

# ECON 3400 - Introduction to Economic and Business Statistics

## Assignment V

Due : 06/28/2020

### *Instructions*

- ❖ Use the data at the following link to answer all the questions:  
<https://www.dropbox.com/sh/tgqt34225ochmzr/AAAXwKo3lqB22i29Z7Z-Pmtga?dl=0>
- ❖ Type in your answers to questions, then submit your assignment in a pdf.
- ❖ Submit just one assignment for your study group with all the names of your team members.
- ❖ Google Docs and Sheets will not be accepted.
- ❖ Please read all questions carefully and make sure you understand them before you begin answering.
- ❖ In answering any question, you should not feel bound by anything you have said in answering an earlier question. Be sure to explain your answers thoroughly and show your work.
- ❖ Write legibly and be as concise and precise as possible.
- ❖ You must show all your work. You will not receive full credit for a correct answer if there is no work shown.

## Questions

**Use A5\_CovidInfection.dta data to answer Question 1 and Question 2.**

The data shows Covid-19 infection rates and average income in each Zip codes in New York City. Covid-19 cases data has been obtained from the Health Department and the income data is from the latest American Community Survey (ACS 5-year Estimates) which is prepared by the U.S. Census Bureau. ACSs show information on ancestry, citizenship, educational attainment, income, language proficiency, migration, disability, employment, and housing characteristics of people in the U.S.

**Q1 [60 points]** Answer the following questions.

- a) Produce detailed summary statistics for the Covid-19 infection rate and average income variables in in Stata for Midwood whose primary Zip Codes are 11210 and 11230 and share them in your Word/pdf file. Comment on the results. The code for the summary statistics of Covid-19 infection rate is given below:

**sum Covid\_Infectionrate if Zipcode ==11210 | Zipcode==11230, detail**

- b) Produce detailed summary statistics for the Covid-19 infection rate and average income variables in in Stata for the entire city and share them in your Word/pdf file. Comment on the results.
- c) Create histograms and box plots for both the Covid-19 infection rates and average incomes in in the entire city in Stata and share them in your Word/pdf file. Comment on each graph by using the summary statistics.
- d) For the Covid-19 infection rates in the entire city draw bell-shaped curves and show where
- 50% of data values are located.
  - 68% of data values are located.
  - 95% of data values are located.

In total 3 bell-shaped normal distribution curves along with the upper and lower boundary data values.

- e) As the results in (a) and (b) show, the average Covid-19 infection rate in Midwood neighborhood is 2.58%. The average infection rate for the entire city is 2.28% and the variance is 0.76. Using the Z-score calculate:
- What percent of the Zip Codes has a lower Covid-19 infection rate? (Lower than the average infection rate in Midwood)
  - What percent of the Zip Codes has a higher Covid-19 infection rate? (higher than the average infection rate in Midwood)
  - How many Zip Codes (the actual number, not the percent) have a lower infection rate? (round to the nearest whole number)?
  - How many Zip Codes (the actual number, not the percent) have a higher infection rate? (round to the nearest whole number)?
  - Using a bell-shaped normal distribution curve show the area below and above the infection rate in Midwood.

**Q2 [50 points]** The summary statistics in Q1 shows that the average Covid-19 infection rate in NYC is 2.28% with a standard deviation of 0.87. Suppose that you want to show that income is a major predictor of coronavirus infections. If you choose a sample of Zip Codes where the average income is above \$100,000, the following summary statistics can be obtained:

```
. sum Covid_Infectionrate if Average_Income >100000, detail
```

Covid_InfectionRate				
	Percentiles	Smallest		
1%	.5406	.5406		
5%	.70465	.6755		
10%	.83155	.6941	Obs	60
25%	.96185	.7152	Sum of Wgt.	60
50%	1.3276		Mean	1.597028
		Largest	Std. Dev.	.7819922
75%	2.1109	3.0537		
90%	2.80955	3.3229	Variance	.6115117
95%	3.1883	3.4585	Skewness	.9374583
99%	3.737	3.737	Kurtosis	3.004468

As the above table show, the average infection rate in neighborhoods where the average income is above \$100,000 is 1.59%.

- If you ran a test to determine if higher income levels does reduce the infection rate, what is the null hypothesis?
- Using an alpha level or p-critical of 0.05, use statistics to show if there is a statistically significant difference in the average infection rate in general population and in wealthy neighborhoods in NYC. Show all of your work to prove this and indicate if you would reject your null hypothesis and why or why not.
- Write a paragraph to explain your results to a non-technical audience.
- Write a paragraph to explain your results in a scholarly journal.

**Q3 [30 points]** Your local take-out restaurant claims that their food is delivered in 25 minutes. You decide to test their claim and order food from them 61 times over the next 3 months. Suppose that at the end of the three-month-period, you realize that on average, the food is delivered in 30 minutes with a standard deviation of 8 minutes.

- What is the null hypothesis?
- Would you reject or not reject the null hypothesis? Show your work to support your decision.
- Construct a 95 percent confidence interval for the true value of the delivery time.
- Draw a bell-shaped curved and show where the 95 percent confidence interval for the true value of the delivery time is located.

**Use the National Survey on Drug Use and Health from 2018 (A5\_NSDUH\_2018.dta) to answer Question 4.**

**Q4 [80 points]** The legal drinking age in the United States is 21 years. Many people, however, try alcohol before they turn 21 years old. Use the National Survey on Drug Use and Health from 2018 (A5\_NSDUH\_2018.dta) to test whether the age when Americans first try alcohol (**alctry**) is 21 years. You can find here detailed explanations of each variables. Before you run the test, use the commands **tab alctry** and share your results on your Word file. Notice that there are five categories related to missing data—985 bad data, 991 never used alcohol, 994 don't know, 997 refused, or 998 blank. Next, run the command **sum alctry** and share your results on your Word file. You will notice that the mean age for first trying alcohol is 301, which doesn't make sense. You should also notice that the maximum value is 998. To remove these missing data from your test, include the command **if alctry < 88** at the end of your summary statistics and tabulation commands since 87 was the oldest age reported and share them in your word file.

- a) What is your null hypothesis?
- b) Would you reject or not reject your null hypothesis? Explain your decision using your output. Interpret the results. Share your Stata output in your Word/pdf file.
- c) Explain the 95 percent confidence interval in your output. Draw a bell-shaped curve and show the confidence interval and explain how you test your null-hypothesis.
- d) Calculate the t-test value by using the formula for t-test and the results you have obtained from Stata.
- e) Explain the 95 percent confidence interval for t-table values. Draw a bell-shaped curve and show how you test your null-hypothesis using the t-table values and t-test value you have calculated in (d).
- f) On the basis of your results, write a few sentences that would explain your results to a non-technical audience. Then, write a few sentences to present your results in a scholarly journal.