

AKÜ FEMÜBİD 13 (2013) 021301 (1-10)
DOI:10.5578/fmbd.6177

AKU J. Sci. Eng. 13 (2013) 021301 (1-10)

Araştırma Makalesi / Research Article

Değişken Dönüşümlerinin Kelebek Dağılan Artıklar için Değişen Varyanslılığa Etkileri

Reşit ÇELİK¹, Aydın ERAR²

¹Yıldız Teknik Üniversitesi Fen Ed. Fak. İstatistik Bölümü, İstanbul.

²Mimar Sinan Güzel Sanatlar Üniversitesi İstatistik Bölümü, İstanbul.

e-posta: resitcelik@gmail.com , aydinerar@yahoo.com

Geliş Tarihi:21.06.2013; Kabul Tarihi:29.08.2013

Anahtar kelimeler

Kelebek Dağılan
Artıklar, Değişen
Varyanslılık, Box - Cox
Dönüşümleri, Box -
Tidwell Dönüşümleri,
Çift Yanlı Dönüşümler

Özet

Klasik doğrusal regresyon analizinde varsayımlar gerçekleşmediğinde, ağırlıklandırma ve dönüşümler kullanılarak varsayım bozulmaları giderilmeye çalışılır. Değişen varyanslılık varsayım bozulmalarından biridir ve kelebek dağılan artıklar da değişen varyanslılığın özel bir halidir. Bu çalışmanın amacı, regresyon eşitliğinde yer alan değişkenlere uygulanan dönüşümlerin kelebek dağılan artıklar için etkilerinin incelenmesidir.

The Effects of Variable Transformations to Heteroscedasticity for Butterfly Distributed Residuals

Key words

Butterfly Distributed
Residuals;
Heteroscedasticity; Box
- Cox Transformations;
Box - Tidwell
Transformations;
Transformation Both
Sides .

Abstract

Weighting or transformations are used for fulfill the assumptions if there is any violation of assumptions in linear regression analysis. Heteroscedasticity is one of the violations these assumptions and butterfly-distributed- residuals are particular type of heteroscedasticity. The aim of this study is to examine the effects of transformations variables of regression model for butterfly distributed residuals.

© Afyon Kocatepe Üniversitesi

1. Giriş

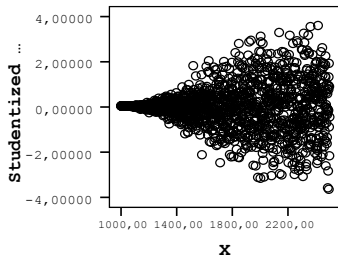
Klasik doğrusal regresyon analizinde, artıkların, normal dağılması, eş varyanslı ve aralarında ilişkisiz olması, bağımsız değişkenlerin bağlantısız olması temel varsayımlardır. Varsayımların bozulması parametrelere ilişkin hipotez testlerini ve güven aralıklarını geçersiz kılar. Bu nedenle, klasik doğrusal regresyon varsayımları bozulduğunda çeşitli araç yöntemler aracılığıyla bozulmaların etkisi giderilmeye çalışılır. Ağırlıklandırma ve dönüşümler bu anlamda kullanılan araç yöntemlerdir (Carroll and Ruppert, 1988, s. 118). Dönüşümler, normal dağılıma uygun hale getirme,

doğrusallaştırma ve artık varyanslarının dengelenmesi gibi amaçlar için kullanılmaktadır. Regresyon analizinde, daha çok regresyon eşitliğinin sol tarafına yani bağımlı değişkene uygulanan dönüşümlere rastlanmaktadır. Bu dönüşümler genel olarak Box and Cox (1964) tarafından önerilen dönüşümler ya da bu dönüşümlerin türevleridir. Uygulamada çok fazla başvurulmasa da kullanılan bir diğer dönüşüm ailesi, yalnızca regresyon eşitliğinin sağ tarafına bir diğer ifadeyle bağımsız değişkenlere uygulanan Box ve Tidwell'in (1962) önerdikleri dönüşümlerdir. Box - Cox ve Box - Tidwell dönüşümleri eş anlı olarak uygulanabilir (Ryan, 1997, s. 196) .

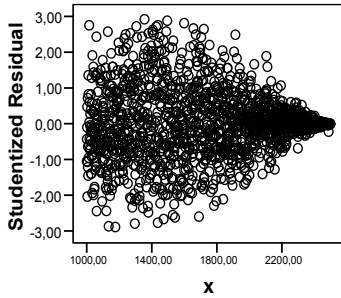
Regresyon analizinde, artıklara ilişkin varyansların, \hat{y} ya da bağımsız değişkenin çeşitli düzeyleri için, istatistiksel olarak farklı olması, 'değişen varyanslılık' olarak tanımlanır. Bu çalışmaya konu olan kelebek dağılan artıklar değişen varyanslılığın özel bir halidir.

Değişen varyanslılığın en önemli nedeni modellemedir. Değişen varyanslılığın saptanmasında, artıkların dikey eksende yer aldığı artık çizimleri kullanılmaktadır.

Değişen varyanslılıkta en çok karşılaşılan durum, artık saçılım görüntüsünün sağa ya da sola megafon biçiminde olduğu durumdur (Şekil 1 ve Şekil 2).

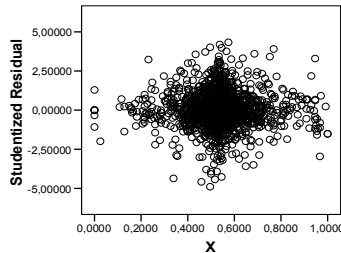


Şekil 1. Sağa Megafon

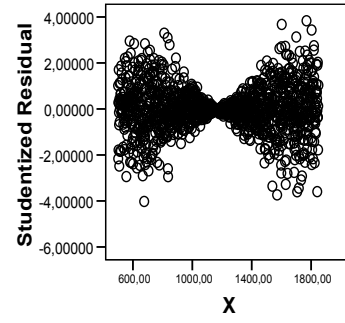


Şekil 2. Sola Megafon

Megafon biçimindeki artık kalıbının yanı sıra "galaksi" (Şekil 3) ve "kelebek" (Şekil 4) şeklinde artık görüntülerine de rastlanmaktadır. Kelebek şeklindeki bir saçılım grafiğinin iki ucu geniş, ortası ise dardır. Galaksi ise bunun tersine orta kısım geniş ve uçlar dardır (Darlington, 1997).



Şekil 3. Galaksi Dağılan Artıklar



Şekil 4. Kelebek Dağılan Artıklar

Megafon biçimindeki şekiller galaksi ya da kelebek biçimindeki diğer iki kalıp kadar istatistiksel çıkarsamayı bozamaz (Darlington, 1997).

Bu çalışmada, regresyon eşitliğinin sol tarafına uygulanan Box -Cox, Arcsin ve Freeman – Tukey, dönüşümlerine kısaca değinildikten sonra regresyon eşitliğinin sağ tarafına uygulanan Box - Tidwell dönüşümleri ve çift yanlı dönüşümler teorik çerçevesiyle verilmiştir. Çalışmanın uygulama bölümünde adı geçen yöntemler, kelebek dağılan artıkları olan regresyon modeline uygulanarak, kelebek dağılan artıklara olan etkileri diğer varsayım kontrolleri ile birlikte değerlendirilmiştir.

2. Box - Cox Dönüşümleri

Regresyon analizinde, bağımlı değişken dönüşümleri, eş varyansla normal dağılan toplamsal artıklar ve bağımsız değişkenlerle temsil edilen basit bir doğrusal model elde etmek için yapılır (Hamasaki and Goto, 2005). En sık başvurulan dönüşüm, Box – Cox'un bağımlı değişkenin dönüşümüdür. Varyansın bağımlı değişkenin artan değerlerine paralel olarak arttığı durumlarda kullanılan bu dönüşüm pozitif değerli değişkenlere uygulanabilmektedir. (Albayrak, 2008). Box - Cox dönüşümü (1) nolu eşitlikteki gibi tanımlanır.

$$y^* = \begin{cases} (Y^\lambda - 1)/\lambda & \lambda \neq 0 \text{ ise} \\ \ln Y & \lambda = 0 \text{ ise} \end{cases} \quad (1)$$

3. Arcsin Dönüşümü

Arcsin dönüşümü, bağımlı değişken değerleri (0,1) aralığında olduğunda uygulanan bir dönüşümdür. Ancak gözlemler (0,1) aralığında değilse (2) nolu eşitlik ile

$$y^* = \frac{Y - \min Y}{\max Y - \min Y} \rightarrow (0,1) \quad (2)$$

şeklinde normalize edildikten sonra arcsin dönüşümü uygulanır (Weisberg, 2005).

Dönüşüm modeli (3) nolu eşitlikteki gibidir.

$$\arcsin(\sqrt{y^*}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (3)$$

4. Freeman - Tukey Dönüşümü

Freeman ve Tukey (1950) tarafından önerilen bu dönüşüm, tüm gözlemler küçük olduğunda uygulanmaktadır. y^* , dönüşümü göstermek üzere,

$$y^* = \sqrt{Y} + \sqrt{Y+1} \quad (4)$$

şeklinde formüle edilir. Bu dönüşüm, Box - Cox (1964) dönüşüm ailesinde $y^* = y^{\lambda=0.5}$ dönüşümüyle benzerdir. Dönüşüm modeli (5) nolu eşitlikteki gibidir.

$$y^* = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (5)$$

5. Box - Tidwell Dönüşümleri

Box ve Tidwell (1962), dönüşüm gerekliliği varsayımıyla X' in uygun bir dönüşümünün belirlenmesi için yöntem geliştirmişlerdir. Bu yöntem yalnızca regresyon eşitliğinin sağ tarafına uygulanır (Ryan, 1997). Basit doğrusal regresyon modeli

$$Y = \beta_0 + \beta X + \varepsilon \quad (6)$$

için koşullu beklenen değer (7) nolu eşitlikteki gibidir.

$$E(Y/X) = \beta_0 + \beta X \quad (7)$$

(7) nolu eşitlik için Box - Tidwell dönüşümü (8) nolu eşitlikteki gibi tanımlanır.

$$E(Y) = \beta_0 + \beta_1 X' \quad (8)$$

Burada X' , (9) nolu eşitlikte tanımlandığı gibidir.

$$X' = \begin{cases} \ln X, & \alpha=0 \text{ için} \\ X^\alpha, & \alpha \neq 0 \text{ için} \end{cases} \quad (9)$$

X' in X' için bir başlangıç tahmini olarak alınması yöntemin hareket noktasıdır. Bu tahmin ve $E(Y)$ 'nin $\alpha_0 = 1$ komşuluğunda Taylor serisi açılımı (10) nolu eşitlikte verildiği gibidir.

$$E(Y) = f(\alpha_0 = 1) + (\alpha - \alpha_0) \left\{ \frac{\partial [E(Y)]}{\partial \alpha} \right\}_{\substack{x=x \\ \alpha=\alpha_0}} \quad (10)$$

Burada $f(\alpha_0 = 1) = \beta_0 + \beta_1 X$ ve

$$E(Y) = \beta_0 + \beta_1 X + (\alpha - 1) \left\{ \frac{\partial [E(Y)]}{\partial \alpha} \right\}_{\substack{x=x \\ \alpha=\alpha_0}} \text{ dir. (10) nolu}$$

eşitlikte türevler için zincir kuralı uygulandığında 3. terim (11) nolu eşitlikteki gibi olur.

$$\left\{ \frac{\partial [E(Y)]}{\partial \alpha} \right\}_{\substack{x=x \\ \alpha=\alpha_0}} = \left\{ \frac{\partial [E(Y)]}{\partial x'} \right\}_{x=x} \left\{ \frac{\partial x'}{\partial \alpha} \right\}_{\alpha=\alpha_0} \quad (11)$$

(11) nolu eşitliğin sağ tarafı

$$X' = X^\alpha \rightarrow \partial X' / \partial \alpha = X^\alpha \ln X,$$

$$\alpha = \alpha_0 = 1 \rightarrow \partial X' / \partial \alpha = X \ln X$$

ve $\partial [E(Y)] / \partial x = \beta_1$ şeklinde açılırsa

$$\left\{ \frac{\partial [E(Y)]}{\partial \alpha} \right\} = \beta_1 X \ln X \text{ elde edilir. Elde edilenler}$$

(10) nolu eşitlikte yerine yazılırsa $E(Y)$ için Taylor serisi, $E(Y) = \beta_0 + \beta_1 X + (\alpha - 1)\beta_1 X \ln X$ şekline dönüşür. Burada $(\alpha - 1)\beta_1 = \gamma$ ve $Z = X \ln X$ yazılırsa (12) nolu eşitlik elde edilir.

$$E(Y) = \beta_0 + \beta_1 X + \gamma Z \quad (12)$$

(12) nolu eşitlik çoklu regresyon modelidir. (12) nolu eşitlikte, γ parametresinin sıfır olması dolayısıyla α parametresinin "1" olması, eşitliğin sağ tarafı için dönüşümün gerekli olmadığı anlamına gelir (Ryan, 1997). Bu çoklu regresyonun çözümünden elde edilen $\hat{\gamma}$ ve dönüşümün dikkate alınmadığı $\hat{Y} = \hat{\beta}_0^* + \hat{\beta}_1^* X'$ den $\hat{\beta}_1^*$ parametre

tahmincileri kullanılarak $\hat{\alpha}_i = \frac{\hat{\gamma}}{\hat{\beta}_1^*} + \hat{\alpha}_{i-1}$ ifadesine

ulaşılır. Dönüşüm gerekli değildir hipotezinin karşılığı $\hat{\alpha}_0 = 1$ 'dir. Burada $X' = X^{\hat{\alpha}_i}$ alınarak süreç istenen yakınsamaya ulaşıncaya kadar yinelenir.

$a_j = (\hat{\gamma}_j / \hat{\beta}_{1,j+1}^*)$ olmak üzere (j+1)'nci iterasyon için $x'_{j+1} = (x_i^{\alpha_{j-1}})^{a_j}$ olarak elde edilir. Benzer şekilde için şekilde $\alpha_j = a_j \alpha_{j-1}$ dir. Box - Tidwell algoritması yalnızca pozitif değişkenlere uygulanabilir (Bagnall, 2006)

6. Çift Yanlı Dönüşümler

Regresyon eşitliğinin bozulmaması için bazen her iki yanına dönüşüm uygulanabilmektedir. Çift yanlı dönüşümün geleneksel kullanımı doğrusal olamayan modellerin doğrusallaştırılmasıdır. Doğrusallaştırma dönüşümleri en küçük karelerin etkinliğini çok azaltarak asimetriye ya da değişen varyanslılığa neden olur. Kimi zaman doğrusallaştırma şiddetli değişen varyanslılığa ve fiziksel olarak mümkün olmayan bir parametre tahminine neden olur (Carroll and Ruppert, 1988). Çift yanlı dönüşümlerde Box - Cox dönüşüm parametresi λ , $[-2, 2]$ aralığında keyfi olarak seçildikten sonra Box - Tidwell dönüşüm parametrelerinin tahminleri λ 'ya bağlı olarak yapılır:

$$y^* = \beta_0^* + \beta_1^* X_1 + \beta_2^* X_2 + \dots + \beta_p^* X_p, y^* = \begin{cases} y^\lambda, & \lambda \neq 0 \\ \ln y, & \lambda = 0 \end{cases} \quad (13)$$

$$y^* = \theta_0 + \theta_1 X_1 + \gamma_1 X_1 \ln X_1 + \dots + \theta_p X_p + \gamma_p X_p \ln X_p \quad (14)$$

$$\hat{\alpha}_i = \frac{\hat{\gamma}_i}{\hat{\beta}_i^*} + \hat{\alpha}_{i-1} \quad (15)$$

Burada p, modeldeki değişken sayısı ve i iterasyon sayısıdır. Box - Cox dönüşüm parametresi seçildikten sonra (14) ve (15) nolu eşitliklerin iteratif kullanımı ile Box - Tidwell dönüşüm parametreleri tahmin edilerek (16) nolu eşitlikte ifade edilen

$$y^\lambda = \beta_0 + \beta_1 X_1' + \beta_2 X_2' + \dots + \beta_p X_p', X_i' = \begin{cases} \ln X_i, & \hat{\alpha}_i = 0 \text{ için} \\ X_i^{\hat{\alpha}_i}, & \hat{\alpha}_i \neq 0 \text{ için} \end{cases} \quad (16)$$

çift yanlı dönüşüm modeline ulaşılır. Elde edilen çift yanlı dönüşüm modeli ile en küçük kareler analizi gerçekleştirilir. Burada eşitliğin sağ tarafına dönüşüm gerekli değildir hipotezinin karşılığı olan $\hat{\alpha}_{i0} = 1$ dir. Eşitliğin sol tarafına dönüşüm gerekli değildir hipotezi $\lambda = 1$ dir. $a_{ij} = (\hat{\gamma}_{ij} / \hat{\beta}_{i,j+1}^*)$ olmak üzere (j+1)'nci iterasyon için $x'_{i(j+1)} = (x_i^{\alpha_{j-1}})^{a_j}$ olarak elde edilir. Benzer şekilde $\alpha_{ij} = a_{ij} \alpha_{i(j-1)}$ dir.

$X_i' = X_i^{\hat{\alpha}_i}$ alınarak süreç istenen yakınsamaya ulaşıncaya kadar yinelenir.

7. Uygulama

Partiküller madde (PM) terimi, havada bulunan katı partiküller ve sıvı damlacıkları ifade eder. İnsan faaliyetleri sonucu ve doğal kaynaklardan, doğrudan atmosfere karışır. Atmosferde diğer kirleticiler ile reaksiyona girerek PM oluşturur ve atmosfere verilirler. Katı ve sıvı partiküllerin boyutları geniş bir aralığa yayılır. Sağlığa konu olan partiküller, aerodinamik çapı $10 \mu\text{m}$ 'nin altındaki partiküllerdir. Bu boyut aralığındaki partiküller, solunum sistemi içine girerek birikim yapabilir. $2.5 \mu\text{m}$ den daha küçük partiküller ince partiküller olarak adlandırılır. İnce partikül kaynakları, tüm yanma prosesleri ve bazı endüstriyel prosesleri içerir. $2.5-10 \mu\text{m}$. aralığındaki partiküller, kaba partiküller olarak adlandırılır. Kaba partikül kaynakları ise kırma, öğütme işlemleri, yollardan kalkan tozlardır (Int. Kynk 1).

Uygulama için Sivas iline ilişkin Ocak 1990-Aralık 2010 aralığında aylık ortalama partiküller madde, kükürt ve sıcaklık ölçüm değerleri kullanılmıştır. Analizler için SAS 9.2, SPSS 15.0 ve Minitab 14 Demo programları kullanılmıştır.

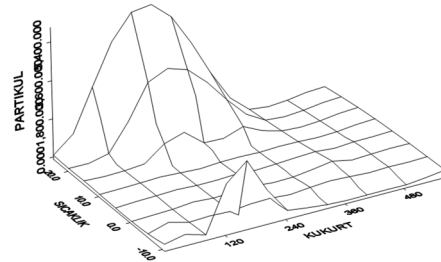
Partiküller madde miktarı bağımlı değişken; taşıt sayısı, sıcaklık, doğal gaz kullanımı ve kükürt miktarı bağımsız değişken varsayılarak modelleme yapılmıştır (Tablo 1). Modelleme yapılırken gözlemler aylar bazında yıllara göre sıralanmıştır. Sıcaklığa ilişkin ortalama aylık veriler devlet meteoroloji müdürlüğünden, kükürt miktarı, taşıt sayısı ve partiküller maddeye ilişkin aylık ortalama veriler ise TÜİK kayıtlarından elde edilmiştir (Int Kynk 2). Taşıt sayısına ilişkin TÜİK kayıtlarının başlangıç yılı 1994 Temmuz; diğer değişkenlere ilişkin kayıtların başlangıcı ise Ocak 1990 dır.

$$PM = \beta_0 + \beta_1 * sıcaklık + \beta_2 * kükürt + \beta_3 * tasit + \beta_4 * dgaz + \varepsilon$$

Tablo 1. Tam Küme Parametre Tahminleri ve Model Seçimi

Gözlem Sayısı	253				
Kullanılan Gözlem Sayısı	199				
Boş Gözlem Sayısı	54				
R-Kare	0.6400				
Değişken	DF	Tahmin	St. Hata	t Değeri	Pr > t
Kesişim	1	66.26640	8.83994	7.50	<.0001
sıcaklık	1	-2.71338	0.30537	-8.89	<.0001
Kükürt	1	0.29889	0.05462	5.47	<.0001
Taşıt Sayısı	1	0.00020002	0.00012476	1.60	0.1105
Dgaz	1	0.41247	4.58087	0.09	0.9283
Model	R-Kare	AIC	BIC	SBC	Modeldeki Değişkenler
1	0.5682	1341.5573	1342.9369	1348.14396	sıcaklık
2	0.4649	1384.2607	1384.8293	1390.84730	Kükürt
3	0.0114	1506.4059	1505.1154	1512.99253	dgaz
4	0.0040	1507.9049	1506.5962	1514.49155	taşı_t_sa
5	0.6351	1310.0485	1312.1215	1319.92843	Kükürt sıcaklık
6	0.5813	1337.4362	1338.7008	1347.31614	sıcaklık dgaz
7	0.5703	1342.6133	1343.7302	1352.49318	sıcaklık taşı_t_sa
8	0.4852	1378.5538	1378.6985	1388.43374	Kükürt dgaz
9	0.4751	1382.4287	1382.4744	1392.30864	Kükürt taşı_t_sa
10	0.0167	1507.3407	1504.8144	1517.22062	dgaz taşı_t_sa
11	0.6400	1309.4071	1311.6116	1322.58035	Kükürt sıcaklık taşı_t_sa
12	0.6352	1312.0182	1314.1168	1325.19141	Kükürt sıcaklık dgaz
13	0.5844	1337.9638	1339.0479	1351.13704	sıcaklık dgaz taşı_t_sa
14	0.4934	1377.3467	1377.0323	1390.51991	Kükürt dgaz taşı_t_sa
15	0.6400	1311.3988	1313.6552	1327.86533	Kükürt sıcaklık dgaz tasita

Seçilen başlangıç modelinde taşıt sayısı ve doğal gaz kullanımı istatistiksel olarak anlamlı çıkmamıştır. Diğer yandan ileri seçim (forward) ile Akaike Bilgi Kriteri (Akaike Information Criterion, **AIC**), Bayes Bilgi Kriteri (Bayes Information Criterion, **BIC**) ve Schwarz Bilgi Kriteri (Schwarz Information Criterion, **SBC**) kriterleri açısından modeller değerlendirildiğinde **model 5** ile **model 11**'in değerleri bir birine çok yakındır. AIC ve BIC kriterleri açısından "Sıcaklık Kükürt ve Taşıt Sayısı" değişkenlerinin yer aldığı **model 11**; SBC kriteri açısından **model 5** seçilmelidir. Öte yandan taşıt sayısına ilişkin kayıtlar 1994 Temmuz ayından itibaren mevcuttur. Dolayısıyla taşıt sayısı değişkenini içeren **model 11**'in seçilmesi durumunda, 54 tane gözlemin vereceği bilgiden yoksun kalınacaktı. Bu nedenle uygulamanın bundan sonraki sürecinde taşıt sayısı ve doğal gaz kullanımı değişkenleri analiz dışı $PM = \beta_0 + \beta_1 * Sıcaklık + \beta_2 * Kükürt + \varepsilon$ modeli ile çalışılmıştır. Modelin 3 boyutlu uzayda çizimi şekil 5'te verilmiştir.

**Şekil 5.** Partiküller Maddenin Kükürt ve Sıcaklıkla Değişimi

Box-Tidwel dönüşümü pozitif değerli değişkenlere uygulanabilmektedir. Veri kümesinde yer alan iki bağımsız değişkenden biri olan sıcaklık Celsius cinsinden ölçüldüğünden negatif değerler içermektedir. Bu nedenle bu değişken bir başka sıcaklık ölçü birimi olan Fahrenheit ölçeğine $F^0 = 1.8 * C^0 + 32$ formülü ile dönüştürülerek pozitif olması sağlanmıştır.

Bu çalışmada, Student türü artıkların sıklık çizimleriyle normal dağılıma uygunluk, Student türü artıkların aylara karşı çizimi ile değişen varyanslılık ve aykırı değerler değerlendirilmiştir. Değişen varyanslılığın belirlenmesi için ayrıca White χ^2

testine de başvurulmuştur. Durbin – Watson (D – W) D testi ile 1’nci mertebeden otokorelasyon test edilmiştir. Varyans büyüme çarpanı (VIF) ile bağımsız değişkenler arasındaki çoklu bağlantı test edilmiştir. F testi, düzeltilmiş belirginlik katsayısı, tahminin standart hatası ve değişim katsayısı kriterleri ile model yeterliliği değerlendirilmiştir.

7.1. En Küçük Kareler Analizi Sonuçları

Veri kümesine uygulanan en küçük kareler (EKK) sonuçları Tablo 2’de verilmiştir.

7.2. Freeman - Tukey Dönüşümü Sonuçları

Bağımlı değişken değerlerine Freeman - Tukey dönüşümü uygulandıktan sonra gerçekleştirilen en küçük kareler analizi sonuçları Tablo 2 Sütun 3’te verilmiştir.

7.3. Arcsin Dönüşümü Sonuçları

Bağımlı değişken değerleri normalize edildikten sonra arcsin dönüşümü uygulanmıştır. Elde edilen sonuçlar Tablo 2 Sütun 4’de verilmiştir.

7.4. Çift Yanlı Dönüşümler

Veri kümesi için çift yanlı dönüşüm modeli *Partiküller Madde = Y* →

$$Y^\lambda = \beta_0 + \beta_1(\text{Sıcaklık})^{\alpha_1} + \beta_2(\text{Kükürt})^{\alpha_2} + \varepsilon \text{ dir.}$$

Uygulamada Box - Cox parametresi λ , $[-2; 2]$ aralığında keyfi olarak seçildikten sonra Box - Tidwell dönüşüm parametreleri iteratif olarak tahmin edilerek Box - Cox başlangıç ve Box - Tidwell final parametrelerinin tahminleri için değişkenler tek tek dönüştürüldükten sonra analizler gerçekleştirilmiştir.

Box - Cox dönüşüm parametresi

$\lambda = -2 \rightarrow y^* = Y^{-2}$ değeri için Box - Tidwell dönüşüm parametrelerinin tahminleri 2 iterasyon sonucunda $\hat{\alpha}_1 = 5.04359$ $\hat{\alpha}_2 = 9.09289$ olarak elde edilmiştir. Regresyon eşitliğinde yer alan değişkenler, Box - Cox ve Box - Tidwell dönüşüm parametreleri için

$$\frac{1}{Y^2} = \beta_0 + \beta_1(\text{Sıcaklık})^{5.04359} + \beta_2(\text{Kükürt})^{9.09289} + \varepsilon$$

şeklinde dönüştürüldükten sonra en küçük kareler

uygulanmıştır. Sonuçlar Tablo 2’de verilmiştir.

Box - Cox dönüşüm parametresi

$\lambda = -1 \rightarrow y^* = Y^{-1}$ değeri için Box - Tidwell dönüşüm parametrelerinin tahminleri 2 iterasyon sonucunda $\hat{\alpha}_1 = 3.02862$ $\hat{\alpha}_2 = -1.19162$ olarak elde edilmiştir. Regresyon eşitliğinde yer alan değişkenler, Box - Cox ve Box - Tidwell dönüşüm parametreleri için

$$\frac{1}{Y} = \beta_0 + \beta_1(\text{Sıcaklık})^{3.02862} + \beta_2(\text{Kükürt})^{-1.19162} + \varepsilon$$

şeklinde dönüştürüldükten sonra en küçük kareler uygulanmıştır. Sonuçlar Tablo 2’de verilmiştir.

Box - Cox dönüşüm parametresi $\lambda = -0.5$ için Box - Tidwell dönüşüm parametrelerinin tahminleri 3 iterasyon sonucunda

$$\hat{\alpha}_1 = 3.10624 \quad \hat{\alpha}_2 = -1.75587 \text{ olarak elde edilmiştir.}$$

Regresyon eşitliğinde yer alan değişkenler, Box - Cox ve Box - Tidwell dönüşüm parametreleri için

$$\frac{1}{\sqrt{Y}} = \beta_0 + \beta_1(\text{Sıcaklık})^{3.10624} + \beta_2(\text{Kükürt})^{-1.75587} + \varepsilon$$

şeklinde dönüştürülerek en küçük kareler uygulanmıştır. Analiz sonuçları Tablo 2’de verilmiştir.

Box - Cox dönüşüm parametresi

$\lambda = 0 \rightarrow y^* = \ln y$ için Box - Tidwell dönüşüm parametrelerinin tahminleri 10 iterasyon sonucunda $\hat{\alpha}_1 = 2.16363$ $\hat{\alpha}_2 = 0.65477$ olarak elde edilmiştir. Regresyon eşitliğinde yer alan değişkenler, Box - Cox ve Box - Tidwell dönüşüm parametreleri için

$$\ln Y = \beta_0 + \beta_1(\text{Sıcaklık})^{2.16363} + \beta_2(\text{Kükürt})^{0.65477} + \varepsilon$$

şeklinde dönüştürüldükten sonra en küçük kareler uygulanmıştır. Analiz sonuçları Tablo 2’de verilmiştir.

Box - Cox dönüşüm parametresi $\lambda = 0.5$ için Box - Tidwell dönüşüm parametrelerinin tahminleri 10 iterasyon sonucunda $\hat{\alpha}_1 = 1.04008$ $\hat{\alpha}_2 = 0.84877$

olarak elde edilmiştir. Regresyon eşitliğinde yer alan değişkenler, Box - Cox ve Box - Tidwell dönüşüm parametreleri için

$$\sqrt{Y} = \beta_0 + \beta_1(\text{Sıcaklık})^{1.04008} + \beta_2(\text{Kükürt})^{0.84877} + \varepsilon$$

şeklinde dönüştürüldükten sonra en küçük kareler uygulanmıştır. Analiz sonuçları Tablo 2’de verilmiştir.

Box - Cox dönüşüm parametresi $\lambda = 1$ için elde edilen Box - Tidwell dönüşüm parametrelerinin tahminleri 7 iterasyon sonucunda $\hat{\alpha}_1 = 0.09726$, $\hat{\alpha}_2 = 1.07993$ olarak elde edilmiştir. Regresyon eşitliğinde yer alan değişkenler, Box - Cox ve Box - Tidwell dönüşüm parametreleri için $Y = \beta_0 + \beta_1(Sıcaklık)^{0.09726} + \beta_2(Kükürt)^{1.07993} + \varepsilon$ şeklinde dönüştürüldükten sonra küçük kareler uygulanmıştır. Analiz sonuçları Tablo 2'de verilmiştir.

Box - Cox dönüşüm parametresi $\lambda = 2$ için elde edilen Box - Tidwell dönüşüm parametrelerinin tahminleri 8 iterasyon sonucunda $\hat{\alpha}_1 = -1.03085$, $\hat{\alpha}_2 = 1.58288$ olarak elde edilmiştir. Regresyon eşitliğinde yer alan değişkenler, Box - Cox ve Box - Tidwell dönüşüm parametreleri için $Y^2 = \beta_0 + \beta_1(Sıcaklık)^{-1.03085} + \beta_2(Kükürt)^{1.58288}$ şeklinde dönüştürüldükten sonra en küçük kareler uygulanmıştır. Analiz sonuçları Tablo 2'de verilmiştir.

Tablo 2. Toplu Analiz Sonuçları

	En Küçük Kareler Analizi Sonuçları	Freeman – Tukey Dönüşümü Sonuçları	Arcsin Dönüşümü Sonuçları	Çift Yanlı Dönüşüm Sonuçları (Box – Cox ve Box – Tidwell Dönüşümleri)						
				$Y^\lambda = \beta_0 + \beta_1(\text{Sıcaklık})^{\alpha_1} + \beta_2(\text{Kükürt})^{\alpha_2} + \varepsilon$						
				$\lambda = -2$ $\hat{\alpha}_1 = 5.04359$ $\hat{\alpha}_2 = 9.09289$	$\lambda = -1$ $\hat{\alpha}_1 = 3.02862$ $\hat{\alpha}_2 = -1.19162$	$\lambda = -0.5$ $\hat{\alpha}_1 = 3.10624$ $\hat{\alpha}_2 = -1.75587$	$\lambda = 0$ $\hat{\alpha}_1 = 2.16363$ $\hat{\alpha}_2 = 0.65477$	$\lambda = 0.5$ $\hat{\alpha}_1 = 1.04008$ $\hat{\alpha}_2 = 0.84877$	$\lambda = 1$ $\hat{\alpha}_1 = 0.09726$ $\hat{\alpha}_2 = 1.07993$	$\lambda = 2$ $\hat{\alpha}_1 = -1.03085$ $\hat{\alpha}_2 = 1.58288$
$\hat{\beta}_0 (S_{\hat{\beta}_0})$	112.91914 (8.87678)	23.20026 (0.95672)	0.61465 (0.03555)	0.00001634 (0.000127)	0.00729 (0.00147)	0.07451 (0.00474)	4.29724 (0.10463)	10.80629 (0.49695)	687.16163 (67.30487)	-5331.59 (1552.83)
$\hat{\beta}_1 (S_{\hat{\beta}_1})$	-1.33722 (0.15216)	-0.18517 (0.01640)	-0.00608 (0.0006093)	1.3637E-12 (1.26E-13)	4.949E-9 (3.52E-10)	0.00000328 (2.03E-7)	-0.0001245 (0.000012)	-0.07197 (0.00696)	-438.01844 (45.68181)	452875 (73289)
$\hat{\beta}_2 (S_{\hat{\beta}_2})$	0.40456 (0.02218)	0.03314 (0.00239)	0.00130 (0.0000888)	-7.73E-30 (9.65E-29)	-3.84349 (1.70167)	-0.20357 (0.07123)	0.02778 (0.00298)	0.04345 (0.00312)	0.24363 (0.01331)	3.77213 (0.17980)
$t_{\hat{\beta}_0}; t_{\hat{\beta}_1}; t_{\hat{\beta}_2}$	12.72; -8.79; 18.24	24.25; -11.29; 13.87	17.29; -9.98; 14.65	0.13; 10.81 -0.08	4.95; 14.05; -2.26	15.71; 16.12; -2.86	41.07; -10.15; 9.31	21.75; -10.34; 13.94	10.21; -9.59; 18.31	-3.43; 6.18; 20.98
$F_{Model}; p$	418.40; <.0001	350.99; <.0001	338.96; <.0001	59.51 <.0001	98.82; <.0001	130.84; -2.86	241.13; <.0001	353.07 <.0001	428.00 <.0001	355.82 <.0001
Tahminin Standart Hatası	31.52893	3.39812	0.12625	0.00146	0.01558	0.04184	0.47386	1.70203	31.25427	9837.8
\bar{R}^2	0.7681	0.7353	0.7284	0.3171	0.4370	0.5075	0.6559	0.7364	0.7722	0.7379
VIF	1.42876	1.42876	1.42876	1.01714	1.01644	1.07013	0.865	1.50773	37.46130	1.21257
Değişim Katsayısı	37.79050	19.87688	29.16247	154.86106	68.96035	30.30818	11.47981	1.70203	1.39445	87.59576
D-W D İstatistiği; 1. Mert. Otokor.	1.249; 0.359	1.024; 0.478	1.172; 0.400	0.749; 0.626	0.683; 0.659	0.626; 0.684	0.865; 0.565	1.040; 0.471	1.239; 0.360	1.541; 0.187
White $\chi^2; p$	14.33; 0.0136	1.68; 0.8917	6.87; 0.2308	36.62; <.0001	50.74; <.0001	45.83; <.0001	40.89; <.0001	1.51; 0.9115	13.86; 0.0166	9.80; 0.0811
Student türü artıkların aylara karşı çizimi										
Student türü Artıkların Normal Dağılıma Uygunluk Çizimi										

En küçük kareler analizi sonuçları, istatistik kriterler açısından incelendiğinde, kelebek dağılan artıklar dolayısıyla değişen varyanslılık dışında bir sorun görülmemektedir.

Freeman - Tukey karekök dönüşümü için regresyon analizi sonuçlarına bakıldığında, kelebek dağılan artıklar sorunun en küçük kareler analizi sonuçlarına göre büyük oranda düzeldiği görülmektedir. Diğer kriterler incelendiğinde ise dönüşümün otokorelasyona neden olduğu söylenebilir.

Bağımlı değişkenin arcsin dönüşümü sonucu elde edilen regresyon analizi sonuçları incelendiğinde, değişen varyanslılığın düzeldiği söylenebilir. Buna karşılık dönüşüm, Studentize artık değeri 5 olan bir aykırı değer üretmiştir. Diğer model değerlendirme kriterleri incelendiğinde ise istatistiksel olarak anlamlı bir bozulma görülmemektedir.

Çift yanlı dönüşümlerin en küçük kareler analiz sonuçları incelendiğinde, yalnızca $\lambda = 0.5$ düzeyi için değişen varyanslılık sorunun kısmen ortadan kalktığı görülmektedir. Buna karşılık söz konusu düzey için bir otokorelasyon sorununun ortaya çıktığı görülmektedir. Diğer çift yanlı dönüşümlerin analiz sonuçları incelendiğinde $\lambda = -2$, $\lambda = -1$ $\lambda = -0.5$ düzeyleri için normal dağılıma uymayan artıklar, galaksi dağılan artıklar dolayısıyla değişen varyanslılık, otokorelasyon, düşük belirginlik katsayısı sorunları gözlenmiştir. Ek olarak $\lambda = -2$ düzeyi için regresyon sabiti ve kükürt değişkeni istatistiksel olarak anlamsız hale gelmiştir. $\lambda = 0 \rightarrow \ln Y$ dönüşümü sonucunda, değişen varyanslılık ve otokorelasyon sorunları ortaya çıkmıştır. $\lambda = 1$ düzeyi için kelebek dağılan artıklar dağılan artıklar dolayısıyla değişen varyanslılık, $\lambda = 2$ düzeyi için kelebek dağılan artıklar ve çok büyük standart hata gibi istenmeyen sonuçlar ortaya çıkmıştır.

8. Sonuç

Bu çalışmada ana amaç, artıklar kelebek biçiminde dağıldığında hangi dönüşümlerin daha uygun olabileceği araştırılmıştır. Bununla birlikte uygulanan dönüşümlerin yalnızca kelebek dağılan artıklara etkileri değerlendirilmemiş, normal dağılıma uygunluk, otokorelasyon, çoklu bağlantı ve aykırı değerler sorunlarına etkileri de dikkate alınmıştır.

Çalışmada, aynı veri kümesine 4 başlıkta toplam 9 analiz uygulanmıştır. Regresyon eşitliğinin yalnızca sol tarafına uygulanan Tukey - Freeman karekök dönüşümü ve ara bir dönüşümden sonra uygulanan arcsin dönüşümü, varyansı dengeleme açısından işe yaramıştır. Buna karşılık, Tukey-Freeman dönüşümü otokorelasyona, arcsin dönüşümü tek aykırı değere neden olmuştur.

Çift yanlı dönüşümlerde, Kelebek dağılan artıklar için $[-2,2]$ aralığında seçilen Box - Cox parametresinin her biri için iteratif olarak elde edilen Box - Tidwell dönüşüm parametrelerinin tahminlerinin değişkenlere uygulanmasından sonra gerçekleştirilen en küçük kareler sonucunda yalnızca bir düzey için varyanslar kısmen dengelenebilmiştir. Diğer 6 düzey için varyansların dengelenememesi bir yana dönüşümler, artık dağılımında bozulmaya, otokorelasyona ve aykırı değerlere neden olmuştur.

Bu çalışmada dikkat çeken bir konu da değişen varyanslılığın belirlenmesinde kullanılan White χ^2 testinin kelebek dağılan artıklar durumunda yanlış sonuç verebilmesidir. $\lambda = 2$ için uygulanan çift yanlı dönüşüm sonucunda gerçekleştirilen en küçük kareler analizi Student türü artıklarının aylara karşı çiziminde açık bir biçimde kelebek dağılan artık kalıbı olmasına karşılık, White χ^2 testine göre değişen varyanslılık sorunu yoktur.

Çalışmanın sonucunda, özellikle çift yanlı dönüşümlerin uygulanabilirlik açısından oldukça zor olduğu görülmüştür. Uygulanan dönüşümler genel anlamda monoton dönüşümlerdir. Kelebek dağılan artıkları olan bu veri kümesi için arcsin ve Freeman - Tukey dönüşümleri, değişen varyanslılığı dengelemede, çift yanlı dönüşümlere göre daha iyi sonuç vermiştir.

Kaynaklar

- Albayrak, A. S. (2008). Değişen Varyans Durumunda En Küçük Kareler Tekniğinin Alternatifi Ağırlıklı Regresyon Analizi ve Bir Uygulama. *Afyon Kocatepe Üniversitesi, İ.İ.B.F. Dergisi*.
- Bagnall, A., Whittley, I., Studley, M., Pettipher, M., Tekiner, F., Bull, L. (2006). Variance Stabilizing Regression Ensembles for Environmental Models, *International Joint Conference on Neural Networks, Sheraton Vancouver BC, Canada July 16-21*.
- Box, G. E. P. and Tidwell, P. W. (1962). Transformation of the Independent Variables. *Technometrics*, 4, 531-550.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, series B*, 26, 211–246.
- Carroll, R.J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*, Chapman and Hall.
- Darlington, R. B. (1997). Correcting the standard errors of regression slopes for heteroscedasticity.
- Freeman, M. and Tukey, J. (1950). Transformations related to angular and the square root. *Ann. Math. Stat.*, 21, 607-611.
- Hamasaki, T. and Goto, M. (2005). Power-Transformation Model to Additivity in Regression, *Behaviormetrika*, Vol.32, No.2, 111–125.
- Ryan, T. P. (1997). *Modern Regression Methods*, Wiley.
- Weisberg, S. (2005). *Applied Linear Regression*, Third Edition, Willey.

İnternet Kaynakları

- 1-<http://www.psych.cornell.edu/darlington/heterosc.htm>
- 2-<http://www.rshm.saglik.gov.tr/hki/hkindex/index5.htm>.
- 3-<http://www.tuik.gov.tr/cevredagitimapp/hava.zul>