Group Homework 5

STAT011-S23

Due: 4/28/23

Introduction and Purpose

The data set you will analyse in this homework is your choice! You can choose from any of the previous data sets we've used in the first four group homework assignments. The purpose of this assignment is for you to practice using all the software analysis techniques you've learned this semester. (If you are doing Option B or C for your final project, this should help refresh you memory on all the different ways you can analyze your data.) This assignment will be graded for completion (do you have the required components) not correctness.

Instructions

Choose which data set you would like to analyze. If you are using Excel, you can download the data sets from our Stat 11 Github Data page. Do this by right clicking on the link "View Raw" and save the link with the name `XXX.csv`.

If you are using R, you can import any of our previous data sets with the following commands:

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

```
EnvoyAir_flights <- read.delim(
        "https://raw.githubusercontent.com/ProfSuzy/Stat11/main/Data/EnvoyAir_flights.txt",
        sep=",")
```

```
arthritis <- read.delim(
        "https://raw.githubusercontent.com/ProfSuzy/Stat11/main/Data/arthritis.csv",
        sep=",")
```

```
titanic <- read.delim(
        "https://raw.githubusercontent.com/ProfSuzy/Stat11/main/Data/titanic.csv",
        sep=",")
```

```
gardasil <- read.delim(
        "https://raw.githubusercontent.com/ProfSuzy/Stat11/main/Data/gardasil_data.txt",
        sep="\t")
```

You are encouraged to work with your classmates (in particular, with your final project group mates) 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 #5.

Problem 1

(a) Choose a numeric variable to describe. Create a histogram to accompany your description.

(b) Choose a categorical variable to describe. Create a table to accompany your description.

Problem 2

Choose a numeric and a categorical variable (these can be the same as the ones you created in Problem 1). Describe any apparent relationships between these two variables and create bar charts to accompany your description.

Problem 3

(a) Choose two numeric variables. Describe any apparent or suspected relationships between these two variables and create a scatterplot to accompany your description.

(b) Determine if a linear regression model could appropriately describe the relationship between these numeric variables.

Problem 4

Define your population of interest and define a relevant proportion related to this population. Conduct a hypothesis test about this unknown proportion using the data. (Don't repeat Problem 5 from GHW #4.) Make sure your answer includes:

1) the definition of your parameter and your significance level;
2) your null and alternative hypotheses;
3) an assessment of the required assumptions and conditions;
4) the calculated test statistic;
5) the p-value and conclusion of your test in context.

Problem 5

Define your population of interest and define a relevant population mean that you will estimate with a confidence interval. (Don't repeat Problem 1 from GHW #4.) Make sure your answer includes:

1) the definition of your parameter and your confidence level;
2) the sample mean;
3) an assessment of the required assumptions and conditions;
4) the critical value corresponding to your confidence level;
5) the standard error of your sample mean;
6) the lower and upper bounds of your interval interpreted in context.