Team Number: 5

Team Captain: Sriharsha Aitharaju

Team Members: Daniel Rodriguez

Kzzy Centeno

Robert Law

Fernando Sosa

Activity on

PART I (20 Points) Programming

Problem 1.1 (8 Points) Read the EXCLE file "COVID_08312020.csv"

Problem 1.2 (8 Points) Produce a scatter plot using "TotalCases" and "TotalDeaths" and impose a loess line on the top of the data.

Problem 1.3 (8 Points) Produce a scatter plot using "ToTCases_1M" and "TotDeath_MPOP" and impose a loess line on the top of the data.

Problem 1.4 (8 Points) Produce a table with the following summary statistic including minimum, mean, median, variance, standard deviation, maximum, and skewness for the following five variables "ToTCases_1M", "TotDeath_MPOP", "TotalCases", "TotalDeaths", and "TotalTested". **(Note: Display only three decimal place)**

Problem 1.5 (8 Points) Obtain both the Spearman correlation and the Pearson correlation between the following variables "ToTCases_1M", "TotDeath_MPOP", "TotalCases", "TotalDeaths", and "TotalTested".

PART II (10 Points) Fill In Blank

1. Suppose that $\{x_1, x_2, x_3, \cdots, x_n\}$ be a set of data and $x_{(15)} = 5$, $x_{(16)} = 7$, and $x_{(17)} = 8$, the median of this data set is _____**6**_____ if n = 30 and the median of this data is _____**7**_____ if n = 31.
2. Suppose that $\{x_1, x_2, x_3, \cdots, x_n\}$ be a set of data and $\sum_{i=1}^{n} \left(x_i - \underline{x}\right)^2 = 100$ and n = 26, the sample variance of this data set is _____**4**_____ .
3. The points at distances 1.5 times of IQR (Inner Quartile Range) from each hinge mark the ___ ___**inner**_____ fences of the data set.
4. Tom is interested in finding out the salary of students graduated from UCF in the past three years. He collected data from one thousand students graduated from UCF. The data he collected including their major, their graduation year, their gender, their salary, and their GPA. Tom's study is a _____**Classification**_____ with _____**1000**_____ observations and _____**5**_____ predictors.
5. Jennifer has a data set to perform an analysis; however, you cannot find any response variable in this set of data. The analysis performed by Jennifer should be a (supervised learning / **non-supervised learning**).
6. Steve fit a model on a set of data. After perform data exploration analysis, he decided to assume that the data come from normal population and the relationship between the response variable and a set of predictors should be linear. The analysis perform by Steve should be (**parametric analysis** / nonparametric analysis / cluster analysis).
7. Lori likes to know the relationship between a given predictor and the response variable. Lori is interested in (prediction / **inference**) problem.