Lab1

Diala Ezzeddine

January 15, 2019

Contents

1 Exploring and understanding data

1.1 Explore the structure of data

We are going to explore the **"usedcars.csv"** dataset, which contains actual data about used cars recently advertised for sale on a popular website. You can access this data from Canvas. Look for the Data file in files or the dataset in Modules Week2.

1. How you can load this dataset into R studio?
2. How the dataset is organized? explain what you see in this data.
3. Are all those features clearly stated?

1.2 Exploring numeric variables

Let's look deeper into the three numerical variables in the usedcars dataset: **year**, **price** and **mileage**.

1. Find the central and spread measures of those 3 features.
2. What can you conclude from the results of the year variable?
3. Discuss the mean and median results of the year and mileage features.
4. What is the range, IQR, 30th, 60th, 90% and 99% percentiles of the price variable?
5. Create the price and mileage boxplots and histograms. Can you conclude the distribution for each variable?
6. Compute the variance and standard deviation of the price and mileage variables.
7. Find the interval of advertised prices of the 95% of the cars.

1.3 Exploring categorical variables

Let's look deeper into the three categorical variables in the usedcars dataset: **model**, **color** and **transmission**.

1. Examine those 3 variables separately.
2. Find the central measure of those 3 features.
3. Can you find the proportion table of the model?
4. Display the percentage table with 2 decimal places of the color variable.

1.4 Exploring relationships between two variables

Let's look deeper into the following two questions:

a. Does the **price** data imply that we are examining only economy-class cars or are there also luxury cars with high mileage?

b. Do relationships between the model and color data provide insight into the type of cars we are examining?

1. What we should do to answer question a.? Explain your result.
2. What we should do to answer question b.? add the chi-squared test to your result. Explain your result.