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 بر اساس یک رابطه خطی با یک متغیر مستقل x پیش بینی می شود اما در رگرسیون چندگانه یک متغیر وابسته y را بر اساس یک رابطه خطی با چند متغیر مستقل پیش بینی می کنیم

مستقل: $\underset{\sim}{X} = [x_1 \dots x_k]$ وابسته: y

لذا در این حالت مدل کلی :

$$y = B_0 + B_1 x_1 + \dots + B_k x_k + \varepsilon$$

وقتی به مدل، مدل خطی می گوییم که نسبت به B ها خطی باشد. برای برآورد $\hat{B}$ و تشکیل خط رگرسیون نیاز به داشتن نمونه می باشد. فرض کنید نمونه ای به حجم n در دست می باشد

$$\begin{array}{lccccccl} & & & x_1 & & x_k & & \\ \text{فرد اول} & y_1 & x_{10}=1 & x_{11} & \dots & x_{1k} & \to & y_1 = B_0 + B_1 x_{11} + \dots + B_k x_{1k} + \varepsilon_1 \\ & \vdots & \vdots & \vdots & & \vdots & & \vdots \\ \text{فرد n ام} & y_n & x_{n0}=1 & x_{n1} & \dots & x_{nk} & \to & y_n = B_0 + B_1 x_{n1} + \dots + B_k x_{nk} + \varepsilon_n \\ & & B_0 & B_1 & \dots & B_k & & \end{array}$$

که به صورت ماتریسی می توان نوشت :

$$\underset{\sim}{y} = X\underset{\sim}{B} + \underset{\sim}{\varepsilon}$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ \vdots & & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} B_0 \\ \vdots \\ B_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

فرضیات اساسی رگرسیون چندگانه :

1) $E(\underset{\sim}{y}) = X\underset{\sim}{B}$ یا $E(\underset{\sim}{\varepsilon}) = 0$

2) $Var(\underset{\sim}{\varepsilon}) = E[(\underset{\sim}{\varepsilon} - E(\underset{\sim}{\varepsilon}))(\underset{\sim}{\varepsilon} - E(\underset{\sim}{\varepsilon}))'] = \sigma^2 I_n \to$ واریانس ثابت خطاها دو به دو ناهمبسته

$$\frac{\partial x'Ax}{\partial x} = 2Ax$$

$$\frac{\partial x'Ay}{\partial x} = Ay$$

Subject:
Year. Month. Date.

و طبق فرضیات فوق مشابه رگرسیون خطی ساده برآوردهای حداقل مربعات قابل محاسبه اند :

$$\sum \varepsilon_i^2 = \varepsilon'\varepsilon = [\overbrace{Y}^{XB+\varepsilon} - \overbrace{E(Y)}^{XB}]'[Y - E(Y)] = (Y - XB)'(Y - XB)$$

$$= Y'Y - 2B'X'Y + B'X'XB$$

باید $\hat{B}$ ای به دست آوریم که عبارت فوق را مینیمم کند پس:

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

$$\Rightarrow X'X\hat{B} = X'Y \longrightarrow$$ دستگاه معادلات نرمال (1)

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

$$r[X'X \quad \underbrace{X'Y}_{XB}] = r(X'X)$$

$X'X$ ترکیب خطی از ←

~~$X'XB$~~

$Y = AX \qquad A^{+} = [A \quad Y] \qquad r(A^{+}) = r(A)$

یعنی چه $(X'X)$ یک ماتریس نادژه باشد یا نباشد معادله جواب دارد

اگر $X'X$ نادژه باشد :
بخواهد

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

زیرا در حالت A ،
$r(AA') = r(A'A)$
$= r(A') = r(A)$

$$\xrightarrow{\text{اگر بخواهد}} r(X) = k+1$$

نادژه باشد باید ستون رتبه کامل باشد

~~$r(X'X) = k+1$~~ $\qquad r(X'X) = k+1 \qquad r(X) = r(X') = k+1$

$r(X'X) \le \underbrace{\cancel{\min(r(X), r(X'))}}_{k+1} \qquad \Rightarrow n \ge k+1$

پس برای اینکه $X'X$ نادژه باشد؟ باید X یک ماتریس ستونی کامل باشد یا به عبارتی $n \ge k+1$

و هنگامی $X'X$ دژه خواهد بود که $n < k+1$ و وارون آن وجود ندارد

Subject:
Year. Month. Date. ( )

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

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

به راحتی می توان نشان داد :

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

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

قضیه گوس-مارکف Gauss_markov

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

برآوردگر :

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

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

مثال ص 18 سیرل.

$E(X'AX) = tr(A\Sigma) + \mu'A\mu$

چرا ستون‌های رتبه کامل نیستند غیر مرکزی

روش حداقل مربعات تعمیم یافته :

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

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

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

در مدل کلی : $\quad \underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

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

$$\Rightarrow \underset{\sim}{y}^{*} = X^{*}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}^{*}.$$

به طوری که $var(y^{*}) = I$ سپس کافیست $(y^{*} - X^{*}\underset{\sim}{\beta})'(y^{*} - X^{*}\underset{\sim}{\beta})$ مینیمم شود و β را برآورد کنیم که مشابه حالت قبلی که واریانس ثابت بود :

$$\tilde{\underset{\sim}{\beta}} = (X^{*\prime}X^{*})^{-1}X^{*\prime}\underset{\sim}{y}^{*}$$

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

به طوری که :

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

روش ماکسیمم درستنمایی : فرض کنید

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

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

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

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

$\underset{\sim}{\beta}$ ای که عبارت فوق را ماکسیمم می‌کند $\underset{\sim}{\beta}$ ای است که توان را min می‌کند که در OLS داریم

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

اگر $var(Y) = \Sigma$

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

Subject:
Year. Month. Date. ( )

اگر $\text{var}(\underline{Y}) = \Sigma$:

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

روش GLS :

$$\underline{\tilde{B}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underline{Y}$$

روش بدست آوردن برآوردگر BLUE برای ترکیب خطی از بردار $\underline{B}$ (مثل $\underline{t}'\underline{B}$):

فرض $\underline{Y} = X\underline{B} + \underline{\varepsilon}$ (بردار پاسخ)، $E(\underline{Y}) = X\underline{B}$ و $\text{var}(\underline{Y}) = \Sigma$ است می‌خواهیم. برای $\underline{t}'\underline{B}$ برآوردی بیابیم که

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

(L)inear خطی بودن — (U)nbiasedness ناریبی — minimum variance (B)est

برای اینکه برآوردگر خطی باشد باید ترکیبی خطی از مشاهدات $\underline{Y}$ باشد. برآوردگر را به صورت $\underline{\lambda}'\underline{Y}$ در نظر می‌گیریم، که در آن $\underline{\lambda}_{1\times n}$ یک بردار سطری از مرتبه n داریم:

$$\widehat{\underline{t}'\underline{B}} = \underline{\lambda}'\underline{Y}$$

می‌خواهیم ناریب باشد پس

$$E(\widehat{\underline{t}'\underline{B}}) = \underline{t}'\underline{B}$$

$$\Rightarrow E(\widehat{\underline{t}'\underline{B}}) = E(\underline{\lambda}'\underline{Y}) = \underline{\lambda}'E(\underline{Y}) = \underline{\lambda}'X\underline{B} \overset{\text{می‌خواهیم}}{=} \underline{t}'\underline{B}$$

$$\Rightarrow \underline{\lambda}'X = \underline{t}' \quad \star$$

از طرفی واریانس هم مینیمم باشد

$$\text{var}(\widehat{\underline{t}'\underline{B}}) = \text{var}(\underline{\lambda}'\underline{Y}) = \underline{\lambda}'\text{var}(\underline{Y})\underline{\lambda} = \underline{\lambda}'\Sigma\underline{\lambda}$$

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

Subject:
Year. Month. Date. ( )

را پیدا می‌کنیم که واریانس فوق را min کند: (با ضرایب لاگرانژ 2θ)

$$Q(\lambda, \theta) = \lambda' \Sigma \lambda - 2\theta'(X'\lambda - t)$$

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

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

اگر طرفین را در X ضرب می‌کنیم

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

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

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

و در نتیجه از ۱ و ۲:

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

و یک برآوردگر خطی برای ترکیب خطی از β:

$$\lambda' Y = t' \underbrace{(X'\Sigma^{-1}X)^{-1} X' \Sigma^{-1} Y}_{\text{همان GLS برای } \beta \text{ است.}} = t'\hat{\beta}$$

مقایسه بین: ~~برآورد واریانس~~ (σ^2)

~~برآورد~~ $V(E(\hat{Y}))$ و $V(\hat{Y})$ ~~با استفاده از~~:

دیدیم مدل به صورت $Y = X\beta + \varepsilon$ به طوری که تحت فرضیات اساسی

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

از طرفی برآورد β:

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

(۲۷)

Subject:
Year. Month. Date. ( )

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

و : ~~دقت کنید~~

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

$Y = XB + \varepsilon$

$\downarrow$

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

چرا : $Cov(X\hat{B},\hat{\varepsilon}) = 0$: چون

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

$$= [I - X(X'X)^{-1}X']\underset{\sim}{y} = [I - P_x]\underset{\sim}{y}$$

که به راحتی می‌توان نشان داد P_x متقارن و خودتوان است همچنین $(I - P_x)$ هم خودتوان است.

از طرفی

~~$X'[I - X(X'X)^{-1}X'] = 0 \Rightarrow X'[I-P_x]\underset{\sim}{Y} = X'\hat{\varepsilon} = 0$~~

$$\begin{cases} X\hat{B} = \overbrace{X(X'X)^{-1}X'}^{A}\underset{\sim}{y} = A\underset{\sim}{Y} \\ \hat{\varepsilon} = \underbrace{[I - X(X'X)^{-1}X']}_{B}\underset{\sim}{Y} = B\underset{\sim}{Y} \end{cases} \rightarrow$$

چون Y نرمال $B\underset{\sim}{Y}$, $A\underset{\sim}{Y}$ مستقلند اگر

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

~~یا با استفاده از نشان داد : $Cov(X\hat{B}, Y - X\hat{B}) = 0$~~

پس از ② : و از قوانین هم می‌توان نتیجه گرفت

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

در مقایسه با ① :

$$\underbrace{var(\hat{y})}_{\text{پیش‌بینی فردی}} > \underbrace{var(\hat{E}(y))}_{\text{پیش‌بینی جمعی}}$$

(از تک تک مشاهدات) (میانگین)

Subject:
Year. Month. Date. ( )

برآورد σ^2 :

$\hat{\underset{\sim}{\varepsilon}} = (I - P_x)\underset{\sim}{Y}$

برگردیم به محاسبه برآورد واریانس ~~خطا~~ می دانیم

پس

$$\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}} = [\underset{\sim}{Y} - \underbrace{\hat{E}(\underset{\sim}{Y})}_{\substack{X\hat{\beta} \\ X(X'X)^{-1}X'Y}}]'[\underset{\sim}{Y} - \underbrace{\hat{E}(\underset{\sim}{Y})}_{X\hat{\beta}}]$$

$$= \underset{\sim}{Y}'[I - P_x][I - P_x]\underset{\sim}{Y} = \underset{\sim}{Y}'[I - P_x]\underset{\sim}{Y}$$

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

از طرفی می دانیم

$$E(\underset{\sim}{Y}'A\underset{\sim}{Y}) = tr(A\Sigma) + \underset{\sim}{\mu}'A\underset{\sim}{\mu}$$

بنابراین

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

همچنین می دانیم :

$$tr(I - P_x) = r(I - P_x) = n - \underbrace{(K+1)}_{r(X)}$$

$$(X_{n\times K+1}\underbrace{(X'X)^{-1}_{(K+1)(K+1)}}_{\text{ماتریس نامنفرد}}X'_{(K+1)\times n})_{n\times n} \Rightarrow r(XX') = r(X) = K+1$$

لذا به طور کلی داریم :

$$E(\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}}) = [n - r(X)]\sigma^2 \Rightarrow E\Big(\underbrace{\hat{\underset{\sim}{\varepsilon}}'\hat{\underset{\sim}{\varepsilon}}/_{n-r(X)}}_{\text{برآوردی نااریب برای } \sigma^2}\Big) = \sigma^2$$

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

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

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

Subject:
Year. Month. Date. -( )

$$Y'Y - Y'P_xY$$

اگر $(X'X)$ نادژه باشد ، $r(X) = k+1$

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

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

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

تمرین: نشان دهید اگر فرض نرمالیتی برقرار باشد:

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

پارامتر غیر مرکزی $\mu'A\mu = 0$

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

$(r(A), \frac{1}{2}\mu'A\mu)$ $\frac{1}{2\sigma^2}\beta'X'(I-P_x)X\beta = \cancel{\beta'X'X\beta - \beta'X'P_xX\beta - \beta'X'X(X'X)^{-1}X'X\beta} = 0$; SSE

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	$Y'Y$	$MSR/MSE = F$

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

می‌توان نشان داد SSR و SSE مستقلند:

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

② طبق قضیه مستقلند ← $AB = 0$

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

همچنین طبق قضیه توزیع فرم درجه دوم داریم:

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

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

خواهیم داشت:

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

و بنابراین از ①، ②، ③:

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

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

PAPCO

④ $SST = Y'Y$ $\quad Y'Y/\sigma^2 \sim \chi^{2\prime}(n, \beta'X'X\beta/2\sigma^2)$

Subject :

Year . Month . Date . ( )

زیر سیوله صفحه 6 خ

حل با انحراف از میانگین ~~(مدل مرکزی)~~

با در نظر گرفتن افراز زیر برای X $\longrightarrow$

$$X = [\underset{\sim}{1} \quad X_1] \longrightarrow \underset{\sim}{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1} \qquad X_1 = \begin{array}{c} \\ 1 \\ \vdots \\ n \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & \cdots & x_k \end{array} \\ \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix}_{n\times k} \\ \begin{array}{ccc} \bar{x}_{.1} & \cdots & \bar{x}_{.k} \end{array} \end{array}$$

از طرفی داریم :

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

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

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

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

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

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

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

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

طبق رابطه 48 صفحه 27 سرل :

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

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) :

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

۲۸

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

۳۳۵۴۲۷۱

Subject:
Year. Month. Date. ( )

که S را می‌توان به صورت زیر نوشت:

$$S = x_c' x_c \quad , \quad x_c = X_1 - \underset{\sim}{1}\, \bar{x}' = \underbrace{(I - \tfrac{1}{n} J)}_{H} X_1 = H X_1$$

و با جایگذاری x_c در $\hat{B}_1$:

$$X_1' y - n\bar{y}\bar{x} = x_c' y$$

و لذا

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

خواهد شد. و خواهیم داشت:

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

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

و نشان دهید:

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

$Cov(\bar{y}, \hat{b}) - \bar{x}' Cov(\hat{b}, \hat{b})$

ترکیب خطی از y

از طرفی داشتیم

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

با جایگذاری $X = [1 \ \ X_1]$ داریم: $[1 \ \ X_1]' y = \begin{bmatrix} n\bar{y} \\ X_1' y \end{bmatrix}$

$$= y'y - [\underbrace{\bar{y} - \hat{B}'\bar{x}}_{\hat{B}_0} \quad \hat{B}_1'] \begin{bmatrix} n\bar{y} \\ X_1' y \end{bmatrix}$$

$$= \underbrace{y'y - n\bar{y}^2}_{} - \hat{B}_1'(X_1' y - n\bar{y}\bar{x})$$

$y' y$ $\quad y'(I - \tfrac{1}{n} J) y = y'y - n\bar{y}$

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

$y_c' y_c = \cancel{SSM - n\bar{y}^2}$

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

$SSR_m = SSR - SSM$

$y' \boxed{X_c' X_c)^{-1} X_c'} y$

که در آن $= SSM$

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

۹۴۴ ۹۹ ۵۷۸۴۷

$[\ ][\ ] \quad y \, [\ \ ]$

Subject:
Year. Month. Date. ( )

$X = [1 \quad X_1]$ $\quad x_c = (I_n - \frac{1}{n} J_n) X_1 = H X_1$ $\quad$ خلاصه: $\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1' \bar{x} \\ \hat{\beta}_1 = (x_c' x_c)^{-1} x_c' y \\ \quad = (X_1' H X_1)^{-1} X_1' H y \end{cases}$

$SSM \Rightarrow ( )$ $\quad SST_m \Rightarrow SST - SSM$ $\quad SSR_m \Rightarrow SSR - SSM$

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

الگو نمی توان نشان داد:

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

$[\ ][\ ] = [\ ]$ $\quad \Rightarrow \frac{SSM}{\sigma^2} \sim \chi^{2'}(1, (1' X \beta)^2 / 2n\sigma^2)$

و همچنین داریم:

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

$$\frac{SSR_m}{\sigma^2} \sim \chi^{2'}(r(x_c' x_c), \beta_1' x_c' x_c \beta_1 / 2\sigma^2)$$
$$(\cancel{r}\ k, \beta_1' x_c' x_c \beta_1 / 2\sigma^2)$$

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

$r(x_c' x_c) = r(x_c')$

$r(A\Sigma) = r(I)_{k\times k}$

در نهایت جدول آنالیز واریانس را می توان به صورت زیر نوشت:

منبع	df	SS	MS
میانگین	1	$n\bar{y}^2 = SSM = y' n^{-1} 1 1' y$	$SSM = MSM$
رگرسیون	K	$\hat{\beta}_1' x_c' y = SSR_m$	$SSR_m / k = MSR_m$
Cfm → corrected for mean	~~n-k-1~~		
خطا	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) $F(M) = \frac{MSM}{MSE}$ $\quad F(R_m) = \frac{MSR_m}{MSE}$

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

Subject:
Year. Month. Date. ( )

② $F(R_m) = \dfrac{SSR_m/k}{SSE/n-k-1} \sim F'(k, n-k-1, \underset{\sim}{B}_1' x_c' x_c \underset{\sim}{B}_1 / 2\sigma^2)$

منبع	df	SS	MS
Regression (c f m)	k	$SSR_m = \hat{\underset{\sim}{B}}_1' x_c' \underset{\sim}{y}$	MSR_m
error	$n-k-1$	$SSE = \underset{\sim}{y}'\underset{\sim}{y} - n\bar{y}^2 - \hat{\underset{\sim}{B}}_1' x_c' \underset{\sim}{y}$	MSE
Total	$n-1$	$SST_m = \underbrace{\underset{\sim}{y}'\underset{\sim}{y}}_{SST} - \underbrace{n\bar{y}^2}_{SSM} = \underset{\sim}{y}_c' \underset{\sim}{y}_c = \underset{\sim}{y}'(I - \frac{1}{n}J)\underset{\sim}{y} = \underset{\sim}{y}' H \underset{\sim}{y}$	

$F(R_m) = \dfrac{MSR_m}{MSE} \sim F'(k, n-k-1, \underset{\sim}{b}' x_c' x_c \underset{\sim}{b} / 2\sigma^2)$

نکته: تحت فرض $H_0 : \underset{\sim}{B} = \underset{\sim}{0}$ ، $F(R) \geqslant F_{k+1, n-k-1} \Rightarrow RH_0$ (در جدول آنالیز واریانس مرکزی شود)

نکته ۲: تحت فرض $H_0: E(\bar{y}) = 0$ ، (در جدول مرکزی نکته ۱) ⟵ پارامتر غیر مرکزی صفر می‌شود

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

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

نکته ۳: تحت فرض $H_0: \underset{\sim}{B}_1 = \underset{\sim}{0}$:

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

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

مثال صفحه ۱۰۶ سیرل

نکته ۴: فاصله اطمینان برای هر ترکیب خطی از ضرایب رگرسیونی به صورت زیر می‌باشد

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

Subject:
Year. Month. Date. ( )

نکته: با داشتن برآورد ضرایب رگرسیونی و مقادیر متغیرها توضیحی می توان یک y جدید را پیش بینی کرد

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

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 : $\left(\frac{SSE}{\chi^2_{n-k-1,\alpha/2}} , \frac{SSE}{\chi^2_{n-k-1,1-\alpha/2}} \right)$ $\quad \frac{SSE}{\sigma^2} \sim \chi^2_{n-k-1}$ نکته ۴

بررسی فرضهای خطی :

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

$r(K'\underset{\sim}{\beta} \;\; \underset{\sim}{m}) = r(K'\underset{\sim}{\beta})$ $\quad K'_{s\times k+1}\, \underset{\sim}{\beta}_{k+1\times 1} \to (K'\beta)_{s\times 1}$

که K در حالت کلی یک ماتریس و در حالت ساده یک بردار است و برای سازگار بودن $K'\underset{\sim}{\beta} = \underset{\sim}{m}$

$K'_{s\times k+1}$ باید سطری رتبه کامل باشد. $r(\underset{s\times s}{K'K}) = r(K) = r(K') = S$

دیدیم : $\underset{\sim}{y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I)$ $\quad$ با به $(K'K)^{-1}$ وجود دارد ۱

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

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

اکنون باید توزیع فرم درجه دوم به صورت :

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

طبق قضیه $\quad Q/\sigma^2 \sim \chi^{2'}\left(S, \underbrace{(K'\hat{\underset{\sim}{\beta}} - \underset{\sim}{m})'(K'(X'X)^{-1}K)^{-1}(K'\hat{\underset{\sim}{\beta}} - \underset{\sim}{m})/2\sigma^2}_{\delta}\right)$

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

$A\Sigma \Rightarrow$ خودتوان

PAPCO

Subject:
Year. Month. Date. ( )

$$rank(K') = s \Rightarrow r(K'(X'X)^{-1}K) = s \longrightarrow K'(X'X)^{-1}K \rightarrow$$ ناویژه است و معکوس آن وجود دارد :نکته

: H_0 تحت فرض

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

چون σ^2 معمولاً نامعلوم است دنبال برآوردگرش می‌گردیم:

لازم است

$$\hat{\varepsilon} = y - \hat{E}(y) = y - X(X'X)^{-1}X'y = (I - P_x)y$$

$$SSE = \hat{\varepsilon}'\hat{\varepsilon} = y'(I - P_x)y \qquad MSE = \frac{SSE}{n-k-1}$$

$$rank(X) = rank(P_x) = k+1 \qquad rank(I - P_x) = n-k-1$$

: در نتیجه

$$(2)\quad \frac{SSE}{\sigma^2} \sim \chi^2_{n-k-1}$$ دیدیم $$\boxed{\hat{\sigma}^2 = \frac{SSE}{n-k-1}}$$

: 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 مستقلند

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

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

چون $X'[I - X(X'X)^{-1}X']$ و $[I - X(X'X)^{-1}X']X$ صفرند داریم $K'\hat{\beta} = m$

$$SSE = \underbrace{[y - XK(K'K)^{-1}m]'}_{X'}\underbrace{[I - X(X'X)^{-1}X']}_{A}\underbrace{[y - XK(K'K)^{-1}m]}_{X}$$

$$Q = (K'(X'X)^{-1}X'y - m)'(K'(X'X)^{-1}K)^{-1}(K'(X'X)^{-1}X'y - m)$$

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

Subject:

Year. Month. Date. ( )

$$Q = [\underbrace{y - XK(K'K)^{-1}m}_{x'}]' \underbrace{X(X'X)^{-1}K[K'(X'X)^{-1}K]^{-1}K'(X'X)^{-1}X'}_{B}$$

$$\underbrace{[y - XK(K'K)^{-1}m]}_{X} \qquad AB = 0$$

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

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

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

وقتی H_0 درست نباشد

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

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

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

$$LS'X'XSL'B^\circ + K\theta = LS'X'y - LS'X'mpm'$$

PAPCO