

ECON 3400 – Introduction to Economic and Business Statistics

Assignment III

Due : 10/27/2020

Instructions

- ❖ Use the data at the following Excel File (A3) to answer all the questions:
<https://www.dropbox.com/sh/5rszjlpokp4erh6/AAA7slkaLpaNh4U1khk7EOs3a?dl=0>
- ❖ Type in your answers to questions, and submit your assignment in a pdf and your Excel File.
- ❖ 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

Q1[20 points]

Use the worksheet **Covid_Cases_by Age_Race** in A3 Excel File to answer this question.

The data on the worksheet is a perfect representative sample of the Health Department dataset that contains data on the number of cases of the Coronavirus Disease 2019 (COVID-19) by different **age** groups in New York City. The data has been retrieved from <https://www1.nyc.gov/site/doh/covid/covid-19-data.page>.

Using the **Patient Age Group** variable:

- a) Prepare a frequency table and share it in your Word/pdf file.
- b) Prepare a relative frequency table and share it in your Word/pdf file.
- c) Prepare Pareto Charts for both Frequency Table and Relative Frequency Tables share it in your Word/pdf file.
- d) Comment on the tables and explain your results.
- e) Do your tables and charts suggest a disproportionate burden of the illness among age groups? Explain.

Q2[20 points]

Use the worksheet **Covid_Cases_by Age_Race** in A3 Excel File to answer this question.

The data on the worksheet is a perfect representative sample of the Health Department dataset that contains data on the number of cases of the Coronavirus Disease 2019 (COVID-19) by different **race** groups in New York City. Data on people identified as other categories, including Native American/Alaska Native or multi-racial, are not provided here. The data has been retrieved from <https://www1.nyc.gov/site/doh/covid/covid-19-data.page>.

Using the **Patient Race Group** variable:

- a) Prepare a frequency table and share it in your Word/pdf file.
- b) Prepare a relative frequency table and share it in your Word/pdf file.
- c) Prepare Pareto Charts for both Frequency Table and Relative Frequency Tables, nicely title them and share them in in your Word/pdf file.
- d) Comment on the tables and explain your results.
- e) Do your tables reveal continued deep disparities by race? Explain.

Q3[160 points]

Use the worksheet **Covid_IncomeLevel** in A3 Excel File to answer this question.

The data on the worksheet shows the total number of Covid-19 cases, 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.

a) For Covid Infection Rate (% of Population by Zip code) and Average Income variables, calculate the following summary statistics and share them with a table in your Word/pdf file. [35 points]

- i.** Mean
- ii.** Median
- iii.** Minimum (with corresponding Zip code and neighborhood)
- iv.** Maximum (with corresponding Zip code and neighborhood)
- v.** Range
- vi.** First Quartile
- vii.** Second Quartile
- viii.** Third Quartile
- ix.** Interquartile
- x.** 10th, 20th, 25th, 30th, 40th, 50th, 60th, 70th, 75th, 80th, 90th, 95th, and 99th Percentiles
- xi.** Variance
- xii.** Standard Deviation
- xiii.** Coefficient Variation
- xiv.** Skewness
- xv.** Kurtosis
- xvi.** Covariance
- xvii.** Correlation Coefficient

b) Interpret and comment on the above summary statics. [35 points]

c) Prepare a scatter plot to show the correlation between Income and Convid-19 Infection Rate variables and share them in your Word/pdf file. Make sure that Income and Covid Infection variables on the horizontal and vertical axes, respectively. Comment on it using the correlation coefficient you have obtained in (a) [20 points]

- d) Prepare Histograms for Income and Covid Infection Rate variables, nicely title them and share them in your Word/pdf file. Are they negatively or positively skewed? Do they verify the results above in (a)? Explain in detail. **[20 points]**
- e) Using the related formulas and the results you have obtained in (a) calculate the following statistics:
- i. Mean,
 - ii. Variance,
 - iii. Standard Deviation,
 - iv. Coefficient of Variation,
 - v. Skewness,
 - vi. Kurtosis and
 - vii. Correlation Coefficients for both variables are obtained. **[30 points]**
- f) Calculate the Z-scores for both variables and check if there is any extreme values? If so, where do they come from? Which neighborhoods are creating these extreme values? Show them in a table and explain. **[20 points]**