

InClassAssignment1(Group of two)
CS160-02
Introduction to Data Science
Spring 2023

Working on Techniques for Analyzing Data

Instructions: Complete the following activities for this project.

1. Create a new GitHub repository named Assignment1_XXX, where XXX are your initials.
2. Using excel (to generate the result) and word documents (type answers and paste the results) work on the following questions and submit your work using **pdf** format.

- a. What are the differences between data analysis and data analytics?

Data analysis is a more hands-on approach typically observing things that have already happened to explain the question being researched. Data analytics is broader than data analysis and it includes data analysis. Typically analytics creates models from the data to predict the future.

- b. Comment on variable types of Murder, Assault, and urban pop.

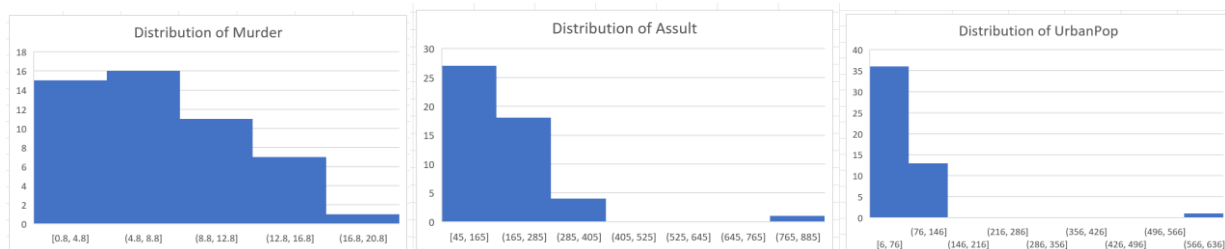
All of the variables are independent and continuous. They are all ratio data because all of them have an absolute zero. The variable type of the states is categorical, and the rest are numerical.

- c. What is the difference between interval and ratio data?

Interval data does not have an absolute zero meaning the values can go above and below zero. In contrast, ratio data has an absolute zero meaning that the values cannot go below zero.

- d. What is descriptive analysis? Represent the data of Murder, Assault, and urban pop. Comment on the distribution.

Descriptive analysis is the first kind of data analysis performed on a data set, and it focuses on three main areas: frequency distribution, measures of centrality, and dispersion of distribution.



All three distributions are right skewed indicating that the mean and median of the data are both greater than the mode, and the mean is greater than the median. All three also appear to have an outlier, especially the assault and urbanpop data.

- e. What is a measure of dispersion? Calculate the interquartile range of those three variables

Measures of dispersion represent the spread of the data. Common measures of dispersion include the range, interquartile range, variance, and standard deviation.

	Murder	Assault	UrbanPop
IQR	7.18	140.00	24.50
Min	0.80	45.00	6.00
Q1	4.08	109.00	53.25
Q2/Median	7.25	159.00	66.00
Q3	11.25	249.00	77.75
Max	17.40	879.00	570.00

- f. What is the measure of centrality? Find the measurement of centrality: mean, median, mode

Measures of centrality describe the center of the distribution. Common measures include the mean, median, and mode.

	Murder	Assault	UrbanPop
Mean	7.79	182.18	74.20
Median	7.25	159.00	66.00
Mode	13.20	120.00	80.00

- g. What are diagnostic analytics? Find diagnostic analysis for pair of variables.

Diagnostic analytics are used to determine why something happened, the most common technique used is correlation.

Correlation		
Murder with Assault	Murder with UrbanPop	Assault with UrbanPop
65%	-19%	-14%

- Using the instructions provided by GitHub, create a git repository named **DS160InClassAssignment**, and push your pdf file to it. Each of you needs to submit your work.

Submission:

Paste a link to your GitHub repository in the area provided for this assignment and submit it by class time.