ÇOKLU REGRESYONDA GÜVEN ARALIKLARI

Regresyon Katsayılarının Güven Aralıkları

$y_i \;\; (i = 1, 2, ..., n)$ gözlemlerinin, $\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$ ortalama ve σ^2 varyansıyla birbirinden bağımsız ve normal dağıldığı varsayılsın. Herhangi bir β_j regresyon katsayısının marjinal dağılımı, β_j ortalamalı ve $\sigma^2 C_{jj}$ varyanslı normal dağılımdır.

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \quad , \quad j = 0, 1, ..., k \qquad \textbf{(2.26)}$$

istatistiklerinin her biri $n - p$ serbestlik dereceli t dağılır. C_{jj}, $(X'X)^{-1}$ matrisinin j. köşegen elemanıdır.

β_j , $j = 0, 1, ..., k$ regresyon katsayısı için yüzde $100(1-\alpha)$ güven aralığı,

$$\hat{\beta}_j - t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \le \beta_j \le \hat{\beta}_j + t_{\alpha/2, n-p} \sqrt{\hat{\sigma}^2 C_{jj}} \qquad \textbf{(2.27)}$$

olarak kullanılır.

Örnek 2.8 Teslim Süresi Verileri

Örnek 2.1'de β_1 parametresi için %95 güven aralığı,

$$\hat{\beta}_1 - t_{0.025, 22} \sqrt{\hat{\sigma}^2 C_{11}} \le \beta_1 \le \hat{\beta}_1 + t_{0.025, 22} \sqrt{\hat{\sigma}^2 C_{11}}$$
$$1.61591 - (2.074)(0.17073) \le \beta_1 \le 1.61591 + (2.074)(0.17073)$$
$$1.26181 \le \beta_1 \le 1.97001$$

olarak elde edilir.

$$x_0 = \begin{bmatrix} 1 \\ x_{01} \\ x_{02} \\ \cdots \\ x_{0k} \end{bmatrix} \qquad \text{ve} \qquad \hat{y}_0 = x_0 ' \hat{\beta} \qquad \textbf{(2.28)}$$

olmak üzere, $x_{01}, x_{02}, ..., x_{0k}$ noktasında **ortalama yanıt için yüzde $100(1-\alpha)$ güven aralığı,**

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 x_0'(X'X)^{-1}x_0} \le E(y/x_0) \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 x_0'(X'X)^{-1}x_0} \quad \textbf{(2.29)}$$

Örnek 2.9 Teslim Süresi Verileri

Örnek 2.1'deki teslim süresi verileri kullanıldığında $x_0 = \begin{bmatrix} 1 \\ 8 \\ 275 \end{bmatrix}$ olmak üzere, bu nokta için ortalama teslim süresinin ortalama güven aralığı,

$$19.22 - 2.074\sqrt{0.56794} \ \le \ E(y/x_0) \ \le \ 19.22 + 2.074\sqrt{0.56794}$$

$$17.66 \le E(y/x_0) \le 20.78$$

olarak elde edilir.

Regresyon Katsayılarının Eş Zamanlı Güven Aralıkları

- Aynı örneklem verisi kullanılarak güven ve ön kestirim aralıkları oluşturulmak istendiğinde aralık kestirimlerinin **bütün kümesi** için **aynı anda** geçerli bir güven katsayısı belirlemek gerekir. Tümünün $1-\alpha$ olasılıkla aynı anda doğru olduğu güven aralıkları ya da ön kestirim aralıkları kümesi, **eş zamanlı** ya da **bileşik güven**(ya da bileşik ön kestirim) aralıkları olarak adlandırılır.

Tüm β parametreleri için yüzde $100(1-\alpha)$ **bileşik güven bölgesi**,

$$P\left\{ \frac{\left(\hat{\beta}-\beta\right)' X'X \left(\hat{\beta}-\beta\right)}{pMS_{\text{Res}}} \ \le \ F_{\alpha,p,n-p} \right\} = 1-\alpha \quad \textbf{(2.30)}$$

olarak kullanılır. Bu eşitsizlik eliptik şekilli bir bölgeyi tanımlar.

Basit doğrusal regresyon için yukarıda belirtilen eşitsizlik aşağıdaki biçime indirgenir.

$$\frac{n(\hat{\beta}_0-\beta_0)^2 + 2\sum_{i=1}^{n} x_i(\hat{\beta}_0-\beta_0)(\hat{\beta}_1-\beta_1) + \sum_{i=1}^{n} x_i^2(\hat{\beta}_1-\beta_1)^2}{2MS_{\text{Res}}} \le F_{\alpha,2,n-2}$$

- Doğrusal regresyon modelinde parametrelerin eş zamanlı aralık kestirimlerini elde etmek için bir genel yaklaşım daha vardır. Bu aralıklar aşağıdaki ifade kullanılarak oluşturulabilir:

$$\hat{\beta}_j \pm \Delta\, se(\hat{\beta}_j) \quad , \quad j = 0,1,\dots,k \quad \textbf{(2.31)}$$

Burada Δ, tüm aralıkların doğru olma olasılığı elde edilecek şekilde seçilmelidir. Δ'yı seçmek için birkaç yöntem kullanılabilir. Bu yöntemlerden biri, **Bonferroni güven aralığı'**dır. Bu yaklaşımda, $\Delta = t_{\alpha/2p,n-p}$ olarak alınır.

Bonferroni güven aralığı,

$$\hat{\beta}_j \pm t_{\alpha/2p,n-p} se(\hat{\beta}_j) \quad , \quad j = 0,1,...,k \tag{2.32}$$

olup güven katsayısı olarak $1-\alpha$ yerine $1-\alpha/p$'yi kullanmaktadır.

Örnek 2.10 Roket Yakıtı Verileri

β_0 ve β_1 parametreleri için %90 bileşik güven aralığı,

$$\hat{\beta}_0 - t_{0.0125,18} se(\hat{\beta}_0) \le \beta_0 \le \hat{\beta}_0 + t_{0.0125,18} se(\hat{\beta}_0)$$
$$2627.822 - (2.445)(44.184) \le \beta_0 \le 2627.822 + (2.445)(44.184)$$
$$2519.792 \le \beta_0 \le 2735.852$$

$$\hat{\beta}_1 - t_{0.0125,18} se(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{0.0125,18} se(\hat{\beta}_1)$$
$$-37.154 - (2.445)(2.889) \le \beta_1 \le -37.154 + (2.445)(2.889)$$
$$-44.218 \le \beta_1 \le -30.090$$

olarak elde edilir.

- Δ'nın seçiminde kullanılan tek yöntem, Bonferroni yöntemi değildir. Diğer yaklaşımlar, Δ'yı

$$\Delta = \left(2F_{\alpha,p,n-p}\right)^{1/2}$$

olarak alan **Scheffe S yöntemi** ve

$$\Delta = u_{\alpha,p,n-p}$$

olarak alan **maksimum modül t yöntemidir.** Burada $u_{\alpha,p,n-p}$, her biri $n-2$ serbestlik dereceli iki bağımsız student-t raslantı değişkenlerinin maksimum mutlak değer dağılımının üst α kuyruk noktasıdır.

*** Genellikle Bonferroni aralıkları, Scheffe aralıklarından daha kısadır; maksimum modül t aralıkları ise Bonferroni aralıklarından daha kısadır.

YENİ GÖZLEMLERİN ÖNKESTİRİMİ

$x_0^{'} = [1, x_{01}, x_{02}, ..., x_{0k}]$ iken $x_{01}, x_{02}, ..., x_{0k}$ noktasında **gelecekteki bir** y_0 **gözleminin nokta kestirimi,**

$$\hat{y}_0 = x_0^{'}\hat{\beta} \quad \textbf{(2.33)}$$

olmak üzere bu gözlem için **yüzde 100(1** $-\alpha$**) önkestirim aralığı,**

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1 + x_0^{'}(X'X)^{-1}x_0} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2(1 + x_0^{'}(X'X)^{-1}x_0} \quad \textbf{(2.34)}$$

olarak kullanılır.

Örnek 2.11 Teslim Süresi Verileri

Teslim hacmi $x_1 = 8$ ve servis elemanının yürümüş olduğu mesafe $x_2 = 275$ olduğunda ortalama teslim süresi için %95 önkestirim aralığı, $x_0^{'} = [1, \ 8, \ 275]$, $\hat{y}_0 = x_0'\hat{\beta} = 19.22$ ve $x_0'(X'X)^{-1}x_0 = 0.05346$ olmak üzere,

$$19.22 - 2.074\sqrt{10.6239(1 + 0.05346)} \le y_0 \le 19.22 + 2.074\sqrt{10.6239(1 + 0.05346)}$$

ve % 95 önkestirim aralığı,

$$12.28 \le y_0 \le 26.16$$

olarak bulunur.

Ödev

Öğrencilerin başarı durumları ile IQ düzeyleri arasında bir ilişki olduğu varsayılıyor.

X_1; Öğrencilerin IQ düzeyleri

X_2; Çalışma süresi

Y; Notlar

olmak üzere,

Not (Y)	75	79	68	85	91	79	98	76
IQ (X_1)	105	110	120	116	122	130	114	102
Çalışma Sür. (X_2)	10	12	6	13	16	8	20	15
X_1X_2	1050	1320	720	1508	1952	1040	2280	1530

Verilerin uyduğu çoklu regresyon modeli

$Y_i = \alpha_0 + \alpha_1 x_{1,i} + \alpha_2 x_{2,i} + \alpha_3 x_{1,i} x_{2,i} + e_i, \;\; i = 1,2,3,\dots,n$ şeklinde olsun.

a) Verilere göre regresyon parametrelerinin en küçük kareler tahmin değerlerini bulup modeli oluşturunuz.
b) Kestrim ve artık değerlerini hesaplayınız.
c) ANOVA tablosunu oluşturunuz. $\alpha = 0{,}05$ düzeyinde modelin anlamlılığını inceleyiniz. R^2 değerini hesaplayıp yorumlayınız.
d) Parametreler için $\alpha = 0{,}05$ düzeyinde hipotez testi yapınız ve güven aralıklarını bulunuz.