Project

The purpose of this project is to apply the concepts of descriptive statistics to a real-life data set and produce a report that you could turn in to your boss or any non-statistician. The report should be organized and formatted well, typed (in a program like Microsoft Word), with numbered pages and a cover page.

**Part A ( 60 marks)**

Select one data set from the attached folder.

For each data set try to answer the following questions:

Part 1: Introduction

Your first part of the paper should introduce the reader to the data set and variables. It should identify the variables

* For the quantitative variable, identify the name of the variable, level of measurement, and the unit of measurement.
* For the qualitative variable identify the name of the variable and category levels.

**Part 2: Summary Statistics**

**For Quantitative variable:**

* List the mean, standard deviation, and 5 number (arrange this information neatly, preferably in a table). Interpret the mean, median, and standard deviation
* Graph a box and whisker plot
* May be done by hand, in Paint (or other graphics program), or in Excel.
* Describe the plot (e.g. skew, IQR, and outliers) and interpret the IQR.
* Graph a Histogram and describe the shape and skew.
* Identify any outliers using the IQR or the z-score approach.

**For Qualitative variable:**

* Create a bar chart. There are 2 options (choose your preferred method). The x-axis must be your qualitative variable. The y-axis can be either be:
* the average value of your quantitative variable for each group (preferred)
* Count (or frequency) for each category
* Create a pie chart (do NOT use 3D effects as this causes distortion).
* Describe the information in each of these (e.g. the most and least common levels).
* Based on your data, determine the appropriate measures for quantitative and qualitative variables.

**Part 3: correlation and regression**

Select the first 10 cases of your data set:

* Compute correlation coefficient for two variables
* Estimate regression equation for two variables

**Part B (40 marks)**

As all world facing a pandemic covid-19, select two countries and aggregate their data from internet for period 1-4-2020 to 30-4-2020.

Write a descriptive statistical report including the following:

* Construct Frequency table to build Histogram for weekly affected, dead and recovered cases.
* Draw ogive
* Compute appropriate statistical measures
* Compare between two selected countries
* Write you view about Covid-19 expectations Regression (Independent variable is time week1 week 2 week3 …. and dependent variable is number of affected cases)

**Best** **Regards**