Assignment 2

Due: Sunday, June 15 by 11:59 pm

(Please submit single pdf document)

Note: Question 5 is optional and ungraded. Work on it only after completing the first four questions.

1. Consider two least-squares regressions

$$y = X_1\tilde{\beta}_1 + \bar{e}$$

and

$$y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{e}$$

Let R_1^2 and R_2^2 be the R-squared from the two regressions, respectively. Show that $R_2^2 \geq R_1^2$.

2. Use the cps09mar data for this question. The data set and the description file is attached under the assignment on Canvas.

   Estimate a log wage regression for the subsample of white male Hispanics. In addition to education, experience, and its square, include a set of binary variables for regions and marital status. For regions, you create dummy variables for Northeast, South and West so that Midwest is the excluded group. For marital status, create variables for married, windowed or divorced, and separated, so that single (never married) is the excluded group.

3. The data Koop_Tobias_subsample is extracted from Koop and Tobias's (2004) study of the relationship between wages and education, ability, and family characteristics. Let X_1equal a constant, education, experience, and ability. Let X_2 contain the mother's education, the father's education, and the number of siblings. Let y be the log of the hourly wage. [Show your regression output]
   a. Compute the least squares regression coefficients in the regression of y on X_1. Report the coefficients.
   b. Compute the least squares regression coefficients in the regression of y on X_1 and X_2. Report the coefficients.
   c. Regress each of the three variables in X_2 on all the variables in X_1 and compute the residuals from each regression. Arrange these new variables in the (15 x 3) matrix X_2^*. What are the sample means of these three variables?
   d. Compute the R^2 for the regression of y on X_1 and X_2. Repeat the computation for the case in which the constant term is omitted from X_1. What happens to R^2?
   e. Compute the adjusted R^2 for the full regression including the constant term.

4. Suppose that you have two independent unbiased estimators of the same parameter θ, say $\hat{\theta}_1$ and $\hat{\theta}_2$, with different variances v_1 and v_2. What linear combination $\hat{\theta} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2$ is the minimum variance unbiased estimator of θ?

Assignment 2
Due: Sunday, June 15 by 11:59 pm
(Please submit single pdf document)

Note: Question 5 is optional and ungraded. Work on it only after completing the first four questions.

5. Analyze the properties of the Least Squares (LS) estimator of the slope coefficient in a simple linear regression model using the Monte Carlo method. Specifically, the true model is given by

$$y_i = 1 + 2x_i + \varepsilon_i$$

where the intercept is 1, and the slope coefficient is 2. The errors ε_i are normally distributed with mean zero and variance 3, $\varepsilon \sim N(0,3)$. The independent variable x is uniformly distributed between 0 and 6, $x \sim U[0,6]$.

Conduct the following experiment: Draw 1000 samples of sizes 25, 50, and 100 from the distributions of the error term and the predictor variable. For each sample, estimate the slope coefficient using the LS estimator. Then, analyze the

a. Unbiasedness, variance, and the shape of the distribution of the LS estimator of the slope coefficient by plotting the distribution for each sample size.
b. Compare the variance of the estimated slope coefficient obtained from the Monte Carlo simulations with the variance obtained using the asymptotic formula and the variance obtained using a bootstrap method.