Group Homework 1

STAT011-S23

Due: X/X/23

Introduction and Purpose

The data set you will analyse in this homework records information from different flights for an airline carrier named Envoy Air in the year 2013. The purpose of this assignment is to practice using software to investigate and describe a rather large data set consisting of 19 variables and 25,037 observations.

Required Tech

Excel

The skills necessary to complete this assignment in Excel are covered in the following seven videos:

- Excel 2016 with Data Analysis Toolpak Introduction to Excel 2016 with Data Analysis Toolpak (1:52)
- Excel 2016 with Data Analysis Toolpak Introduction to Excel 2016 with Data Analysis Toolpak: Common Procedures (3:08)
- Excel 2016 with Data Analysis Toolpak Descriptive Statistics and Confidence Intervals for a Mean (2:57)
- Excel 2016 with Data Analysis Toolpak Histogram (3:08)

RStudio

The skills necessary to complete this assignment in RStudio are covered in the following seven videos:

- R Studio Video Introduction to R and RStudio (1:52)
- R Studio Video Getting Started (3:51)
- R Studio Video Working with Data Objects 1 (3:29)
- R Studio Video Working with Data Objects 2 (5:15)
- R Studio Video Importing Data (4:36)
- R Studio Video Descriptive Statistics (3:33)
- R Studio Video Plotting - Histograms, Bar Charts, Boxplots, Scatterplots (4:08)

Instructions

If you are analyzing this data in Excel you first need to download the data set for HW 1 from our Stat 11 Github Data page. Do this by right clicking on the link "View Raw" and save the link with the name `EnvoyAir_flights.csv`. This may take a few moments as it is a large data set.

If you are analyzing this data in RStudio, you will import the data with the following command

```
EnvoyAir_flights <- read.csv("https://raw.githubusercontent.com/ProfSuzy/Stat11/main/Data/EnvoyAir_flig
```

The data object is called `EnvoyAir_flights`.

Once you have access to the data set, complete all parts of the six problems in this assignment. You are encouraged to work with your classmates on this assignment but you must hand in your own, unique write up of the solutions. In a Word document, clearly label each problem's solution. Most solutions will include graphics which can be copied from Excel or RStudio and pasted into your solution document. All solutions require a written component. When you are ready to submit your assignment, save the Word document as a PDF and upload it to the Moodle link for Group Homework #1.

Problem 1

Answer the following questions about you data set called `myData`:

1. What constitutes an observational unit?
2. What are the different variables being collected?
3. Which of the variables are quantitative and which are categorical?
4. What kind of relationship would you expect (if any) between the variables `dep_delay` (departure delay) and `arr_delay` (arrival delay)?

Problem 2

Perform a one variable exploratory data analysis by providing of the variable `distance` by creating:

1. A labeled histogram (with a reasonable number of bins);
2. A labeled box plot;
3. A five-number summary of the minimum, maximum, lower 25% quantile, lower 75% quantile, and the mean.

What story do these analyses tell about the variable `distance`?

Problem 3

For the variable `origin` perform a one variable exploratory data analysis by creating:

1. A labeled bar/frequency chart;
2. A table of frequencies for different airports of origin.

What story do these analyses tell about the variable `origin`?

Problem 4

Generate a scatter plot of the variables `dep_delay` (departure delay) and `arr_delay` (arrival delay). What story does this scatter plot tell about any apparent relationship between these variables? Is this consistent with your answer from Problem 1?

Problem 5

Does the variable `air_time` look like it could be from a Normal (Gaussian) population? Provide a yes or no answer accompanied by a histogram and 1-2 sentences interpreting this histogram.

Problem 6

What is the standard deviation of the variable `air_time`? How many standard deviations away from the mean `air_time` is a flight that takes 60 hours?