır. Tanımlama hatasını test edebilmek için tanımlama hatasından etkilenmeyen ikinci bir tahminciye ihtiyaç vardır. Böyle bir tahminci ise sıfır hipotezi altında asimptotik etkin olmayacaktır (Hausman 1978).

Hausman test istatistiği için eşitlik;

$$m = T\hat{Q}^{-1}var[\hat{Q}]^{-1}\hat{q}$$

olarak yazılır. Sıfır hipotezi altında, tanımlama hatası yok iken K bilinmeyen parametre sayısını göstermek üzere m test istatistiğinin dağılımı  $\chi^2_k$  dağılımına uygundur (Hausman 1978).

#### 4. BULGULAR

Çalışmaya toplam 18 değişken dahil edilmiş ve bu değişkenlere ilk önce adimsal regresyon analizi uygulanarak, modele girecek değişkenler belirlenmiştir daha sonra bu değişkenler ile standart regresyon analizi yapılmıştır (Tablo 2). Dokuz değişkene ilaveten, modele girmeyen değişkenler “aracı değişken” olarak alınmış ve İki aşamalı en küçük kareler regresyon analizi yapılarak analiz sonuçları Tablo 2'de özetlenmiştir.

**Tablo 2.**Standart regresyon analizi ile 2AEKK analizi istatistikleri (Dokuz değişken)

| Değişkenler           | Standart regresyon |       | 2AEKK          |       |
|-----------------------|--------------------|-------|----------------|-------|
|                       | $b \pm S_b$        | p     | $b \pm S_b$    | p     |
| Sabit                 | 48.655±1.430       | 0.001 | 60.040±35.877  | 0.095 |
| Yetişkin ölümü        | 0.015±0.002        | 0.001 | 0.053±0.070    | 0.447 |
| Okullaşma             | 0.911±0.140        | 0.001 | 3.705±1.479    | 0.013 |
| HIV/AIDS              | 0.318±0.033        | 0.001 | 0.088±0.419    | 0.834 |
| Gelir kompozisyonu    | 8.403±2.222        | 0.001 | 19.880±74.155  | 0.789 |
| Alkol                 | 0.397±0.090        | 0.001 | 0.308±1.932    | 0.874 |
| Difteri               | 0.039±0.009        | 0.001 | 0.035±0.062    | 0.565 |
| İncelik 5-9 yaş arası | 0.175±0.040        | 0.001 | 0.230±0.212    | 0.278 |
| BKI                   | 0.120±0.028        | 0.001 | 0.088±1.021    | 0.931 |
| Genel harcama         | 0.312±0.105        | 0.001 | 0.812±1.650    | 0.623 |
|                       | $R^2 = \%79.3$     |       | $R^2 = \%27.8$ |       |

Tablo 2'de görüldüğü üzere standart regresyon analizi sonucunda modele dahil edilen bütün değişkenlere ilişkin regresyon katsayıları istatistik olarak anlamlı bulunmuştur ( $p < 0.001$ ). Bu değişkenlerle kurulan modele ilişkin belirleme katsayısı %79.3 olarak bulunmuştur.

İki aşamalı en küçük kareler regresyonu yönteminde ise  $R^2 \%27.8$  olarak bulunmuştur. Standart regresyon analizinde istatistik olarak önemli bulunan regresyon katsayılarından yalnızca okullaşma oranına ait regresyon katsayısı, İki aşamalı en küçük kareler regresyon yönteminde de istatistik olarak anlamlı bulunmuş diğer katsayılar ise anlamlı bulunmamıştır.

**Tablo 3.**Standart regresyon analizi ile 2AEKK analizi istatistikleri (Yedi deęişken)

| Deęişkenler           | Standart regresyon |       | 2AEKK          |       |
|-----------------------|--------------------|-------|----------------|-------|
|                       | $b \pm S_b$        | p     | $b \pm S_b$    | p     |
| Sabit                 | 47.688±1.312       | 0.001 | 80.095±34.664  | 0.022 |
| Yetişkin ölümü        | 0.015±0.002        | 0.001 | 0.061±0.042    | 0.155 |
| Okullaşma             | 0.977±0.142        | 0.001 | 0.390±5.232    | 0.941 |
| HIV/AIDS              | 0.326±0.034        | 0.001 | 0.431±0.433    | 0.321 |
| Gelir kompozisyonu    | 12.192±2.135       | 0.001 | 7.209±29.117   | 0.805 |
| Alkol                 | 0.390±0.009        | 0.001 | 0.115±0.228    | 0.613 |
| Difteri               | 0.424±0.093        | 0.001 | 0.480±0.914    | 0.600 |
| İncelik 5-9 yaş arası | 0.128±0.038        | 0.001 | 0.225±0.259    | 0.385 |
|                       | $R^2 = \%77.6$     |       | $R^2 = \%16.9$ |       |

Aracı deęişken olarak 7 deęişken modele dahil edilerek regresyon analizi yapılmış ve analiz sonuçları Tablo 3'te özetlenmiştir. Tablo 3'te görüldüğü üzere standart regresyon analizi sonucunda modele dahil edilen bütün deęişkenlere ait regresyon katsayıları istatistik olarak anlamlı bulunmuştur. Bu deęişkenlerle kurulan modele ilişkin belirleme katsayısı ( $R^2$ ) %77.6 olarak bulunurken, İki aşamalı en küçük kareler regresyon yönteminde %16.9 olarak bulunmuştur.

**Tablo 4.**Standart regresyon analizi ile 2AEKK analizi istatistikleri (Beş deęişken)

| Deęişkenler        | Standart regresyon |       | 2AEKK          |       |
|--------------------|--------------------|-------|----------------|-------|
|                    | $b \pm S_b$        | p     | $b \pm S_b$    | p     |
| Constant (Sabit)   | 50.523±1.242       | 0.001 | 22.503±237.817 | 0.925 |
| Yetişkin ölümü     | 0.017±0.002        | 0.001 | 0.007±0.173    | 0.967 |
| Okullaşma          | 0.954±0.145        | 0.001 | 5.055±32.721   | 0.877 |
| HIV/AIDS           | 0.355±0.035        | 0.001 | 0.750±2.158    | 0.728 |
| Gelir kompozisyonu | 9.589±2.175        | 0.001 | 93.564±178.796 | 0.601 |
| Alkol              | 0.034±0.010        | 0.001 | 0.235±1.575    | 0.882 |
|                    | $R^2 = \%86.6$     |       | $R^2 = \%3.3$  |       |

Aracı deęişken olarak 5 deęişken modele dahil edilerek regresyon analizi yapılmış ve analiz sonuçları Tablo 4'te özetlenmiştir. Tablo 4'te görüldüğü üzere standart regresyon analizi sonucunda modele dahil edilen bütün deęişkenlere ait regresyon katsayıları istatistik olarak anlamlı bulunmuştur. Bu deęişkenlerle kurulan modele ilişkin belirleme katsayısı ( $R^2$ ) %86.6 olarak bulunurken, İki aşamalı en küçük kareler regresyon yönteminde %3.3 olarak bulunmuştur.

**Tablo 5.**Standart regresyon analizi ile 2AEKK analizi istatistikleri (Üç değişken)

| Değişkenler      | Standart regresyon |       | 2AEKK         |       |
|------------------|--------------------|-------|---------------|-------|
|                  | $b \pm S_b$        | p     | $b \pm S_b$   | p     |
| Constant (Sabit) | 53.012±1.194       | 0.001 | 14.608±31.172 | 0.640 |
| Yetişkin ölümü   | 0.019±0.002        | 0.001 | 0.010±0.063   | 0.879 |
| Okullaşma        | 1.504±0.105        | 0.001 | 5.328±2.960   | 0.073 |
| HIV/AIDS         | 0.358±0.036        | 0.001 | 0.741±1.048   | 0.480 |
|                  | $R^2 = \%85.2$     |       | $R^2 = \%5.8$ |       |

Aracı değişken olarak 3 değişken modele dahil edilerek regresyon analizi yapılmış ve analiz sonuçları Tablo 5’te özetlenmiştir. Tablo 5’te görüldüğü üzere standart regresyon analizi sonucunda modele dahil edilen bütün değişkenlere ait regresyon katsayıları istatistik olarak anlamlı bulunmuştur. Bu değişkenlerle kurulan modele ilişkin belirleme katsayısı ( $R^2$ ) %85.2 olarak bulunurken, İki aşamalı en küçük kareler regresyon yönteminde %5.8 olarak bulunmuştur.

**Tablo 6.**Standart regresyon analizi ile 2AEKK analizi istatistikleri

| Değişkenler           | Standart regresyon |       | 2AEKK         |       |
|-----------------------|--------------------|-------|---------------|-------|
|                       | $b \pm S_b$        | p     | $b \pm S_b$   | p     |
| Constant (Sabit)      | 57.989±1.700       | 0.001 | 76.533±26.981 | 0.005 |
| Beş yaş altı ölüm     | 0.003±0.002        | 0.128 | 0.112±0.092   | 0.226 |
| Kızamık               | 5.418±0.001        | 0.025 | 0.001±0.001   | 0.446 |
| Hepatit B             | 0.030±0.016        | 0.065 | 0.432±0.658   | 0.512 |
| Çocuk Felci           | 0.052±0.016        | 0.002 | 0.444±0.645   | 0.492 |
| İncelik 1-9 yaş arası | 0.197±0.80         | 0.015 | 1.039±2.263   | 0.646 |
| GSYİH                 | 0.001±0.001        | 0.001 | 0.001±0.006   | 0.929 |
| Popülasyon            | 3.888±0.001        | 0.294 | 4.075±0.00    | 0.236 |
| Genel harcama         | 1.034±0.186        | 0.001 | 1.237±1.710   | 0.470 |
|                       | $R^2 = \%25$       |       | $R^2 = \%1.4$ |       |

Modele dahil edilen aracı değişkenler ile tahmin edici değişkenler yer değiştirilerek regresyon analizi yapılmış ve sonuçlar Tablo 6’ da özetlenmiştir. Tablo 6’ da görüldüğü üzere, standart regresyon analizinde Hepatit B ve Popülasyon değişkenleri dışında diğer değişkenlere ait regresyon katsayıları istatistik olarak önemli bulunmakla birlikte, belirleme katsayısı %79.3’ten %25’e düşmüştür. Benzer şekilde, iki aşamalı en küçük kareler regresyonunda ise belirleme katsayısı %27.8’den %1.4’e düşerken, regresyon sabiti dışında, değişkenlere ait katsayılar istatistik olarak anlamlı bulunmamıştır.

## 5. TARTIŞMA VE SONUÇ

Çalışmaya dahil edilen 18 değişkenin yarısının aracı değişken olarak modele dahil edilmesi ile yapılan standart regresyon analizinde, yaklaşık %79 belirleme katsayısı elde edilirken, İki aşamalı en küçük kareler regresyon yönteminde bu değer %27.8'e düşmüştür. Diğer yandan standart regresyon analizinde tüm değişkenlere ait katsayılar istatistik olarak anlamlı bulunurken, İki aşamalı en küçük kareler regresyon yönteminde yalnızca “Okullaşma” değişkeni istatistik olarak anlamlı bulunmuştur. Değişken sayısının 7, 5 ve 3 olarak azaltılması ile kurulan modellerde benzer durum gözlenmiştir. Açıklayıcı değişken sayısının azaltılması ile önemli bir değişim göstermeyen bu farklılık, aracı değişkenlerin modele dahil edilmesine atfedilebilir. Zira, standart regresyon analizi modellerinde bağımlı değişkene ait hata terimlerinin bağımsız değişkenler ile ilişkisiz olması gerektiği varsayımı bulunmaktadır. Bu varsayımın sağlanmadığı durumda, standart regresyon analizi yöntemine ait tahminlerde yanlılığı gidermek üzere İki aşamalı en küçük kareler regresyonu aracı değişken kullanmaktadır.

Bound ve ark. (2001), modelin doğrusal olması ve açıklayıcı değişkenlerde ölçüm hatası olması durumunda, aracı değişkenlerin modele dahil edilmesinin iyi sonuçlar verdiğini belirtmektedir. Benzer şekilde Schennach (2007), açıklayıcı değişkenlerde ölçüm hatası olması durumunda, aracı değişken kullanımının, bu problemin çözümünde en yaygın kullanılan seçenek olduğunu belirtmiştir.

İçsellik problemi (endogeneity problem) olduğunda, standart regresyon analizi tahminleri tutarsız (inconsistent) olmakta yani tahmin değerleri popülasyon parametresine yakınsamamakta ve tahminler yanlı olmaktadır. İçsellik problemi ise eşzamanlı nedensellik (simultaneouscausality), açıklayıcı değişkenlerin hata terimi ile korelasyonlu olması, açıklayıcı değişkenlerde ölçüm hatası olması ve değişkenlerin modelden atılması ile ortaya çıkmaktadır (Bascle ve Paris, 2008).

Bağımlı değişkeni etkileyen açıklayıcı değişkenlerden herhangi birisinin, diğer açıklayıcı değişkenler ile yüksek korelasyonlu olması gerekçesiyle modelden çıkarılması durumunda, içsellik problemi ortaya çıkmakta ve standart regresyon analizi tahminleri yanlı olabilmektedir (Wooldridge, 2006). Bu problemin çözümü için aracı

değişken kullanımı, (diğer bir ifade ile 2AEKK yöntemi) uygun bir seçenek olmaktadır (Bascle ve Paris, 2008).

Eşzamanlı nedensellik (simultaneous causality) modelde iki yönlü etki yani feed-back halkası olduğu durumda, diğer bir ifade ile aynı değişkenin hem sebep değişkeni hem de sonuç değişkeni olarak modelde yer alması durumunda ortaya çıkmaktadır. Böylece, açıklayıcı değişkenler ile hata terimleri korelasyonlu olmakta ve standart regresyon analizi tahminleri yanlı ve tutarsız olmaktadır. Bu durumda aracı değişken kullanımı, diğer bir ifade ile 2AEKK yöntemi uygun bir seçenek olmaktadır (Bascle ve Paris, 2008).

Ebbes (2004), modele dahil edilecek herhangi bir aracı değişkenin zayıf aracı değişken olup olmadığının test edilemediğini belirterek, gözlenen herhangi bir aracı değişkenin gerçek dışsal değişken olmaması durumunda, 2AEKK tahminlerinin yanlı olacağını ifade etmiştir. Bu durumda aracı değişkenin, içsel değişken üzerinden dolaylı etkisinin yanı sıra, bağımlı değişkene olan doğrudan etkisinin de alınması gerektiğini vurgulamıştır. Benzer şekilde Ebbes (2004), içsel aracı değişkenin, regresyon eşitliğinde hata terimi ile korelasyonlu olan değişken olduğunu belirterek, bu durumda, adı geçen değişkenin bağımlı değişkene doğrudan etkisinin olduğunu vurgulamış ve hata terimi ile korelasyonlu olacağından şüphe edilen herhangi bir aracı değişkenin, ana regresyon modeline dahil edilmesi gerektiğini ve ardından, bağımlı değişken üzerine etkisinin olup olmadığını test etmede Wald testinin kullanılması gerektiğini ifade etmiştir.

Oczkowski (2003), 2AEKK yönteminin, Maksimum Olasılık yöntemine göre numerikoptimizasyon algoritmalarını gerektirmediğini ve doğrusal olmayan interaksiyon etkilerini belirleyebildiğini vurgularken, Maksimum Olasılık tahmin edicilerinin bütün ilişkiler için 2AEKK yöntemine göre daha etkin olduğunu belirtmiştir.

Nagler (1999), 2AEKK yönteminin ilk aşamasında X için yapılan tahminlerin zayıf olması durumunda ikinci aşamada yapılan tahminlerin kesinlik (precision) bakımından yeterli olmayacağını ve tahminlerin Standart regresyon analizi tahminleri ile aynı olacağını vurgulamıştır.

İki aşamalı en küçük kareler regresyon yöntemi, özellikle yapısal eşitlik modellerinde kullanılmakta ve standart regresyon analizi yönteminin genişletilmiş hali olarak düşünülmektedir. Bağımlı değişkene ait hata terimlerinin bağımsız değişkenler ile ilişkili olması durumunda diğer bir ifade ile standart regresyon analizi yönteminin varsayımlarının sağlanmaması durumunda kullanılan analiz yöntemidir. Bowden ve Turkington (1984). Bununla birlikte modelde feed-back ya da geri bildirim halkası bulunması durumunda da iki aşamalı en küçük kareler regresyonu kullanılabilir. Yapısal eşitlik modellerinde en yüksek olabilirlik yöntemi kullanıldığında, katsayıların tahmini için alternatif yöntem olarak da kullanılabilir.

Sonuç olarak, veri setinde içsellik problemi, bağımsız değişkenlerin hatalı olarak ölçülmüş olması ve modelde feed-back halkasının (aynı değişkenin hem sebep değişkeni hem de sonuç değişkeni olması) olması durumunda iki aşamalı en küçük kareler regresyon yönteminin, standart regresyon analizi yöntemine alternatif olarak kullanılabilir.



## KAYNAKLAR

- Angrist J, Graddy K, Imbens G. The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Model swith an Application to the Demand for Fish. *Rev Econ Stud.* 2000;67:499-527.
- Blackburn M, Neumark D. Unobserved Ability, Eciency Wages, and Interindustry Wage Dierentials , *QJ Econ*, 1992;107:1421-36.
- Card D. Using Geographic Variation in College Proximityto Estimatethe Return to Schooling in Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp 1995:125-85.
- Christophides LN, Grant EK, Swidinsky R. Toronto: University of Toronto Press. 1995:201-22.
- Emine Ç. Eşanlı denklem modeli ile VAR modelinin öngörü başarısı açısından karşılaştırılması: Türkiye Örneği, Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Anabilim Dalı Yüksek Lisans Tezi, Aralık. 2007:15-6.
- Graddy K. Testing for Imperfect Competition at the Fulton Fish Market. *Rand J Econ* 1995; 26:75-92.
- Greene WH. *Econometric Analysis*. PrenticeHall, New Jersey, 2000;5(4):125-56.
- Gülizar SY, Sacit HK. Kamu Sağlık Harcamalarının Etkililiği: Panel Veri Analizi, Adnan Menderes Üniversitesi. *Sosyal Bilimler Enstitüsü Dergisi.* 2012;3(2):85-110.
- Harun T, Sabiha O. Harcama vergi geliri hipotezi Türkiye örneği, *İktisadi ve İdari Bilimler Dergisi.* 2006;20(2):15-22.
- Hausmann JA. Specification Test in Econometrics. *Econometrica* 1978;46(6):98-125.
- İnternet Kaynağı 1. Eşanlı Denklemler Modellerin Çözüm Yöntemler I: Matrisiz Çözüm: Dolaylı EKKY, 2 Aşamalı EKKY, Sınırlı Bilgi ile Eçby [İnternet]. 2019 [ET: 10.06.2019]. Erişim adresi: <http://debis.deu.edu.tr/userweb//hamdi.emec/bolum17.pdf>
- İnternet Kaynağı 2. [İnternet]. 2019 [ET: 10.04.2019]. Erişim adresi: <http://mlr.cs.umass.edu/ml/machine-learning-databases/thyroiddisease/hypothyroid>.
- Klein LR. *A Text Book of Econometrics*, Prentice Hall, Inc, New Jersey 1974:85-99.
- Koutsoyiannis A. *Theory of Econometrics*, Second Edition, Harper Row Publishers, Ontario, 1977:124-40.
- Koutsoyiannis A, *Ekonometri Kuramı*, Verso Yayıncılık, Ankara, 1989:76-125.
- Marquardt DW. An Algorithm For Least-Squares Estimation of Nonlinear Parameters. *J Soc Ind App Math.* 1963;11:431-41.

Montgomery DC, Peck EA. Introduction to Linear Regression Analysis. John Wiley&Sons, Inc, Canada. 1992:85-120.

Mroz TA. The Sensitivity of an Empirical Model of Married Womens Hours of Work to Economic and Statistical Assumptions. *Econometrica*. 1987;55:765-99.

Oczkowski E. 2SLS estimation of Structural Equation Models [İnternet]. 2018 [ET: 05.03.2019]. Erişim adresi: <http://csusap.csu.edu.au/~eoczkwows/home.htm>

Pazarlıoğlu MV ve Akkaya Ş. *Ekonometri II*, Erkan Matbaacılık, İstanbul. 1998:324-6.

Raymond HM. *Classical and Modern Regression with Applications*, 2.Ed. Boston: PWS-Kent. 1990:95-110.

Sanford W. *Applied Linear Regression*, John Wiley&Sons, Ltd, Newyork, 2.Ed. 1985:65-106.

Tatlıdil H. *Uygulamalı Çok Değişkenli İstatistiksel Analiz*. Cem Ofset Ltd. Şti. Ankara. 1960;11-85.

Vural A. *Aykırı Değerlerin Regresyon Modellerine Etkileri ve Sağlam Kestiriciler*. Marmara Üniversitesi Sosyal Bilimler Enstitüsü. Yüksek Lisans Tezi, İstanbul 2007.

Bascle G, Paris HEC. Controlling for endogeneity with instrumental variables in strategic management research, *Strategic Organization* 2008;6(3): 285-327.

Bound J, Brown C, Mathiowetz N. Measurement Error in Survey Data, JJ. Heckman and EE. Leamer (eds) *Handbook of Econometrics*, Amsterdam: North-Holland 2001:370-843.

Schennach SM. Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models, *Econometrica*, 2007;75(1):201-39.

Wooldridge JM. *Introductory Econometrics: A Modern Approach*, 3.Ed. Mason, OH: Thomson-South Western, 2006:35-85.

Ebbes Peter. Latent instrumental variables. Unpublished Ph. D. thesis, University of Groningen, Groningen, 2004:85-110.

Roger J. Bowden Darrell A. *Turkington Instrumental variables* Cambridge University press Cambridge, 1984:100-232.

Oczkowski E. Two stage Least squares (2SLS) and Structural Equation Models (İnternet). 2003 [ET: 05.06.2019]. <http://csusap.csu.edu.au/eoczkwows/home.htm>.

Nagler J. Notes on Simultaneous Equations and Two Stage Least Squares Estimates, Class Notes. New York University, New York, NY. April 1999:1-10.

## ÖZGEÇMİŞ

Fırat DEMİR, Hakkari/Yüksekova'da 1988'de doğdu. İlkokul eğitimini Altınoluk köyünde, orta ve lise öğrenimini Yüksekova'da tamamladı. 2014 yılında Van Yüzüncü Yıl Üniversitesi, Sağlık Yüksekokulu, Hemşirelik Bölümü'nden mezun oldu. 2015 yılında Sağlık Bilimleri Üniversitesi Van Bölge Eğitim ve Araştırma Hastanesinde göreve başladı. 2015 yılında Van Yüzüncü Yıl Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı'nda yüksek lisans öğrenimine başladı. Evli ve iki çocuk babasıdır.



## EKLER

### EK 1. Etik Kurul Raporu



T.C.  
YÜZÜNCÜ YIL ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ




### BİYOİSTATİSTİK ANABİLİM DALI KURUL KARARI


15.02.2019 tarihinde toplanan YYÜ Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı Kurulu; Yüksek Lisans Öğrencisi Fırat DEMİR'in, "İki Aşamalı En Küçük Kareler Yöntemi ve Uygulaması" başlıklı Yüksek Lisans tez çalışmasının konusu itibariyle; bir klinik araştırma olmaması, yöntem incelemesi olarak, serbest erişimli kaynaklardan sağlanacak verileri kullanacak olması, elde edilecek sonuçlar için yalnızca istatistik değerlendirmeleri ve yorumlamaları içerecek olması gibi nedenlerle,

Etik Kurulla doğrudan ilişkisinin olmadığı kanaatine varmıştır.

  
Prof. Dr. Sıddık KESKİN



  
Prof. Dr. Hüseyin Avni ŞAHİN

  
Dr. Öğr. Üyesi Can ATEŞ

  
Dr. Öğr. Üyesi Sinemis ÇETİN DAĞLI

## EK 1.Tez Orijinallik Raporu

### EK 1.Tez Orijinallik Raporu

|  |   |  |
|--|---|--|
|   | <p>T.C.<br/>VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ<br/>Sağlık Bilimleri Enstitüsü</p> |   |
| <b>YÜKSEK LİSANS TEZİ ORJİNALLİK RAPORU</b>  |   |  |
|  |   | Tarih:17/06/2019Tez  |
| <p><b>Başlığı/Konusu:</b><br/><b>İki Aşamalı En Küçük Kareler Yöntemi</b></p> <p>Yukarıda başlığı/konusu belirlenen tez çalışmamın Kapak sayfası, Giriş, Ana bölümler ve Sonuç bölümlerinden oluşan toplam <b>11 sayfa</b>lık kısmına ilişkin, 17/06/2019 tarihinde şahsım tarafından <b>Turnitin</b> intihal tespit programından aşağıda belirtilen filtreleme uygulanarak alınmış olan orijinallik raporuna göre, tezimin <b>benzerlik oranı %10 (on)</b>'dur.</p> <p>Uygulanan filtreler aşağıda verilmiştir:</p> <ul style="list-style-type: none"><li>- Kabul ve onay sayfası hariç,</li><li>- Teşekkür hariç,</li><li>- İçindekiler hariç,</li><li>- Simge ve kısaltmalar hariç,</li><li>- Gereç ve yöntemler hariç,</li><li>- Kaynakça hariç,</li><li>- Alıntılar hariç,</li><li>-Tezden çıkan yayınlar hariç,</li><li>- 7 kelimeden daha az örtüşme içeren metin kısımları hariç (Limit match size to 7 words)</li></ul> <p>Van Yüzüncü Yıl Üniversitesi Lisansüstü Tez Orijinallik Raporu Alınması ve Kullanılmasına İlişkin Yönergeyi inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim. Gereğini bilgilerinize arz ederim.</p> <p style="text-align: right;"><br/>Firat DEMİR</p> |   |  |
| <b>Öğrencinin Adı Soyadı</b>   | :   | Firat DEMİR  |
| <b>Anabilim Dalı</b>   | :   | Biyoistatistik AD  |
| <b>Öğrenci No</b>  | :   | 159302024  |
| <b>Programı</b>  | :   | <input checked="" type="checkbox"/> Yüksek Lisans <input type="checkbox"/> Doktora   |
| <b>DANIŞMAN ONAYI</b><br>Prof. Dr. Sıddık KESKİN<br>UYGUNDUR<br>  |   | <b>ENSTİTÜ ONAYI</b><br>Hacı SAHİN AÇYDINYURT<br>UYGUNDUR<br> |