Answer Key Midterm I : Fall 04

Question 1

(i) $\sum x_i = 5$, $\sum y_i = 3$, $\sum x_i^2 = 39$, $n = 6$

$\sum y_i^2 = 49$, $\sum x_i y_i = 30$, $\bar{x} = 0.833$

$\bar{y} = 0.5$

(ii) $$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{30 - (6)(0.833)(0.5)}{39 - 6(0.833)^2} = \frac{27.501}{34.837}$$

$\Rightarrow \hat{\beta}_1 = 0.7894$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 0.5 - (0.7894)(0.833)$

$= -0.1575$

(iii) $SST = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2$

$= 49 - 6(0.5)^2 = 47.5$

$SSE = \hat{\beta}_1^2 \sum (x_i - \bar{x})^2$

$= (0.7894)^2 (34.837) = 21.709$

$$R^2 = \frac{SSE}{SST} = \frac{21.709}{47.5} = 0.457$$

(iv) $$\hat{\sigma}^2 = \frac{RSS}{n-2} = \frac{(47.5 - 21.709)}{4}$$

$$= \frac{25.791}{4} = 6.4478$$

$$\widehat{Var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum(x_i - \bar{x})^2} = \frac{6.4478}{34.837} = 0.1851$$

$$se(\hat{\beta}_1) = \sqrt{0.1851} = 0.4302$$

$H_0: \beta_1 = 1$
$H_1: \beta_1 < 1$

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{0.7894 - 1}{0.4302} = \frac{-0.2106}{0.4302}$$

$$= -0.4895$$

From the list of critical values, we obtain $c_{5\%}$, at 4 d.f. $= 2.077$.

Since $t > -c$ $\Rightarrow$ fail to reject the null at the 5% level of significance

(V) 95% CI for β_1

$$= \left[\hat{\beta}_1 \pm c_{\alpha/2} \cdot se(\hat{\beta}_1)\right] \quad , \alpha = 5\%$$

$c_{2.5\%}$ at α d.f $= 2.49$

$$\Rightarrow \left[0.7894 \pm (2.49)(0.4302)\right]$$

$$= \left[0.7894 \pm 1.0712\right] = \left[-0.2818, 1.8606\right]$$

Question 2

$$\ln \hat{y} = \underset{(0.113)}{8.71} + \underset{(0.005)}{0.14S} + \underset{(0.009)}{0.023N} \quad , R^2 = 0.37$$

(i) Holding N constant, 1 additional yr of S $\Rightarrow \hat{y} \uparrow$ by 14%

Holding S constant, 1 additional yr of N $\Rightarrow \hat{y} \uparrow$ by 2.3%

(ii) $H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$ $\quad \alpha = 5\%$

$$t_{\hat{\beta}_1} = \frac{0.14}{0.005} = 28 \quad ; \quad C_{2.5\%} \text{ at } 40 \text{ d.f.} = 2.329$$

Since $|t| > c \Rightarrow$ reject Ho at ~~the~~ 5% level of significance.

(iii) 95% CI for β_1

$$= \left[\hat{\beta}_1 \pm c_{2.5\%} \cdot se(\hat{\beta}_1)\right]$$

$$= \left[0.14 \pm (2.329)(0.005)\right] = \left[0.14 \pm 0.0116\right]$$

$$= \left[0.1284,\ 0.1516\right]$$

Under the null, $\beta_1 = 0$ is not in this CI
$\Rightarrow$ reject H_0 against $H_1: \beta_1 \neq 0$ at the 5% level of significance.

This ~~is the~~ same as our conclusion in (ii).

(iv) $H_0: \beta_2 = 0$, $H_1: \beta_2 > 0$

$$\Rightarrow t_{\hat{\beta}_2} = \frac{0.023}{0.009} = 2.55$$

$$\Rightarrow \text{p-value} = P(t > 2.55) \text{ at } \infty \text{ df}$$
$$= 0.00735$$

(v) $$\ln \hat{y} = 8.71 + 0.14S + 0.023N \quad (A)$$

$$\ln \hat{y} = \hat{\alpha}_0 + \hat{\alpha}_1 S = 8.98 + 0.19S \quad (B)$$

$$\hat{N} = \hat{\delta}_0 + \hat{\delta} S \quad (C)$$

The result in (iv) implies that $\hat{\beta}_2$ is statistically different from zero at the 5% level of significance.

Hence, in estimating model B instead of model A, your friend has excluded a relevant variable from his model. That is, model B is underspecified.
Consequence of excluding the relevant variable N from model B :-
- $\hat{\alpha}_1$ is a biased estimate of β_1, in particular, the bias in $\hat{\alpha}_1$ is positive ($\hat{\beta}_2 > 0$ and $\delta > 0$).

In essence, the estimate of the slope coefficient in model B, $\hat{\alpha}_1$, overestimates the true effect of years of schooling on earnings.

$$\hat{\alpha}_1 = \hat{\beta}_1 + \hat{\beta}_2 \hat{\delta}$$

$$\Rightarrow \hat{\delta} = \frac{\hat{\alpha}_1 - \hat{\beta}_1}{\hat{\beta}_2} = \frac{0.19 - 0.14}{0.023} = 2.1739$$

Result identical (i.e. $\hat{\alpha}_1 = \hat{\beta}_1$) if :-
(i) N and S are uncorrelated $\Rightarrow \hat{\delta} = 0$
(ii) $\beta_2 = 0$, that is N does not appear in the true model, (A).

(vi) $\ln Y = \beta_0 + \beta_1 S + \beta_2 N + u$

$H_0: 2\beta_1 + 4\beta_2 = 0$;

Set $\theta = 2\beta_1 + 4\beta_2$ $\Rightarrow H_0: \theta = 0$

$H_1: \theta \neq 0$

$\beta_1 = \frac{1}{2}\theta - 2\beta_2$ or $H_1: \theta > 0$

$$\Rightarrow \ln Y = \beta_0 + \theta(\tfrac{1}{2}S) + \beta_2(N - 2S) + u$$

$$\Rightarrow \ln Y = \beta_0 + \theta Z_1 + \beta_2 Z_2 + u$$

Regress $\ln y$ on a constant, z_1 and z_2, and test whether θ is statistically different from zero.

$\Rightarrow$ reject null if $|t_{\hat{\theta}}| > c_{2.5\%}$ at 40 d.f.

(vii) $c = F(2,40)$ at $5\% = 3.23$

$H_0: \beta_1 = \beta_2 = 0$

H_A: not H_0

restricted model : $\ln y = \beta_0 + u$

$$F = \frac{R^2/q}{(1-R^2)/n-k-1} = \frac{0.37}{0.63} \times \frac{40}{2} = 11.746$$

$\Rightarrow$ reject H_0 at $\alpha = 5\%$ because $F > C$.

(viii) $H_0: \beta_1 = \beta_2 = 0$ $\quad c = F(2,96)$ at 5%

H_1: not H_0

unrestricted model : $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$

restricted model : $y = \beta_0 + \beta_3 X_3 + u$

run both models in excel to obtain SSR or R^2.

Then $$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/n-k-1}$$

reject H_0 at $\alpha = 5\%$ if $F > C$.