(۳۱)

Subject:
Year. Month. Date. ( )

~~اول روش درستنمایی [illegible] گفته شود؟~~

برآورد تحت شرایط صحت فرض H_0 :

$H_0 : K'\beta = m$

$\Omega_0 = \{\beta : K'\beta = m\}$

$\Omega = \{\beta : \beta \in R^{k+1}\}$

اکنون هدف بدست آوردن برآورد پارامتر تحت فرض H_0 به صورت می باشد

تحت H_0 می خواهیم برآورد β° را برای β بدست آوریم که با مینیمم کردن مجموع مربعات خطا :

$$(y - X\beta^\circ)'(y - X\beta^\circ) \quad \star$$

تحت $K'\beta^\circ = m$ ، β° را بدست می آوریم.

$\begin{pmatrix} K' \\ L' \end{pmatrix}$ K $\beta^\circ = K'^{-1}m$

از لاگرانژ داریم :

$$(y - X\beta^\circ)'(y - X\beta^\circ) + 2\theta'(K'\beta^\circ - m) + 2\lambda'(L'\beta^\circ - \delta)$$

$$2(\theta' + \lambda')\left((K' + L')\beta^\circ - (m + \delta)\right)$$

مشتق گیری از تابع لاگرانژ نسبت به β° و θ داریم :

$$\begin{cases} X'X\beta^\circ + K\theta = X'y \Rightarrow \beta^\circ = (X'X)^{-1}(X'y - K\theta) & ① \\ K'\beta^\circ = m & ② \end{cases}$$

با توجه به ① و $\hat{\beta} = (X'X)X'y$:

$$\beta^\circ = \hat{\beta} - (X'X)^{-1}K\theta \Rightarrow K'\beta^\circ = K'\hat{\beta} - K'(X'X)^{-1}K\theta \overset{②}{=} m$$

$L'\beta = \delta$

$$\Rightarrow \theta = [K'(X'X)^{-1}K]^{-1}(K'\hat{\beta} - m)$$

③

$$\Rightarrow \beta^\circ = \hat{\beta} - (X'X)^{-1}K[K'(X'X)^{-1}K]^{-1}(K'\hat{\beta} - m)$$

تمرین: که در اینجا می توان نشان داد (با جایگذاری β° در ★) :

مجموع مربعات خطا $= SSE + Q$

آزمون فرض H_0 به روش نسبت درستنمایی:

تمرین:

$$\Omega_0 = \{\beta ,\ K'\beta = m\}$$

$$H_0 : K'\beta = m$$

$$\Omega = \{\beta ,\ \beta \in R^{k+1}\}$$

$$Y \sim N(X\beta , \sigma^2 I) \qquad L(\Omega) = \frac{1}{|2\pi\sigma^2 I|^{1/2}} \exp\{(Y - X\beta)'(Y - X\beta) \cdot \frac{-1}{2\sigma^2}\}$$

$$\overset{\text{دیدیم}}{\Longrightarrow} \hat{\beta} = (X'X)^{-1}X'Y$$

تمرین: نشان دهید $\hat{\sigma}^2 = $ (برآوردگر اریب) درستنمایی $\hat{\sigma}^2 = \frac{1}{n}(Y - X\hat{\beta})'(Y - X\hat{\beta})$

$$\max L(\Omega) = L(\hat{\sigma}^2, \hat{\beta}) = \frac{1}{(2\pi\hat{\sigma}^2 I)^{n/2}} \exp\{-n/2\}$$

تمرین $\max L(\Omega_0) = L(\hat{\sigma}_0^2, \hat{\beta}_0) = \frac{1}{(2\pi\hat{\sigma}_0^2)^{n/2}} \exp\{-n/2\}$

$\hat{\sigma}_0^2 = \frac{1}{n}(SSE + Q)$ ، $\hat{\beta}_0$: رابطه ③

$$-\tfrac{1}{2}(Y - X\beta_0)'(Y - X\beta_0) / \tfrac{SSE+Q}{n}$$

$$\Lambda = \frac{\max L(\Omega_0)}{\max L(\Omega)} \Rightarrow \Lambda^{2/n} = \frac{SSE}{SSE + Q} = \frac{1}{1 + Q/SSE}$$

آماره نسبت درستنمایی

که توزیع $\frac{n-k-1}{s} \cdot \frac{Q}{SSE} \overset{H_0}{\sim} F(s, n-k-1)$

$$Q = (K'\hat{\beta} - m)'(K'(X'X)^{-1}K)^{-1}(K'\hat{\beta} - m)$$

$$F(H) = \frac{Q/s}{SSE/n-k-1}$$

برای حالت خاص فرض : $K'\beta = m$

1) $H_0 : \beta = 0$ $\qquad m = 0 \qquad K' = I \qquad s = k+1$

$$F(H) = \frac{\hat{\beta}'X'X\hat{\beta}}{(k+1)\hat{\sigma}^2} = \frac{SSR}{r(X)} \cdot \frac{n - r(X)}{SSE}$$

$$\overset{H_0}{\sim} F_{r(X),\, n - r(X)} \qquad r(X) = k+1$$

$$\underset{\sim}{B}^{\circ} = \hat{\underset{\sim}{B}} - (X'X)^{-1}((X'X)^{-1})^{-1}\hat{\underset{\sim}{B}} = \underset{\sim}{0}$$ در این حالت

۲) $H_0 : \underset{\sim}{B} = \underset{\sim}{B}_0$ $\quad K' = I \quad S = k+1 \quad \underset{\sim}{m} = \underset{\sim}{B}_0$

پس $[K'(X'X)^{-1}K]^{-1} = X'X$

در این حالت

$$F(H) = \frac{(\hat{\underset{\sim}{B}} - \underset{\sim}{B}_0)' X'X(\hat{\underset{\sim}{B}} - \underset{\sim}{B}_0)}{(k+1)\hat{\sigma}^2}$$

$\overset{H_0}{\sim} F_{r(X), n-r(X)}$ که تحت H_0 می توان نشان داد

۳) $H_0 : \underset{1\times k+1}{\underset{\sim}{\lambda}'\underset{\sim}{B}} = \overset{\text{عدد}}{m}$ $\quad K' = \underset{\sim}{\lambda}' \Rightarrow S = 1, \quad \underset{\sim}{m} = m$

$$F(H) = \frac{(\underset{\sim}{\lambda}'\hat{\underset{\sim}{B}} - m)'(\overbrace{\underset{\sim}{\lambda}'(X'X)^{-1}\underset{\sim}{\lambda}}^{\text{اسکالر}})^{-1}(\underset{\sim}{\lambda}'\hat{\underset{\sim}{B}} - m)}{\hat{\sigma}^2}$$

$\lambda'\underset{\sim}{B} = m$ تحت H_0

$\underset{\sim}{\lambda}'\hat{\underset{\sim}{B}} \sim N(\underset{\sim}{\lambda}'\underset{\sim}{B}, \lambda'(X'X)^{-1}\lambda\sigma^2) \Rightarrow$

$$= \frac{(\underset{\sim}{\lambda}'\hat{\underset{\sim}{B}} - m)^2}{\underset{\sim}{\lambda}'(X'X)^{-1}\underset{\sim}{\lambda}\hat{\sigma}^2} \overset{H_0}{\sim} F_{1, n-r(X)}$$

t^2

$$\Rightarrow \sqrt{F(H)} \sim t_{n-r(X)}$$

۴) $H_0 : \underset{\sim}{\beta}_q = \underset{\sim}{0}$ i.e $\beta_i = 0 \ \forall i = 0, 1, 2, \dots, q-1 \quad q < K$

q تا از β صفر

$$\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{q-1} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \longrightarrow \beta_i = 0 \ \forall i = 0, \dots, q-1$$

$K' = [\, \underset{q}{I_q} \quad \underset{\sim}{0}^{p} \,], \quad \underset{\sim}{m} = \underset{\sim}{0}, \quad S = q$ در این صورت

$$\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{q-1} \end{bmatrix} = \underset{q\times(k+1)}{\begin{bmatrix} 1 & & & 0 & \cdots & 0 \\ & \ddots & & & & \\ & & 1 & & & \end{bmatrix}} \underset{k+1\times 1}{\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{q\times 1}$$

$p + q = k+1$

Subject: $Q=(K'\hat{\beta}-m_0)'(K'(X'X)^{-1}K)^{-1}(K'\hat{\beta}-m_0)$

Year. Month. Date. ( )

~~$(\hat{\beta}_q)' T^{-1}_{qq} \hat{\beta}_q$~~

$$\beta = \begin{bmatrix} \beta_q \\ \beta_p \end{bmatrix} \begin{matrix} \to q\times 1 \\ \to p\times 1 \end{matrix} \qquad p+q=k+1$$

سپر ادامه حالت

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_q \\ \hat{\beta}_p \end{bmatrix} \quad \text{و} \quad (X'X)^{-1} = \begin{bmatrix} T_{qq} & T_{qp} \\ T_{pq} & T_{pp} \end{bmatrix}$$

و بنابراین:

$$[K'(X'X)^{-1}K]^{-1} = T^{-1}_{qq} \quad \text{و} \quad K'\hat{\beta} = \hat{\beta}_q$$

و در این حالت داریم:

$$F(H) = \hat{\beta}_q' T^{-1}_{qq} \hat{\beta}_q / q\hat{\sigma}^2 \sim F_{q,\, n-k-1}$$

تحت فرض صفر به صورت $H_0: \beta_q = 0$ برآورد بردار β به صورت زیر است:

از β°

$$\tilde{\beta} = \hat{\beta} - (X'X)^{-1} \begin{bmatrix} I_q \\ 0 \end{bmatrix} T^{-1}_{qq} (\hat{\beta}_q - 0)$$

$\beta^\circ = \hat{\beta} - (X'X)^{-1}K[K'(X'X)^{-1}K]^{-1}(K'\hat{\beta}-m)$

$$= \hat{\beta} - \begin{bmatrix} T_{qq} \\ T_{pq} \end{bmatrix} T^{-1}_{qq} \hat{\beta}_q = \begin{bmatrix} \hat{\beta}_q \\ \hat{\beta}_p \end{bmatrix} - \begin{bmatrix} \hat{\beta}_q \\ T_{pq} T^{-1}_{qq} \hat{\beta}_q \end{bmatrix}$$

$K' = [I_q' \ 0]$

$m = 0$

$$= \begin{bmatrix} 0 \\ \hat{\beta}_p - T_{pq} T^{-1}_{qq} \hat{\beta}_q \end{bmatrix}$$

PAPCO

Subject :
Year . Month . Date . ( )

جدول آنالیز واریانس برای مدلهای تقلیل یافته :

در عمل ممکن است علاوه بر آنکه محدودیتهایی برای انتخاب مدل در نظر می گیریم محدودیتهایی هم برای فضای پارامتری در نظر بگیریم پس می توان به دو صورت مسئله زیر سنتوری را داشته باشیم مدل کامل و مدل تقلیل یافته :

مدل کامل $\left[\underset{\sim}{y} = X\underset{\sim}{B} + \underset{\sim}{\varepsilon} \qquad E(\underset{\sim}{y}) = X\underset{\sim}{B} \qquad var(\underset{\sim}{y}) = \sigma^2 I\right]$.

Full model

مدل تقلیل یافته $[\underset{\sim}{y} = X\underset{\sim}{B} + \varepsilon \;,\; E(\underset{\sim}{y}) = X\underset{\sim}{B} \;,\; var(y) = \sigma^2 I \;,\; K'\underset{\sim}{B} = \underset{\sim}{m}]$

Reduced model

برای مثال فرض کنید مدل کامل به صورت :

$$\underset{\sim}{y} = B_0 + B_1 \underset{\sim}{X_1} + B_2 \underset{\sim}{X_2} + \underset{\sim}{\varepsilon}$$

که در n نمونه :

$$y_i = B_0 + B_1 x_{i1} + B_2 x_{i2} + \varepsilon_i$$

اکنون تحت فرض $H : B_1 = B_2$ مدل تقلیل یافته به صورت

$$y = B_0 + B_1(X_1 + X_2) + \varepsilon$$

در نمونه

$$y_i = B_0 + B_1(x_{i1} + x_{i2}) + \varepsilon_i$$

که همان طور دیدیم وقتی تحت فرض $k'\underset{\sim}{B} = \underset{\sim}{m}$ برآورد پارامترها را بدست آوردیم برآورد پارامتر تغییر کرد

$$\underset{\sim}{B}^{\circ} = \underset{\sim}{\hat{B}} - (X'X)^{-1}K[K'(X'X)^{-1}K]^{-1}(K'\underset{\sim}{\hat{B}} - \underset{\sim}{m})$$

$$SSE' = SSE + Q$$

اما در حالتی که $H_0 : B_1 = B_2$:

اما مجموع مربعات کل SSy تغییری نمی کند چون به B ها بستگی ندارد

$$SSy = y'y - n\bar{y}^2$$

Subject:
Year. Month. Date. ( )

اما اگر فرض کنیم $H_0: B_1 = B_2 + 4$ یعنی

$\underset{\sim}{k}' = [0 \ \ 1 \ \ -1]$

$m = 4$

$$y = B_0 + B_1 X_1 + B_2 X_2 + \varepsilon \rightarrow full$$

$$\xRightarrow{\text{تحت فرض}} \underset{\sim}{y} = B_0 + (B_2+4)X_1 + B_2 X_2 + \varepsilon$$

$$\Rightarrow y - 4X_1 = B_0 + B_2(X_1 + X_2) + \varepsilon$$

در این حالت S_y هم تفاوت می کند.

~~حالت اول $K'B - m = 0$~~ فرض کلی $k'\underset{\sim}{B} = \underset{\sim}{m}$ را در نظر بگیرید

پس اگر $\underset{\sim}{y} = X\underset{\sim}{B} + \underset{\sim}{\varepsilon}$ ، $K'\underset{\sim}{B} = \underset{\sim}{m}$ ، $E(\underset{\sim}{y}) = X\underset{\sim}{B}$, $Var(y) = \sigma^2 \underset{\sim}{I}$.

اکنون L' را طوری در نظر می گیریم که ماتریس

$$\underset{(k+1)(k+1)}{R} = \begin{bmatrix} K' \\ L' \end{bmatrix} \begin{array}{l} \rightarrow s \times k+1 \\ (k+1-s) \times (k+1) \end{array}$$

ماتریس دلخواه

رتبه کامل باشد و $R^{-1} = [P \ \ S]$ ← معکوس وجود دارد

و مدل را بازنویسی می کنیم:

یعنی L طوری انتخاب می کنیم که از k' مستقل باشد روی رتبه تاثیر نگذارد

$$\underset{\sim}{y} = X \underbrace{R^{-1}R}_{I} \underset{\sim}{B} + \underset{\sim}{\varepsilon}$$

$$= X[P \ \ S]\begin{bmatrix} \overbrace{K'\underset{\sim}{B}}^{m} \\ \underbrace{L'\underset{\sim}{B}}_{\gamma} \end{bmatrix} + \underset{\sim}{\varepsilon}$$

$$\left(\underset{(k+1-s)\times k+1}{L'} \underset{\sim}{B}_{(k+1)\times 1} \right)$$

(*)

$$\Rightarrow \overbrace{\underset{\sim}{y} - XP\underset{\sim}{m}}^{y^*} = XSL'\underset{\sim}{B} + \underset{\sim}{\varepsilon} \quad (k+1-s)\times 1$$

(k+1-s) ترکیب از B ها داریم LIN هستند.

L' دلخواه است و ماتریسهای زیادی می توانیم تعریف کنیم، شرایط را برقرار کند

لذا L' منحصر به فرد نیست.

$(K'\underset{\sim}{B})_{s\times 1}$

$(L'\underset{\sim}{B})_{(k+1-s)\times 1}$

*. اگر $m=0$ روی S_{yy} تاثیری ندارد ولی اگر $m \neq 0$ با توجه به رابطه (*)

S_{yy} تغییر می کند

PAPCO

درحالت خاص $\underset{\sim}{y} = XSL'\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ ، $K'\underset{\sim}{\beta} = \underset{\sim}{0}$

و

$SSE' = SSE + Q$ تقلیل یافته

همچنین داریم : $Q = y'y - SSE - y'y + SSE + Q$ ‏ $y'y$ اضافه کنیم

$= SSR - \underbrace{[y'y - (SSE+Q)]}_{SSR'}$

Reduced

$$\left\{\begin{array}{l} \overbrace{(SSE+Q)}^{\text{Residual}(R)} + \overbrace{(SSR-Q)}^{\text{Reduction}(R)} = SST \\ \underbrace{SSE}_{\text{residual}(f)} + \underbrace{SSR}_{\text{Reduction}(f)} = SST \end{array}\right.$$

Full

جدول آنالیز واریانس در این حالت ($K'\underset{\sim}{\beta} = \underset{\sim}{0}$) به صورت زیر می‌باشد

منبع	df	SS
رگرسیون	$k+1$	SSR
فرضیه	s	Q
مدل تقلیل یافته	$k+1-s$	$SSR-Q$
خطا	$n-k-1$	SSE
کل	n	SST

که قبلاً دیدیم

① از طرفی داریم $\frac{Q}{\sigma^2} \sim \chi^{2'}\left(s, \underbrace{\underset{\sim}{\beta}'K(K'(X'X)^{-1}K)^{-1}K'\underset{\sim}{\beta}}_{\delta}/2\sigma^2\right)$

② $(\overbrace{y'y - SSE}^{SSR})/\sigma^2 \sim \chi^{2'}(k+1, \underset{\sim}{\beta}'X'X\underset{\sim}{\beta}/2\sigma^2)$

$\hat{\beta}'$ ‏ $y'X(X'X)^{-1}X'y$

از ① و ② خواهیم داشت $y'(X(X'X)^{-1}K(K'(X'X)^{-1}K)^{-1}K'(X'X)^{-1}X')y$

$$\frac{SSR-Q}{\sigma^2} \sim \chi^{2'}\left(k+1-s, \underset{\sim}{\beta}'(X'X - K(K'(X'X)^{-1}K)^{-1}K')\underset{\sim}{\beta}/2\sigma^2\right)$$

$y'(I - X(X'X)^{-1}X')y$

و مستقل از $\frac{SSE}{\sigma^2}$ می‌باشد

و بنابراین :

برای بررسی معنی داری مدل کامل

$$F(R) = \frac{SSR/k+1}{SSE/n-k-1} \sim F'\left(r(X),\ n-r(X),\ \frac{1}{2\sigma^2}\underset{\sim}{B}'X'X\underset{\sim}{B}\right) \xrightarrow[\underset{\sim}{B}=\underset{\sim}{0}]{H_0\ \text{تحت}} F\left(\underbrace{r(X)}_{k+1},\ \underbrace{n-r(X)}_{k+1}\right)$$

$$F(H) = \frac{Q/s}{SSE/n-k-1} \sim F'(s,\ n-k-1,\ \delta) \xrightarrow[K'\underset{\sim}{B}=0]{H_0\ \text{تحت}} \sim F(s,\ n-k-1)$$

برای بررسی فرض $H_0: K'\underset{\sim}{B} = 0$

$$\underset{\sim}{F(SH)} = \frac{(SSR-Q)/(k+1-s)}{SSE/n-k-1} \sim F'(k+1-s,\ n-k-1,\ \delta')$$

sub Hypothesis $L'\underset{\sim}{B} = \underset{\sim}{0}$

$$\xrightarrow[\text{for } L'\underset{\sim}{B}=\underset{\sim}{0}]{K'\underset{\sim}{B}=\underset{\sim}{0}\ \text{تحت}} \sim F(k+1-s,\ n-k-1)$$

برای بررسی زیرفرض $L'\underset{\sim}{B} = \underset{\sim}{0}$

- حالت خاص دیگری که ممکن است متداول باشد به صورت $\underset{\sim}{B}_q = \underset{\sim}{0}$ است که در فرض می $K' = [I_q \ \ 0]$ و $\underset{\sim}{m} = \underset{\sim}{0}$ در نظر گرفته شده است

$$K' = \left[\underbrace{\begin{matrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{matrix}}_{q\times q} \ \underbrace{\begin{matrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{matrix}}_{q\times p}\right]_{q\times \underbrace{(p+q)}_{k+1}} \qquad q \le k$$

q تا از $\underset{\sim}{B}$ ها صفر باشند

$$\underset{\sim}{B}'_q = [B_0 \ \ B_1 \ \ldots\ldots \ B_{q-1}]$$

که در این حالت داریم

$$F(H) = \frac{\hat{B}'_q T_{qq}^{-1} \hat{B}_q}{q\hat{\sigma}^2}$$

در جدول آنالیز واریانس در این حالت به صورت زیر خواهد شد.

PAPCO

مثال صفحه ۱۲۱ (Searl) :

y	x_1	x_2	x_3
8	2	1	4
10	-1	2	1
9	1	-3	4
6	2	1	2
12	1	4	6

$r(X)=3$

$(X'X)_{3\times3}$

بدون جمله ثابت

$$(X'X)^{-1}=\begin{bmatrix}11 & 3 & 21\\ 3 & 31 & 20\\ 21 & 20 & 73\end{bmatrix}^{-1}=\begin{bmatrix}0.2145 & 0.0231 & -0.0680\\ 0.0231 & 0.0417 & -0.0181\\ -0.0680 & -0.0181 & 0.0382\end{bmatrix}$$

$\underset{\sim}{y}'\underset{\sim}{y}=SST=425 \qquad X'\underset{\sim}{y}=\begin{bmatrix}39\\55\\162\end{bmatrix} \qquad \hat{\underset{\sim}{B}}'=[-1.39 \quad 0.27 \quad 2.54]$

منبع	df	SS
رگرسیون	3	372.9
خطا	2	52.1
کل	5	425

$$F(R)=\frac{372.9/3}{52.1/2}<F_{3,2}\simeq 19$$

فرض $H_0: B_1=B_2+4$

$k'=[1 \quad -1 \quad 0] \qquad m=4 \qquad s=1$

$$Q=(k'\hat{\underset{\sim}{B}}-m)'(k'(X'X)^{-1}k)^{-1}(k'\hat{\underset{\sim}{B}}-m)$$

$$=(\hat{B}_1-\hat{B}_2-4)\left((1\ -1\ 0)(X'X)^{-1}\begin{pmatrix}1\\-1\\0\end{pmatrix}\right)^{-1}(\hat{B}_1-\hat{B}_2-4)$$

$=152.2$

Subject :

Year . Month . Date . ( )

$$F(H) = \frac{Q/s}{SSE/n-k} = \frac{\frac{152.2}{1}}{\frac{52.1}{2}} = 5.8$$

مدل را تحت فرض بازنویسی می کنیم $H_0: B_1 = B_2 + 4$

$$y = B_1 x_1 + B_2 x_2 + B_3 x_3 + \varepsilon \overset{\text{تحت } H_0}{=} (B_2+4)x_1 + B_2 x_2 + B_3 x_3 + \varepsilon$$

$$\Rightarrow \overbrace{y - 4x_1}^{y^*} = B_2(x_1 + x_2) + B_3 x_3 + \varepsilon \quad (*)$$

$y - xpm = xSL'B + \varepsilon$

که می دانیم در این حالت SST تغییر می کند :

y^* : 0 14 5 -2 8

$$SST_{new} = y^{*\prime} y^* = 289$$

مجموع مربعات تشکیل یافته $SSE' = \underbrace{SSE}_{2} + \underbrace{Q}_{1} = 152.2 + 52.1 = 204.3$

پس جدول آنالیز واریانس در این حالت

	df	ss
رگرسیون (مدل تشکیل یافته)	2	84.7
خطا (مدل تشکیل یافته)	3	204.3
کل	5	289

برای مدل جدید (*)

$\bar{B}$ را بدست آورید.

با فرض کنید فرض $H_0: B_1 = 0$ مدنظر باشد :

$m = 0 \quad K' = [1 \ 0 \ 0]$

$$Q = \overbrace{\hat{B}_1^2}^{(-1.39)^2} / (\underbrace{K'(X'X)^{-1}K}_{0.2145})^{-1} = 8.9$$

و در این حالت می دانیم SST تغییر می کند .

	df	ss
رگرسیون کامل	3	372.9
فرض	1	8.9
مدل تشکیل یافته	2	364
خطا	2	52.1
کل	5	425

$\bar{B}$ را بدست آورید.

PAPCO

$R^{-1}R = [P \quad S]\begin{bmatrix} K' \\ L' \end{bmatrix} = [PK' + SL'] = I$

Subject:
Year. Month. Date. ( )

$PK' = P(I - SL')$

منبع	d.f	SS
رگرسیون (full)	$k+1$	$SSR = \hat{\beta}' X' y$
تحت فرض $\beta_q = 0$	q	$Q = \hat{\beta}_q' T_{qq}^{-1} \hat{\beta}_q$
رگرسیون (Reduced)	$k+1-q$	$SSR' = SSR - Q$
خطا	$n-k-1$	$SSE = SST - SSR$
کل	n	$SST = y'y$

که قبلاً دیدیم : $T_{qq}^{-1} = [K'(X'X)^{-1}K]^{-1}$ (مثال ص 121)

محاسبه R^2 در رگرسیون چندگانه :

ضریب تعیین برای مدل کلی و بدون تصحیح میانگین از رابطه $R^2 = \frac{SSR}{SST}$ محاسبه می‌شود

و برای مدل تصحیح شده بصورت

$$R^2 = \frac{SSR_m}{SST_m} = \frac{\hat{\beta}' X' y - n\bar{y}^2}{y'y - n\bar{y}^2}$$

در هر صورت R^2 عددی بین صفر و یک است و جذر آن همبستگی ساده بین y ها مشاهده شده

و $\hat{y}$ های برازش شده را می‌دهد . یعنی

$$R = r_{y\hat{y}}$$

نکته : R^2 را می‌توان براساس ماتریس ضریب همبستگی نمونه‌ها بصورت زیر بیان کرد

$$P = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \vdots & \ddots & & \\ \rho_{p1} & & & 1 \end{bmatrix}$$

$$\boxed{R^2 = r_{yx}' R_{xx}^{-1} r_{yx}}$$

بدون

$$R_{yx} = \begin{bmatrix} 1 & r_{yx}' \\ r_{yx} & R_{xx} \end{bmatrix} = \left[\begin{array}{c|ccc} 1 & r_{y1} & r_{y2} \cdots & r_{yk} \\ \hline r_{1y} & 1 & r_{12} \cdots & r_{1k} \\ \vdots & & \ddots & \\ r_{ky} & r_{k1} & & 1 \end{array}\right]$$

ضریب همبستگی نمونه برای $(y, x_1 \ldots x_k)$

ماتریس ضریب همبستگی بین x ها

PAPCO

$$r_{y1} = \frac{S_{y1}}{\sqrt{S_{yy}^2 S_{11}^2}} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_{i1} - \bar{x}_1)}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2}}$$ مثلاً

or

$$r_{12} = \frac{S_{12}}{\sqrt{S_{11}^2 S_{22}^2}}$$

نکته: چون در رگرسیون چندگانه با افزایش k ممکن است R^2 بزرگ شود در صورتی که ممکن است مدل واقعاً خوب نباشد برای حل این مشکل ضریب تعیین تعدیل شده R_a^2 به صورت زیر پیشنهاد شده است:

$$R_a^2 = \frac{(R^2 - \frac{k}{n-1})(n-1)}{n-k-1} = \frac{(n-1)R^2 - k}{n-k-1}$$

بررسی اعتبار مدل برازش شده در رگرسیون چندگانه:

مدل به صورت $y = X\beta + \varepsilon$ با فرض‌های $E(\varepsilon) = 0$ و $Cov(\varepsilon) = \sigma^2 I$ در نظر گرفته می‌شود. به طوری که فرض کردیم X ثابت و رتبه ستونی کامل است.

(فصل ۴ مونتگمری) فصل ۹ رگرسیون کامل خوانده شود. فرضیاتی که معمولاً مورد آزمون قرار می‌گیرند:

Normality (نرمالیتی), Homoscedasticity (هم‌واریانسی), Autocorrelation (خودهمبستگی) (بین خطاهای مدل همبستگی وجود دارد)

multicollinearity
هم‌خطی (بین متغیرهای مستقل همبستگی وجود دارد)

روش‌های معمول: مانده‌های خطای استاندارد شده و رسم نمودار P.P برای نرمالیتی و خطای استاندارد شده و $\hat{y}$ را رسم می‌کنند برای بررسی هم‌واریانسی.

از آماره دوربین واتسون برای بررسی وجود خودهمبستگی بین باقیمانده‌ها استفاده می‌شود که بین صفر و ۴ است اگر حول ۲ باشد نشانه عدم وجود خودهمبستگی [illegible] باشد.

مقدار کمتر از ۲ همبستگی پیاپی مثبت (مقدار باقیمانده مثبت برای یک مشاهده نشان مثبت برای باقیمانده مشاهده دیگر را افزایش می‌دهد) و مقدار بیشتر از ۲ همبستگی پیاپی منفی را در باقیمانده‌ها نشان می‌دهد.

$$d = \sum_{i=2}^{n} (e_i - e_{i-1})^2 / \sum_{i=1}^{n} e_i^2$$

نرم‌افزار R : / Lmtest / dwtest

منظور از هم‌خطی که معمولاً باید در رگرسیون چندگانه مورد بررسی قرار گیرد میزان ارتباط خطی بین متغیر مستقل است. معمولاً هم‌خطی در مدل‌های رگرسیونی چندگانه وجود دارد و شدت هم‌خطی باید بررسی شود و وجود هم‌خطی بالا، باعث نقض فرضیات اساسی خواهد شد.

برای تشخیص هم‌خطی معمولاً رگرسیون معنی‌دار و R^2 بزرگ است اما مشاهده می‌شود که متغیرهای ضرایب معنی‌دار نشدند. اگر یک متغیر از مدل خارج یا اضافه شود و R^2 تغییر زیادی نداشته باشد آن متغیر مستعد ایجاد هم‌خطی است. دو معیار tolerance و VIF هم برای بررسی هم‌خطی استفاده می‌شود.

$$tolerance = 1 - R_j^2 \qquad VIF = \frac{1}{tolerance}$$

ضریب تعیین رگرسیون بین متغیر مستقل j ام

if tol < 0.2 or VIF > 10 ⟹ هم‌خطی وجود دارد

همچنین وجود داده‌های پرت و یا مشاهدات مؤثر در رگرسیون چندگانه مورد بررسی علاقه است.

این گونه مشاهدات تأثیر زیادی بر روی برآورد پارامترها دارند و برای تشخیص آنها از:

ص ۲۳۶ بخ

(دور از مرکز x باشد تأثیرگذاری خطا) Cook's ⟶ مشاهده مؤثر ⟹ برازش نادرست داده‌ها

(اعتبارسنجی) PRESS ⟶ مشاهده دورافتاده ⟹ مدل‌ها با PRESS کوچکتر: ارجح‌تر

Box-plot

فواصل کوک $D_i = \dfrac{(\hat{\beta}_{(i)} - \hat{\beta})' X'X (\hat{\beta}_{(i)} - \hat{\beta})}{(k+1)\hat{\sigma}^2}$

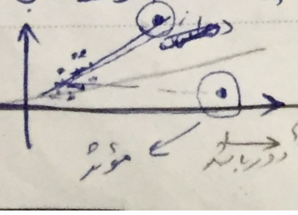
$(y_i, \underline{x}_i')$ ⟵ اثر حذف این نقطه بر $\hat{\beta}$ مورد بررسی قرار می‌گیرد

سپس n فاصله کوک قابل محاسبه است

⟹ هرچه D_i بزرگ‌تر باشد $D_i = \dfrac{r_i^2}{k+1}\left(\dfrac{h_{ii}}{1-h_{ii}}\right)$

نشان‌دهنده $(y_i, \underline{x}_i')$ تأثیر زیادی بر $\hat{\beta}$ دارد

یا فاصله ماهالانوبیس.



مؤثر ⟵ دور باشد از مرکز x یا غیرمعمول باشد در جهت y دورافتاده

Subject:
Year. Month. Date. ( )

مدلهای خطی رتبه ناقص :

فصل V ← Practical Regression and anova using R. J. Faraway 2002

~~[illegible]~~ پروژه تعریف شود در ترم سوم (چندگانه کامل)

تشخیص مدل مناسب ، بررسی مدل ---

وکتا لسفارسی R آقای موسوی ص ۱۱۶

جدول 7-3 ۱۸۳ ~~۱۷۱~~ ص → داده‌ی نجار بتن → وزن - سعیدی

رنجبر - لاش

جدول 7-5 ~~۱۸۴~~ ص داده‌ای اجاره زمین → رای - غضنفری

~~[illegible]~~ جدول - مونتگمری → احمدخانی - کاظم‌خانی - ترابی
پیوست B-5

مونتگمری - جدول B-6 پیوست → جاهدی - الکوئی لواف - جعفرزاده ; نوبخت - معصومی

مونتگمری جدول B-2 پیوست → امیری - علی‌نژاد اعتدای