- Another way of testing for the apparent lack of fit is a formal F test which can be used provided we have more than 1 observation for @ least one value of ...

eg. $x_1 = 20 \Rightarrow y_1 = 86$
$x_2 = 40 \Rightarrow y_2 = 78$
$x_3 = 40 \Rightarrow y_3 = 84$
$x_4 = 20 \Rightarrow y_4 = 33$
$x_5 = 60 \Rightarrow y_5 = 64$

here for x=20 we have two observations + for for x=40 as well

-In this case we can divide SSE into 2 prts:
① Pure experimental error
② error due to lack of fit

i.e. SSE = SSPE + SSLF
df $n-2 = \Sigma(n-1) + n-2 - \Sigma(n-1)$

eg. in our example

$$\text{df for SSPE} = \sum_{i}^{k}(n_i-1) = \overset{=2-1}{1} + \overset{=2-1}{1} + 0 = 2$$

∴ since n=5, k = # of levels of x's
n_i = # of observations for each level of x

d.f. for SSE = $n-2 = 5-2 = 3$
· SSLF = $n-2-\Sigma(n_i-1) = 3-2 = 1$

where if there are n_i observations @ the i^{th} level of the independent variable ⇒

$\sum_{j=1}^{n_i}(y_j-\bar{y}_i)^2$ measures the pure experimental error and SSPE $= \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$

and hence SSLF = SSE − SSPE

eg. $x_1 = 20$ $y_1 = 86, 33$ ⇒ $\bar{y}_1 = 59.5$
$x_2 = 40$ $y_2 = 78, 84$ ⇒ $\bar{y}_2 = 81$
$x_3 = 60$ $y_3 = 64$ ⇒ $\bar{y}_3 = 64$

and hence $\sum_{i=1}^{n} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = (86-59.5)^2 + (33-59.5)^2 +$
$+ (78-81)^2 + (84-81)^2 + (64-64)^2 = 1422.5$

and SSE $= S_{yy} - \hat{\beta}_1 S_{xy} = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$

where $S_{yy} = \sum_i y_i^2 - \frac{(\sum_i y_i)^2}{n} = 25721 - \frac{(345)^2}{5} = 1916$

$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 7600 - \frac{(180)^2}{5} = 1120$

$S_{xy} = \sum xy - \frac{(\sum x_i)(\sum y_i)}{n} = 12700 - \frac{(180)(345)}{5} = 280$

$\Rightarrow S_{yy} - \frac{S_{xy}^2}{S_{xx}} = 1916 - \frac{(280)^2}{1120} = 1846$

∴ SSE $= \underline{\underline{1846}}$ y ⇒ SSLF = SSE − SSPE
$= 1846 - 1422.5 = \underline{\underline{423.5}}$

SSPE $= \underline{\underline{1422.5}}$

- To test for lack for of fit, i.e.
  H_0: a linear model is appropriate
  H_a: " " " is not "

- test stat: $F = \frac{MSSLF}{MSSPE} = \frac{SSLF/[n-2-\sum(n_i-1)]}{SSPE/[\sum(n_i-1)]}$

it follows $F(n-2-\Sigma(n_i-1), \Sigma(n_i-1)) = F(df, df_2)$

$\therefore$ We reject H_0 if $F > F(df, df_2)_\alpha$

---

eg in our example

$\alpha = .05$

$$F = \frac{SSLF/df_1}{SSPE/df_2} = \frac{423.5/1}{1422.5/2} = \frac{423.5}{711.25} = .5954$$

- We reject H_0 if $F > F(1,2)_{.05} = 18.51$

$\therefore$ do not reject H_0 and conclude that @ 5% level of significance no evidence to indicate that a linear model is not appropriate

Possible Remedies for Assumption violations

- non-normality + unequal variances:

  - transformation of the response variable y of the form: $\sqrt{y}$, $\ln y$ or $\frac{1}{y}$ may solve both problems (usually they occur together)

- non-linearity of Regression f^n (and independence of y_i's)

  - try adding an x_i^2 term to the model (usually they occur together)

Steps in a SLR Analysis

① Plot a scatter diagram as a preliminary check. If no obvious non-linearities, go to step 2

② State the SLR model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ along w/ the assumptions required for estimation + prediction.

③ Use sample data to find the least squares fitted line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

④ Perform a residual analysis to check for violations of assumptions. If satisfied go to step 5, otherwise try remedies stated above (ie. going to step 2 + starting over)

⑤ Test usefulness of model for prediction purposes:
$H_0: \beta_1 = 0$ vs $H_a: \beta_1 \neq 0$, also ~~calc~~ calc.
$$r^2 = \frac{SSR}{TSS}$$

⑥ If $H_0: \beta_1 = 0$ rejected in step 5, use model for prediction + estimation

Chap. 13 - Multiple Linear Regression

very often we can find that the response variable y is related to more than one predictor variable,

eg.

y = selling price of the house ← dependent variable
x_1 = house size
x_2 = age of house
x_3 = neighborhood
} → independent prediction variables measured w/out error

ect.

in this case our model would become

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

↑ so-called "first-order model"

note:

-this order of the model is determined by the highest exponent of the predictor variables

eg. $$\left.\begin{aligned} &y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \beta_3 x_3 + \varepsilon \\ &y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \cancel{\beta_4 x_2 x_3} + \varepsilon \end{aligned}\right\} \text{second-order models}$$

cross product term (or so-called interaction term)

The basic steps followed in an SLR model are also applicable to MLR

Assumptions: $y = \beta_0 + \beta_1 x_1 + \dots \beta_K x_K + \varepsilon$

① $\varepsilon \sim N(0, \sigma^2)$ for any given set of $x_1, \dots, x_2$

② $x_1 \dots x_2$ are observed w/out error

-need not be independent of each other

eg. x_1 = hrs of sleep

$x_2 = x_1^2$ ect.

③ ε_i's are independently distributed w/ mean zero and common variances σ^2

Note: assumptions 1 + 3 are equivalent to y_i's

- y_i's are indep + normally distributed w/ $E(y) = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K$ and com. variances for all x's

- Linear model means linear in parameters (β's)

i.e. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \varepsilon$ is linear in β's

but $y = \beta_0 + \beta_1 x_1 + \beta_2^2 x_2 + \varepsilon$ - not linear in β's

or $y = \beta_0 e^{\beta_1 x} + \varepsilon$ ← not linear in β's

* Tues Feb. 6 - Test 1 8:30am - Chap. 12 SLR

Chap. 13 MLR - no scatter plot

Review

MLR - $y = \beta_0 + \beta_1 x_1 + \dots \beta_K X_K + \mathcal{E}$

Assump.'s $y \sim N(E(y), \sigma^2) \forall X_1, X_2 \dots X_K$

$\mathcal{E}(y) = \beta_0 + \beta_1 X_1 + \dots \beta_K X_K$ const $\Leftrightarrow \mathcal{E} \sim N(0, \sigma^2) \forall X_1 \dots X_2$

Interpretation of Parameters in MLR (hold only for 1st order models when X_i's are independent of each other.)

β_0 = y-intercept - as an SLR, it represents the portion of $\mathcal{E}(y)$ when each $X = 0$

$SLR = \beta_0 + \beta_1 X + \mathcal{E}$

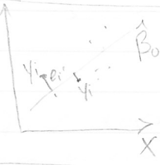
$\beta_1 \dots \beta_K$ - partial slopes

where $\beta_j (j = 1 \dots K)$ represents the chng in $\mathcal{E}(y)$ for a unit incrs of x_j, holding other X's constant.

Least square method

- similarly, as in SLR case, parameters $\beta_0, \beta_1 \dots \beta_K$ are unknown + hence need to be estimated by $\hat{\beta}_0, \hat{\beta}_1 \dots \hat{\beta}_K$

SLR



$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$= \sum [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$

$y = \beta_0 + \beta_1 x_1 + \dots \beta_K X_K + \mathcal{E}$; $\mathcal{E} \sim N(0, \sigma^2)$ / $\forall X_1, X_2 \dots X_K$

or $\mathcal{E}(y) = \beta_0 + \beta_1 x_1 + \dots \beta_K X_K$

- estimate $\mathcal{E}(y)$ by using $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_K X_K$

$SSE = \sum_{i=1}^{n} e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_K X_K)]^2$

$\rightarrow$ CONT

$$\frac{\partial SSE}{\partial \hat{\beta}_0} \overset{!}{=} 0 \quad \frac{\partial SSE}{\partial \hat{\beta}_1} \overset{!}{=} 0 \ \ldots \ \frac{\partial SSE}{\partial \hat{\beta}_K} = 0$$

- Solving $(K+1)$ normal equations in $(K+1)$ unknowns will give us solution for $\hat{\beta}_0, \hat{\beta}_1 \ldots \hat{\beta}_K$

- When working w/ matrices, situation is much easier

Bold are matrices

<u>MLR model</u>: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

and $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, where

vector of responses $\rightarrow$ $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times1}$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}_{K+1}$ $\leftarrow$ vector of parameters

vector parameter estimates $\rightarrow$ $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}_{(K+1)\times1}$, vector of error $\rightarrow$ $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times1}$

matrix of constants or design matrix $\rightarrow$ $\mathbf{X} = \begin{bmatrix} \overset{(\text{for } \beta_0)}{1} & \overset{(\beta_1)}{X_{11}} & \overset{(\beta_2)}{X_{12}} & \cdots & \overset{(\beta_K)}{X_{1K}} \\ 1 & X_{21} & X_{22} & \cdots & X_{2K} \\ - & - & - & - & - \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nK} \end{bmatrix}_{n\times(K+1)}$

$(K+1)$ normal equations can be written down as:

$$(\mathbf{X}^T\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y} \quad / \cdot (\mathbf{X}^T\mathbf{X})^{-1}$$

$$\boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}, \mathbf{X}^T\mathbf{Y}}$$

NOTE: $A + B = B + A$.
$A \cdot B \neq B \cdot A$

(Ex) Suppose we want to predict the selling prc. (Y) of a house based on its size (X_1) + # of bedrooms (X_2)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \Rightarrow \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\mathbf{Y} = \begin{bmatrix} 25 \\ 30 \\ 31 \end{bmatrix}_{3\times1}$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}_{3\times1}$, $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}_{3\times1}$

$$\mathbf{X} = \begin{array}{ccc} \text{prc} & \text{size} & \text{bed} \\ \hline 1 & 6 & 3 \\ 1 & 5 & 3 \\ 1 & 8 & 4 \end{array}$$

$\rightarrow$ cont'h

$\mathbf{X}^T\mathbf{X} \Rightarrow$ "transpose" = swap columns + rows col #1 $\quad$ col #1

$(1)(1)+(1)(1)+(1)(1)=3 \qquad (6)(1)+(5)(1)+(8)(1)=19 \qquad (3)(1)+(3)(1)+(4)(1)=10$

$(1)(6)+(1)(5)+(1)(8)=19 \qquad (6)(6)+(5)(5)+(8)(8)=125 \qquad (3)(6)+(3)(5)+(4)(8)=65$

$(1)(3)+(1)(8)+(1)(4)=15 \qquad (6)(3)+(5)(8)+(8)(4)=90 \qquad (3)(3)+(3)(8)+(4)(4)=49$

$$X^T=\begin{bmatrix}1&1&1\\6&5&8\\3&3&4\end{bmatrix}_{3\times3}$$

1st row 1st col. (add products)
1st row 2nd col.
1st row 3rd col.

$$X^TX=\begin{bmatrix}1&1&1\\6&5&8\\3&3&4\end{bmatrix}_{\textcircled{3}\times3}\cdot\begin{bmatrix}1&6&3\\1&5&8\\1&8&4\end{bmatrix}_{3\times\textcircled{3}}=\begin{bmatrix}3&19&10\\19&125&65\\10&65&34\end{bmatrix}_{3\times3}$$

size of result

$$\underbrace{A=\left\{\begin{array}{ccc|ccc}a_{11}&a_{12}&a_{13}&1&0&0\\a_{21}&a_{22}&a_{23}&0&1&0\\a_{31}&a_{32}&a_{33}&0&0&1\end{array}\right\}=\cdots=\left\{\begin{array}{ccc|}1&0&0\\0&1&0\\0&0&1\end{array}\right.}_{\text{Inverse}}\underbrace{\begin{array}{ccc}b_{11}&b_{12}&b_{13}\\b_{21}&b_{22}&b_{23}\\b_{31}&b_{32}&b_{33}\end{array}}_{B=A^{-1}}$$

$$(X^TX)^{-1}=\left\{\begin{matrix}25&4&-15\\4&2&-5\\-15&-5&14\end{matrix}\right\}_{3\times3}$$

$$X^TY=\begin{bmatrix}1&1&1\\6&5&8\\3&3&4\end{bmatrix}_{\textcircled{3}\times3}\cdot\begin{bmatrix}25\\30\\31\end{bmatrix}_{3\times\textcircled{1}}=\begin{bmatrix}86\\548\\289\end{bmatrix}_{3\times1}$$

$$\hat{\beta}=(X^TX)^{-1}\cdot X^TY=$$

$$=\begin{bmatrix}25&4&-15\\4&2&-5\\-15&-5&14\end{bmatrix}_{\textcircled{3}\times3}\cdot\begin{bmatrix}86\\548\\289\end{bmatrix}_{3\times\textcircled{1}}=\begin{bmatrix}7\\-5\\16\end{bmatrix}_{3\times1}=\hat{\beta}\text{ i.e. }\begin{aligned}\hat{\beta}_0&=7\\\hat{\beta}_1&=-5\\\hat{\beta}_2&=16\end{aligned}$$

least squares lines | regression line | best fitting line $\hat{y}=7-5x_1+16x_2$

$\longrightarrow$ CONT

If we want to predict the pr. of a new house w/ size $x_1 = 3$ + # of bed.'s $x_2 = 5$
$\Rightarrow \hat{y} = 7 - 5(3) + 16(5) = \underline{72}$

---

using derivatives not used

$$* \left\{ \begin{array}{l} \text{N.E.'s} \quad \sum y_i = m\hat{\beta}_0 + \hat{\beta}_1 \sum x_i + \dots + \hat{\beta}_K \sum x_{iK} \\ \sum x_i y_i = \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2 + \dots + \hat{\beta}_K \sum x_i x_K \\ \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \end{array} \right.$$

Residual Analysis

- after fitting the regression equation, we should check for any violations of the model and/or its assumptions
- as in SLR we plot $\hat{y}_i$ vs $e_i \rightarrow$ to check for violations of independence of y's

x_1 vs e_i, x_2 vs. e_i, ... x_K vs $e_i \Rightarrow$ to check for violations of constant variance of y's for every value of $x_1, x_2 \dots x_K$

- histogram of errors $\rightarrow$ to check for any violations of normality

Testing for the Significance of the model

ie testing for a linear relationship btwn y + @ least one of the x's

$y = \beta_0 + \beta_1 x_1 \dots + \beta_K x_K + \mathcal{E}$

must use F test because >1 parameter

$H_0: \beta_1 = \beta_2 \dots \beta_K = 0$ (no linear relationship) ; α
H_a: at least one $\beta_i \neq 0$ (linear)

Need to run anova

i.e. TSS = SSR + SSE

Computational formulas

$$TSS = \sum y^2 - \frac{(\sum y)^2}{n} = y^T y - \frac{(\sum y)^2}{n}$$

$$SSR = \hat{\beta}^T (X^T y) - \frac{(\sum y)^2}{n}$$

$$SSE = TSS - SSR = y^T y - \hat{\beta}^T (X^T y)$$

memorize

$df_{TSS} = n-1$

$df_{SSR} = K$ (# of parameters in the model associated w/ X's)

$df_{SSE} = n-(K+1)$ (n - total # of parameters in the model, including β_0)

$$MSR = \frac{SSR}{K}$$

$$MSE = \frac{SSE}{n-(K+1)} \Rightarrow F = \frac{MSR}{MSE}$$

$H_0 = \beta_1 = \beta_2 = \ldots = \beta_K = 0$ (no linear relationship) ; α

$H_a = \beta$'s $\neq 0$ (linear relationship)

because more than 1 parameter ↑

$F = \frac{MSR}{MSE}$ RR - Rej H_0 if $F > F_{\alpha}(\underset{MSR}{\overset{\text{par. ass. w/ X's}}{K}}, \underset{MSE}{\overset{\text{tot. par.}}{n-(K+1)}})$

TSS = SSR + SSE

Conclusion: linear rel. betwn Y + @ least one X

ANOVA Table

Source of variation	df	SS	MS	F
Regression	K	SSR	$MSR = \frac{SSR}{K}$	F
error	n-(K+1)	SSE	$MSE = \frac{SSE}{n-(K+1)}$	
Tot	n-1			

$TSS = \Sigma y_i^2 \frac{(\Sigma y_i)^2}{n} = y^T y - \frac{(\Sigma y_i)^2}{n}$

$SSR = \hat{\beta}^T (x^T y) - \frac{(\Sigma y_i)^2}{n}$

$SSE = TSS - SSR$

SAS in SAS

```
proc Reg;
 model y = X1 X2 X3 ... Xk / XPX  I;
run;
```

$\underbrace{XPX}$ $x^T x$, $x^T y$, $y^T y$

$\overset{\uparrow}{I}$ $(x^T x)^{-1}$

Inferences Concerning a set of β's

-Consider model

$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$

$\xleftarrow{}$ -complete or full model

-suppose we suspect that the "real" model should be the one w/ less variables (ie we suspect that some of the β's should be = 0)

e.g. Full model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$
we suspect the variables x_2 + x_3 do not contribute to the model

i.e $H_0: \beta_2 = \beta_3 = 0$; α
H_a: @ least one of β's ≠ 0

-if we don't reject H_0 then we can remove X_2 + X_3 from the model and obtain new model $y = \beta_0 + \beta_1 x_1 + \varepsilon$ → "reduced model"

- in general

<u>full model</u>: $y = \beta_0 + \beta_1 x_1 + \dots \beta_g x_g + \beta_{g+1} + x_{g+1} + \dots \beta_K x_K + \varepsilon$

<u>reduced model</u>: $y = \beta_0 + \beta_1 x_1 + \dots \beta_g x_g + \varepsilon \; ; \; g < K$

$H_0 : \beta_{g+1} = \beta_{g+2} = \dots = \beta_K = 0$

H_a: @ least of the β's $\neq 0$

<u>Full</u>: K x's, (K+1) par.s

<u>Red</u>: g x's, (g+1) par.s

- we need to obtain 2 ANOVA tables ⇒

SSR_F^{full} and $SSR_R^{reduced}$

SSE_F^{full} $\quad$ $SSE_R^{reduced}$

<u>Test statistic</u> $F_{drop} = \dfrac{MSR_{drop}}{MSE_{full}} = \dfrac{[SSE_R - SSE_F] / [df_{SSE_R} - df_{SSE_F}]}{SSE_F / df_{SSE_F}}$

$$= \frac{[SSE_R - SSE_F] / [n-(g+1) - (n-(K+1))]}{SSE_F / n-(K+1)} = \frac{(SSE_R - SSE_F / (K-g)}{SSE_F / n-(K+1)}$$

(# of β's equal 0)

<u>R.R.</u> We reject H_0 if $F_{drop} > F_{\alpha ; (K-g), n-(K+1))}$

(same as F_{drop} - can use either

or equivolantley

$$F_{part} = \frac{[SSR_F - SSR_R] / [df_{SSR_F} - df_{SSR_R}]}{MSE_F}$$

$$= \frac{[SSR_F - SSR_R] / (K-g)}{SSE_F / n-(K+1)}$$

<u>R.R.</u> - Reject H_0 if $F_{part} > F_{\alpha}(K-g, n-(K+1))$

models { Fi $\quad TSS = \overset{\uparrow}{\boxed{SSR_F}} + SSE_F \downarrow$

Ri $\quad TSS = SSR_R \downarrow + SSE_R \uparrow$

<u>Rem</u> - Always working w/ SSR or SSE, not both

Ⓔⓧ. Chemist

y = weight loss

x_1 = time

x_2 = humidity of air

3rd order, interaction

Full: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \beta_4 x_1 x_2 + \beta_5 x_1 x_2^2 + \varepsilon$

Reduced: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^x + \varepsilon$

i.e. $H_0: \beta_4 = \beta_5 = 0$; α

H_a: @ least 1 of β's $\neq 0$

always > SSR_R

from SAS: $SSR_f = 32.04217$ df = 5

$SSE_f = 0.42700$ df = 6 [n-(k+1)]

and

$SSR_R = 31.72583$ df = 3

always > SSE_f ← $SSE_R = 0.74333$ df = 8 [12-4]

test stat: $F_{drop} = \dfrac{[SSE_R - SSE_f]/[df_{SSE_R} - df_{SSE_f}]}{MSE_f}$

$= \dfrac{(0.74333 - .42700)/(8-6)}{SSE_f/df_{SSE_f}} = \dfrac{.31633/2}{.42700/6} = \dfrac{.158165}{.0711666666}$

$= \underline{2.222459037}$

RR - we rej. H_0 if $F_{drop} > F_\alpha(2;6)$

or equivantley

$F_{part} = \dfrac{[SSR_f - SSR_R]/[df_{SSR_f} - df_{SSR_R}]}{MSE_f}$

$= \dfrac{[32.04217 - 31.72583]/(5-3)}{SSE_f/df_{SSE_f}} = \dfrac{(\quad .31634/2}{0.42700/6} =$

$= \dfrac{.158165}{.0711666666} = 2.22245903 7$

R.R - Rej. $F_{part} > F_{\alpha}(\overset{\text{= \# of par.'s to be removed}}{2}, 6)$

Inferences about a single parameter

(i.e.) testing contribution of 1 x)

Remember $(X^TX)^{-1} = V$

$$\begin{array}{ccccc} v_{00} & v_{01} & v_{02} & \dots & v_{0K} \\ v_{10} & v_{11} & v_{12} & \dots & v_{1K} \\ \vdots & & & & \vdots \\ v_{K0} & v_{K1} & v_{K2} & \dots & v_{KK} \\ \downarrow & \downarrow & \downarrow & & \downarrow \\ \hat{\beta}_0 & \hat{\beta}_1 & \beta_2 & & \beta_K \end{array}$$

$\Rightarrow var.(\hat{\beta}_j) = \overset{\text{the diagonal}}{v_{jj}} \cdot \sigma^2$ - σ^2 unknown ⇒ we use $s^2 = mse \Rightarrow \widehat{var\,\beta_j} = v_{jj}\,mse$

v_{jj} → $(j+1)^{th}$ diagonal element of $(X^TX)^{-1}$

use t when 1 parameter being tested

Ⓘ t test

$H_0: \beta_j = 0$; α $\Rightarrow t = \frac{\hat{\beta}_j - \overset{=0}{\beta_j}}{\sqrt{var(\hat{\beta}_j)}} = \frac{\hat{\beta}_j}{\sqrt{v_{jj} \cdot mse}}$

$H_a: \beta_j \neq 0$

R.R. - Reject H_0 if $t > t_{\alpha/2; n-(K+1)}$ or $t < -t_{\alpha/2; n-(K+1)}$

Ⓘⓘ F part (or F drop)

full model: $y = \beta_0 + \beta_1 x_1 + \dots + \beta_{j-1} x_{j-1} + \beta_{j+1} x_{j+1} + \dots + \beta_K x_K + \varepsilon$

red: $y = \beta_0 + \beta_1 x_1 + \dots + \beta_{j-1} x_{j-1} + \beta_{j+1} x_{j+1} + \dots \beta_K x_K + \varepsilon$

test-stat $F_{part} = \frac{[SSR_F - SSR_R] / [df_{SSR_F} - df_{SSR_R}]}{mse_F}$

$= \frac{[SSR_F - SSR_R] / (K - (K-1))}{SSE_F / df_{SSE_F}} =$

$= \frac{[SSR_F - SSR_R] / 1}{SSE_F / n-(K+1)}$ or equivalently,

F drop

R.R. We rej. H_0 if $F_{part} > F_\alpha(1, n-(k+1))$

$$F_{drop} = \frac{[SSE_R - SSE_F]/[df_{SSE_R} - df_{SSE_F}]}{mSE_F}$$

$$= \frac{[SSE_R - SSE_F]/[n-k-(n-(k+1))]}{SSE_F / df_{SSE_F}} =$$

$$\frac{[SSE_R - SSE_F]/1}{SSE_F / n-(k+1)} =$$

R.R. Rej. H_0 if $F_{drop} > F_\alpha(1, n-(k+1))$

NOTE: $t^2_{(n-(k+1))} = \frac{\hat{\beta}_j^2}{v_{jj} \cdot mSE} = F_{part}(1, n-(k+1)) = F_{drop}(1, n-(k+1))$

if $H_0: \beta_j = 0$ vs $H_a: \beta_j \neq 0$

Full:

reduced:

$H_0: \beta_{g+1} = \dots = \beta_k = 0$

H_a: @ least one of β's $\neq 0$; α

test stat: $F_{drop} = \frac{[SSE_R - SSE_F]/[df_{SSR} - df_{SSE_F}]}{mSE_F}$

or equivalently

$F_{part}\ \frac{[SSR_F - SSR_R]/[\underbrace{df_{SSR_F} - df_{SSR_R}}_{k-g}]}{mSE_F}$

R.R. We reject H_0 if F_{drop} (or F_{part}) $> F_\alpha(k-g; n-(k+1))$

one t

$H_0: \beta_j = 0$

$H_a: \beta_j \neq 0$; α

$var(\hat{\beta}_j) = v_{jj}\sigma^2 \Rightarrow \hat{var}(\hat{\beta}_j) = \underset{\downarrow}{v_{jj}} mSE$

$(j+1)^{th}$ diagonal elemnt of $(X^TX)^{-1}$

t-test: $t = \frac{\hat{\beta}_j}{\sqrt{\hat{var}(\hat{\beta}_j)}} = \frac{\hat{\beta}_j}{\sqrt{v_{jj} mSE}}$