BÖLÜM 12

DEĞİŞKEN DÖNÜŞÜMLERİ

GİRİŞ

- Bağımlı veya bağımsız değişken dönüşümlerinin en küçük kareler regresyonunda kullanışlı olduğu birçok durum vardır.

- Bölüm 10'da en küçük karelerdeki problemlerden bazıları için olası bir çözüm olarak bağımlı değişken dönüşümü önerilmişti.

- Bu bölümde, bağımsız değişken üzerindeki dönüşümler dahil olmak üzere dönüşüm yapmanın nedenleri ve uygun dönüşümü yapmak için kullanılan metotlar tam olarak tartışılacaktır.

- Genelleştirilmiş en küçük kareler ve ağırlıklandırılmış en küçük kareler dönüştürülmüş bağımlı değişken üzerinden klasik en küçük kareler olarak görülebildiğinden dolayı bu bölüme dahil edilmiştir.

DÖNÜŞÜM YAPMANIN NEDENLERİ

- Regresyonda değişken dönüşümü yapmak için üç temel neden vardır.

- Bağımlı değişkendeki dönüşümlerin normal dağılmama ve hatalardaki değişen varyans için mümkün çözüm olduğu Bölüm 10'da gösterilmiştir.

- Dönüşüm yapmanın üçüncü nedeni bağımlı ve bağımsız değişkenler arasındaki ilişkileri basitleştirmektir.

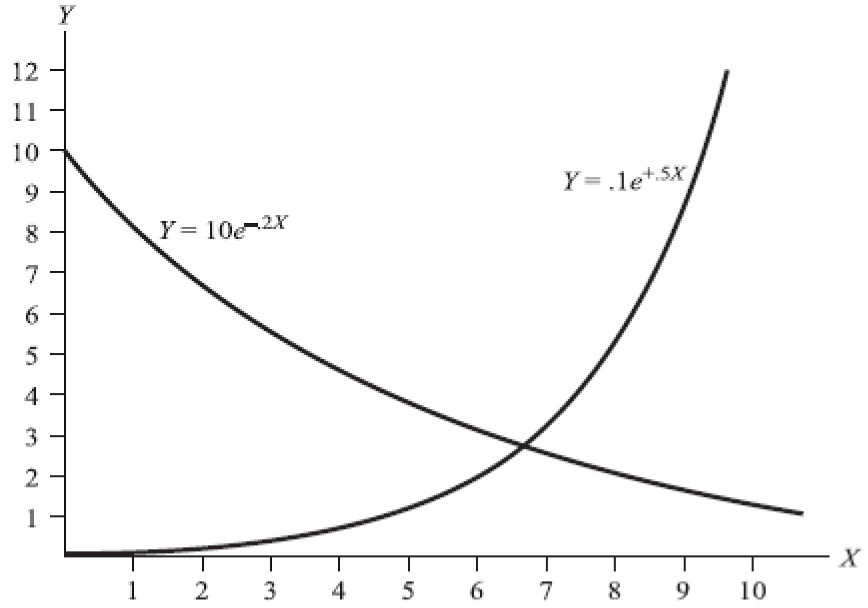
İLİŞKİLERİ BASİTLEŞTİRMEK İÇİN DÖNÜŞÜMLER

- İlişkileri basitleştirmek için dönüşümlerin ele alındığı iki durumu ayırt etmek yararlı olacaktır.

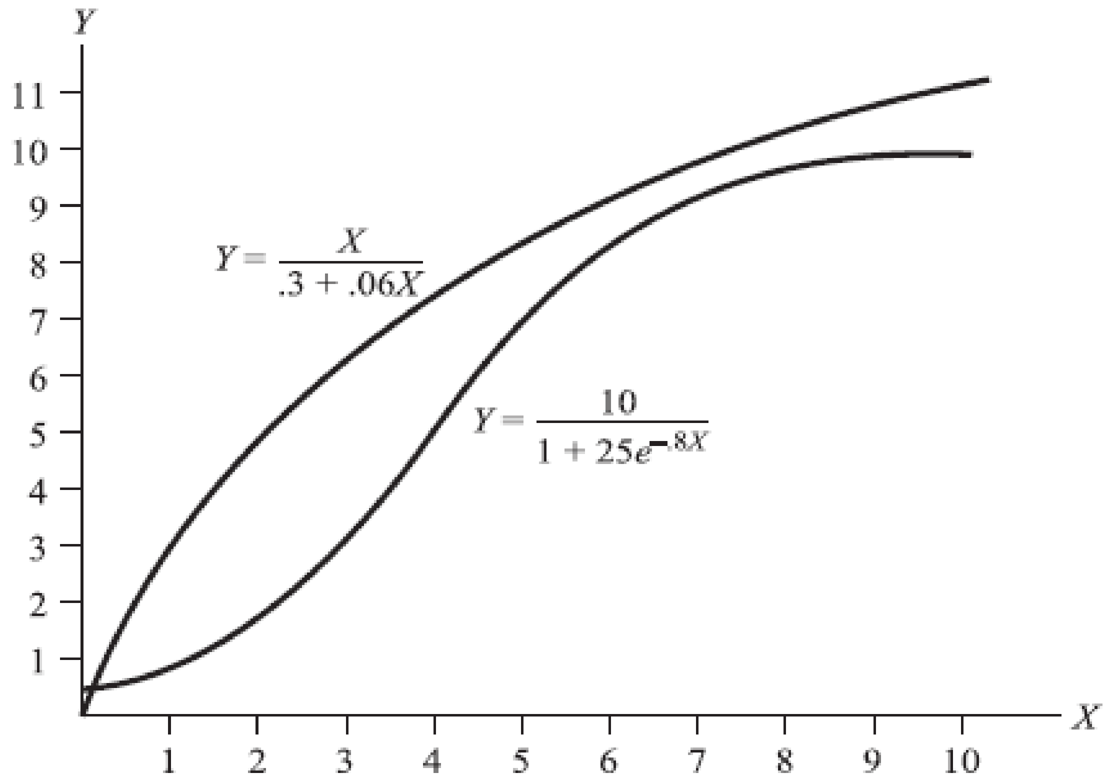
- Birinci durumda, modelin olması gereken biçimi hakkında önsel bir fikir yoktur.

- Amaç tercihen doğru şeklinde en basit formda gözlemlenen ilişkilerin temsil edilmesine izin veren bağımlı ve bağımsız değişkenlerin matematiksel biçimini ampirik olarak belirlemektir.

- Model parametrelerde doğrusal olmaktadır; sadece değişkenler ifade edildiği biçimde dikkate alınmaktadır.



ŞEKİL 12.1. *Üstel büyüme eğrisi ve üstel azalma eğrisinin örnekleri.*



ŞEKİL 12.2. *Ters polinom model ve lojistik modelin örnekleri.*

TABLO 12.1. *Tuzlu su baskınına 0, 72 veya 144 saat maruz kalan çam fidelerinin iğnesinin klorür içeriği (kuru ağırlık yüzdesi). Sekiz genetik ailenin her birinin dokuz fidesi tamamen rassal deney tasarımında kullanıldı. (Veri seti, S. B. Land, Jr. 1973 Doktora Tezinden, N. C. State University, izinde kullanılmıştır.)*

	Tuzlu su baskını saati								
Aile	0			72			144		
1	.36	.47	.30	3.54	4.35	4.88	6.13	6.49	7.04
2	.32	.63	.51	4.95	4.45	1.50	6.46	4.35	2.18
3	.00	.43	.72	4.26	3.89	6.54	5.93	6.29	9.62
4	.54	.70	.49	3.69	2.81	4.08	5.68	4.68	5.79
5	.44	.42	.39	3.01	4.08	4.54	6.06	6.05	6.97
6	.55	.57	.45	2.32	3.57	3.59	4.32	6.11	6.49
7	.20	.51	.27	3.16	3.17	3.75	4.79	5.74	5.95
8	.31	.44	.84	2.80	2.96	2.04	10.58	4.44	1.70

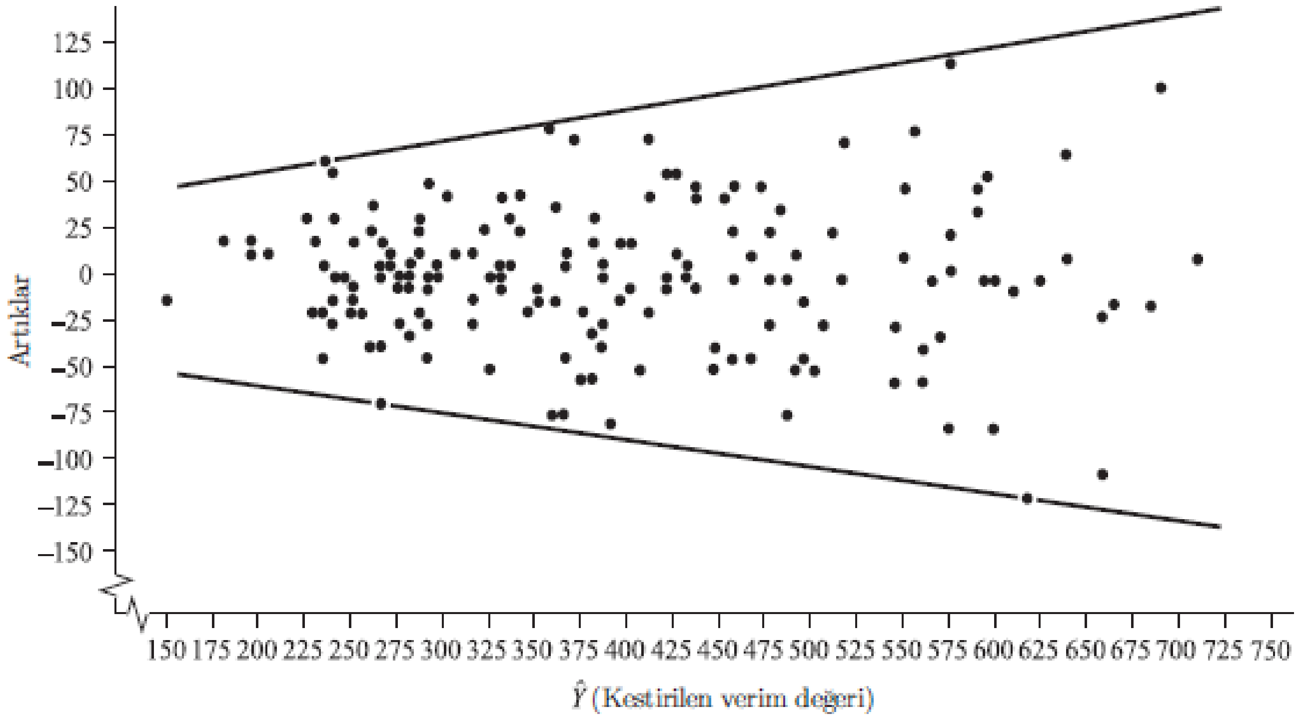
VARYANSI SABİTLEŞTİRMEK İÇİN DÖNÜŞÜMLER

Varyans ve ortalama normal olasılık dağılımında bağımsızdır. Tüm diğer ortak dağılımlarda ortalama ve varyans arasında doğrudan bir bağlantı vardır. Örneğin, sıklıkla sayma verileri ile ilişkili bir dağılımda, Poisson dağılımında varyans ortalamaya eşittir. Ortalamaya karşı Poisson varyansının şekli bir eğimli düz bir çizgi olacaktır. Binom dağılımlı rassal değişkenin varyansı $np(1-p)$ ve ortalaması np dir. Binomun ortalamaya karşı varyansının şekli $p=0$ ve $p=1$'de sıfır ve $p=1/2$'de maksimum varyans gösterecektir. Ki-kare dağılımının varyansı ortalamasının iki katına eşittir. Poisson gibi, bu ortalama ve varyans arasında doğrusal fakat daha dik bir eğimli ilişkidir. Muhtemelen, rassal değişken normal dağılımlı olmadığında varyansların değişen olması beklenecektir.

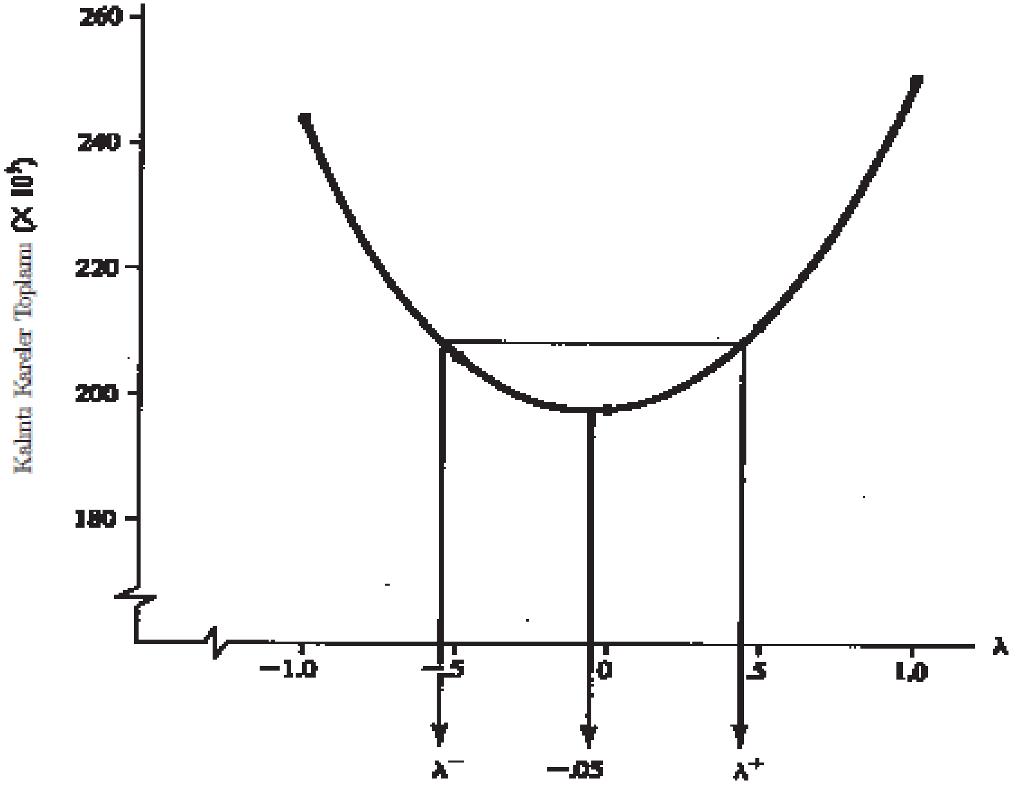
NORMALLİĞİ İYİLEŞTİMEK İÇİN DÖNÜŞÜMLER

- Normalliği iyileştirmek için dönüşümlere ilişkilerin basitleştirmesi veya varyansın sabitleştirilmesindeki dönüşümlerden daha az öncelik verilmektedir.

- En küçük kareler tahmini kendi başına normalliği gerektirmemesine ve normallikten ayrılışların ciddi olmadığı bilinmesine rağmen (Bartlett, 1947), normallikle ilgilenmek için yeterince neden vardır (bk. Kısım 10.2).



ŞEKİL 12.3. *Soya veriminin ozon ve sülfüt dioksitinin etkileri hakkındaki dört deneyinin birleştirilmiş analizinden (dönüştürülmemiş) $\widehat{Y}_i$'ye karşı e_i'nin grafiği.*



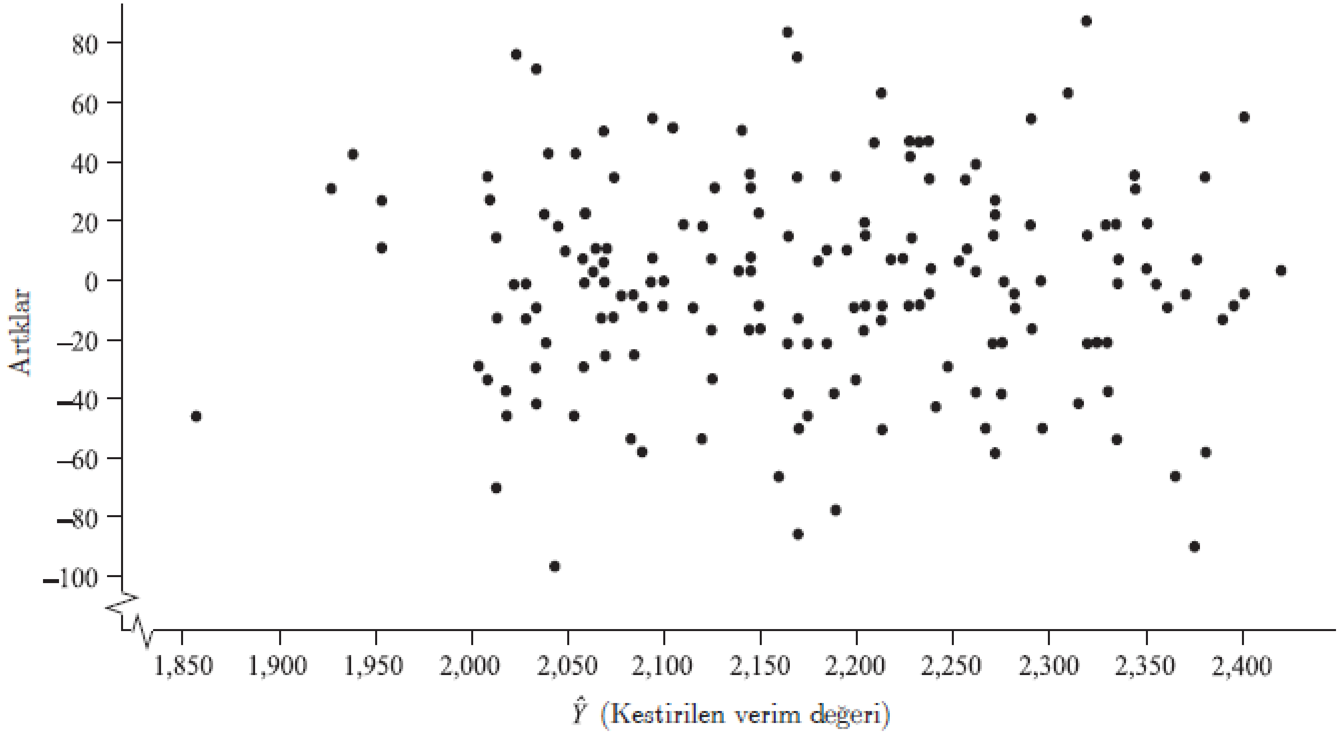
ŞEKİL 12.4. *Soya fasulyesi deneyinde Box-Cox dönüşümü için λ'ya karşılık çizilen artık kareler toplamı. λ'nın uygun % 95 güven aralığının tahminin üst ve alt sınırları sırasıyla λ^- ve λ^+ ile gösterilmiştir.*

GENELLEŞTİRİLMİŞ EN KÜÇÜK KARELER

- Homojen varyans varsayımını sağlamayan bağımlı değişkeni kullanmak için gerekli olan veya en azından arzu edilen durumlar olacaktır.

- Varyansı sabitleştirmek için gerekli dönüşüm Y ve X arasındaki ilişkiyi yok ettiğinden veya artıkların toplanırlılığı ve normal dağılmasını yok ettiğinden dolayı arzu edilemiyor.

- Varyansı yeterince sabitleştiremeyen bir dönüşüm veya değişen varyansın kaldığı bir ilişkiyi basitleştirmek için yapılan bir dönüşüm olabilir.



ŞEKİL 12.5. *Soya veriminin ozon ve sülfüt dioksitinin etkileri hakkındaki dört deneyinin birleştirilmiş analizinden logaritmik dönüşüm sonrası $\widehat{Y}_i$'ye karşı e_i'nin çizimi*

AĞIRLIKLANDIRILMIŞ EN KÜÇÜK KARELER

Doğrusal modelin

$$\begin{aligned} \mathbf{Var}(\boldsymbol{\epsilon}) &= \boldsymbol{V}\sigma^2 \\ &= \text{Diag}(\, a_1^2 \quad a_2^2 \quad \cdots \quad a_n^2\,)\,\sigma^2. \end{aligned}$$

ile birlikte

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{12.10}$$

olduğu varsayılsın. ϵ_i ve Y_i'nin varyansı $a_i^2\sigma^2$ ve tüm kovaryanslar sıfırdır.

Rassal değişken bir sabit ile çarpıldığı zaman rassal değişkenin varyansı değişir:

$$\begin{aligned} \sigma^2(cZ) &= \text{Var}(cZ) = c^2\text{Var}(Z) \\ &= c^2[\sigma^2(Z)], \end{aligned} \tag{12.11}$$

burada c bir sabittir. Şayet sabit Z'nin standart sapmasının karşılıklı oranı olarak seçilirse, $c = k/\sigma(Z)$, yeniden ölçeklendirilmiş değişkenin varyansı k^2 dir:

$$\sigma^2(cZ) = \left(\frac{k}{\sigma(Z)}\right)^2 \sigma^2(Z) = k^2. \tag{12.12}$$

GENELLEŞTİRİLMİŞ EN KÜÇÜK KARELER

Genelleştirilmiş en küçük kareler ϵ'nun keyfi pozitif tanımlı varyans-kovaryans matrisine $\mathbf{Var}(\epsilon) = \boldsymbol{V}\sigma^2$ izin vermek için klasik doğrusal modeli genişletiyor. Köşegen elemanlarının eşit ve köşegen dışı elemanların sıfır olması gerekli değildir. Bu pozitif tanımlı koşul gözlemlerin doğrusal fonksiyonunun pozitif varyanslı olacağı uygun bir varyans matrisi sağlamaktadır. Ağırlıklandırılmış en küçük kareler gibi, dönüştürülmüş model $\mathbf{Var}(\epsilon^*) = \boldsymbol{I}\sigma^2$'nin en küçük kareler varsayımlarını sağlaması için doğrusal dönüşüm Y'ye yapılmalıdır.

TABLO 12.2. *AR(1) modelinde β_0'ın genelleştirilmiş en küçük kareler tahminine göre klasik en küçük kareler tahmininin göreli etkinliği.*

n	.1	.3	.5	.7	.9	.95
25	.999	.993	.978	.947	.897	.909
50	1.000	.996	.988	.968	.906	.887
75	1.000	.997	.992	.977	.923	.890
100	1.000	.998	.994	.982	.936	.899

TABLO 12.3. *100 çam ağacının 54 inçteki çapına kütük çapının oranının logaritmasının DBH sınıfı tarafından ortalaması $\overline{Y}_{ij.}$, burada $Y_{ijk} = \ln(SD_{ijk}) - \ln(DBH_{ik})$. $c = .1$ için bağımsız değişken $X_j = 54^c - SHt^c$'nin değerleri son satırda gösterildi. (Veri B.J. Rawlings den, yayınlanmamış).*

DBH	*No.*	*Kütük Boyu (Yerden İnç.)*					
(*in.*)	*Ağaçlar*	2	4	6	8	10	12
6	4	.3435	.3435	.2715	.1438	.0719	.0719
8	16	.3143	.2687	.2548	.2294	.1674	.1534
10	42	.2998	.2514	.2083	.1733	.1463	.1209
12	26	.3097	.2705	.2409	.1998	.1790	.1466
14	9	.2121	.1859	.1597	.1449	.1039	.1039
16	3	.2549	.2549	.1880	.1529	.1529	.1529
X_j		.4184	.3415	.2940	.2590	.2313	.2081