

Econometrics
Dr. Todd Richard Yarbrough

Problem Set #1

Completed problem sets will include an R log file, attached to the back of your typed answers. Be sure to include any graphs or figures you're asked to generate.

1. Download the "AgeHourlyEarnings" data from the google drive folder in the "StockWatson" data folder. This data contains the joint distribution of age (*Age*) and average hour earnings (*AHE*) for 25 to 34 year old full time workers in 2012 with an education level that exceeds a high school diploma. Use this joint distribution to answer the following questions.

- a. Generate a histogram and compute the probability density function for the distribution of *Age*.
 - b. What is the mean of *AHE* for each value of *Age*? That is, what is the conditional mean of *AHE* given each age level?
 - c. Compute and create a scatter plot of the mean of *AHE* versus *Age*. Are average hourly earnings and age seemingly related? Explain.
 - d. Using the Law of Iterated Expectations, compute the mean of *AHE*.
 - e. What is the Variance of *AHE*?
 - f. What is the Covariance of *AHE* and *Age*.
 - g. What is the Correlation between *AHE* and *Age*?
 - h. Comment on your answers to parts f and g with respect to the plot you created in part c.
-

2. Download the "CPS9212" data from the google drive folder in the "StockWatson" data folder. The data contains information on full-time workers, ages 25-34, with a high school diploma or B.A./B.S. as their highest degree. Use these data to answer the following questions.

- a. Compute the sample mean for average hourly earnings (*AHE*) in 1992 and 2012.
 - b. Compute the sample standard deviation for *AHE* in 1992 and 2012.
 - c. Construct a 95% confidence interval for the population means of *AHE* in 1992 and 2012.
 - d. Construct a 95% confidence interval for the change in population mean of *AHE* between 1992 and 2012.
 - e. Construct 95% confidence intervals for the mean of *AHE* for high school graduates, the mean of workers with a college degree, and the difference between the two means. Do this just for 2012.
-

3. Suppose that a researcher, using data on class size and average test scores from 100 third-grade classes, estimates the OLS regression:

$$\hat{y} = 520.4 - 5.82x, R^2 = 0.08, u = 11.5$$

where \hat{y} is estimated test score, and x is class size.

- a. A classroom has 22 students. What is the regression's prediction for that classroom's average test score?
- b. Last year a classroom had 19 students, and this year it has 23 students. What is the regression's prediction for the change in the classroom average test score?
- c. The sample average class size across the 100 class rooms is 21.4. what is the sample average of the test scores across the 100 classrooms?
- d. What is the sample standard deviation of test scores across the 100 classrooms?