MC Answers

1. D
2. B
3. A
4. C
5. A
6. D
7. C
8. D
9. D
10. A
11. B
12. D

FINAL 2014 - SA

S.1) $\bar{X} = \dfrac{3+4+6}{3} = 4.\bar{3}$

$\bar{Y} = \dfrac{2+3+4}{3} = 3$

$$\hat{\beta}_1 = \frac{(3-4.\bar{3})(2-3)+(4-4.\bar{3})(3-3)+(6-4.\bar{3})(4-3)}{(3-4.\bar{3})^2+(4-4.\bar{3})^2+(6-4.\bar{3})^2}$$

$$= \frac{1.\bar{3}+1.\bar{6}}{1.\bar{7}+0.\bar{1}+2.\bar{7}} = \frac{3}{4.\bar{6}} = 0.64$$

S.2) $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 3 - 0.64(4.\bar{3}) = 0.21$

$\hat{u}_1 = 2 - (0.21 + 0.64(3)) = -0.14$
$\hat{u}_2 = 3 - (0.21 + 0.64(4)) = 0.21$
$\hat{u}_3 = 4 - (0.21 + 0.64(6)) = -0.07$

So, $SSR = -0.14^2 + 0.21^2 + -0.07^2 = 0.06$
and, $TSS = (2-3)^2 + (3-3)^2 + (4-3)^2 = 2$

So, $R^2 = 1 - \dfrac{0.06}{2} \approx 0.97$

5.3) $\frac{\Delta \widehat{wage}}{\Delta \, exper} = 1.52$

So, wage is predicted to increase by 3.04 when exper increases by 2.

5.4) Situation (ii) is likely worse. In this case, the estimator for $s.e.(\hat{\beta}_1)$ is biased and inconsistent. Inference may be wrong.

In situation (i), the estimator is inefficient, but still unbiased and consistent.

Since the loss of unbiasedness and consistency is usually considered worse than the loss of efficiency, situation (ii) is worse.

5.5) A possible consequence is that $\hat{\beta}_1$ is biased. Omitted variable bias will occur if the following two conditions are met:

① X_2 causes Y
② X_2 and X_1 are correlated

5.6) $H_0: \beta_1 = 0$ | Reject at too large sig.: -2
$H_A: \beta_1 \neq 0$

$t = \frac{-2.28}{0.52} = -4.39$

According to the standard normal $dist^n$, the p-val associated with this t-stat is 0.00001. We reject.

5.7) Including the same variable twice in the regression means that there is perfect multicollinearity, and that the OLS estimates can't be calculated. OLS is neither unbiased or biased - the estimators are not defined.

5.8) The R^2 of a regression will always decrease when variables are dropped from the regression. Hence, the observed decrease in R^2 is not a basis for model selection.

5.9) When we include irrelevant regressors, we use up sample information that could otherwise be used elsewhere: OLS will be inefficient, but still unbiased and consistent.

When we exclude relevant regressors, our model can suffer from omitted variable bias. (OLS will also be inconsistent). This situation is worse than the former.

5.10) Population model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$

Null hypothesis: $H_0: \beta_2 = 0$

Restricted model: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

Long answer

a) It does not. The intercept would be the earnings of an individual with age = 0. This is clearly outside the range of the data, and doesn't make sense theoretically.

b) The estimated effect is 0.44.

The 95% confidence interval is:

$$0.44 \pm 1.96(0.03) = 0.38,\ 0.50$$

c) When an estimate changes significantly from model to model, there is evidence of O.V.B.

We can test whether or not the change from (1) to (2) is significant:

$H_0: \beta_1 = 6.86$
$H_A: \beta_1 \neq 6.86$

$$t = \frac{6.49 - 6.86}{0.18} = -2.05$$

We reject the null at 5% significance (we suspect model (1) is suffering from O.V.B.).

d) In both of the models in which age^2 is included, the estimated coefficient is statistically significant at the 1% level. We reject the null of linearity.

e) $\widehat{earnings} = -21.80 + 6.86(0) - 3.15(0)$

$+ 2.05(26) - 0.03(26^2) = 11.22$

f) We can test to see if the coefficient on female × bach is statistically different from zero, in model (4) or (5):

$$t = \frac{-0.31}{0.36} = -0.89.$$

We fail to reject the null at the 10% sig. level.

g) For women: $0.53 - 0.22 = 0.31$

For men: 0.53

H_0: the coefficient on female × age $= 0$
H_A: not H_0

Given that the estimated coefficient is significant at the 1% level, we reject the null that there is no difference in the effect of age on earnings between men and women.

h) $R^2_u = 0.1898$ $\quad q = 2,\ n = 7985,$
$R^2_R = 0.1364$ $\quad K_u = 3$

$$F = \frac{(0.1898 - 0.1364)/2}{(1-0.1898)/(7985-3-1)} = 263.01$$

i) R^2 always increases when a regressor is added. The $\bar{R}^2$ decreased because the fit of the model did not improve enough to overcome the penalty induced by adding a regressor.