

Ch. 12-2 sta 3164

①

Multiple Regression with K independent variables
model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K + \epsilon$$

There are many software programs that provide calculations to obtain least squares estimates. The output of such programs typically has a list of variables with the estimates of $\beta_0, \beta_1, \beta_2, \dots, \beta_K$ and other information along with the analysis of regression model.

ANOVA has

SS_{Total}

SSR (sum of squares of ^{regression} ~~regression~~)

SSE (sum of squares of errors or residuals)

$$SS_{Tot} = \sum (Y_i - \bar{Y})^2$$

$$SSE = \sum (Y_i - \hat{Y}_i)^2$$

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2$$

$$SS_{Tot} = SSR + SSE$$

ANOVA table looks like this

(2)

Source	df	SS	MS	F	Prob > F
Regression	K	SSR	MSR	$\frac{MSR}{MSE}$	
Error	$n - k - 1$	SSE	MSE		X
Total	$n - 1$	SS _{tot}			

All calculations are done by the software programs and our purpose is to look for information in the output for the estimates of β coefficients, to write down the prediction equation, estimate σ^2 , confidence interval estimates of β 's, Test of hypothesis of β 's etc.

For this purpose we will discuss the output for example 12.5 on page 637 and other such examples of outputs in this chapter.

1) Est of $\sigma^2 =$ MSE from ANOVA table

2) Estimates of $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots$ in output along with $SE\hat{\beta}_0, SE\hat{\beta}_1, SE\hat{\beta}_2$ etc

for discrete
$$t = \frac{\hat{\beta}_i - \beta_i}{SE\hat{\beta}_i} \quad \text{with } n-k-1 \text{ df}$$

CI for β_i : $\hat{\beta}_i \pm t_{\alpha/2} SE\hat{\beta}_i, \text{ df} = n-k-1$
similarly t test for β_i

Sec 12.4

Test for $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$
 H_a : at least one $\beta_i \neq 0$

p 647

T.S. $F = \frac{MSR}{MSE} \quad (k, n-k-1)$
in ANOVA table.

We need Two ANOVA's here

Test for (a subset of $(k-g)$ β 's) $= 0$
 $H_0: \beta_{g+1} = \beta_{g+2} = \dots = \beta_k = 0$
T.S. $F = \frac{SSR(\text{complete}) - SSR(\text{Reduced})}{(k-g)} \div \frac{SSR(\text{complete})}{n-k-1}$

Sec 12.5

p 653

df num = $k-g$
df den = $n-k-1$

R^2 = Coefficient of determination

$$R^2 = \frac{SST_{\text{tot}} - SSE}{SST_{\text{tot}}} = \frac{SSR}{SST_{\text{tot}}}$$

measure

→ proportion of variation in y
due to its relationship with
explanatory variables

or

proportion of variation in y
explained by these independent
variables.

— This is what we commonly use
to see if the model is a good
fit.

(5)

Estimation of β 's & Prediction equation

for ex 12.4 PG 37

$k=2$

model is $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + E$

In the output locate the estimates

$$\hat{\beta}_0 = 0.667; \hat{\beta}_1 = 1.317; \hat{\beta}_2 = -8.000$$

\Rightarrow Prediction equation

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

$$\hat{Y} = 0.667 + 1.317 X_1 - 8.000 X_2$$

The prediction for Y when $X_1 = 6.5$, $X_2 = 0.35$

$$\hat{Y} = 0.667 + 1.317(6.5) - 8.000(0.35)$$

$$\hat{Y} = 6.428$$

Estimate of $\sigma^2 = \text{MSE} = 0.14944$

from ANOVA table

6

Conf. Int for β_i 's p. 650

$$\hat{\beta}_i \pm t_{\alpha/2} s_{\hat{\beta}_i}, \text{ df} = n - k - 1$$

$$s_{\hat{\beta}_i} = \text{SE of } \hat{\beta}_i$$

Ex. 12.13

95% Conf Int for β_3 : for data in

Ex. 12.12

For data in Ex 12.12 p. 649, $\text{df} = 17$
locate $\hat{\beta}_3 = .26528$, $s_{\hat{\beta}_3} = \text{SE } \hat{\beta}_3 = 0.10127$

$$1 - \alpha = .95 \quad \frac{\alpha}{2} = .025 \quad t_{.025}(\text{df} = 17) = 2.110$$

$$\text{CI} = .26528 \pm 2.110(0.10127) = (.05160 \leq \beta_3 \leq .478)$$

Similarly other Conf. Int.

P. 651 Test of hyp. about β_i 's (individual β 's)
(one sided or two sided tests)

- 1) $H_0: \beta_i = 0$ 2) $H_0: \beta_i = 0$ 3) $H_0: \beta_i = 0$
- $H_a: \beta_i > 0$ $H_a: \beta_i < 0$ $H_a: \beta_i \neq 0$

$$\downarrow \text{Test stat. } t = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}, \text{ df} = n - k - 1 \downarrow$$

$$\text{RR: } t > t_{\alpha} \quad t < -t_{\alpha} \quad |t| > t_{\alpha/2}$$

Ex 12.15 ✓ ^{p652} Ref to output 12.14 (p651) ^{Ex}

Here $n=54, k=4 \rightarrow 12.14$ $p=647$

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$ Two sided

$\alpha = .10$

$\hat{\beta}_1 = .01291$

$\hat{\sigma}_{\hat{\beta}_1} = .00283$

T.S. $t = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}}, df = 54 - 4 - 1 = 49$

RR: $|t| > t_{\frac{\alpha}{2}} \Rightarrow |t| > t_{.05} = 1.677$

Calculated $t = \frac{.01291}{.00283} = 4.562$

Conc: $|t| = |4.562| > 1.677 \rightarrow$ Reject H_0

[When we reject H_0 it implies that ~~that~~ Variable X_1 provides additional predictive power in the presence of the other variables in the equation)