

PAPER DETAILS

TITLE: REGRESYON ANALİZİNDE DOĞRUSALA DÖNÜSTÜRME YÖNTEMLERİ VE BİR UYGULAMA

AUTHORS: Selahattin YAVUZ

PAGES: 165-179

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/30223>

REGRESYON ANALİZİNDE DOĞRUSALA DÖNÜŞTÜRME YÖNTEMLERİ VE BİR UYGULAMA

Selahattin YAVUZ^(*)

Özet: Regresyon analizinde verilere en iyi matematiksel modeli uydurmak şüphesiz çok önemlidir. Bu model doğrusal bir model olabildiği gibi çoğu kez doğrusal olmayabilir. Doğrusal olmayan modeller herhangi bir dönüşümle doğrusal yapılabilen ve herhangi bir dönüşümle doğrusal yapılamayan modeller olmak üzere ikiye ayrılır. Bu çalışmada herhangi bir dönüşümle doğrusal yapılabilen modeller üzerinde durulmuştur. Doğrusal olmayan bu modellerden logaritmik modeller, hiperbolik modeller ve kareköklü modeller bazı basit matematiksel dönüşümler yapılarak doğrusal hale dönüştürülmüştür. Dönüşümlerle doğrusal hale dönüştürülen bu modellere, 1960-2006 yılları aralığında temel sağlık hizmetleri istatistiklerinden “Türkiye’de bir diş hekimine düşen kişi sayısı” verileri uygulanmış olup en iyi model veya modeller tahmin edilmiştir.

Anahtar Kelimeler: Regresyon, Doğrusal Olmayan Model, Doğrusala Dönüştürme Yöntemi

Abstract: It is very important to adapt the best mathematicial model fort he data in regression analysis. Most of the time this model may be nonlinear as well as linear. Nonlinear models can be sorted in two groups such as models which can be changed into linear models by any kind of transformation and models which can not be changed into linear models by any kind of transformation. In this study models which can be changed into linear models by any kind of transformation has been examined. The models such as logarithmic model, hyperbolic model and square root model have been transformed to linear by using same easy mathematical transformation. The best model or models have been predicted by using the data one of the basic health statistics between 1960 and 2006 “number of people for a dendist in Turkey”.

Key Words: Regression, Nonlinear Model, Method of Transformation to Linear

I. Giriş

Regresyonda bir bağımlı değişken ve bir ya da daha fazla sayıda bağımsız değişken vardır. Bir bağımlı değişken ve bir bağımsız değişken varken bu iki değişken arasındaki ilişkiyi belirleyecek model her zaman doğrusal olmayabilir. Fakat birçok araştırmacı iki değişken arasında doğrusal ilişki varmış gibi değişkenler için doğrusal regresyon modeli kurar. Araştırmacıların doğrusal regresyon modelini tercih etmelerinin en önemli nedenlerinden biri kuşkusuz doğrusal regresyon modelinin parametrelerini tahmin etmenin diğer modellere göre daha kolay olmasındandır. Eğer gerçekten bu iki değişken arasında doğrusal ilişkiden söz edilemiyorsa veya doğrusal ilişki çok zayıf ise bu iki değişken için kurulan doğrusal model ve bu model için yapılan testler ve analizler yanıltıcı olabilecektir. Hâlbuki en basit şekilde bu iki değişken için

^(*)Öğr.Gör.Dr.Kahramanmaraş Sütçü İmam Üniversitesi Elbistan MYO

elde edilen veriler bir serpm diyagramında gösterildiğinde bu iki değişken arasındaki ilişkinin doğrusal değil de eğrisel olabileceği açıkça görülebilecektir.

Doğrusal regresyon yönteminin, verilere en iyi doğru ya da yüksek dereceden polinom veya bilinen fonksiyonların doğrusal birleşimlerinden oluşan fonksiyonları uydurmada çok yararlı teknikler ortaya koyduğu bir gerçektir. Ancak bağımlı ve bağımsız değişkenler arasındaki ilişkinin doğrusal olmadığı durumlarla da karşılaşmak mümkündür. Doğrusal olmayan bu modellerden bazıları basit matematiksel dönüşümlerle doğrusallaştırılabilir. Doğrusala hale dönüştürülen bu modeller ile ilgili testler ve analizler aynen doğrusal regresyon analizinde olduğu gibi yapılır (Johnston, 1981: 162).

II. Regresyon Analizi

Regresyon analizi, aralarında sebep-sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi, o konu ile ilgili tahminler (estimation) ya da kestirimler (prediction) yapabilmek amacıyla regresyon modeli olarak adlandırılan matematiksel bir model ile karakterize eden bir istatistik analiz tekniğidir. Ayrıca matematiksel model bulunduktan sonra bu modelin yeterli olup olmadığını kontrol etmek için gerekli olan analizler ve testlerdir. (Şahinler, 2000: 57 73).

Herhangi bir dönüşümle doğrusal kılınabilen modellere geçmeden önce doğrusal regresyon analizi ve ilgili konular kısaca incelenecektir.

A. Doğrusal Regresyon Analizi

Birçok istatistik çalışmada olduğu gibi, regresyon analizinde de anakütle verilerinin tümü yerine bu anakütleden seçilen örnek verileriyle analiz yapılır. Daha sonra elde edilen sonuçlar anakütledeki ilişkinin tahmininde kullanılır. Anakütle ve örnek verileriyle yapılan istatistik araştırmalarda tekniklerinin uygulanmasında farklılık yoktur. Ancak teknikler uygulandıktan sonra örnekleme teorisinden yararlanılarak anakütle parametrelerinin testleri ve tahminleri yapılır (Orhunbilge, 1996: 11).

Y bağımlı değişken ve X bağımsız değişken olmak üzere bu iki değişken arasındaki doğrusal regresyon modeli,

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad (1)$$

biçimindedir. Burada α ve β modelin parametreleri, ε_i ise şansa bağlı hata değerini göstermektedir. Bu model örnek veriler için,

$$Y_i = a + bX_i + e_i \quad (2)$$

olarak ifade edilir. Burada a ve b sırasıyla α ve β parametrelerinin tahminleyicileri olan istatistiklerdir (Yıldız ve Bircan, 1999: 218).

- En Küçük Kareler Metodu İle Parametrelerin Tahmini

Bir serpm diyagramındaki verilere en iyi uyan doğruyu bulmak için veri noktalarının doğrusal regresyon denkleminde olan sapmalarının, diğer bir deyişle gerçek Y değerleri ile doğru üzerinde yer alan teorik $\hat{Y}$ değerleri

arasındaki farklar olan hataların kareleri toplamının minimize eden doğru seçilir (Gürsaka, 2002: 311). Bu durum hata kareleri toplamının minimum olması anlamına gelir. Yani,

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - a - bX_i)^2 = \min. \quad (3)$$

olması gerekir. Bu yöntemle en küçük kareler (EKK) metodu denilmektedir. Bu fonksiyonun minimum olabilmesi için a ve b parametrelerine göre birinci dereceden türevleri bulunup sıfıra eşitledikten sonra gerekli sadeleştirme işlemleri yapıldığında,

$$\begin{aligned} \sum Y_i &= na + b \sum X_i \\ \sum X_i Y_i &= a \sum X_i + b \sum X_i^2 \end{aligned} \quad (4)$$

normal denklemler adı verilen denklem sistemi elde edilir. Bu denklem sistemi sayesinde a ve b parametreleri kolaylıkla bulunabilir (Karagöz, 1999: 198). Bu parametreler bulunduğu regresyon modeli,

$$\hat{Y} = a + bX \quad (5)$$

biçiminde yazılır.

- Tahminin Standart Hatası

Regresyon analizinde gerçek ve tahmini Y değerleri arasındaki farklar, hataları meydana getirmektedir. Bağımlı değişken ile bir tek bağımsız değişken kullanıldığında tahminin standart hatası,

$$S_{YX} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}} \quad (6)$$

eşitliği ile bulunur (Kohler, 1988: 591).

- Katsayıların Anlamlılık Testi

Bir bağımlı ve bir bağımsız değişkenli model için α ve β parametreleri test edilecektir. Bilindiği gibi α 'nın kestirimi a olan regresyon sabiti ve b ise β 'nin kestirimi olup regresyon katsayısıdır. Regresyon modelindeki regresyon katsayısına ilişkin yapılan test, regresyon doğrusunun anlamlılığını da test etmektedir (Yüzer, 2004: 264). Parametrelerin testi için hipotezler,

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

biçiminde oluşturulur. Katsayıların anlamlılık testi için t dağılışı kullanılır. Test istatistiği,

$$t_h = \frac{b - \beta}{S_b}, \quad \left[S_b = \sqrt{\frac{S_{YX}^2}{\sum (X_i - \bar{X})^2}}; \quad S_{YX}^2 = \frac{\sum (Y_i - \hat{Y})^2}{n - 2} \right] \quad (7)$$

eşitliği ile hesaplanır. Tek yönlü testte t_c ile gösterilen tablo değeri $t_{\alpha(n-2)}$, iki yönlü testte ise $t_{\alpha/2(n-2)}$ olarak belirlenir. Yapılan hesaplamalar sonucunda $t_h > t_c$ ise H_0 hipotezi reddedilir (Yıldız ve Bircan, 1999: 219).

- Belirlilik Katsayısı

Regresyonda amaçlardan birinin bağımlı değişkendeki varyansın kaynağını araştırmaktır. Regresyon denkleminin verilere olan uyumunun sağlanıp sağlanmadığının bir göstergesi de açıklanabilen değişimin toplam değişime olan oranıdır. Bu oran belirlilik katsayısı olarak adlandırılır ve bağımlı değişkendeki değişimin ne kadarının bağımsız değişken tarafından açıklanabildiğini gösterir. Korelasyon katsayısının karesine eşit olan belirlilik katsayısı R^2 olmak üzere,

$$0 \leq R^2 \leq 1$$

biçimindedir (Yüzer, 2004: 272).

Bağımsız değişkenin bağımlı değişkenin varyansının yüzde kaçını açıkladığının göstergesi olan belirlilik katsayısı (determinasyon katsayısı),

$$R_{YX}^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} \quad \text{veya} \quad R_{YX}^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} \quad (8)$$

eşitliklerinden herhangi biri ile bulunabilir. Bu katsayı regresyonda önemli bir kavramdır. Bu değer 1'e yakın olması bağımlı değişkendeki varyansın büyük bir kısmını modeldeki bağımsız değişkenin açıkladığı sonucuna varılır (Ünver, 1996: 283).

- Doğrusal Hale Dönüştürülebilen Modeller

Logaritmik Dönüşümlü Modeller

İşletmecilik ve iktisat uygulamalarında doğrusal modelin daha sık rastlanan bir alması log- doğrusal modeldir. Bu kalıpta doğrusal ilişkinin, değişkenlerin kendi aralarında değil logaritmaları arasında olduğu ileri sürülür. Logaritmik dönüşümler, tam logaritmik dönüşümler ve yarı logaritmik dönüşümler olarak ikiye ayrılır. Tam logaritmik dönüşümde X bağımsız değişken ve Y bağımlı değişken olmak üzere her iki değişkenin logaritması alınır. Bu model,

$$Y = aX^b \quad (9)$$

biçiminde olsun. Her iki tarafın logaritması alınarak bu eğrisel ifade,

$$\log Y = \log a + b \log X \quad (10)$$

biçiminde doğrusal hale dönüştürülmüş olur (Newbold, 2002: 622). Burada $\log Y = Y^*$, $\log a = a_1$ ve $\log X = X^*$ denildiğinde model,

$$Y^* = a_1 + bX^* \quad (11)$$

doğrusal regresyon denklemine dönüşmüş olur. Bu modelin normal denklemleri,

$$\begin{aligned}\sum \log Y &= n \log a + b \sum \log X \\ \sum \log Y \log X &= \log a \sum \log X + b \sum (\log X)^2\end{aligned}\quad (12)$$

olarak elde edilir. Regresyon denkleminin standart hatası,

$$S_{\log Y \log X} = \sqrt{\frac{\sum (\log Y - \log Y')^2}{n-2}} \quad (13)$$

biçimindedir. Korelasyon katsayısı ise,

$$R_{\log Y \log X} = \sqrt{1 - \frac{\sum (\log Y - \log Y')^2}{\sum (\log Y - \log Y)^2}} \quad (14)$$

olarak ifade edilir (Orhunbilge, 1996: 61).

Yarı logaritmik dönüşümde ise bağımlı veya bağımsız değişkenden sadece birinin logaritması alınır. Diğer değişken olduğu gibi modele girer. Bu modeller ise

$$\log Y = a + bX \text{ ve } Y = a + b \log X \quad (15)$$

biçimindedir. Bu modellerin normal denklemleri, denklemin standart hatası ve korelasyon katsayısı ise tam logaritmik dönüşümde olduğu gibi bulunabilir.

Hiperbolik (Ters) Dönüşümlü Modeller

Hiperbolik dönüşümler, tam hiperbolik dönüşüm ve yarı hiperbolik dönüşüm olarak ikiye ayrılır. Tam hiperbolik dönüşümde x bağımsız değişken ve y bağımsız değişken olmak üzere her iki değişkenin tersi alınır. Ve 1/X ve 1/Y olarak tanımlanır. Böylece,

$$\frac{1}{Y} = a + \frac{b}{X} \quad (16)$$

modeli elde edilir. Burada $\frac{1}{Y} = Y^*$ ve $\frac{1}{X} = X^*$ olarak seçildiğinde model

klasik doğrusal model olan $Y^* = a + bX^*$ modeline dönüşmüş olur. Bu modelin normal denklemleri ise,

$$\begin{aligned}\sum \frac{1}{Y} &= na + b \sum \frac{1}{X} \\ \sum \left(\frac{1}{Y} \frac{1}{X}\right) &= a \sum \frac{1}{X} + b \sum \left(\frac{1}{X}\right)^2\end{aligned}\quad (17)$$

olarak elde edilir. Regresyon denkleminin standart hatası,

$$S_{\frac{1}{Y} \frac{1}{X}} = \sqrt{\frac{\sum (1/Y - 1/Y')^2}{n-2}} \quad (18)$$

biçimindedir. Korelasyon katsayısı ise,

$$R_{\frac{1}{Y} \frac{1}{X}} = \sqrt{1 - \frac{\sum (1/Y - 1/Y')^2}{\sum (1/Y - \overline{1/Y})^2}} \quad (19)$$

olarak ifade edilir (Orhunbilge, 1996: 62).

Yarı hiperbolik dönüşümde ise bağımlı veya bağımsız değişkenden sadece birinin tersi alınır. Diğer değişken olduğu gibi modele girer. Bu modeller ise,

$$\frac{1}{Y} = a + bX \text{ ve } Y = a + \frac{b}{X} \quad (20)$$

biçimindedir. Bu modellerin normal denklemleri, denklemin standart hatası ve korelasyon katsayısı ise tam hiperbolik dönüşümde olduğu gibi bulunabilir.

Ayrıca bağımlı değişkenin logaritmali ve bağımsız değişkenin ise tersi alınmış hali olan model bir model daha vardır. Bu modele logaritmali ters model denir. Bu model,

$$\log Y = a - b\left(\frac{1}{X}\right) \quad (21)$$

biçimindedir. Burada $\log Y = Y^*$ ve $1/X = X^*$ olarak seçildiğinde doğrusal model olarak bilinen $Y^* = a - bX^*$ modeline dönüşür (Akkaya, 2000: 220).

2.2.3. Kareköklü Dönüşümlü Modeller

Kareköklü dönüşümler, tam kareköklü dönüşüm ve yarı kareköklü dönüşüm olarak ikiye ayrılır. Tam kareköklü dönüşümde x bağımsız değişken ve y bağımsız değişken olmak üzere her iki değişkenin karekökü alınır. Ve $\sqrt{X}$ ve $\sqrt{Y}$ olarak tanımlanır. Böylece,

$$\sqrt{Y} = a + b\sqrt{X} \quad (22)$$

modeli elde edilir. Bu modelde $\sqrt{Y} = Y^*$ ve $\sqrt{X} = X^*$ olarak seçildiğinde model doğrusal model olan $Y^* = a + bX^*$ modeline dönüşmüş olur. Bu modelin normal denklemleri,

$$\begin{aligned} \sum \sqrt{Y} &= na + b \sum \sqrt{X} \\ \sum \sqrt{Y} \sqrt{X} &= a \sum \sqrt{X} + b \sum X \end{aligned} \quad (23)$$

olarak elde edilir. Regresyon denkleminin standart hatası,

$$S_{\sqrt{X}\sqrt{Y}} = \sqrt{\frac{\sum (\sqrt{Y} - \sqrt{Y'})^2}{n-2}} \quad (24)$$

biçimindedir. Korelasyon katsayısı ise,

$$R_{\sqrt{Y}\sqrt{X}} = \sqrt{1 - \frac{\sum (\sqrt{Y} - \sqrt{Y'})^2}{\sum (\sqrt{Y} - \overline{\sqrt{Y}})^2}} \quad (25)$$

olarak ifade edilir (Orhunbilge, 1996: 63).

Yarı kareköklü dönüşümde ise bağımlı veya bağımsız değişkenden sadece birinin karekökü alınır. Diğer değişken olduğu gibi modele girer. Bu modeller ise,

$$\sqrt{Y} = a + bX \text{ ve } Y = a + b\sqrt{X} \quad (26)$$

biçimindedir. Bu modellerin normal denklemleri, denklemin standart hatası ve korelasyon katsayısı ise tam kareköklü dönüşümde olduğu gibi bulunabilir.

III. Uygulama

Doğrusal ve doğrusallaştırılabilen modeller için Türkiye’de yıllar itibariyle bir diş hekimine düşen kişi sayısı (nüfus) verileri kullanılmıştır. 1960-2006 yılları arasında bir diş hekimine düşen kişi sayısı (nüfus) verileri Tablo-1’de verilmiştir.

Tablo 1: Yıllar İtibariyle Türkiye’de Bir Diş Hekimine Düşen Nüfus

Yıllar	Diş Hekimine Düşen Nüfus	Yıllar	Diş Hekimine Düşen Nüfus	Yıllar	Diş Hekimine Düşen Nüfus
1960	19718	1976	7607	1992	5454
1961	20649	1977	7015	1993	5375
1962	18605	1978	6247	1994	5290
1963	16257	1979	6200	1995	5269
1964	17180	1980	6279	1996	5068
1965	16123	1981	6707	1997	5026
1966	14923	1982	6123	1998	4855
1967	14581	1983	6166	1999	4660
1968	14106	1984	6033	2000	4213
1969	11386	1985	6057	2001	4319
1970	10885	1986	6116	2002	4070
1971	10297	1987	6120	2003	3922
1972	9800	1988	5573	2004	3924
1973	8898	1989	5418	2005	3637
1974	9144	1990	5341	2006	3428
1975	7932	1991	5390		

Kaynak: <http://www.saglik.gov.tr/TR/BelgeGoster.aspx>

Tüm modeller için kullanım kolaylığı bakımından sıkça kullanılmakta olan bağımsız değişken olan yıllar yerine 1, 2, 3, 4,47 değerleri kullanılmıştır. Doğrusallaştırılabilen modellere geçmeden önce veriler doğrusal modele uygulanacaktır. Tüm modellerin tahmini ve gerekli olan analiz sonuçları Minitab (13.1) istatistiksel yazılım programıyla bulunmuştur.

Doğrusal Model

Doğrusal modelin bulunması için bağımlı ve bağımsız değişkenler hesaplamalara aynen girmiştir. Bunun sonucunda bulunan doğrusal regresyon denklemi ve her parametrenin standart hatası,

$$\hat{Y} = 15567 - 305X$$

$$(658,6) \quad (23,89)$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 2222$ ve belirlilik katsayısı ise $R^2 = 78,4$ olarak bulunmuştur. Parametrelerin testi için,

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

hipotezleri oluşturulur. Katsayıların anlamlılık testi için t dağılışı kullanılır. Test istatistiği,

$$t_h = \frac{b - \beta}{S_b} = \frac{-305 - 0}{23,89} = -12,77$$

olarak bulunur. %1 anlamlılık seviyesinde tablo değeri $t_{2;45} = 2,680$ olarak bulunur. Hesaplanan değeri tablo değerden büyük ($|t_h| = 12,77 > t_{tab} = 2,680$) olduğu için H_0 hipotezi red edilir. Yani ilişkinin anlamlı olduğuna karar verilir. Bir bağımsız değişkenli analizde tek parametre (β) için test yapılır.

Regresyon denkleminin bir bütün olarak anlamlılığını test etmek için varyans analizinden yararlanılır. Varyans analizi için gerekli hesaplamalar Tablo-2'de verilmiştir.

Tablo 2: Varyans Analizi Tablosu

Değişim Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Ortalama Kare	F	p
Regresyon	1	805593225	805593225	163,22	0,000
Hata	45	222109360	4935764		
Toplam	46	1027702585			

F testi için oluşturulacak hipotezler,

H_0 : Değişkenler arasında doğrusal ilişki yoktur.

H_1 : Değişkenler arasında doğrusal ilişki vardır.

biçimindedir.

%1 anlamlılık seviyesinde F tablo değeri $F_{1;45} = 7,22$ olarak bulunur. Tablo-2'de görüldüğü gibi hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler arasında doğrusal ilişki vardır. Ayrıca %1 anlamlılık seviyesinde $p = 0,000$ değeri de anlamlı bulunmuştur.

Regresyonda önemli bir kavram olan belirlilik katsayısının $R^2 = 78,4$ olarak bulunması doğrusal ilişkinin var olduğu fakat bu ilişkinin kuvvetli olmadığını göstermektedir. Bunun için doğrusal olmayan bir modelin araştırılıp tahmin için kullanılması daha anlamlıdır.

Logaritmik Modeller

Bu model için bağımlı ve bağımsız değişkenler yerine bunların logaritmaları alınarak modele girmiştir.

Bunun sonucunda bulunan doğrusal regresyon denklemi ve her parametrenin standart hatası,

$$\log Y = 4,54 - 0,541 \log X$$

$$(0,0309) \quad (0,0234)$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 0,06108$ ve belirlilik katsayısı ise $R^2 = 92,2$ olarak bulunmuştur.

Parametrelerin testi için hipotezler,

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

biçiminde oluşturulur. Katsayıların anlamlılık testi için t dağılışı kullanılır. Test istatistiği,

$$t_h = \frac{b - \beta}{S_b} = \frac{-0,541 - 0}{0,0234} = -23,12$$

olarak bulunur. %1 anlamlılık seviyesinde tablo değeri $t_{2,45} = 2,680$ olarak bulunur. Hesaplanan değeri tablo değerden büyük ($|t_h| = 23,12 > t_{tab} = 2,680$) olduğu için H_0 hipotezi red edilir. Yani ilişkinin anlamlı olduğuna karar verilir.

Varyans analizi için gerekli hesaplamalar Tablo-3'de verilmiştir.

Tablo 3: Varyans Analizi Tablosu

Değişim Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Ortalama Kare	F	p
Regresyon	1	1,9873	1,9873	532,65	0,000
Hata	45	0,1679	0,0037		
Toplam	46	2,1552			

F testi için oluşturulacak hipotezler şu biçimdedir:

H_0 : Değişkenler arasında logaritmik ilişki yoktur.

H_1 : Değişkenler arasında logaritmik ilişki vardır.

%1 anlamlılık seviyesinde F tablo değeri $F_{1,45} = 7,22$ olarak bulunur. Tablo-3'de görüldüğü gibi hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler arasında logaritmik ilişki vardır. Ayrıca %1 anlamlılık seviyesinde $p = 0,000$ değeri de anlamlı bulunmuştur.

Yarı logaritmik model için bağımlı ve bağımsız değişkenlerden biri aynen diğeri ise logaritması alınarak modele girmiştir.

Yarı logaritmik modellerden biri için bulunan regresyon denklemi ve her parametrenin standart hatası,

$$\hat{\log Y} = 4,22 - 0,0151X$$

$$(0,01916) \quad (0,0006952)$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 0,06465$ ve belirlilik katsayısı ise $R^2 = 91,3$ olarak bulunmuştur. Parametreler için test istatistiği,

$$t_h = \frac{b - \beta}{S_b} = \frac{-0,0151 - 0}{0,0006952} = -21,72$$

olarak bulunur. %1 anlamlılık seviyesinde tablo değeri $t_{2,45} = 2,680$ olarak bulunur. Hesaplanan değer tablo değerden büyük ($|t_h| = 21,72 > t_{tab} = 2,680$) olduğu için β katsayısının anlamlı olduğuna karar verilir.

Ayrıca bu model için F değeri 470,68 olarak bulunmuştur. %1 anlamlılık seviyesinde F tablo değeri $F_{1,45} = 7,22$ olarak bulunur. Hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler arasında yarı logaritmik ilişki vardır. Ayrıca analiz sonucunda bulunan $p = 0,000$ değeri de %1 anlamlılık seviyesinde modelin anlamlı olduğunu göstermektedir.

Diğer yarı logaritmik model için bulunan regresyon denklemi ve her parametrenin standart hatası,

$$\hat{Y} = 23390 - 11983 \log X$$

$$(544,8) \quad (412,7)$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 1076$ ve belirlilik katsayısı ise $R^2 = 94,9$ olarak bulunmuştur. Parametreler için test istatistiği,

$$t_h = \frac{b - \beta}{S_b} = \frac{-11983 - 0}{412,7} = -29,04$$

olarak bulunur. %1 anlamlılık seviyesinde tablo değeri $t_{2,45} = 2,680$ olarak bulunur. Hesaplanan değer tablo değerden büyük ($|t_h| = 29,04 > t_{tab} = 2,680$) olduğu için β katsayısının anlamlı olduğuna karar verilir.

Ayrıca bu model için F değeri 842,87 olarak bulunmuştur. %1 anlamlılık seviyesinde F tablo değeri $F_{1,45} = 7,22$ olarak bulunur. Hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler arasında yarı logaritmik ilişki vardır. Ayrıca analiz sonucunda bulunan $p = 0,000$ değeri de %1 anlamlılık seviyesinde modelin anlamlı olduğunu göstermektedir.

Hiperbolik (Ters) Modeller

Bu model için bağımlı ve bağımsız değişkenler yerine bunların tersleri alınarak modele girmiştir.

Bunun sonucunda bulunan doğrusal regresyon denklemi ve her parametrenin standart hatası,

$$\begin{aligned} (1/Y) &= 0,000176 - 0,000231(1/X) \\ (0,00000907) \quad (0,0000487) \end{aligned}$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 0,0000535$ ve belirlilik katsayısı ise $R^2 = 33,4$ olarak bulunmuştur. Belirlilik katsayısı çok küçük olduğundan model anlamlı görülmemektedir. Bu durumda diğer test ve analizlerin incelenmesine gerek yoktur.

Yarı hiperbolik model için bağımlı ve bağımsız değişkenlerden biri aynen diğeri ise tersi alınarak modele girmiştir.

Yarı hiperbolik modellerden biri için bulunan regresyon denklemi ve her parametrenin standart hatası,

$$\begin{aligned} (1/Y) &= 0,000043 + 0,000005X \\ (0,0000037) \quad (0,00000013) \end{aligned}$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 0,00001253$ ve belirlilik katsayısı ise $R^2 = 96,4$ olarak bulunmuştur. Parametreler için test istatistiği,

$$t_h = \frac{b - \beta}{S_b} = \frac{-0,000005 - 0}{0,00000013} = -38,46$$

olarak bulunur. %1 anlamlılık seviyesinde tablo değeri $t_{2,45} = 2,680$ olarak bulunur. Hesaplanan değer tablo değerden büyük ($|t_h| = 38,46 > t_{tab} = 2,680$) olduğu için β katsayısının anlamlı olduğuna karar verilir.

Ayrıca bu model için F değeri 1188,77 olarak bulunmuştur. %1 anlamlılık seviyesinde F tablo değeri $F_{1,45} = 7,22$ olarak bulunur. Hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler arasında yarı hiperbolik ilişki vardır. Ayrıca analiz sonucunda bulunan $p = 0,000$ değeri de %1 anlamlılık seviyesinde modelin anlamlı olduğunu göstermektedir.

Diğer yarı hiperbolik model için bulunan regresyon denklemi ve her parametrenin standart hatası,

$$\begin{aligned} Y &= 6138 + 22286(1/X) \\ (523,0) \quad (2814) \end{aligned}$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 3089$ ve belirlilik katsayısı ise $R^2 = 58,2$ olarak bulunmuştur. Belirlilik katsayısının küçük çıkması modelin anlamlı olmadığını göstermektedir. Bu durumda diğer test ve analizlerin incelenmesine gerek yoktur.

Kareköklü Modeller

Bu model için bağımlı ve bağımsız değişkenler yerine bunların karekökleri alınarak modele girmiştir.

Bunun sonucunda bulunan doğrusal regresyon denklemi ve her parametrenin standart hatası,

$$\hat{\sqrt{Y}} = 155 - 14,4\sqrt{X}$$

$$(2,604) (0,5315)$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 5,737$ ve belirlilik katsayısı ise $R^2 = 94,2$ olarak bulunmuştur. Parametrelerin testi için,

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

hipotezleri oluşturulur. Katsayıların anlamlılık testi için t dağılışı kullanılır. Test istatistiği,

$$t_h = \frac{b - \beta}{S_b} = \frac{-14,4 - 0}{0,5315} = -27,09$$

olarak bulunur. %1 anlamlılık seviyesinde tablo değeri $t_{2,45} = 2,680$ olarak bulunur. Hesaplanan değeri tablo değerden büyük ($|t_h| = 27,09 > t_{tab} = 2,680$) olduğu için H_0 hipotezi red edilir. Yani ilişkinin anlamlı olduğuna karar verilir.

Varyans analizi için gerekli hesaplamalar Tablo-4’de verilmiştir.

Tablo 4: *Varyans Analizi Tablosu*

Değişim Kaynağı	Serbestlik Derecesi	Kareler Toplamı	Ortalama Kare	F	p
Regresyon	1	24276	24276	737,50	0,000
Hata	45	1481	33		
Toplam	46	25757			

F testi için oluşturulacak hipotezler şu biçimdedir:

H_0 : Değişkenler arasında kareköklü ilişki yoktur.

H_1 : Değişkenler arasında kareköklü ilişki vardır.

%1 anlamlılık seviyesinde F tablo değeri $F_{1;45} = 7,22$ olarak bulunur. Tablo-4’de görüldüğü gibi hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler arasında kareköklü ilişki vardır. Ayrıca %1 anlamlılık seviyesinde $p = 0,000$ değeri de anlamlı bulunmuştur.

Yarı köklü model için bağımlı ve bağımsız değişkenlerden biri aynen diğeri ise karekökü alınarak modele girmiştir.

Yarı kareköklü modellerden biri için bulunan regresyon denklemi ve her parametrenin standart hatası,

$$\hat{Y} = 21326 - 2820\sqrt{X}$$

$$(679,9) (138,8)$$

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 1498$ ve belirlilik katsayısı ise $R^2 = 90,2$ olarak bulunmuştur. Parametreler için test istatistiği,

$$t_h = \frac{b - \beta}{S_b} = \frac{-2820 - 0}{138,8} = -20,32$$

olarak bulunur. %1 anlamlılık seviyesinde tablo değeri $t_{2,45} = 2,680$ olarak bulunur. Hesaplanan değer tablo değerden büyük ($|t_h| = 20,32 > t_{tab} = 2,680$) olduğu için β katsayısının anlamlı olduğuna karar verilir.

Ayrıca bu model için F değeri 412,92 olarak bulunmuştur. %1 anlamlılık seviyesinde F tablo değeri $F_{1,45} = 7,22$ olarak bulunur. Hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler arasında yarı kareköklü ilişki vardır. Ayrıca analiz sonucunda bulunan $p = 0,000$ değeri de %1 anlamlılık seviyesinde modelin anlamlı olduğunu göstermektedir.

Diğer yarı kareköklü model için bulunan regresyon denklemi ve her parametrenin standart hatası,

$$\sqrt{\hat{Y}} = 126 - 1,59X$$

(2,714) (0,09844)

olarak bulunmuştur. Ayrıca regresyon denkleminin standart hatası $S = 9,154$ ve belirlilik katsayısı ise $R^2 = 85,4$ olarak bulunmuştur. Parametreler için test istatistiği,

$$t_h = \frac{b - \beta}{S_b} = \frac{-1,59 - 0}{0,09844} = -16,15$$

olarak bulunur. %1 anlamlılık seviyesinde tablo değeri $t_{2,45} = 2,680$ olarak bulunur. Hesaplanan değer tablo değerden büyük ($|t_h| = 16,15 > t_{tab} = 2,680$) olduğu için β katsayısının anlamlı olduğuna karar verilir.

Ayrıca bu model için F değeri 262,35 olarak bulunmuştur. %1 anlamlılık seviyesinde F tablo değeri $F_{1,45} = 7,22$ olarak bulunur. Hesaplanan F değeri tablo değerden büyük çıktığı için değişkenler arasında yarı kareköklü ilişki vardır. Ayrıca analiz sonucunda bulunan $p = 0,000$ değeri de %1 anlamlılık seviyesinde modelin anlamlı olduğunu göstermektedir.

IV.Sonuç

Doğrusal olmayan modeller herhangi bir dönüşümle doğrusal kılınabilen ve herhangi bir dönüşümle doğrusal kılınamayan modeller olmak üzere ikiye ayrılır. Herhangi bir dönüşümle doğrusal kılınabilen bu modellerden logaritmik modeller, hiperbolik modeller ve kareköklü modeller bazı basit matematiksel dönüşümler yapılarak doğrusal hale dönüştürülmüştür. Dönüşümlerle doğrusal hale dönüştürülen bu modellere, 1960-2006 yılları aralığında temel sağlık hizmetleri istatistiklerinden “Türkiye’de bir diş hekimine düşen kişi sayısı” verileri uygulanmış olup en iyi model veya modeller tahmin edilmiştir. Bu modeller ve ilgili sonuçlar şunlardır:

- ✓ $\hat{Y} = 15567 - 305X$ doğrusal model için belirlilik katsayısı $R^2 = 78,4$ bulunmuş olup t ve F testleri anlamlı çıkmıştır.
- ✓ $\log \hat{Y} = 4,54 - 0,541 \log X$ tam logaritmik model için belirlilik katsayısı $R^2 = 92,2$ bulunmuş olup t ve F testleri anlamlı çıkmıştır.
- ✓ $\log \hat{Y} = 4,22 - 0,0151X$ yarı logaritmik model için belirlilik katsayısı $R^2 = 91,3$ bulunmuş olup t ve F testleri anlamlı çıkmıştır.
- ✓ $\hat{Y} = 23390 - 11983 \log X$ yarı logaritmik model için belirlilik katsayısı $R^2 = 94,9$ bulunmuş olup t ve F testleri anlamlı çıkmıştır.
- ✓ $(1/\hat{Y}) = 0,000176 - 0,000231(1/X)$ tam hiperbolik model için belirlilik katsayısı $R^2 = 33,4$ bulunmuştur. Belirlilik katsayısı çok küçük olduğundan bu model için diğer testleri incelemeye gerek kalmamıştır.
- ✓ $(1/\hat{Y}) = 0,000043 + 0,000005X$ yarı hiperbolik model için belirlilik katsayısı $R^2 = 96,4$ bulunmuş olup t ve F testleri anlamlı çıkmıştır.
- ✓ $\hat{Y} = 6138 + 22286(1/X)$ yarı hiperbolik model için belirlilik katsayısı $R^2 = 58,2$ bulunmuştur. Belirlilik katsayısı çok küçük olduğundan bu model için diğer testleri incelemeye gerek kalmamıştır.
- ✓ $\sqrt{\hat{Y}} = 155 - 14,4\sqrt{X}$ tam kareköklü model için belirlilik katsayısı $R^2 = 94,2$ bulunmuş olup t ve F testleri anlamlı çıkmıştır.
- ✓ $\hat{Y} = 21326 - 2820\sqrt{X}$ yarı kareköklü model için belirlilik katsayısı $R^2 = 90,2$ bulunmuş olup t ve F testleri anlamlı çıkmıştır.
- ✓ $\sqrt{\hat{Y}} = 126 - 1,59X$ yarı kareköklü model için belirlilik katsayısı $R^2 = 85,4$ bulunmuş olup t ve F testleri anlamlı çıkmıştır.

Sonuç olarak ilk model olan doğrusal modelin t ve F testleri anlamlı olmasına rağmen doğrusal olmayan modellere göre R^2 değeri oldukça küçük olduğundan doğrusal olmayan modellerin tahmin için kullanılması daha anlamlıdır. Tahmin için kullanılması gereken doğrusal olmayan model veya modeller, belirlilik katsayısı (R^2) yüksek ve t testi ile F testi anlamlı olan modellerdir. Bunlar ise sırasıyla: $(1/\hat{Y}) = 0,000043 + 0,000005X$ yarı hiperbolik modeli, $\hat{Y} = 23390 - 11983 \log X$ yarı logaritmik modeli, $\sqrt{\hat{Y}} = 155 - 14,4\sqrt{X}$ tam kareköklü modeli, $\log \hat{Y} = 4,54 - 0,541 \log X$ tam

logaritmik modeli, $\log \hat{Y} = 4,22 - 0,0151X$ diğer yarı logaritmik modeli, $\hat{Y} = 21326 - 2820\sqrt{X}$ yarı kareköklü modeli biçimindedir.

Kaynaklar

- Akaya, Ş. (2000), Ekonometri-I, Anadolu Matbaacılık, Dördüncü Baskı, İzmir.
- Gujarati, D. (2001), Temel Ekonometri, Çeviren: Ümit Şenesen, Literatür Yayıncılık, İstanbul.
- Gürsaka, N. (2002), Bilgisayar Uygulamalı İstatistik II, Alfa Yayınları, İstanbul.
- Johnston, J. (1981), Ekonometrik Metodlar, Çeviren: Yüksel İşyar, Atatürk Üniversitesi Basımevi.
- Karagöz, M. (1999), İstatistik Yöntemleri, Dördüncü Baskı, Malatya.
- Kohler, H. (1988), Statistics for Business and Economics, U.S.A.:Scott, Foresman and Company.
- Newbold, P. (2002), İşletme ve İktisat için İstatistik, Çev: Ümit Şenesen, Literatür Yayınları, Yayın No: 44, İstanbul.
- Orhunbilge, N. (1996), Uygulamalı Regresyon ve Korelasyon Analizi, İstanbul Üniversitesi İşletme Fakültesi Yayını, Yayın No:267, İstanbul.
- Şahinler, S. (2000), “En Küçük Kareler Yöntemi ile Doğrusal Regresyon Modeli Oluşturmanın Temel Prensipleri”, Mustafa Kemal Üniversitesi Ziraat Fakültesi Dergisi, Sayı:5, ss.57 73.
- Türker, Eyüp S. ve Can, E. (1998), Bilgisayar Uygulamalı Sayısal Analiz Yöntemleri, Değişim Yayınları, İstanbul.
- Ünver, Ö. (1996), Uygulamalı İstatistik Yöntemler, İkinci Baskı, Siyasal Kitabevi, Ankara.
- Yıldız, N. ve Bircan, H. (1999), İstatistiğe Giriş, Aktif Yayınevi, İkinci Baskı, Erzurum.
- Yüzer, Ali F. (2004), İstatistik, Anadolu Üniversitesi Yayını, Yayın No:1448, Eskişehir.