

# LMTH 2014 – Quantitative Reasoning II: Research Methods and Data Visualization Assignment V

Due : 05/02/2021

## Instructions

- ❖ Use the data at the following link to answer all the questions:  
<https://www.dropbox.com/sh/b7vpm3vgizmrse3/AADpJHNDEGhdK4g4iCwoGKFUa?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.**

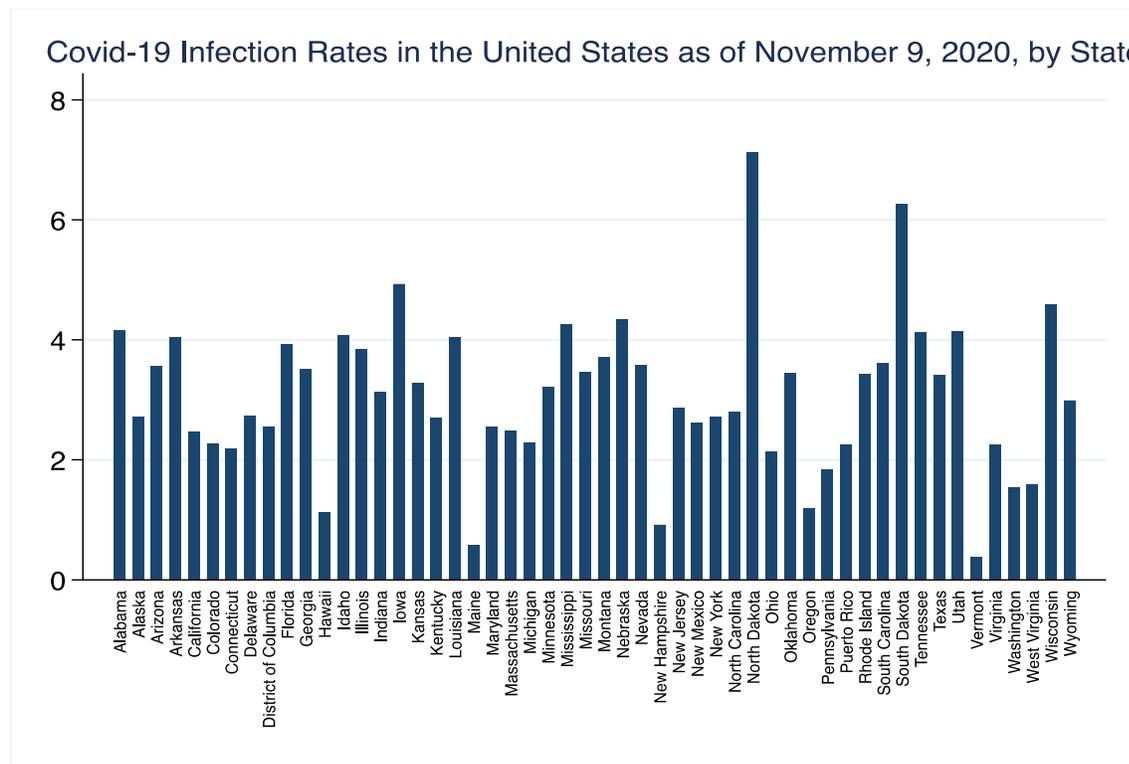
The data shows Covid-19 infection rates and average income in each Zip codes in New York City. Covid-19 cases data has been obtained from the Health Department and the income data is from the latest American Community Survey (ACS 5-year Estimates) which is prepared by the U.S. Census Bureau. ACSs show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

**Q1 [70 points]** The summary statistics below shows that the average Covid-19 infection rate in NYC is 2.93%. Suppose that you want to

show that the policies employed in NYC led to a significant decrease in Covid-19 infection rates compared to those in other states presented by the graph below. We know that the average Covid-19 infection rate in the entire country is 3.07 but don't know the variance.

```
. sum Covid_Infectionrate, d
```

Covid_InfectionRate				
Percentiles	Smallest			
1%	1.18792	.75894		
5%	1.39849	1.18792		
10%	1.58584	1.20688	Obs	177
25%	2.16962	1.32891	Sum of Wgt.	177
50%	2.9851		Mean	2.934688
			Std. Dev.	.9707436
75%	3.6339	4.93264		
90%	4.13196	4.94757	Variance	.9423431
95%	4.48611	5.05906	Skewness	-.0103007
99%	5.05906	5.13408	Kurtosis	2.186829



- What is your research question and null hypothesis?
- 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). Explain how your results support your conclusions in (b)
- f) On the basis of your results, write a few sentences to present your results in a scholarly journal.

**Use A5\_US\_Unemployment.dta data to answer Question 2.**

**Q2 [80 points]** Suppose that you believe that the average unemployment rates among men and women in the US is the same. Thus, you want to test this by using A5\_US\_Unemployment.dta data which presents the data on unemployment rates by gender in the US between 1990-2019. The data has been obtained by OECD data platform at the following link: <https://stats.oecd.org/index.aspx?queryid=54743#>

- a) In Stata generate a table that shows the mean, the standard deviation, and the sample size of the unemployment rate among men and women. Format the table so that there are two digits to the right of the decimal point. You should share this table on your word file along with the codes necessary to generate it. Comment on your findings in the table.
- b) Suppose that you want to determine if there is a significant difference in the yearly average unemployment rates between men and women since 1990. What would be your null hypothesis? What would be your null hypothesis about the variance?
- c) Can you reject both null hypotheses? Use statistics to support your conclusions and explain each results.
- d) Use the information obtained from Stata to calculate a 95 percent confidence interval of the mean difference and show it on a bell-shaped curve. Do your results verify the Stata outputs (the confidence interval values on the forth row called **diff**). How can you evaluate your null hypothesis by using these confidence interval values.
- e) Using the t-test formula and Stata results calculate the t-statistics and check if it is equal to the one you have obtained from Stata.
- f) Draw a bell-shaped curve for the t-table values and t-statistic and evaluate your null-hypotehis.

- g) Calculate the Cohen's  $d$  both in Stata and manually to examine the effect size. Comment on the results.
- h) Write the results of your findings for a non-technical audience.
- i) Write the results of your findings for a journal article.