

Econ 310 Final Exam, **Part 3**: Spring 2020

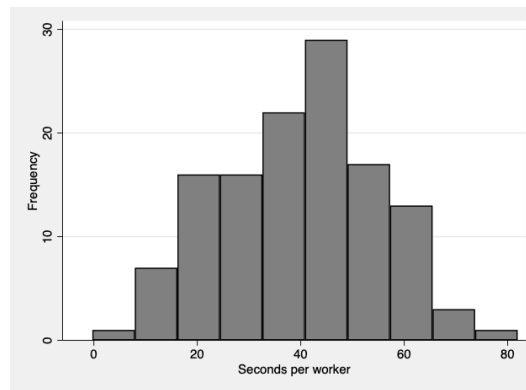
4 May 2020

8. (10 points) The S&P 500 is a stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the United States. The average return on the stocks in the S&P 500 over the past 90 years has been 9.8% and the standard deviation has been about 15%.
- (a) Suppose that you take a simple random sample of 60 companies from the S&P 500. You calculate an average return of $\bar{X}$ for this sample. What is the distribution of $\bar{X}$?
 - (b) Continuing from part (a), for a simple random sample of 60 companies, what is the probability that the average return in this sample will be lower than 10.2%?
9. (20 points) You are a consultant hired by a factory that is considering adopting a new robot that can replace some of the current human workers. It is known that the robot can complete the assignment in 40 seconds and that this duration is constant and does not vary. Over one week, you collect data about the time that it takes the human workers to complete the same assignment. With 125 observations, you find that the mean time it takes the human workers to complete the assignment is actually 40 seconds as well. The population variance is 16^2 seconds.
- (a) You suspect the true mean amount of time that it takes the workers to complete the assignment is in fact 42 seconds. What are the null and alternative hypotheses you should be testing? Describe what a Type I error would mean in this case. Test the hypothesis that you've put forward and calculate the associated p-value.

Now assume that the variance wasn't known after all, but the sample standard deviation is 15 seconds.

- (b) The factory manager looks at your findings and says that the sample is not representative and that in fact it takes workers on average more than 40 seconds to complete the assignment. How do you respond? Support your claim with the full hypothesis test, including the hypothesis you are testing, the test statistic, and your conclusion. You can assume a significance level of 5%.

- (c) The factory manager also claims that, even if the mean is actually 40 seconds, if the variance of the time that it takes the workers to complete the assignment is greater than 17^2 seconds, investing in the robot might be worth the sunk cost. Can you conclude that the variance is greater than 17^2 seconds at the 1% significance level? Support your claim with the full hypothesis test, including the hypothesis you are testing, the test statistic, and your conclusion.
- (d) After you've answered the manager's claims, you forgot that you have not actually looked at a histogram of the sample. You draw the following histogram.



Do your answers change in light of this histogram?

10. (20 points) You run the following regression on 189 observations collected using a random sample to examine which factors affect the weight of babies at birth:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

where

- y is the birth weight in grams
- x_1 is number of times the mother visited the physician during the 1st trimester
- x_2 is mother's weight at last menstrual period in kilograms
- x_3 is the age of the mother in years

You obtain the following results:

Variable	Estimate b_i	Standard error s_{b_i}
x_1	15.660	51.103
x_2	4.124	1.757
x_3	7.356	10.282
b_0	2225.513	301.536

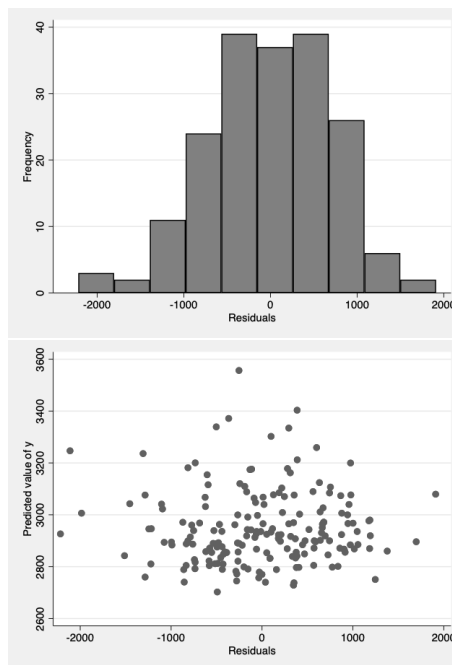
- (a) Interpret the estimate $b_1 = 15.660$.
- (b) Calculate and interpret a 95% confidence interval for β_3 .

In addition, you also obtain the following results from the estimation of the model:

$$SSR = 3822396$$

$$SSE = 96092902$$

- (c) Calculate and interpret the value of the coefficient of determination.
- (d) Test the validity of the model at a significance level of 5%. What is the null hypothesis and the alternative? What is the test statistic? What is your conclusion?
- (e) You obtain the two following graphs from the estimation results: A histogram of the residuals $\hat{\varepsilon}$, and a scatter plot of the residuals $\hat{\varepsilon}$ and the predicted values $\hat{y}$.



What do these two graphs tell you about the necessary conditions that must hold in order for the regression analysis to be valid?