Assignment 2

AAE 636: Applied Econometric Analysis I

Due: Beginning of lecture on Tuesday, October 1, 2019

Please show your work on all problems. Please also be considerate to the TA by writing neatly and having well-organized code. Include all relevant Stata or R output; do not include irrelevant output. You are encouraged to work together but must write up your problem sets individually.

The dataset for this assignment is called WAGE 2.DTA (wage2.RData) and is available on the textbook website.[1] The data are from a 1992 QJE paper by M. Blackburn and D. Neumark. "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials." This assignment is intended to get you started with Stata or R. The next assignment will use the same dataset and include more questions that require the use of Stata or R.

The data set includes observations for workers. The variables we will use for this assignment are the following:

Variable	Description
EDUC	Years of education
AGE	Age
EXPER	Years of work experience
TENURE	Years of employment with the current employer (job tenure)
WAGE	Monthly wages
MARRIED	An indicator for marital status (1 = married, 0 = single)

1. Confidence intervals

   Using the years of education variable, create the following four categories:

   i.) high school dropout (i.e. less than 12 years of eduction)

   ii.) high school graduate (exactly 12 years)

   iii.) some college (13-15 years)

   iv.) college graduate (16+ years)

   a. BY HAND (i.e. show all steps involved to create), calculate a 95% confidence interval for the average monthly wages for **high school graduates** (exactly 12 years of education)

[1] https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

b. Explain in words what the result you found in a. tells us about the true expected value of average monthly wages for **high school graduates**

c. BY HAND, calculate a 99% confidence interval for the average monthly wages for **college graduates**

2. Hypothesis Testing

   BY HAND, test the following hypothesis at the 5% level of significance. Be SURE to state your decision rule for each test

   a. The mean monthly wages of **college graduates** are $1,100.

   b. The mean monthly wages of **married** individuals with **some college** are $1000.

3. Calculating p-values

   a. BY HAND, determine the probability that you would observe the sample mean monthly wages of **high school graduates** that you find in the data if the true mean monthly wages of **high school graduates** is $900.

   b. BY HAND, determine the probability that you would observe the sample mean monthly wages of **unmarried** individuals that you find in the data if the true mean monthly wages of **unmarried** individuals is $750.

4. Relationship between wages and experience for high school graduates. **Restrict your analysis for this question to high school graduates**

   a. Plot wages against experience (put experience on the horizontal axis and wages on the vertical axes). Is the correlation between wages and experience positive or negative? Does the correlation appear to be strong or weak? Explain.

   b. What are the sample means and standard deviations of wages and of experience?

   c. What is the covariance between wages and experience? What is the correlation between wages and experience?

   d. What is the covariance between tenure and experience? The correlation?

   e. We are interested in estimating the regression model

$$Wage = \beta_0 + \beta_1 Exper + u$$

   Using the results you found in B. and C., calculate BY HAND the OLS regression coefficients for the intercept and slope terms in a regression of wage on experience. (i.e., compute these values using the formulas and the above results, not using the output of the `regress` command in Stata (`lm` command in *R*))

f. What does the estimated coefficient, $\hat{\beta}_1$, tell you about the relationship between wages and experience?

g. Given what you found in D. about the relationship between *tenure* and *exper*, do you expect your estimates to be causal? Be explicit in your reasoning.

h. Using the Stata output generated by estimating the above regression model, what is the estimated variance of the error term, $\hat{\sigma}^2$.

i. BY HAND, compute the estimates of the variances of $\hat{\beta}_0$ and $\hat{\beta}_1$ by inserting the estimate $\hat{\sigma}_2$ from Part F. into the appropriate formulas. (Hint: The other values needed for these formulas can be generated using `summarize` in Stata and `summary` in *R*.)

j. BY HAND, create a 95% confidence interval for the slope parameter (i.e., use the appropriate confidence interval formula along with the values you have computed above).

5. Relationship between wages and experience for college graduates. **Restrict your analysis for this question to college graduates**.

a. What are the OLS estimates of the intercept and slope terms for the regression of wages on experience?

b. Test the null hypothesis that the intercept parameter is zero at $\alpha = 0.10$ level of confidence. Be sure to state your decision rule.

c. Test the null hypothesis that the slope parameter is zero at $\alpha = 0.01$ level of confidence. Be sure to state your decision rule.

d. What are the SST, SSE, and SSR?

e. Calculate the r^2 for this regression using your results from d. Show your work, including what formula you used (do not just take the r^2 from the regression output). Does your result seem low or high to you? Why do you think you found such an r^2?

f. Use the `summarize` command in Stata with the detail option (`summary` command in *R*) to determine the values of experience at the 25^{th}, 50^{th}, and 75^{th} percentiles of the experience distribution (among college graduates only). What are your best guesses of the expected value of wages for college graduates whose years of experience are at the three values you just found?

6. Regression Analysis (By Hand) In a random sample of people on a street corner, four people were asked their age and annual earnings.

Subject #	Age	Annual Earnings
1	40	60
2	35	45
3	25	40
4	50	55

Annual earnings in thousands

From this sample, estimate the effect that age has on annual earnings. So we wish to estimate the model

$$Earnings = \beta_0 + \beta_1 Age + u$$

BY HAND (i.e. do not just show the final answer calculated using a statistics package such as Stata or R),

a. Calculate the OLS estimator for β_1.

b. Calculate the OLS estimator for β_0.

c. Calculate the r^2.

d. Calculate the F statistic.

e. Calculate the OLS estimator for σ^2.

f. Find the standard error for $\hat{\beta}_1$.

g. Verify that the averages of actual and predicted earnings are the same.

h. Verify that the sum of the residuals equals zero.

i. Verify that the residuals are uncorrelated with the regressor, Age.

j. Verify that the residuals are uncorrelated with predicted earnings.

7. Wooldridge 2.2: In the simple regression model, $Earnings = \beta_0 + \beta_1 Age + u$, suppose that $\mathbb{E}(u) \neq 0$. Letting $\alpha_0 = \mathbb{E}(u)$, show that the model can always be rewritten with the same slope, but a new intercept and error, where the new error has zero expected value.

8. Wooldridge 2.7: Consider the relationship between savings, sav, and income, inc

$$save = \beta_0 + \beta_1 inc + u, \qquad u = \sqrt{inc} \cdot e$$

where e is a random variable with $\mathbb{E}(e) = 0$ and $Var(e) = \sigma_e^2$. Assume that e is independent of inc.

i. Show that $\mathbb{E}(u|inc) = 0$, so that the key zero conditional mean assumption (Assumption SLR.4) is satisfied. [Hint: If e is independent of inc, then $\mathbb{E}(e|inc) = \mathbb{E}(e)$].

ii. Show that $Var(u|inc) = \sigma_e^2$, so that the homoskedasticity assumption (Assumption SLR.5) is violated. In particular, the variance of sav increases with inc. [Hint: $Var(e|inc) = Var(e)$, if e and inc are independent].

iii. Provide a discussion that supports the assumption that the variance of savings increases with family income.