BELİRTME KATSAYISI

x bağımsız değişkenini göz önünde bulundurmadan y'deki değişkenliğin ölçüsü, SS_T ve x göz önünde bulundurulduğunda y'deki değişkenliğin ölçüsü, SS_R olmak üzere; **belirtme katsayısı**,

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{\mathrm{Res}}}{SS_T} \qquad \textbf{(1.43)}$$

eşitliği kullanılarak elde edilir.

- R^2, x bağımsız değişkeni tarafından açıklanan y'deki değişkenliğin oranıdır.
- $0 \leq SS_{\mathrm{Res}} \leq SS_T$ olduğundan $0 \leq R^2 \leq 1$'dir.
- R^2'nin 1'e yakın değerleri y'deki değişkenliğin büyük kısmının regresyon modeliyle açıklandığını belirtmektedir.

Örnek 1.1'deki roket itici verileri için oluşturulan regresyon modeli için,

$$R^2 = \frac{SS_R}{SS_T} = \frac{1,527,334.95}{1,693,737.60} = 0.9018$$

olup kesme dayanımındaki değişkenliğin % 90.18'i model tarafından açıklanmaktadır.

- R^2, regresyon doğrusunun eğiminin büyüklüğünü ölçmez. Ayrıca doğrusal modelin uygunluğunu da ölçmez. y ve x, doğrusal olmayan bir biçimde ilişkili olsalar bile R^2 genellikle büyük olacaktır.

- Delta yöntemi kullanılarak doğrusak regresyonda beklenen R^2 değerinin yaklaşık,

$$E(R^2) = \frac{\beta_1^2 S_{xx} / n-1}{\dfrac{\beta_1^2 S_{xx}}{n-1} + \sigma^2}$$

REGRESYONUN KULLANIMINA İLİŞKİN BAZI DÜŞÜNCELER

Regresyonun sözü edilmesi gereken bazı yanlış kullanımları vardır:

1) Regresyon modelleri, modeli uydurmak için kullanılan bağımsız değişkenler aralığının ötesinde dış değer bulmak için kullanılmaktadır.

2) x değerlerinin eğilimi en küçük kareler uyumunda önemli bir rol oynamaktadır.

3) Aykırı değerler verilerin geri kalanından ciddi biçimde farklılaşan gözlemlerdir. Bu değerler en küçük kareler uyumunu ciddi biçimde bozabilirler.

4) Değişkenler arasındaki neden-sonuç ilişkileri regresyonla belirlenmemelidir.

5) Regresyonun bazı uygulamalarında, y 'yi kestirmek için gereken x bağımsız değişken değeri bilinmemektedir.

ORİJİNDEN GEÇEN REGRESYON

Kesim noktasız regresyon modeli, genellikle kimyasal üretim süreçleri ve diğer üretim süreçlerinden elde edilen verileri analiz etmek için kullanılmaktadır. Örneğin, işlem sıcaklığı sıfır olduğunda, kimyasal bir sürecin ürünü de sıfır olmaktadır.

Kesim noktasız regresyon modeli,

$$y = \beta_1 x + \varepsilon \qquad \textbf{(1.44)}$$

olup "n" sayıda (y_i, x_i), $i = 1, 2, ..., n$ gözlemi için en küçük kareler fonksiyonu,

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

eşitliği ile tanımlanır.

Tek normal denklem,

$$\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \qquad \textbf{(1.45)}$$

ile elde edilir. Eğimin en küçük kareler kestiricisi,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \qquad \textbf{(1.46)}$$

olup bu kestirici, β_1 parametresi için **yansız** bir kestiricidir.

Uydurulan regresyon modeli,

$$\hat{y} = \hat{\beta}_1 x \qquad \textbf{(1.47)}$$

olup σ^2 'nin kestiricisi ise

$$\hat{\sigma}^2 = MS_{\mathrm{Res}} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - \hat{\beta}_1 \sum_{i=1}^{n} y_i x_i}{n-1} \qquad \textbf{(1.48)}$$

eşitliği kullanılarak hesaplanır.

Hatalar üzerinde normallik varsayımı yapılarak hipotezler test edilebilir ve kesim noktasız model için güven ve önkestirim aralıkları oluşturulabilir. β_1 **üzerindeki yüzde** $100(1-\alpha)$ **güven aralığı,**

$$\hat{\beta}_1 - t_{\alpha/2,n-1}\sqrt{\frac{MS_{\mathrm{Res}}}{\sum_{i=1}^{n} x_i^2}} \le \beta_1 \le \hat{\beta}_1 + t_{\alpha/2,n-1}\sqrt{\frac{MS_{\mathrm{Res}}}{\sum_{i=1}^{n} x_i^2}} \qquad \textbf{(1.49)}$$

olup $x = x_0$ olduğunda $E(y/x_0)$ **yanıt ortalaması üzerindeki yüzde** $100(1-\alpha)$ **güven aralığı** da aşağıdaki gibidir :

$$\hat{\mu}_{y/x_0} - t_{\alpha/2,n-1}\sqrt{\frac{x_0^2 MS_{\mathrm{Res}}}{\sum_{i=1}^{n} x_i^2}} \le E(y/x_0) \le \hat{\mu}_{y/x_0} + t_{\alpha/2,n-1}\sqrt{\frac{x_0^2 MS_{\mathrm{Res}}}{\sum_{i=1}^{n} x_i^2}} \qquad \textbf{(1.50)}$$

$x = x_0$ **olduğunda** y_0 **gelecek gözlemi üzerindeki yüzde** $100(1-\alpha)$ **önkestirim aralığı** şöyledir:

$$\hat{y}_0 - t_{\alpha/2,n-1}\sqrt{MS_{\mathrm{Res}}\left(1+\frac{x_0^2}{\sum_{i=1}^{n} x_i^2}\right)} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-1}\sqrt{MS_{\mathrm{Res}}\left(1+\frac{x_0^2}{\sum_{i=1}^{n} x_i^2}\right)} \qquad \textbf{(1.51)}$$

Hem (1.50) güven aralığı hem de (1.51) önkestirim aralığı x_0 arttıkça genişlemektedir. Eğer $\beta_0 = 0$ varsayımı kesim noktalı modelle reddedilemiyorsa bu uyum kesim noktasız model kullanılarak iyileştirilebilir.

- Artık kareler ortalaması uyumun kalitesini ölçmenin iyi bir yoludur. En küçük artık kareler ortalamasına sahip model, regresyon doğrusu etrafındaki y'nin varyansının kestirimini minimuma indirdiği için en iyi uyum olmaktadır.

Kesim noktası olan model için R^2,

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = \frac{regresyon\ ile\ açıklanan\ y'deki\ değişim}{y'deki\ toplam\ gözlenen\ değişim}$$

eşitliği kullanılarak hesaplanırken orjinden geçen model için ise

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} \hat{y}_i^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

olmak üzere,

$$R_0^2 = \frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} y_i^2}$$

eşitliği ile hesaplanır. Bu istatistik, orijin(sıfır) etrafındaki değişkenliğin regresyon tarafından açıklandığını göstermektedir.

Orijinden geçen model için R^2 tanımlamanın farklı yolları vardır. Bunlardan biri şöyledir:

$$R_0^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Ancak, $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ 'nin büyük olduğu durumlarda, R^2 negatif olabilir.

Örnek 1.8 Raf Stoklama Verileri

Bir satıcının bir bakkal dükkanının rafını meşrubatlar ile doldurması için gereken süre ve stoklanan ürün paket sayısı Tablo 1.10'da verilmiştir.

TABLO 1.7 Örnek 1.8 İçin Raf Stoklama Verileri

Süre, y (dakika)	**Stok Sayısı, x**
10.15	25
2.96	6
3.00	8
6.88	17
0.28	2
5.06	13
9.14	23
11.86	30
11.69	28
6.04	14
7.57	19
1.74	4
9.38	24
0.16	1
1.84	5

Kesim noktasının olmadığı modeldeki eğim,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} = \frac{1841.98}{4575} = 0.4026$$

olarak hesaplanır.

Verilere uydurulan model,

$$\hat{y} = 0.4026x$$

olup bu model için artık kareler ortalaması, $MS_{\mathrm{Res}} = 0.0893$ ve $R^2 = 0.9883$'tür. $H_0 : \beta_1 = 0$ hipotezi için test istatistiği $t_0 = 91.13$ (P değeri 8.02 × 10^{-21}) olarak elde edilmiştir. Bu özet istatistikler orijinden geçen model için göze çarpan bir yetersizlik göstermemektedir.

Karşılaştırma amacıyla verilere kesim noktalı model de uydurulabilir.

$$\hat{y} = -0.0938 + 0.4071x$$

$H_0 : \beta_0 = 0$ hipotezini test etmek için kullanılan t istatistiği $t_0 = -0.65$ 'tir ve anlamlı değildir. Bu durum kesim noktasız modelin daha üstün bir uyum olduğunu göstermektedir.

Kesim noktası olan model için artık kareler ortalaması $MS_{\mathrm{Re}s} = 0.0931$ ve $R^2 = 0.9997$ 'dir. Kesim noktası olmayan model için $MS_{\mathrm{Re}s}$ değeri önceki modelden daha küçük olduğu için orijinden geçen modelin daha üstün olduğu sonucuna varılabilir. R^2 istatistikleri doğrudan karşılaştırılamaz.