Practice MLR

1. In a study relating college grade point average to time spent in various activities, you distribute a survey to several students. The students are asked how many hours they spend each week in four activities: studying, sleeping, working, and leisure. Any activity is put into one of the four categories, so that for each student, the sum of hours in the four activities must be 168. In the model

$$GPA = \beta_0 + \beta_1 study + \beta_2 sleep + \beta_3 work + \beta_4 leisure + u,$$

does it make sense to hold *sleep*, *work*, and *leisure* fixed, while changing study?

Answer: No, since *study*, *sleep*, *work*, and *leisure* are perfectly collinear. Moreover, *study* + *sleep* + *work* + *leisure* = 168. We cannot change *study*, without changing at least one of the other categories.

2. Which of the following can cause the OLS estimator, i.e., $\hat{\beta}$, to be biased?
(i) Heteroskedasticity.
(ii) Omitting an important variable.
(iii) A sample correlation coefficient of .95 between two independent variables both included in the model.

Answer: An omitted variable that affects the dependent variable and is correlated with the included explanatory variables can cause bias. The homoskedasticity assumption, plays no role in unbiasedness of the OLS estimators. Further, high (but not perfect) collinearity between the explanatory variables in the sample does not affect the assumptions needed for unbiasedness. However, it does inflate the corresponding standard errors. Thus, our answer is only (ii).

3. The following equation describes the median housing price in a community in terms of amount of pollution (*nox* for nitrous oxide) and the average number of rooms in houses in the community (*rooms*):

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 rooms + u.$$

(i) What is the interpretation of β_1? Explain in terms of percentage changes in *nox* and *price*.

Answer: β_1 is the elasticity of *price* with respect to *nox*. Thus, a 1% increase in *nox* is associated with a β_1% change in *price*.

(ii) Does the simple regression of log(*price*) on log(*nox*) produce an unbiased estimator of β_1? Explain in terms of omitted variables bias.

Answer: This is unlikely due to omitted variables bias. For example, quality of houses could be such an unobserved characteristic. Better quality houses may have more rooms and be located in neighborhoods with less pollution.

4. Use the data in HPRICE1 to estimate the model

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u,$$

where *price* is the price of a house in thousands of dollars; *sqrft* represents the size of a house in square feet; *bdrms* denotes the number of bedrooms.

(i) Write out the results in an equation form. While it is sufficient to report the $\hat{\beta}$ estimates, you may also paste the Stata results.

Answer: Estimated equation:

$$\widehat{price} = -19.32 + 0.128sqrft + 15.198bdrms.$$

```
. reg price sqrft bdrms

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(2, 85)        =     72.96
       Model |  580009.152         2  290004.576   Prob > F        =    0.0000
    Residual |  337845.354        85  3974.65122   R-squared       =    0.6319
-------------+----------------------------------   Adj R-squared   =    0.6233
       Total |  917854.506        87  10550.0518   Root MSE        =    63.045

------------------------------------------------------------------------------
       price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       sqrft |   .1284362   .0138245     9.29   0.000     .1009495    .1559229
       bdrms |   15.19819   9.483517     1.60   0.113    -3.657582    34.05396
       _cons |    -19.315   31.04662    -0.62   0.536    -81.04399      42.414
------------------------------------------------------------------------------
```

(ii) What is the estimated increase in price for a house with one more bedroom, holding square footage constant?

Answer: Holding square footage constant, price increases by 15.198, i.e., $15,198.

(iii) What percentage of the variation in price is explained by square footage and number of bedrooms?

Answer: About 63.2%.

(iv) The first house in the sample has *sqrft* = 2,438 and *bdrms* = 4. Find the predicted selling price for this house from the OLS regression line.

Answer: The predicted price is –19.315 + 0.128(2,438) + 15.198(4) = 353.541, or $353,541.

(v) The actual selling price of the first house in the sample was $300,000 (so *price* = 300). Find the residual for this house.

Answer: From part (iv), the estimated value of the home based only on square footage and number of bedrooms is $353,541. The actual selling price was $300,000 and the residual is -53541.