BÖLÜM 1 : BASİT DOĞRUSAL REGRESYON

BASİT DOĞRUSAL REGRESYON MODELİ

y yanıt değişkeni ile doğrusal ilişkiye sahip tek bir x bağımsız değişkeninin bulunduğu model, **basit doğrusal regresyon modeli** olarak ele alınmaktadır. Bu model,

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad \textbf{(1.1)}$$

olarak tanımlanmaktadır. Burada β_0 kesim noktası ve β_1 eğim olmak üzere bilinmeyen sabitlerdir ve ε' da rasgele hata bileşenidir. Hataların sıfır ortalamaya ve bilinmeyen varyansa sahip oldukları ve ilişkisiz oldukları varsayılır.

- x değişkeninin her olası değerine karşılık y değişkeni, bir olasılık dağılımına sahiptir. Bu dağılımın ortalaması,

  $$E(y/x) = \beta_0 + \beta_1 x \qquad \textbf{(1.2a)}$$

  varyansı,

  $$Var(y/x) = Var(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2 \qquad \textbf{(1.2b)}$$

  olup y'nin varyansı x'in değerine bağlı olmasa da y'nin ortalaması, x'in doğrusal bir fonksiyonudur. Hatalar ilişkisiz olduğundan yanıtlar da ilişkisizdir.

β_0 ve β_1 parametreleri **"regresyon katsayıları"** olarak adlandırılır. β_1, **eğimi** yani x'teki bir birim değişiklikle elde edilen y'nin dağılımının ortalamasındaki değişikliği verir. β_0 ise **kesim noktası** olup $x = 0$ olduğunda y değişkeni dağılımının ortalamasını vermektedir.

PARAMETRELERİN EN KÜÇÜK KARELER KESTİRİMİ

β_0 ve β_1 bilinmeyen parametreler olup örneklem verileri kullanılarak kestirilmeleri gerekmektedir. $(y_1, x_1), (y_2, x_2), ..., (y_n, x_n)$ olmak üzere "n" sayıda veri çifti olduğu varsayılsın.

β_0 ve β_1'in Kestirimi

β_0 ve β_1 parametrelerini kestirmek için **"En Küçük Kareler Yöntemi"** kullanılır. Bu yöntemle β_0 ve β_1, y_i gözlemleri ile regresyon doğrusu arasındaki farkların karelerinin toplamı minimum olacak şekilde kestirilir.

Denklem (1.1) kullanılarak "n" sayıda veri çifti için,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , \quad i = 1, 2, ..., n \tag{1.3}$$

denklemi yazılabilir. Denklem (1.1), bir **kitle regresyon modeli** iken bu denklem ise bir **örneklem regresyon modelidir**.

En küçük kareler ölçütü,

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{1.4}$$

olup β_0 ve β_1 için en küçük kareler kestiricileri, $\hat{\beta}_0$ ve $\hat{\beta}_1$ aşağıdaki koşulları sağlamalıdır.

$$\frac{\partial S}{\partial \beta_0}\bigg|_{\hat{\beta}_0, \hat{\beta}_1} = -2\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{\partial S}{\partial \beta_1}\bigg|_{\hat{\beta}_0, \hat{\beta}_1} = -2\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

Bu koşulların sağlanması durumunda,

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\ \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} y_i x_i \end{aligned} \tag{1.5}$$

denklemleri elde edilir. Bu denklemler, **en küçük kareler normal denklemleri** olarak da adlandırılmaktadır. Normal denklemlerin çözümü sonrasında; $\hat{\beta}_0$ ve $\hat{\beta}_1$, sırasıyla kesim noktası ve eğimin **en küçük kareler kestiricileri** elde edilir.

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \tag{1.6}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \qquad \textbf{(1.7)}$$

Burada, $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ ve $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, y ve x'in ortalamalarıdır.

Oluşturulan basit doğrusal regresyon modeli,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \qquad \textbf{(1.8)}$$

olup bu denklem, belli bir x için y'nin ortalamasına ilişkin bir nokta kestirimini vermektedir. x'in düzeltilmiş kareler toplamı, S_{xx}, x_i ve y_i'nin çapraz çarpımlarının düzeltilmiş toplamı, S_{xy} ;

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = \sum_{i=1}^{n} (x_i - \overline{x})^2 \qquad \textbf{(1.9)}$$

$$S_{xy} = \sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n} = \sum_{i=1}^{n} y_i (x_i - \overline{x}) \qquad \textbf{(1.10)}$$

olmak üzere, Denklem (1.7) ;

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \qquad \textbf{(1.11)}$$

olarak tanımlanabilir.

y_i'nin gözlenen değeri ile buna karşılık gelen kestirilmiş (uydurulmuş) $\hat{y}_i$ değeri arasındaki fark **"artık"** olarak adlandırılır. Matematiksel olarak $i.$ artık,

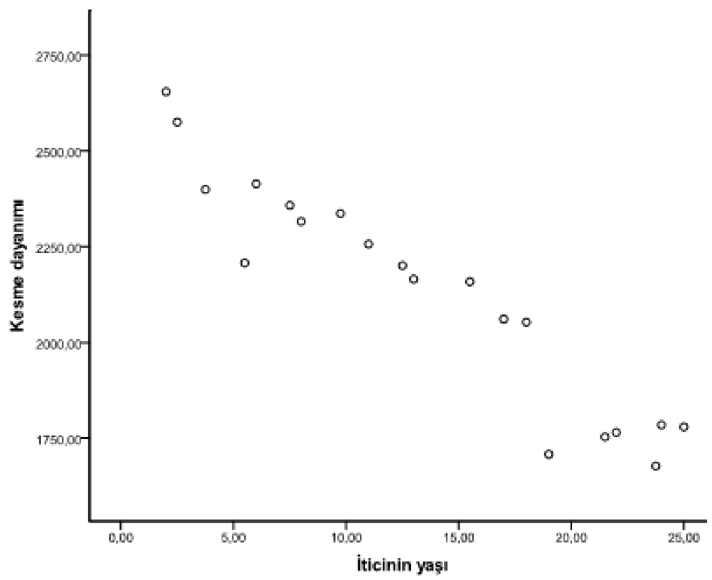
$$\begin{aligned} e_i &= y_i - \hat{y}_i \\ &= y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) \quad , \quad i = 1, 2, \ldots, n \end{aligned} \qquad \textbf{(1.12)}$$

Örnek 1.1. Roket Yakıtı Verileri

Bir roket motoru, ateşleme iticisi ve takviye iticisinin birlikte bir metal kaplama içinde birleştirilmesiyle üretilmektedir. İki itici güç arasındaki bağın kesme dayanımı önemli bir nitelik özelliğidir. Kesme dayanımının, takviye itici grubunun hafta olarak yaşıyla ilgili olduğundan şüphelenilmektedir. İlgili itici grubunun kesme dayanımı ve yaşı üzerine yirmi gözlem toplanmış ve Tablo 1.1 'de verilmiştir. Şekil 1.1'deki dağılım grafiği, kesme dayanımı ve itici yaşı arasında güçlü bir istatistiksel ilişki olduğunu ve $y = \beta_0 + \beta_1 x + \varepsilon$ doğrusal modeli geçici varsayımının mantıklı olduğunu göstermektedir.

Tablo 1.1. Örnek 1.1 'in verileri

Gözlem, i	Kesme Dayanımı, y_i (psi)	İticinin Yaşı x_i (Hafta)
1	2158.70	15.50
2	1678.15	23.75
3	2316.00	8.00
4	2061.30	17.00
5	2207.50	5.50
6	1708.30	19.00
7	1784.70	24.00
8	2575.00	2.50
9	2357.90	7.50
10	2256.70	11.00
11	2165.20	13.00
12	2399.55	3.75
13	1779.80	25.00
14	2336.75	9.75
15	1765.30	22.00
16	2053.50	18.00
17	2414.40	6.00
18	2200.50	12.50
19	2654.20	2.00
20	1753.70	21.50



Şekil 1.1. İticinin yaşına karşı kesme dayanımının dağılım grafiği, örnek 1.1.

$$S_{xx} = \sum_{i=1}^{n} x_i^{2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 4677.69 - \frac{71422.56}{20} = 1106.56$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n20} = 528492.64 - \frac{(267.25)(42627.15)}{20} = -41112.65$$

olmak üzere,

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-41112.65}{1106.56} = -37.15$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = 2131.3575 - (-37.15)(13.3625) = 2627.82$$

olup Tablo 1.2. kullanılarak en küçük kareler uyumu,

$$\hat{y} = 2627.82 - 37.15x$$

eşitliği şeklinde elde edilir. İticinin yaşına bağlı olarak iticinin kesme dayanımındaki haftalık ortalama azalma 37.15 birimdir. 2627.82 kesim noktası, üretimden hemen sonra bir itici grubunun kesme dayanımını temsil etmektedir.

TABLO 1.2. Örnek 1.1 İçin Veriler, Uydurulan Değerler ve Artıklar

Gözlenen Değer, y_i	Uydurulan Değer, $\hat{y}_i$	Artık, e_i
2158.70	2051.94	106.76
1678.15	1745.42	– 67.27
2316.00	2330.59	– 14.59
2061.30	1996.21	65.09
2207.50	2423.48	– 215.98
1708.30	1921.90	– 213.60
1784.70	1736.14	48.56
2575.00	2534.94	40.06
2357.90	2349.17	8.73
2256.70	2219.13	37.57
2165.20	2144.83	20.37
2399.55	2488.50	– 88.95
1799.80	1698.98	80.82
2336.75	2265.58	71.17
1765.30	1810.44	– 45.14
2053.50	1959.06	94.44
2414.40	2404.90	9.50
2200.50	2163.40	37.10
2654.20	2553.52	100.68
1753.70	1829.02	– 75.32
$\sum y_i = 42627.15$	$\sum \hat{y}_i = 42627.15$	$\sum e_i = 0.00$

TABLO 1.3 Örnek 1.1'in Minitab Regresyon Çıktısı

Regression Analysis

The regression equation is
Strength = 2628 – 37.2 Age

Predictor	Coef	StDev	T	P
Constant	2627.82	44.18	59.47	0.000
Age	– 37.154	2.889	– 12.86	0.000

S = 96.11 R-Sq = % 90.2 R-Sd(adj) = % 89.6

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1527483	1527483	165.38	0.000
Error	18	166255	9236		
Total	19	1693738			

Ödev

12 kişiye ilişkin kan basıncı (Y) ve yaşları (X) aşağıdaki tabloda verilmiştir.

Yaş	56	42	72	36	63	47	55	49	38	42	68	60
Kan Basıncı	147	125	160	118	149	128	150	145	115	140	152	155

a) Yukarıdaki problemde verilen verilere göre kullanılacak doğrusal regresyon modeli ile ilgili parametrelerin tahminlerini bulunuz.

b) Bulduğunuz tahminlere göre bir kestirim denklemi elde ediniz ve bu denklemden $\widehat{Y_i}$,$\hat{e}_i$ tahmin değerlerini bulunuz.