Subject :

Year . Month . Date . ( )

رگرسیون چندگانه

در بعضی مطالعات بررسی ارتباط خطی بین یک متغیر پاسخ و چند متغیر مستقل است در رگرسیون خطی ساده دیدیم که یک متغیر وابسته Y براساس یک رابطه خطی با یک متغیر مستقل X پیش بینی می‌شود اما در رگرسیون چندگانه یک متغیر وابسته Y را براساس یک رابطه خطی با چند متغیر مستقل پیش بینی می‌کنیم

مستقل وابسته

$$\underset{\sim}{X} = [X_1 \dots X_K] \qquad Y$$

لذا در این حالت مدل کلی :

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K + \varepsilon$$

وقتی به مدل، مدل خطی می‌گوییم که نسبت به βها خطی باشد. برای برآورد $\hat{\beta}$ و تشکیل خطا رگرسیون نیاز به داشتن نمونه می‌باشد. فرض کنید نمونه‌ای به حجم n در دست باشد

$$\begin{array}{lccccc} & & \underset{\sim}{X_0} & \underset{\sim}{X_1} & \cdots & \underset{\sim}{X_K} \\ \text{متغیر اول} & Y_1 & x_{10}=1 & x_{11} & \cdots & x_{1K} \\ Y & \vdots & \vdots & \vdots & & \vdots \\ \text{نمونه ام} & Y_n & x_{n0}=1 & x_{n1} & \cdots & x_{nK} \\ & & \underset{\sim}{\beta_0} & \underset{\sim}{\beta_1} & & \underset{\sim}{\beta_K} \end{array}$$

$$Y_1 = \beta_0 + \beta_1 x_{11} + \dots + \beta_K x_{1K} + \varepsilon_1$$

$$Y_n = \beta_0 + \beta_1 x_{n1} + \dots + \beta_K x_{nK} + \varepsilon_n$$

که به صورت ماتریسی می‌توان نوشت :

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1K} \\ \vdots & & & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nK} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

و فرضیات اساسی رگرسیون چندگانه :

۱) $E(\underset{\sim}{Y}) = X\underset{\sim}{\beta} \quad \& \quad E(\underset{\sim}{\varepsilon}) = 0$

۲) $Var(\underset{\sim}{\varepsilon}) = E[(\underset{\sim}{\varepsilon} - E(\underset{\sim}{\varepsilon}))(\underset{\sim}{\varepsilon} - E(\underset{\sim}{\varepsilon}))'] = \sigma^2 I_n$ ← واریانس ثابت و خطاها از هم مستقل

P4PCO

$\frac{\partial X'AX}{\partial X} = 2AX$

$\frac{\partial X'AY}{\partial X} = AY$

Subject:
Year. Month. Date.

طبق فرضیات فوق شاید در رسیدن خطی ساده برآوردهای حداقل مربعات قابل محاسبه اند:

$$\sum \varepsilon_i^2 = \underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon} = [\underset{\sim}{Y} - E(\underset{\sim}{Y})]'[\underset{\sim}{Y} - E(\underset{\sim}{Y})] = (\underset{\sim}{Y} - X\underset{\sim}{B})'(\underset{\sim}{Y} - X\underset{\sim}{B})$$

$$= \underset{\sim}{Y}'\underset{\sim}{Y} - 2\underset{\sim}{B}'X'\underset{\sim}{Y} + \underset{\sim}{B}'X'X\underset{\sim}{B}$$

برای $\hat{\underset{\sim}{B}}$ ای که راست آدیم، عبارت فوق را مینیمم میکند:

$$\frac{\partial(\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon})}{\partial \underset{\sim}{B}} = -2X'\underset{\sim}{Y} + 2X'X\underset{\sim}{B} = 0$$

$$\Rightarrow X'X\hat{\underset{\sim}{B}} = X'\underset{\sim}{Y} \quad \text{(1) دستگاه معادلات نرمال}$$

درصورت وجود $(X'X)^{-1}$ یک جواب یکتا دارد ولی به هر حال دستگاه معادله فوق سازگار است زیرا

$$r[X'X \;\; \underbrace{X'Y}_{X\underset{\sim}{B}}] = r(X'X)$$

$X'Y$ ← ترکیب خطی از $X'X$

$Y = AX \quad A^+ = [A \;\; Y] \quad r(A^+) = r(A)$

یعنی همیشه $(X'X)$ یک ماتریس نادرجه باشد یا نباشد معادله جواب دارد

اگر $X'X$ نادرجه باشد:

$$X_{n\times(k+1)} \rightarrow \underbrace{(X'X)_{(k+1)(k+1)}}$$

اگر بخواهد $\Rightarrow r(X) = k+1$

نادرجه باشد باید ستونها کامل باشد $r(X'X) = k+1$

$r(X) = r(X') = k+1$

$r(X'X) \leq \min(r(X), r(X')) = k+1$

$\Rightarrow n \geq k+1$

برای A دلخواه،
$r(AA') = r(A'A) = r(A') = r(A)$

پس برای اینکه $X'X$ نادرجه باشد X یک ماتریس ستونی کامل باشد یا عبارتی $n \geq k+1$

وقتی $X'X$ درجه خودش است که $n < k+1$ دراین صورت آن درجه ندارد

Subject :
Year . Month . Date . ( )

از دستگاه معادلات نرمال ① و در صورت وجود $(X'X)^{-1}$:

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$$

به راحتی می‌توان نشان داد :

$$E(\hat{\underset{\sim}{\beta}}) = E((X'X)^{-1}X'\underset{\sim}{y}) = (X'X)^{-1}X'E(\underset{\sim}{y}) = (X'X)^{-1}X'X\underset{\sim}{\beta} = \underset{\sim}{\beta}$$

$$var(\underbrace{\hat{\underset{\sim}{\beta}}}_{A\underset{\sim}{y}}) = \underbrace{(X'X)^{-1}X'}_{A}\underbrace{\sigma^2 I}_{var(\underset{\sim}{y})}\underbrace{X(X'X)^{-1}}_{A'} = \sigma^2(X'X)^{-1}$$

قضیه گوس-مارکف Gauss-Markov

در مدل رگرسیونی $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ ، $E(\underset{\sim}{y}) = X\underset{\sim}{\beta}$ ، $var(\underset{\sim}{y}) = \sigma^2 I$

برآوردگر :

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$$

بهترین برآوردگر نااریب خطی (BLUE) از $\underset{\sim}{\beta}$ را به دست می‌دهد

مثال صفحه بعد.

$E(X'AX) = tr(A\Sigma) + \mu'A\mu$
چرا استفاده نمی‌کنیم؟ نیاز به نرمال

روش حداقل مربعات تعمیم یافته :

فرض کنید مدل رگرسیونی $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ به طوری که $E(\underset{\sim}{y}) = X\underset{\sim}{\beta}$ و

$$var(\underset{\sim}{y}) = \Sigma \quad , \quad |\Sigma| \neq 0$$

در این روش باید $(\underset{\sim}{y} - X\underset{\sim}{\beta})'\Sigma^{-1}(\underset{\sim}{y} - X\underset{\sim}{\beta})$ مینیمم شود

PAPCO

Subject:
Year. Month. Date. ( )

مدل کلی :

$$Y = X\beta + \varepsilon$$

$$\Rightarrow \Sigma^{-1/2} Y = \Sigma^{-1/2} X\beta + \Sigma^{-1/2}\varepsilon$$

$$\Rightarrow Y^* = X^*\beta + \varepsilon^*$$

به طوری که $Var(Y^*) = I$ میباشد $(Y^* - X^*\beta)'(Y^* - X^*\beta)$ مینیمم میشود و β را برآورد میکنیم که به حالت قبل که واریانس ثابت بود :

$$\tilde{\beta} = (X^{*\prime} X^*)^{-1} X^{*\prime} Y^*$$

$$\Rightarrow \tilde{\beta} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1} Y \rightarrow GLS$$ برآوردگر

به طوری که :

$$E(\tilde{\beta}) = \beta \quad , \quad Var(\tilde{\beta}) = (X'\Sigma^{-1}X)^{-1}$$

روش ماکسیمم درستنمایی : فرض کنید

$$Y = X\beta + \varepsilon$$
$$E(Y) = X\beta \quad , \quad V(Y) = \sigma^2 I \quad , \quad Y \sim N(X\beta, \sigma^2 I)$$

$$f(y) = \frac{1}{|2\pi \underbrace{\Sigma}_{\sigma^2 I}|^{1/2}} \exp\{-\tfrac{1}{2}(y - X\beta)'(\sigma^2 I)^{-1}(y - X\beta)\}$$

$$= \frac{1}{|2\pi\sigma^2 I|^{1/2}} \exp\{\frac{-1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\}$$

β ای که عبارت فوق را ماکسیمم می کند β ای است که توان را min می کند که در OLS دیدیم

$$\hat{\beta} = (X'X)^{-1} X'Y$$

~~$f(y) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\{-\frac{1}{2}(y - X\beta)'\Sigma^{-1}(Y - X\beta)\}$~~ $Var(Y) = \Sigma$ که

$\rightarrow GLS$ $\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$

PAPCO

Subject :
Year . Month . Date . ( )

اگر $var(\underset{\sim}{Y}) = \Sigma$:

$$f(\underset{\sim}{y}) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\{-\tfrac{1}{2}(\underset{\sim}{Y}-X\underset{\sim}{\beta})'\Sigma^{-1}(\underset{\sim}{Y}-X\underset{\sim}{\beta})\}$$

و در GLS :

$$\underset{\sim}{\hat{\beta}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underset{\sim}{Y}$$

روش برآورد برآوردگر BLUE آن کسی از بردار $\underset{\sim}{\beta}$ یعنی $(t'\underset{\sim}{\beta})$:

فرض $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ ، $E(\underset{\sim}{Y}) = X\underset{\sim}{\beta}$ و $var(\underset{\sim}{Y}) = \Sigma$ است می‌خواهیم برآوردی بیابیم برای $t'\underset{\sim}{\beta}$

۱) خطی باشد ۲) نااریب باشد ۳) دارای حداقل واریانس

Linear خطی بودن — Unbiasedness نااریبی — minimum variance — Best

برای اینکه برآوردگر خطی باشد باید تابعی خطی از مشاهدات $\underset{\sim}{Y}$ باشد برآوردگر را به صورت $\lambda'\underset{\sim}{Y}$ در نظر می‌گیریم ، که در آن $\underset{1\times n}{\lambda}$ یک بردار ثابت از مرتبه n داریم :

$$\hat{t'\beta} = \lambda'\underset{\sim}{Y}$$

می‌خواهیم نااریب باشد پس

$$E(\hat{t'\beta}) = t'\underset{\sim}{\beta}$$

$$\Rightarrow E(\hat{t'\beta}) = E(\lambda'\underset{\sim}{Y}) = \lambda' E(\underset{\sim}{Y}) = \lambda' X\underset{\sim}{\beta} = t'\underset{\sim}{\beta}$$

می‌خواهیم

$$\Rightarrow \lambda' X = t' \quad \star$$

از طرفی واریانس هم می‌خواهیم کمینه باشد

$$var(\hat{t'\beta}) = var(\lambda'\underset{\sim}{Y}) = \lambda' var(\underset{\sim}{Y})\lambda = \lambda'\Sigma\lambda$$

حال از لاگرانژ استفاده می‌کنیم و با شرط $\star$ ($\lambda' X = t'$) ، λ ای

Subject:
Year Month Date ( )

را پیدا می‌کنیم، واریانس فوق را min کند: (با ضرایب لاگرانژ $2\underset{\sim}{\theta}$)

$$Q(\underset{\sim}{\lambda}, \underset{\sim}{\theta}) = \underset{\sim}{\lambda}'\Sigma\underset{\sim}{\lambda} - 2\underset{\sim}{\theta}'(X'\underset{\sim}{\lambda} - \underset{\sim}{t})$$

$$\rightarrow \frac{\partial Q}{\partial \lambda} = 2\Sigma\underset{\sim}{\lambda} - 2X\underset{\sim}{\theta} = 0 \Rightarrow \Sigma\underset{\sim}{\lambda} = X\underset{\sim}{\theta}$$

$$\Rightarrow \underset{\sim}{\lambda} = \Sigma^{-1}X\underset{\sim}{\theta} \quad (1)$$

ترانهاده طرفین را در X ضرب می‌کنیم

$$\underset{\sim}{\lambda}'X = \underset{\sim}{\theta}'X'\Sigma^{-1}X$$

$$\Rightarrow \underset{\sim}{t}' = \underset{\sim}{\theta}'X'\Sigma^{-1}X$$

$$\Rightarrow \underset{\sim}{\theta}' = \underset{\sim}{t}'(X'\Sigma^{-1}X)^{-1} \quad (2)$$

در نتیجه از ۱ و ۲:

$$\underset{\sim}{\lambda} = \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}\underset{\sim}{t}$$

و یک برآوردگر خطی برای ترکیب خطی از $\underset{\sim}{\beta}$:

$$\underset{\sim}{\lambda}'\underset{\sim}{Y} = \underset{\sim}{t}'\underbrace{(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}}\underset{\sim}{Y} = \underset{\sim}{t}'\hat{\underset{\sim}{\beta}}$$

همان GLS برای β است.

~~برآورد واریانس (σ^2)~~ مقایسه بین:

~~تجزیه واریانس~~ $V(\hat{Y})$ ، $V(E(Y))$ ~~برآورد کمترین~~ :

دیدیم مدل به صورت $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ به طوری که صدق فرضیات آن می‌کند:

$$E(\underset{\sim}{Y}) = X\underset{\sim}{\beta} \quad , \quad Var(\underset{\sim}{Y}) = \sigma^2 I$$

از روش حداقل مربعات برآورد $\underset{\sim}{\beta}$:

$$\hat{E}(\underset{\sim}{Y}) = X\hat{\underset{\sim}{\beta}}$$

PAPCO

Subject :

Year . Month . Date . ( )

$$var(\hat{E}(y)) = var(X\hat{\beta}) = X\, var(\hat{\beta})\, X' = X\sigma^2(X'X)^{-1}X' \quad (1)$$

و: کنید

$$var(\hat{y}) = var(X\hat{\beta} + \hat{\varepsilon}) = var(X\hat{\beta}) + var\hat{\varepsilon} + 0 \quad (2)$$

$Y = X\beta + \varepsilon$

$\hat{Y} = X\hat{\beta} + \hat{\varepsilon}$

چون: $Cov(X\hat{\beta}, \hat{\varepsilon}) = 0$ زیرا

$$\varepsilon = Y - \underbrace{E(Y)}_{X\beta} \rightarrow \hat{\varepsilon} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y$$

$$= [I - X(X'X)^{-1}X']y = [I - P_X]y$$

P_X متقارن و خودتوان است همچنین $(I - P_X)$ هم خودتوان است.

از طرفی

$$X'[I - X(X'X)^{-1}X'] = 0 \Rightarrow X'[I - P_X]y - X'\hat{\varepsilon} = 0$$

$$\begin{cases} X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}_{A}y = Ay \\ \hat{\varepsilon} = \underbrace{[I - X(X'X)^{-1}X']}_{B}y = By \end{cases} \rightarrow By, Ay \text{ مستقلند اگر}$$

$$A\underbrace{\Sigma}_{\sigma^2 I}B' = 0 \rightarrow AB' = 0$$

$Cov(X\hat{\beta}, y - X\hat{\beta}) = 0$: و از خواص توان‌ها

پس از (۲):

$$var(\hat{y}) = X\sigma^2(X'X)^{-1}X' + var(\hat{\varepsilon})$$

و با مقایسه با (۱):

$$var(\hat{y}) > var(\hat{E}(y))$$

پیش‌بینی‌پذیری (واقعی) پیش‌بینی (مجهول)

PqPCO

Subject:
Year. Month. Date. ( )

برآورد σ^2 :

$\hat{\varepsilon} = (I - P_X)\, Y$ می دانیم $\hat{Y} = X\hat{B}$ برگردیم به محاسبه برآورد واریانس

$$\hat{\varepsilon}'\hat{\varepsilon} = [Y - \hat{E}(Y)]'[Y - \hat{E}(Y)]$$

$\hat{E}(Y) = X\hat{B}$, $X\hat{B} = X(X'X)^{-1}X'Y$

$$= Y'[I - P_X][I - P_X]Y = Y'[I - P_X]Y$$

$$= Y'\underbrace{[I - X(X'X)^{-1}X']}_{A}Y = \underbrace{Y'Y}_{SST} - \underbrace{\hat{B}'X'Y}_{SSR} = SSE$$

از طرفی می دانیم:

$$E(Y'AY) = tr(A\Sigma) + \mu'A\mu$$

بنابراین $\Rightarrow$ $E(\hat{\varepsilon}'\hat{\varepsilon}) = \sigma^2 tr(\underbrace{I - X(X'X)^{-1}X'}_{\text{ماتریس خودتوان}}) + \underbrace{\beta'X'(I - X(X'X)^{-1}X')X\beta}_{0}$

همچنین می دانیم: $tr(I - P_X) = r(I - P_X) = n - \underbrace{r(X)}_{(k+1)}$

$r(XX') = r(X) = k+1$ $\quad (X_{n\times k+1}\ \underbrace{(X'X)^{-1}}_{\text{ماتریس نادوره}}\ X'_{(k+1)\times n})_{n\times n}$

لذا به طور کلی داریم:

$$E(\hat{\varepsilon}'\hat{\varepsilon}) = [n - r(X)]\sigma^2 \Rightarrow \sigma^2\ E\left(\underbrace{\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n - r(X)}}_{\text{برآوردی نااریب برای } \sigma^2}\right) = \sigma^2$$

$$\hat{\sigma}^2 = \frac{Y'[I - X(X'X)^{-1}X']Y}{n - r(X)} = \frac{Y'[I - P_X]Y}{n - r(X)}$$

$$= \frac{Y'Y - \hat{\beta}'X'Y}{n - r(X)} = \frac{SSE}{n - r(X)}$$

PAPCO

(۲۷)

$$\hat{\varepsilon}'\hat{\varepsilon} = SSE = Y'[I - X(X'X)^{-1}X']Y = Y'Y - \hat{\beta}'X'Y$$

P_X — SST — SSR

$Y'Y - Y'P_XY$

Subject:

Year. Month. Date. ( )

اگر $(X'X)$ ناوردا باشد $r(X) = k+1$

$X'AX \sim \chi^2_{r(A),\ \frac{1}{2}\mu'A\mu}$ توزیع

$(Y/\sigma)'(I - X(X'X)^{-1}X')(Y/\sigma) \sim \chi^2_{n-k-1}$

$Y \sim N(X\beta, \sigma^2 I) \rightarrow Y/\sigma \sim N(\frac{X\beta}{\sigma}, I)$

تمرین: نشان دهید اگر فرض نرمالیتی برقرار باشد:

$$[Y'(I - P_X)Y]/\sigma^2 \sim \chi^2_{n-r(X)} \quad (1)$$

پارامتر غیرمرکزی $\mu'A\mu = 0$

$r(A\Sigma) = r(A)$

$(r(A), \frac{1}{2}\mu'A\mu)$ $\quad \frac{1}{\sigma^2}\beta X'[I - P_X]X\beta = \frac{1}{\sigma^2}(\beta'X'X\beta - \beta'X'P_X X\beta) = 0$

جدول آنالیز واریانس:

	df	SS	MS
رگرسیون (model)	$rank(X) = k+1$	$Y'P_XY = \hat{\beta}'X'Y$	$SSR/k+1$
Residual خطا	$n-k-1$	$Y'[I - P_X]Y$	$SSE/n-k-1$
Total کل	n	$Y'Y$	$MSR/MSE = F$

$\hat{\beta} = Y'X(X'X)^{-1} \rightarrow SSR = \hat{\beta}'X'Y = Y'P_XY$

ی توکن نگران SSR و SSE مستقلند:

$$SSR = Y'P_XY = Y\underbrace{X(X'X)^{-1}X'}_{A}Y$$

$$SSE = Y'\underbrace{[I - X(X'X)^{-1}X']}_{B}Y$$

(۴) طبق قضیه مستقلند $\rightarrow$ $AB = 0$

همچنین طبق قضیه توزیع مربع کای داریم:

$$SSR/\sigma^2 \sim \chi^{2'}(\underbrace{r(X(X'X)^{-1}X')}_{r(X)}, \frac{1}{2\sigma^2}\underbrace{\beta'X'X(X'X)^{-1}X'X\beta}_{\mu'A\mu}) \quad (3)$$

$\mu = E(Y) = X\beta$

و همراه است:

$$\frac{SSR}{\sigma^2} \sim \chi^{2'}(r(X), \frac{1}{2\sigma^2}\beta'X'X\beta)$$

و بنابراین از (۱)، (۲) و (۳):

$$\frac{SSR/r(X)}{SSE/n-r(X)} \sim F'(r(X), n-r(X), \frac{1}{2\sigma^2}\beta'X'X\beta)$$

$$F(R) = \frac{MSR}{MSE} \sim F'(r(X), n-r(X), \frac{1}{2\sigma^2}\beta'X'X\beta)$$

PdPCO

$SST = Y'Y \quad Y'I/\sigma^2 \sim \chi^{2'}(n, \beta'X'X\beta/2\sigma^2)$ (۴)

Subject :
Year . Month . Date . ( )

رگرسیون چندگانه
حل با انحراف از میانگین (~~مدل رگرسیون~~)

با درنظر گرفتن اثر متغیرهای X

$$X = [\underset{\sim}{1} \; X_1] \rightarrow \underset{\sim}{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1} \qquad X_1 = \begin{matrix} \\ n \end{matrix}\overset{\begin{matrix} x_1 & \cdots & x_k \end{matrix}}{\begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \\ x_{n1} & \cdots & x_{nk} \end{bmatrix}}_{n\times k}$$

ازطرفی داریم : $\overline{x}_{.1} \quad \cdots \quad \overline{x}_{.k}$

$$\underset{\sim}{1}'\underset{\sim}{1} = n \qquad \underset{\sim}{1}'\underset{\sim}{y} = \sum_{i=1}^{n} y_i = n\bar{y} \qquad \underset{\sim}{1}'X_1 = \left[\sum_{i=1}^{n} x_{i1} \; \cdots \; \sum_{i=1}^{n} x_{ik}\right]$$

$$= [n\bar{x}_{.1} \; \cdots \; n\bar{x}_{.k}]$$

$$= n\underset{\sim}{\bar{x}}'$$

$$\underset{\sim}{\bar{x}}' = [\bar{x}_{.1} \; \cdots \; \bar{x}_{.k}]$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = [\underset{\sim}{1} \; X_1]\begin{bmatrix} B_0 \\ \underset{\sim}{B}_1 \end{bmatrix} + \underset{\sim}{\varepsilon}$$

با جایگذاری در برآورد $\hat{\underset{\sim}{B}}$:

$$\hat{\underset{\sim}{B}} = (X'X)^{-1}X'\underset{\sim}{y} = \left[\begin{bmatrix} \underset{\sim}{1}' \\ X_1' \end{bmatrix}[\underset{\sim}{1} \; X_1]\right]^{-1}\begin{bmatrix} \underset{\sim}{1}' \\ X_1' \end{bmatrix}\underset{\sim}{y}$$

$$= \begin{bmatrix} n & n\underset{\sim}{\bar{x}}' \\ n\underset{\sim}{\bar{x}} & X_1'X_1 \end{bmatrix}^{-1}\begin{bmatrix} n\bar{y} \\ X_1'\underset{\sim}{y} \end{bmatrix}$$

طبق رابطه 48 صفحه 27 سیل :

$$= \begin{bmatrix} \frac{1}{n} + \underset{\sim}{\bar{x}}'S^{-1}\underset{\sim}{\bar{x}} & -\underset{\sim}{\bar{x}}'S^{-1} \\ -S^{-1}\underset{\sim}{\bar{x}} & S^{-1} \end{bmatrix}\begin{bmatrix} n\bar{y} \\ X_1'\underset{\sim}{y} \end{bmatrix} = \begin{bmatrix} \bar{y} - \underset{\sim}{\bar{x}}'S^{-1}(X_1'\underset{\sim}{y} - n\bar{y}\underset{\sim}{\bar{x}}) \\ S^{-1}(X_1'\underset{\sim}{y} - n\bar{y}\underset{\sim}{\bar{x}}) \end{bmatrix} \quad (1)$$

که در آن $S = X_1'X_1 - n\underset{\sim}{\bar{x}}\underset{\sim}{\bar{x}}'$

و با تجزیه بردار $\underset{\sim}{B}$ به صورت

$$\underset{\sim}{B} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_k \end{bmatrix} = \begin{bmatrix} B_0 \\ \underset{\sim}{B}_1 \end{bmatrix}$$

و از (1) :

$$\begin{cases} \hat{B}_0 = \bar{y} - \underset{\sim}{\bar{x}}'\hat{\underset{\sim}{B}}_1 \\ \hat{\underset{\sim}{B}}_1 = S^{-1}(X_1'\underset{\sim}{y} - n\bar{y}\underset{\sim}{\bar{x}}) \end{cases}$$

PAPCO

۲۸

Subject:

Year. Month. Date. ( )

$SSR_m = \hat{B}_1' X_c' y = y' X_c (X_c' X_c)^{-1} X_c' y$

۳۲۵۴۲۸۱

که S را می‌توان به صورت زیر نوشت :

$$S = X_c' X_c \quad , \quad X_c = X_1 - \mathbf{1}\bar{x}' = (I - \tfrac{1}{n} J) X_1 = H X_1$$

و با جایگذاری X_c^H در $\hat{B}_1$:

$$x_1' y - n\bar{y}\bar{x} = x_c' y$$

پس، $\hat{B}_1 = (x_c' x_c)^{-1} x_c' y = (X_1' H X_1)^{-1} X_1' H y$

خواهد شد. و خواهیم داشت :

$$var(\hat{B}_1) = (x_c' x_c)^{-1} \sigma^2 = (X_1' H X_1)^{-1} \sigma^2$$

$$var(\hat{B}_0) = \frac{\sigma^2}{n} + \bar{x}' var(\hat{B}_1) \bar{x} = \sigma^2 [\frac{1}{n} + \bar{x}' (x_c' x_c)^{-1} \bar{x}]$$

و نیز داریم :

$$Cov(\hat{B}_0, \hat{B}_1) = -\bar{x}' var(\hat{B}_1)$$

$\bar{y} - \bar{x}'\hat{B}_1$ ، $Cov(\bar{y}, b) - \bar{x}' Cov(\hat{b}, \hat{b})$ ، ترکیب خطی از

$$SSE = y'y - \hat{B}' x' y$$

از طرفی داریم :

$$= y'y - [\bar{y} - \hat{B}_1'\bar{x} \quad \hat{B}_1'] \begin{bmatrix} n\bar{y} \\ x_1' y \end{bmatrix}$$

$\hat{B}_0$ ، $\hat{B}_1$ ، $x' = [1 \ x_1]$ با جایگذاری

$[1 \ x_1]' y = \begin{bmatrix} n\bar{y} \\ x_1' y \end{bmatrix}$

$$= y'y - n\bar{y}^2 - \hat{B}_1' (x_1' y - n\bar{y}\bar{x})$$

$$= y_c' y_c - \hat{B}_1' x_c' y$$

$y_c' y_c = y'(I - \frac{1}{n} J) y = y'y - n\bar{y}^2$ ~~$= SSM - n\bar{y}^2$~~

مجموع مربعات تصحیح شده y ها $= SST_m = SST - SSM$

$SSR_m = SSR - SSM$

$y' [(X_c' X_c)^{-1} X_c'] y$: SSM به دست آمده

$$SSM = n\bar{y}^2 = y' n^{-1} \mathbf{1}\mathbf{1}' y$$

۸۳۸۸۹۹ ۳۳۶

$[\ ][1 \ \ ] = [\ \ \ \ ]$.

PAPCO

$X = [1 \;\; X_1]$ $\quad X_c = \left(I_n - \frac{1}{n} J_n\right) X_1 = H X_1$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1' \bar{x}$

$\hat{\beta}_1 = (X_c' X_c)^{-1} X_c' y$

$= (X_1' H X_1)^{-1} X_1' H y$

خلاصه:

Subject:

Year. Month. Date. ( )

SSM ⇓ SSTm ⇓ SSRm ⇓

SST − SSM SSR − SSM

$n\bar{y}^2 = y' n^{-1} 1 1' y$

الف) نحوی توان نشان داد:

$$\frac{SSM}{\sigma^2} \sim \chi^{2'}\left(r(n^{-1}11'), \beta' X' n^{-1} 11' X \beta / 2\sigma^2\right)$$

$(\beta' X' 1)(1' X \beta)$ 1×1

$$\Rightarrow \frac{SSM}{\sigma^2} \sim \chi^{2'}\left(1, (1' X \beta)^2 / 2n\sigma^2\right)$$

$\hat{\beta}_1'(X_c' X_c)(X_c' X_c)^{-1} X_c' y$

ب) همچنین داریم:

$$\begin{cases} SSR_m = \hat{\beta}_1' X_c' y = \hat{\beta}_1' (X_c' X_c) \hat{\beta}_1 \\ \hat{\beta}_1 \sim N(\beta_1, (X_c' X_c)^{-1} \sigma^2) \end{cases} \Rightarrow \mu = \beta_1 \quad A = X_c' X_c \quad \Sigma = (X_c' X_c)^{-1} \sigma^2$$

$$\frac{SSR_m}{\sigma^2} \sim \chi^{2'}\left(r(X_c' X_c), \beta_1' X_c' X_c \beta_1 / 2\sigma^2\right)$$

$$\left(\overbrace{\quad}^{K}, \beta_1' X_c' X_c \beta_1 / 2\sigma^2\right)$$

$(X_c)_{n\times k}$

$r(X_c' X_c) = r(X_c')$

$r(A\Sigma) = r(I)$ $\quad k\times k$

ج) درنهایت جدول آنالیز واریانس را می‌توان به صورتی زیر نوشت:

منبع	df	SS	MS
میانگین	1	$n\bar{y}^2 = SSM = y' n^{-1} 11' y$	$SSM = MS_M$
رگرسیون	K	$\hat{\beta}_1' X_c' y = SSR_m$	$SSR_m/k = MSR_m$
cfm → corrected for mean	$n-k-1$	$y'y - n\bar{y}^2 - \hat{\beta}_1' X_c' y = SSE$	$SSE/n-k-1$
کل	n	$y'y = SST$	

(2) $F(M) = \frac{MSM}{MSE}$ $\qquad F(R_m) = \frac{MSR_m}{MSE}$

(1) $F_{(M)} = \frac{SSM/1}{SSE/n-k-1} \sim F'(1, n-k-1, (1' X \beta)^2 / 2n\sigma^2)$

PAPCO

Subject :

Year . Month . Date . ( )

$$F(R_m) = \frac{SSR_m/k}{SSE/n-k-1} \sim F'(k, n-k-1, \beta_1' \mathcal{X}_c' \mathcal{X}_c \beta_1 / 2\sigma^2)$$

منبع	dR	SS	MS
Regression (کل)	k	$SSR_m = \hat{\beta}_1' \mathcal{X}_c' y$	MSR_m
error	$n-k-1$	$SSE = y'y - n\bar{y}^2 - \hat{\beta}_1' \mathcal{X}_c' y$	MSE
Total	$n-1$	$SST_m = \underbrace{y'y}_{SST} - \underbrace{n\bar{y}^2}_{SSM} = y_c' y_c = y'(I - \frac{1}{n})y = y'Hy$	

$$F(R_m) = \frac{MSR_m}{MSE} \sim F'(k, n-k-1, \beta' \mathcal{X}' \mathcal{X} \beta / 2\sigma^2)$$

نکته ۱: تحت فرض $H_0: \beta = 0$ (در صدل مرکزی) $F(R) \geq F_{k+1, n-k-1} \Rightarrow RH_0$

نکته ۲: تحت فرض $H_0: E(\bar{y}) = 0$ (در صدل مرکزی) ← پارامتر غیر مرکزی صفر می‌شود

$1'X\beta = 1'E(y) = E(1'y) = nE(\bar{y}) = 0 \Rightarrow F(M) \sim F(1, n-k-1)$

$\Rightarrow F(M) \geq F(1, n-k-1) \Rightarrow RH_0$

نکته ۳: تحت فرض $H_0: \beta_1 = 0$:

$F(R_m) \sim F(k, n-k-1)$

$\Rightarrow F(R_m) \geq F_{n, n-k-1} \Rightarrow RH_0$

مثال صفحه ۱۰۶ سیدل.

نکته ۴: قاعده اخیا برای هر ترکیب خطی ضرایب رگرسیون به صورت زیری باشد

$$q'\beta : q'\hat{\beta} \pm \hat{\sigma} t_{n-k-1} \sqrt{q'(X'X)^{-1}q}$$

PAPCO

Subject:
Year. Month. Date. ( )

نکته: با داشتن برآورد ضرایب رگرسیونی و مقادیر متغیرهای توضیحی می‌توان یک y جدید را پیش‌بینی کرد

$$\hat{E}(y_0) = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \cdots + \hat{\beta}_k x_{0k} = \underset{\sim}{x}_0' \hat{\underset{\sim}{\beta}}$$

و یک فاصله اطمینان برای $\underset{\sim}{x}_0' \underset{\sim}{\beta}$:

$$\underset{\sim}{x}_0' \underset{\sim}{\beta} : \underset{\sim}{x}_0' \hat{\underset{\sim}{\beta}} \pm \hat{\sigma}\, t_{n-k-1,\alpha/2} \sqrt{\underset{\sim}{x}_0' (X'X)^{-1} \underset{\sim}{x}_0}$$

برای σ^2: $\left(\dfrac{SSE}{\chi^2_{n-k-1,\alpha/2}} , \dfrac{SSE}{\chi^2_{n-k-1,1-\alpha/2}} \right)$ نکته: $\frac{SSE}{\sigma^2} \sim \chi^2_{n-k-1}$

بررسی فرضهای خطی:

صورت عمومی انواع فرضهای خطی را می‌توان به صورت $H_0: K'\underset{\sim}{\beta} = \underset{\sim}{m}$ نوشت

که K در حالت کلی یک ماتریس و در حالت ساده یک بردار است و برای سازگاری بودن $K'\underset{\sim}{\beta} = \underset{\sim}{m}$ باید $K'_{s \times k+1}$ دارای رتبه کامل باشد. $r(K'K) = r(K) = r(K') = s$

دیدیم: $\underset{\sim}{Y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I)$ و با توجه به وجود $(K'K)^{-1}$

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1} X' \underset{\sim}{Y} \sim N(\underset{\sim}{\beta}, (X'X)^{-1}\sigma^2)$$

لذا:

$$K'\hat{\underset{\sim}{\beta}} - \underset{\sim}{m} \sim N(K'\underset{\sim}{\beta} - \underset{\sim}{m}, K'(X'X)^{-1}K\sigma^2)$$

تحت H_0 $K'\hat{\underset{\sim}{\beta}} - \underset{\sim}{m} \sim N(0, K'(X'X)^{-1}K\sigma^2)$

اکنون با در نظر گرفتن فرم درجه دوم به صورت:

$$Q = \underbrace{(K'\hat{\underset{\sim}{\beta}} - \underset{\sim}{m})'}_{1\times s} \underbrace{(K'(X'X)^{-1}K)^{-1}}_{A} \underbrace{(K'\hat{\underset{\sim}{\beta}} - \underset{\sim}{m})}_{s\times 1} \qquad K_{s\times k+1}$$

طبق قضیه $Q/\sigma^2 \sim \chi^{2'}(s, \delta)$ که $\delta = (K'\underset{\sim}{\beta} - \underset{\sim}{m})'(K'(X'X)^{-1}K)^{-1}(K'\underset{\sim}{\beta} - \underset{\sim}{m})/2\sigma^2$

$$\begin{cases} A = \frac{1}{\sigma^2}(K'(X'X)^{-1}K)^{-1} \\ A\Sigma = I_{s\times s} \end{cases} \quad \Sigma = \mathrm{var}(K'\hat{\underset{\sim}{\beta}} - \underset{\sim}{m}) = K'(X'X)^{-1}K\sigma^2$$

، $A\Sigma \Rightarrow$ خودتوان

Subject :

Year . Month . Date . ( )

$\text{rank}(K') = s \Rightarrow r(K'(X'X)^{-1}K) = s \longrightarrow K'(X'X)^{-1}K \rightarrow$ ماتریس $s \times s$ است و وارون آن وجود دارد

نکته :

تحت فرض H_0 :

$$\frac{Q}{\sigma^2} \overset{H_0}{\sim} \chi^2_s \qquad (1)$$

که σ^2 مجهول نامعلوم است و باید برآورد کنیم که در این

$$\begin{cases} \hat{\underset{\sim}{\varepsilon}} = \underset{\sim}{Y} - \hat{E}(\underset{\sim}{Y}) = \underset{\sim}{Y} - X(X'X)^{-1}X'\underset{\sim}{Y} = (I - P_X)\underset{\sim}{Y} \\ SSE = \hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}} = \underset{\sim}{Y}'(I - P_X)\underset{\sim}{Y} \qquad MSE = \dfrac{SSE}{n-k-1} \\ \text{rank}(X) = \text{rank}(P_X) = K+1 \qquad \text{rank}(I - P_X) = n-k-1 \end{cases}$$

لازم است

$$(2)\ \frac{SSE}{\sigma^2} \sim \chi^2_{n-k-1}$$ در نتیجه $\hat{\sigma}^2 = \dfrac{SSE}{n-k-1}$: در نهایت

از ۱ و ۲ :

$$F(H) = \frac{Q/s}{SSE/n-k-1} \overset{H_0}{\sim} F(s, n-k-1) \rightarrow F(H) > F_{s,n-k-1} \Rightarrow RH_0$$

تمرین : ثابت کنیم Q و SSE مستقل

$$SSE = \underset{\sim}{Y}'(I - X(X'X)^{-1}X')\underset{\sim}{Y}$$

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

$K'\hat{\underset{\sim}{\beta}} = \underset{\sim}{m}$ صورت داریم

چون $X'[I - X(X'X)^{-1}X']$ ، $[I - X(X'X)^{-1}X']X$

$$SSE = [Y - XK(K'K)^{-1}m]'[I - X(X'X)^{-1}X'][Y - XK(K'K)^{-1}m]$$

X' A X

$$Q = (K'(X'X)^{-1}X'\underset{\sim}{Y} - \underset{\sim}{m})'(K'(X'X)^{-1}K)^{-1}(K'(X'X)^{-1}X'\underset{\sim}{Y} - m)$$

$$K'(X'X)^{-1}X'[Y - XK(K'K)^{-1}m]$$

PqPCO

Subject:

Year. Month. Date. ( )

$$Q = [y - XK(K'K)^{-1}m]' X(X'X)^{-1}k[k'(X'X)^{-1}k]^{-1}k'(X'X)^{-1}X' [y - XK(K'K)^{-1}m]$$

X' B $AB = 0$ X

$var(X) = \sigma^2 I$

$$p(I) = \alpha = p(F_{(H)} > F_{s,n-k-1,\alpha} \mid H_0 \text{ is True})$$

$$p(II) = \beta = p(F_{(H)} < F_{s,n-k-1,\alpha} \mid H_0 \text{ is false})$$

$F'(s,n-k-1,\delta)$

اگر H_0 نادرست باشد،

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

$$\text{power} = 1-\beta = p(F > F_{s,n-k-1,\alpha} \mid H_0 \text{ is false})$$

$$(y - Xpm - XSL'B^\circ)'(y - Xpm - XSL'B^\circ) + 2\theta'(k'B^\circ - m$$

$$LS'X'XSL'B^\circ + k\theta = LS'X'y - LS'X'mp'm'$$

Subject :

Year . Month . Date . ( )

~~اول روش دوم تابع لاگرانژ~~

برآورد تحت قیود خطی فرض H_0 :

$H_0 : K'\underset{\sim}{\beta} = \underset{\sim}{m}$

انگیزه هدف در بدست آوردن برآورد با فرض اینکه فرض H_0 برقرار باشد می‌باشد

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

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

تحت H_0 می‌خواهیم برآورد $\underset{\sim}{\beta}^*$ را برای $\underset{\sim}{\beta}$ بدست آوریم

که باعث کمینه کردن مجموع مربعات خطا :

$$(\underset{\sim}{y} - X\underset{\sim}{\beta}^*)'(\underset{\sim}{y} - X\underset{\sim}{\beta}^*) \quad \star$$

تحت $K'\underset{\sim}{\beta}^* = \underset{\sim}{m}$ ، $\underset{\sim}{\beta}^*$ را بدست می‌آوریم.

از لاگرانژ داریم :

$$(\underset{\sim}{y} - X\underset{\sim}{\beta}^*)'(\underset{\sim}{y} - X\underset{\sim}{\beta}^*) + 2\underset{\sim}{\theta}'(K'\underset{\sim}{\beta}^* - \underset{\sim}{m})$$

مشتق‌گیری از تابع لاگرانژ نسبت به $\underset{\sim}{\beta}^*$ و $\underset{\sim}{\theta}$ داریم :

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

با توجه به $\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$ و (۱) :

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

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

$$\Rightarrow \underset{\sim}{\beta}^* = \hat{\underset{\sim}{\beta}} - (X'X)^{-1}K[K'(X'X)^{-1}K]^{-1}(K'\hat{\underset{\sim}{\beta}} - \underset{\sim}{m}) \quad (3)$$

تمرین : نشان دهید با جایگذاری $\underset{\sim}{\beta}^*$ در ($\star$) :

مجموع مربعات خطا $= SSE + Q$

Subject :

Year . Month . Date . ( )

آزمون فرض H_0 بر روی β کلی (دستگاهی)

مثال :

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

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

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

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

$$\Rightarrow \hat{\beta} = (X'X)^{-1}X'Y$$

تمرین: ثابت کنید (برآوردهای) دستگاهی $\hat{\sigma}^2 = \frac{1}{n}(Y-X\hat{\beta})'(Y-X\hat{\beta})$

$\hat{\sigma}^2 =$ ـــــ

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

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

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

$$-\frac{1}{2}(Y-X\beta^*)'(Y-X\beta^*)/SSE+Q/n$$

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

آماره نسبت درستنمایی

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

که توزیع

$$F(H) = \frac{Q/s}{SSE/n-k-1} \qquad Q = (k'\hat{\beta}-m)'(k'(X'X)^{-1}k)^{-1}(k'\hat{\beta}-m)$$

حالت خاص فرض $k'\beta = m$:

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

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

PAPCO

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

Subject :

Year . Month . Date . ( )

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

۱) $H_0 : \underset{\sim}{B} = \underset{\sim}{B}_0$ $\quad k' = I \quad s = k+1 \quad \underset{\sim}{m} = \underset{\sim}{B}_0$

$\Rightarrow [k'(X'X)^{-1}k]^{-1} = X'X$

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 می‌توان نشان داد

۲) $H_0 : \underset{1\times k+1}{\underset{\sim}{\lambda}'}\underset{\sim}{B} = \underset{\sim}{m}$ $\quad K' = \underset{\sim}{\lambda}' \Rightarrow s = 1 \;,\; \underset{\sim}{m} = m$

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

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

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

t^2

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

۳) $H_0 : \underset{\sim}{B}_q = \underset{\sim}{0}$ i.e $B_i = 0 \;\; \forall i = 0, 1, 2, \dots q-1 \quad q < K$

$\underset{\sim}{B}$ از q صفر

$$\begin{bmatrix} B_0 \\ \vdots \\ B_{q-1} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \longrightarrow B_i = 0 \;\; \forall i = 0, \dots q-1$$

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

$$\begin{bmatrix} B_0 \\ \vdots \\ B_{q-1} \end{bmatrix} = \underset{q\times(k+1)}{\begin{bmatrix} 1 & & & 0 & \cdots & 0 \\ & \ddots & & & & \\ & & 1 & & & \end{bmatrix}} \underset{k+1\times 1}{\begin{bmatrix} B_0 \\ \vdots \\ B_k \end{bmatrix}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{q\times 1}$$

q (over the 1's) $\quad p$ (over the 0's)

$p + q = K + 1$

PdPCO

Subject:

Year. Month. Date. ( )

$$Q=(k'\hat{\beta}-m)'(k'(x'x)^{-1}k)^{-1}(k'\hat{\beta}-m)$$

$$(\hat{\beta}_q)'T_{qq}^{-1}\hat{\beta}_q$$

سپس در این حالت

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

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

و بنابراین :

$$[k'(X'X)^{-1}k]^{-1}=T_{qq}^{-1} \Rightarrow k'\hat{\beta}=\hat{\beta}_q$$

و در انجالت داریم :

$$F(H)=\hat{\beta}_q'T_{qq}^{-1}\hat{\beta}_q/q\hat{\sigma}^2 \sim F_{q,n-k-1}$$

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

$$\tilde{\beta}=\hat{\beta}-(X'X)^{-1}\begin{bmatrix}L_q\\ 0\end{bmatrix}T_{qq}^{-1}(\hat{\beta}_q-0)$$

از $\tilde{\beta}$

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

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

$$k'=[L_q\ 0]$$

$$m=0$$

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

PAPCO

Subject :
Year . Month . Date . ( )

جدول آنالیز واریانس برای مدل تقلیل یافته :

در عمل ممکن است علاوه بر آنکه محدودیتهای برای انتخاب مدل در نظر می‌گیریم محدودیتهای هم برای حضای پارامتری در نظر بگیریم پس می‌توان به دو صورت مسئله برسیم داشته باشیم مدل کامل و مدل تقلیل یافته :

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

Full model

مدل تقلیل یافته $\left[\underset{\sim}{Y} = X\underset{\sim}{\beta} + \varepsilon \;,\; E(\underset{\sim}{Y}) = X\underset{\sim}{\beta} \;,\; var(\underset{\sim}{Y}) = \sigma^2 I \;,\; K\underset{\sim}{\beta} = \underset{\sim}{m} \right]$

Reduced model

برای مثال فرض کنید مدل کامل به صورت :

$$\underset{\sim}{y} = \beta_0 + \beta_1 \underset{\sim}{x_1} + \beta_2 \underset{\sim}{x_2} + \underset{\sim}{\varepsilon}$$

که برداری نوشته :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

اکنون تحت فرض $H : \beta_1 = \beta_2$ مدل تقلیل یافته به صورت

$$y = \beta_0 + \beta_1 (x_1 + x_2) + \varepsilon$$

که برداری نوشته

$$y_i = \beta_0 + \beta_1 (x_{i1} + x_{i2}) + \varepsilon_i$$

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

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

$$SSE' = SSE + Q$$

آماره‌های تست $H_0 : \beta_1 = \beta_2$:

اما مجموع مربعات کل SS_y تغییری نمی‌کند چون به β ها بستگی ندارد

$$SS_y = \underset{\sim}{y}'\underset{\sim}{y} - n\bar{y}^2$$

Subject:
Year. Month. Date. ( )

اما اگر فرض کنیم $H_0: \beta_1 = \beta_2 + 4$ یعنی

$k' = [0 \ 1 \ -1]$

$m = 4$

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \rightarrow$ Full

تحت فرض $\Rightarrow y = \beta_0 + (\beta_2 + 4) x_1 + \beta_2 x_2 + \varepsilon$

$\Rightarrow y - 4x_1 = \beta_0 + \beta_2 (x_1 + x_2) + \varepsilon$

در اینجالت SS_y هم تغییرات می کند.

حالت اول: $K'\beta - m = 0$ فرض کلی $k'\beta = m$ را در نظر میگیریم

اگر $k'\beta = m$ ، $y = X\beta + \varepsilon$ پس $E(y) = X\beta, \ var(y) = \sigma^2 I$

اکنون L' را طوری در نظر می گیریم که ماتریس

ماتریس دلخواه

$$R_{(k+1)(k+1)} = \begin{bmatrix} k' \\ L' \end{bmatrix}_{(k+1-s)\times(k+1)} \rightarrow s \times k+1$$

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

و حال را بازنویسی می کنیم:

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

$$y = X R^{-1} R \beta + \varepsilon$$

$$= X [P \ S] \begin{bmatrix} k'\beta = m \\ L'\beta = \gamma \end{bmatrix} + \varepsilon$$

$$\Rightarrow (*) \quad y - XPm = XSL'\beta + \varepsilon \qquad y^*$$

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

$(k+1-s)\times 1$

L' دلخواه است و ماتریس های زیادی می توانیم تعریف کنیم، شرایط را برآورده کند

لذا L' منحصر به فرد نیست.

$(k'\beta)_{s\times 1}$

$(L'\beta)_{(k+1-s)\times 1}$

* اگر $m = 0$ روی SS_{xy} تاثیری ندارد. حتی اگر $m \neq 0$ باز هم به رابطه (*) توجه کنید.

SS_{11} تغییر می کند

PAPCO

Subject :

Year . Month . Date . ( )

در حالت خاص $\underset{\sim}{y} = X SL' \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ ، $k'\underset{\sim}{\beta} = \underset{\sim}{0}$

و

تبدیل یافته $SSE' = SSE + Q$

همچنین داریم : $Q = y'y - SSE - y'y^{*} + SSE + Q$ اضافه $y'y$

$= SSR - [y'y - (SSE+Q)]$

Reduced Residual (R) Reduction(R) SSR'

$\begin{cases} (SSE + Q) + (SSR - Q) = SST \\ (SSE + SSR = SST \end{cases}$

residual (F) Reduction(F)

Full ←

معدل آنالیز واریانس در این حالت ($k'\underset{\sim}{\beta} = \underset{\sim}{0}$) به صورت زیر می‌شود

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

مثلاً دیدیم

① $\frac{Q}{\sigma^2} \sim \chi^{2'}(S, \underset{\sim}{\beta}' k (k'(x'x)^{-1}k)^{-1} k' \underset{\sim}{\beta} / 2\sigma^2)$ از طرفی داریم

SSR

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

$\underset{\sim}{y}'X(X'X)^{-1}X'\underset{\sim}{y}$ $\underset{\sim}{y}'(X(X'X)^{-1}k(k'(X'X)^{-1}k)^{-1}k'(X'X)^{-1}X')\underset{\sim}{y}$ ②، ① و خواهیم داشت

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

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

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

P4PCO

Subject:

Year, Month, Date, ( )

و بنابراین :

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

برای بررسی معنی داری مدل کامل

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

برای بررسی فرض $H_0: k'\underset{\sim}{\beta}=0$

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

sub Hypothesis $l'\underset{\sim}{\beta}=\underset{\sim}{0}$

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

برای بررسی زیر فرض $l'\underset{\sim}{\beta}=\underset{\sim}{0}$

- حالت خاص دیگری که ممکن است مد نظر باشد به صورت $\underset{\sim}{\beta}_q = \underset{\sim}{0}$ است که در فرض کی $\underset{\sim}{m}=\underset{\sim}{0}$ ، $k' = [I_q \; 0]$ در نظر گرفته شده است

$$k' = \underbrace{\begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix}}_{q\times(p+q)}$$

$q \times q$ ، $q \times p$ ، $k+1$

$q \leq k$ تا از β ها صفر باشد

$$\underset{\sim}{\beta}_q' = [\beta_0 \; \beta_1 \; \cdots \; \beta_{q-1}]$$

که در این حالت دیدیم

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

و جدول آنالیز واریانس در این حالت به صورت زیر خواهد شد :

PAPCO

Subject:
Year. Month. Date. ( )

مثال صفحه 124 (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

$r(X)=3$

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

بدون عدد بایاس

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

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

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

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

فرض $H_0: \beta_1=\beta_2+4$

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

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

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

$$=152.2$$

P4PCO

Subject:
Year. Month. Date. ( )

$$F(H_1) = \frac{QS}{SSE/n-k} = \frac{\frac{152.2}{1}}{\frac{52.1}{2}} = 5.8$$

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

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

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

که می دانیم در این حالت SST تغییری کند: $y - x\beta_m = x_{SL}\beta + \varepsilon$

$$y^* : \quad 0 \quad 14 \quad 5 \quad -2 \quad 8$$

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

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

پس جدول آنالیز واریانس در این حالت

مدل $(*)$

	df	SS
رگرسیون (مدل تقلیل یافته)	2	84.7
خطا (مدل تقلیل یافته)	3	204.3
کل	5	289

$\hat{B}$ را بدست آورید.

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

$$m = 0 \qquad k' = [1\ 0\ 0]$$

$$Q = \hat{B}_1^2 / (k'(x'x)^{-1}k)^{-1} = 8.9$$

$(-1.39)^2$ ، 0.2145

در این حالت می دانیم SST تغییری کند.

	df	SS
رگرسیون کامل	3	372.9
فرض / مدل تقلیل یافته	1	8.9
	2	364
خطا	2	52.1
کل	5	425

$\hat{B}$ را به دست آورید.

PAPCO

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

(1)

Subject:

Year. Month. Date. ( )

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

منبع	d.f	SS
رگرسیون (Full)	$k+1$	$SSR = \hat{B}' X' y$
تحت فرض $B_q = 0$	q	$Q = \hat{B}_q' T_{qq}^{-1} \hat{B}_q$
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 صفحه 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

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

در هر صورت R^2 عددی بین صفر و یک است و جذر آن همبستگی ساده بین y_i ها مشاهده شده و $\hat{y}_i$ که برازش شده را می‌دهد. یعنی

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

نکته: R^2 را می‌توان برای کل ماتریس ضریب همبستگی بصورت زیر بیان کرد $P = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ & \ddots & & \\ \rho_{1p} & & & 1 \end{bmatrix}$

$$R = \begin{bmatrix} 1 & r_{yx}' \\ r_{yx} & R_{xx} \end{bmatrix} = \begin{bmatrix} 1 & r_{y1} & r_{y2} & \cdots & r_{yk} \\ \hline r_{1y} & 1 & r_{12} & \cdots & r_{1k} \\ \vdots & & \ddots & & \\ r_{ky} & r_{k1} & & & 1 \end{bmatrix}$$

R_{yx} ↓ ضریب همبستگی بقیه برای $(y, x_1 \cdots x_k)$

ماتریس ضریب همبستگی بین xها

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

بدون y

Subject:
Year. Month. Date. ( )

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

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 درصورتی 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 $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ با فرض $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ ، $Cov(\underset{\sim}{\varepsilon}) = \sigma^2 I$ در نظر گرفته می‌شود. به طوری که فرض کردیم X ثابت و رتبه ستونی کامل است.

فصل ۹ ریخر کامل خوانده شود. فرضیاتی که معمولاً مورد آزمون قرار می‌دهند: (صفحه ۴ مونتگمری)

Normality (نرمالیتی) ، Homoscedasticity (هموارینس) ، Autocorrelation (خودهمبستگی / بین خطاهای مدل همبستگی وجود ندارد)

multicollinearity (هم‌خطی / بین متغیرهای مستقل همبستگی وجود ندارد)

روش معمول: محاسبه خطای استاندارد شده و رسم نمودار P.P برای نرمالیتی و خطای استاندارد شده و $\hat{y}$ را رسم می‌کنند برای بررسی هم واریانس.

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 می‌باشد.

PAPCO

Subject :

Year . Month . Date . ( )

مقدار کمتر از ۲ همبستگی پیاپی مثبت (مقدار باقیمانده مثبت برای یک دوره نشان مثبت بودن باقیمانده مشاهده دیگر را افزایش می دهد) و مقدار بیشتر از ۲ همبستگی پیاپی منفی را در باقیمانده ها نشان می دهد.

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

R نرم‌افزار : / lmtest / dwtest

منظور از هم خطی که معمولاً باید در رگرسیون چندگانه مورد بررسی قرار گیرد میزان ارتباط خطی بین متغیرهای مستقل است. معمولاً هم خطی در مدلهای رگرسیونی چندگانه وجود دارد و نکته هم خطی باید بررسی شود و وجود هم خطی بالا، باعث نقض فرضیات آن می خواهد شد.

برای تخمین هم خطی معمولاً در رگرسیون معنی دار بودن R^2 نزدیک است اما شاهد می شوید که تنها ضرایب معنی دار نیستند. اگر یک متغیر از مدل خارج یا اضافه شود R^2 تغییر زیادی نداشته باشد آن متغیر مستعد ایجاد هم خطی است. دو معیار tolerance و VIF هم برای بررسی هم خطی استفاده می شود.

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

ضریب تعیین رگرسیون بین متغیر مستقل j ام

if $tol < 0.2$ or $VIF > 10 \Rightarrow$ هم خطی 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 یا آزمون دارند برای تخمین آنها از

Cook's → مشاهده مؤثر ⇒ خطا یا ارزش نادرست داده‌ها (دور از روند X یا تأثیر زیاد روی خطا) صفحه 236 نیمه

PRESS → مشاهده دور افتاده ⇒ مدلها با PRESS کوچکتر ارجح‌تر (خوبی برازش)

Box-plot

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

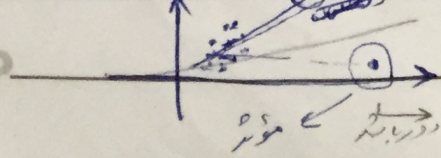
اثر حذف این نقطه بر $\hat{\beta}$ مورد بررسی قرار می‌گیرد

(y_i, X_i') ⟶ سپس n فاصله کوک قابل محاسبه است

⇒ هرچه D_i بزرگتر باشد مشاهده (y_i, X_i') تأثیر زیادی بر $\hat{\beta}$ دارد

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

یا فاصله فاصله لوس



دور افتاده در جهت y نسبت به بقیه داده‌ها

دور افتاده در جهت x ← مؤثر

Subject :
Year . Month . Date . ( )

مدلهای خطی رتبه ناقص :

Practical Regression and anova ← فصل V
using R. J. Faraway 2002
پروژه تعریف شود (در سول جنبه کامل)

تشخیص مدل مناسب ، بررسی مدل ---

دکتر لسفارسی R آقای موسوی ص ۱۱۶

جدول 7-3 ص ۱۸۳ → داده های بارشترین → زلزله - لغزش
رنجر لاین
داده های اجاره بها ص ۱۸۴ → رای - عضویت

جدول 7-5 {
- مونت کارلو - جدول B-5 پیوست → احمدخانی - کاظم فنی - ترابی
- مونت کارلو - جدول B-6 پیوست → نوجوان - محمود حامدی - اکبری لوان جنبه زاده
- مونت کارلو جدول B-2 پیوست → اکبری - علیزاده اعدادی