Practice Questions – Chapter 9
AS.180.334.SP19

Question 1

The R-squared from estimating the model

$$\log(salary) = \beta 0 + \beta 1*\log(sales) + \beta 2*\log(mktval) + \beta 3*profmarg + \beta 4*ceoten + \beta 5*comten + u,$$

is $R^2 = 0.353$ ($n = 177$). When $ceoten^2$ and $comten^2$ are added, $R^2 = 0.375$. Is there evidence of functional form misspecification in this model?

Question 2

Let *math10* denote the percentage of students at a Michigan high school receiving a passing score on a standardized math test. We are interested in estimating the effect of per student spending on math performance. A simple model is

$$math10 = \beta 0 + \beta 1*\log(expend) + \beta 2*\log(enroll) + \beta 3*poverty + u,$$

where *poverty* is the percentage of students living in poverty.

(i) The variable *lnchprg* is the percentage of students eligible for the federally funded school lunch program. Why is this a sensible proxy variable for poverty?

(ii) The table that follows contains OLS estimates, with and without *lnchprg* as an explanatory variable.

Dependent Variable: *math10*		
Independent Variables	**(1)**	**(2)**
log(*expend*)	11.13 (3.30)	7.75 (3.04)
log(*enroll*)	.022 (.615)	−1.26 (.58)
lnchprg	—	−.324 (.036)
intercept	−69.24 (26.72)	−23.14 (24.99)
Observations *R*-squared	428 .0297	428 .1893

© Cengage Learning, 2013

Explain why the effect of expenditures on *math10* is lower in column (2) than in column (1). Is the effect in column (2) still statistically greater than zero?

(iii) Does it appear that pass rates are lower at larger schools, other factors being equal? Explain.

(iv) Interpret the coefficient on *lnchprg* in column (2).

(v) What do you make of the substantial increase in R^2 from column (1) to column (2)?

Question 3

The following equation explains weekly hours of television viewing by a child in terms of the child's age, mother's education, father's education, and number of siblings:

$$tvhours^* = \beta 0 + \beta 1 * age + \beta 2 * age^2 + \beta 3 * motheduc + \beta 4 * fatheduc + \beta 5 * sibs + u.$$

We are worried that *tvhours** is measured with error in our survey. Let *tvhours* denote the reported hours of television viewing per week.

(i) What do the classical errors-in-variables (CEV) assumptions require in this application?

(ii) Do you think the CEV assumptions are likely to hold? Explain.