BASİT DOĞRUSAL REGRESYONDA ARALIK KESTİRİMİ

β_0, β_1 ve σ^2 İçin Güven Aralıkları

Hatalar normal ve bağımsız olarak dağılmışsa o zaman hem $(\hat{\beta}_1 - \beta_1)/se(\hat{\beta}_1)$ hem de $(\hat{\beta}_0 - \beta_0)/se(\hat{\beta}_0)$'nın örnekleme dağılımları $n-2$ serbestlik dereceli t olmaktadır. Dolayısıyla β_1 eğimi üzerindeki ve β_0 kesim noktası üzerindeki $100(1-\alpha)$ güven aralığı (CI) sırasıyla,

$$\hat{\beta}_1 - t_{\alpha/2,n-2}\, se(\hat{\beta}_1) \;\le\; \beta_1 \;\le\; \hat{\beta}_1 + t_{\alpha/2,n-2}\, se(\hat{\beta}_1) \qquad \textbf{(1.35)}$$

$$\hat{\beta}_0 - t_{\alpha/2,n-2}\, se(\hat{\beta}_0) \;\le\; \beta_0 \;\le\; \hat{\beta}_0 + t_{\alpha/2,n-2}\, se(\hat{\beta}_0) \qquad \textbf{(1.36)}$$

olup aynı x değeriyle aynı büyüklükte tekrarlanan örneklemler ele alınıp her bir örneklemden eğim için oluşturulan güven aralıklarının % 95'i β_1'in gerçek değerini içerecektir. Eğer hatalar normal ve bağımsız olarak dağılmışsa; $(n-2)MS_{\text{Res}}/\sigma^2$'nin örnekleme dağılımı, (n-2) serbestlik derecesine sahip ki-kare dağılımı olur.

$$P\left(\chi^2_{1-\alpha/2,\,n-2} \le \frac{(n-2)MS_{\text{Res}}}{\sigma^2} \le \chi^2_{\alpha/2,\,n-2}\right) = 1-\alpha$$

$$P\left(\frac{(n-2)MS_{\text{Res}}}{\chi^2_{\alpha/2,n-2}} \le \sigma^2 \le \frac{(n-2)MS_{\text{Res}}}{\chi^2_{1-\alpha/2,n-2}}\right) = 1-\alpha \qquad \textbf{(1.37)}$$

Örnek 1.5 Roket Yakıtı Verileri

$se(\hat{\beta}_1) = 2.89$ ve $t_{0.025,18} = 2.101$ olmak üzere eğime (β_1) yönelik %95 güven aralığı,

$$\hat{\beta}_1 - t_{0.025,18} se(\hat{\beta}_1) \;\le\; \beta_1 \;\le \hat{\beta}_1 + t_{0.025,18} se(\hat{\beta}_1)$$

$$-\,37.15 \;-(2.101)(2.89) \;\le\; \beta_1 \;\le -\; 37.15 + (2.101)(2.89)$$

$$-\,43.22 \;\le\; \beta_1 \;\le -31.08$$

Bu türden güven aralıklarının % 95'i eğimin gerçek değerini içerecektir.

- Eğer α için farklı bir değer seçilseydi ortaya çıkan güven aralıklarının genişliği farklı olurdu.

  - ✓ β_1 parametresi için % 90 güven aralığı $-42.16 \le \beta_1 \le -32.14$ $\rightarrow$ % 95 güven aralığından daha dardır.

  - ✓ β_1 parametresi için % 99 güven aralığı $-45.49 \le \beta_1 \le -28.81$ $\rightarrow$ %95'ten daha geniştir.

Güven katsayısı $(1-\alpha)$ ne kadar büyükse güven aralığı da o kadar geniş olmaktadır.

σ^2 üzerindeki % 95 güven aralığı;

$$\frac{18(9244.59)}{\chi^2_{0.025,18}} \le \sigma^2 \le \frac{18(9244.59)}{\chi^2_{0.975,18}}$$

$$\frac{18(9244.59)}{31.5} \le \sigma^2 \le \frac{18(9244.59)}{8.23}$$

$$5282.62 \le \sigma^2 \le 20219.03$$

Ortalama Yanıtın Aralık Kestirimi

Regresyon modelinin başlıca bir kullanımı, x bağımsız değişkeninin belli bir değeri için $E(y)$ yanıt ortalamasının kestirilmesidir. x_0, $E(y/x_0)$ ortalama yanıtını kestirmek istediğimiz bağımsız değişkenin değeri olmak üzere kestirilmiş modelden $E(y/x_0)$'in yansız nokta kestiricisi,

$$\widehat{E(y/x_0)} = \hat{\mu}_{y/x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0 \qquad \textbf{(1.38)}$$

$\hat{\mu}_{y/x_0}$'nın varyansı aşağıdaki gibidir :

$$Var(\hat{\mu}_{y/x_0}) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_0) = Var\left[\bar{y} + \hat{\beta}_1(x_0 - \bar{x})\right]$$

$$= \frac{\sigma^2}{n} + \frac{\sigma^2 (x_0 - \bar{x})^2}{S_{xx}} = \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]$$

$Cov(\bar{y}, \hat{\beta}_1) = 0$ olmak üzere,

$$\frac{\hat{\mu}_{y/x_0} - E(y/x_0)}{\sqrt{MS_{\mathrm{Res}}(1/n + (x_0 - \overline{x})^2 / S_{xx})}}$$

denkleminin örnekleme dağılımı (n-2) serbestlik dereceli t'dir. Sonuç olarak $x = x_0$ noktasındaki ortalama yanıt için yüzde 100(1-α) güven aralığı;

$$\hat{\mu}_{y/x_0} - t_{\alpha/2,n-2}\sqrt{MS_{\mathrm{Res}}\left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)} \le E(y/x_0) \le \hat{\mu}_{y/x_0} - t_{\alpha/2,n-2}\sqrt{MS_{\mathrm{Res}}\left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)}$$

(1.39)

$E(y/x_0)$ için güven aralığının genişliği x_0'ın bir fonksiyonudur. $x_0 = \overline{x}$ için aralık genişliği minimumdur ve $|x_0 - \overline{x}|$ arttıkça genişlik de artmaktadır.

Örnek 1.6 Roket Yakıtı Verileri

Örnek 1.1'deki roket yakıtı verileri üzerinden $E(y/x_0)$ için % 95 güven aralığı, $x_0 = \overline{x} = 13.3625$ ve $\hat{y}_0 = \hat{\mu}_{y/x_0} = 2131.40$ olmak üzere,

$$2086.230 \le E(y/13.3625) \le 2176.571$$

olarak elde edilir.

TABLO 1.6 x_0'ın Bazı Değerleri İçin $E(y/x_0)$'nin Güven Sınırları

Alt Güven Sınırı	x_0	Üst Güven Sınırı
2438.919	3	2593.821
2341.360	6	2468.481
2241.104	9	2345.836
2136.098	12	2227.942
2086.230	$\overline{x} = 13.3625$	2176.571
2024.318	15	2116.822
1905.890	18	2012.351
1782.928	21	1912.412
1657.395	24	1815.045

YENİ GÖZLEMLERİN ÖNKESTİRİMİ

Eğer x_0 ilgilenilen bağımsız değişkenin değeri ise y_0 yanıtının yeni değerinin nokta kestirimi,

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \tag{1.40}$$

Gelecekteki bir y_0 gözleminin aralık kestirimi elde edilmek istensin.

$$\psi = y_0 - \hat{y}_0$$

raslantı değişkeninin y_0 gelecek gözlemi $\hat{y}_0$'dan bağımsız olduğu için sıfır ortalama ve

$$Var(\psi) = Var(y_0 - \hat{y}_0) = \sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right]$$

varyanslı ile Normal dağılıma sahiptir. x_0 noktasındaki bir gelecek gözlem için yüzde 100(1-α), önkestirim aralığı,

$$\hat{y}_0 - t_{\alpha/2,n-2}\sqrt{MS_{\text{Res}}\left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-2}\sqrt{MS_{\text{Res}}\left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)} \tag{1.41}$$

Bu önkestirim aralığı, $x_0 = \overline{x}$'de en küçük genişliktedir ve $|x_0 - \overline{x}|$ arttıkça aralık da genişler. (1.41) ve (1.39) karşılaştırıldığında önkestirim aralığı hem uydurulan modeldeki hataya hem de gelecek gözlemlerle ilgili hataya bağlı olduğu için x_0'daki önkestirim aralığı, her zaman x_0'daki güven aralığından geniştir.

Örnek 1.7 Roket Yakıtı Verileri

10 haftalık bir sevk iticisi grubundan üretilen bir roket motorundaki itici kesme dayanımının gelecek değeri üzerindeki % 95 önkestirim aralığı, denklem 1.41 kullanılarak,

$$2256.32 - (2.101)\sqrt{9244.59\left(1 + \frac{1}{20} + \frac{(10 - 13.3625)^2}{1106.56}\right)} \le y_0 \le 2256.32 + (2.101)\sqrt{9244.59\left(1 + \frac{1}{20} + \frac{(10 - 13.3625)^2}{1106.56}\right)}$$

$$2048.32 \le y_0 \le 2464.32$$

10 haftalık bir sevk iticisi grubundan üretilen yeni bir motorun, 2048.32 ve 2464.32 psi arasında bir kesme dayanımına sahip olmasını beklemek olasıdır.

*** Denklem (1.41), $x = x_0$'daki yanıt için "m" sayıda gelecek gözlemin ortalaması üzerinden yüzde 100(1-α) önkestirim aralığını bulmak için genelleştirilebilir. $\overline{y}_0$, $x = x_0$'daki gelecek gözlemlerin ortalaması ve $\overline{y}_0$'nin nokta kestiricisi $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ olmak üzere $\overline{y}_0$ üzerine yüzde 100(1-α) önkestirim aralığı;

$$\hat{y}_0 - t_{\alpha/2,n-2}\sqrt{MS_{\mathrm{Res}}\left(\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}\right)} \le y_0 \le \hat{y}_0 + t_{\alpha/2,n-2}\sqrt{MS_{\mathrm{Res}}\left(\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{S_{xx}}\right)} \tag{1.42}$$

olarak kullanılır.

EĞİM VE KESİM NOKTASI ÜZERİNE HİPOTEZ TESTLERİ

t Testlerinin Kullanılması

$$H_0 : \beta_1 = \beta_{10} \quad , \quad H_1 : \beta_1 \neq \beta_{10} \tag{1.20}$$

hipotezi test edilmek istensin.

*** $\varepsilon_i \sim N(0, \sigma^2)$, $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ ve $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$ olmak üzere; $H_0 : \beta_1 = \beta_{10}$ hipotezi gerçekte doğru ise

$$Z_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\sigma^2 / S_{xx}}}$$

istatistiği, N(0,1) olarak dağılır. Eğer σ^2 bilinseydi hipotezleri test etmek için Z_0 kullanılırdı. Ancak genelde σ^2 bilinmemektedir. Bilinmemesi durumunda, eğer $H_0 : \beta_1 = \beta_{10}$ hipotezi gerçekte doğruysa

$$t_0 = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{MS_{\mathrm{Res}} / S_{xx}}} \tag{1.21}$$

ifadesi, t_{n-2} dağılımına sahiptir. Burada t_0 oranı, $H_0 : \beta_1 = \beta_{10}$ hipotezini test etmek için kullanılan test istatistiğidir.

t_0'ın Denklem (1.21)'den gözlenen değeri, t_{n-2} dağılımının $\alpha / 2$ üst yüzdelik noktasıyla ($t_{\alpha/2,\, n-2}$) karşılaştırılır.

$$|t_0| > t_{\alpha/2,\, n-2} \tag{1.22}$$

Bu durumda H_0 hipotezi (sıfır hipotezi) reddedilir. Alternatif olarak karar vermede p-değeri yaklaşımı da kullanılabilir.

Denklem (1.21)'deki t_0 test istatistiğinin paydası, **kestirilen standart hata** veya **eğimin standart hatası** olarak adlandırılmaktadır.

$$se(\hat{\beta}_1) = \sqrt{\frac{MS_{\mathrm{Res}}}{S_{xx}}} \qquad \textbf{(1.23)}$$

Böylece t_0,

$$t_0 = \frac{\hat{\beta}_1 - \beta_{10}}{se(\hat{\beta}_1)} \qquad \textbf{(1.24)}$$

biçiminde yazılabilir.
Kesim noktası parametresine ilişkin hipotezleri test etmek için de benzer bir işlem kullanılabilir.

$$H_0 : \beta_0 = \beta_{00} \quad , \quad H_1 : \beta_0 \neq \beta_{00} \qquad \textbf{(1.25)}$$

hipotezini test etmek için,

$$t_0 = \frac{\hat{\beta}_0 - \beta_{00}}{\sqrt{MS_{\mathrm{Res}}(1/n + \overline{x}^2 / S_{xx})}} = \frac{\hat{\beta}_0 - \beta_{00}}{se(\hat{\beta}_0)} \qquad \textbf{(1.26)}$$

Burada, $se(\hat{\beta}_0) = \sqrt{MS_{\mathrm{Res}}(1/n + \overline{x}^2 / S_x)}$ **kesim noktasının standart hatasıdır.** Eğer $|t_0|$ > $t_{\alpha/2,\,n-2}$ ise $H_0 : \beta_0 = \beta_{00}$ hipotezi reddedilir.

Regresyonun Anlamlılık Testi

$$H_0 : \beta_1 = 0 \quad , \quad H_1 : \beta_1 \neq 0 \qquad \textbf{(1.27)}$$

hipotezi **regresyonun anlamlılığına** ilişkindir.

- $H_0 : \beta_1 = 0$ hipotezinin **reddedilememesi,** x **ve** y **arasında herhangi bir doğrusal ilişkinin olmadığını** göstermektedir.
- $H_0 : \beta_1 = 0$ **reddedilirse, doğrusal modelin yeterli olduğu** ya da x**'in doğrusal bir etkisi olsa bile** x**'e daha üst dereceden polinom terimlerin eklenmesiyle daha iyi sonuçlar elde edilebileceği** anlamına gelmektedir.

$H_0 : \beta_1 = 0$ hipotezi için,

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

test istatistiği kullanılır. $|t_0| > t_{\alpha/2}, n-2$ ise regresyonun anlamlılığına ilişkin sıfır hipotezi reddedilecektir.

Örnek 1.3 Roket Yakıtı Verileri

Örnek 1.1'deki roket yakıtı regresyon modelinin anlamlılığı test edilmek istensin. $\hat{\beta}_1 = -37.15$ ve $MS_{\mathrm{Res}} = \hat{\sigma}^2 = 9244.59$ olmak üzere **eğimin standart hatası**,

$$se(\hat{\beta}_1) = \sqrt{\frac{MS_{\mathrm{Res}}}{S_{xx}}} = \sqrt{\frac{9244.59}{1106.56}} = 2.89$$

olarak elde edilir.
Test istatistiği,

$$t_0 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = -12.85$$

olup $|-12.85| > t_{0.05/2}, 18 = 2.101$ olduğundan $H_0 : \beta_1 = 0$ hipotezi reddedilir. Kesme dayanımı ile iticinin yaşı arasında doğrusal bir ilişki olduğu söylenebilir.

Ödev

Bir hastanede bulunan 8 çocuğun yaşları (x) ve tedavi görme sayıları aşağıdaki gibidir. Yaş ve tedavi arasında doğrusal bir ilişki olduğu varsayılsın.

Yaş (x)	1	3	4	6	8	9	11	14
Tedavi Sayısı (y)	1	2	4	4	5	7	8	9

a) Verilere göre regresyon parametrelerinin en küçük kareler tahmin değerlerini bulup modeli oluşturunuz.
b) Parametreler için $\alpha = 0{,}05$ düzeyinde hipotez testi yapınız ve güven aralıklarını bulunuz.