

# Sample Quiz 1 solu

*Linda*

*Due: 5:00pm, Jul 19, 2021*

## Contents

<b>1</b>	<b>Overview</b>	<b>2</b>
<b>2</b>	<b>Case study: are Asian cars more efficient?</b>	<b>2</b>
2.1	Question 1: Read data . . . . .	2
2.2	Question 2: EDA . . . . .	3
2.3	Question 3: Comparison . . . . .	5

# 1 Overview

This is a 30 minute, **individual** quiz. It is an open book exercise. Feel free to copy any R-chunks from the lectures, homework. Ask your TF for help if you get stuck somewhere.

Submit the following two files to the canvas

- quiz1.rmd
- quiz1.html or .pdf (Only 1 compiled one)
- We will keep the submission open until 6:00pm

## Objectives:

We focus on the following aspects

- Read data into R
- Make sense out of the data through EDA
- Be able to use R-Markdown
- Be able to submit your work to the canvas
- No need to beautify your R-markdown .html or .pdf.

## 2 Case study: are Asian cars more efficient?

Asian cars, especially Japanese and Korean cars have been loved by families due to their reputation of being fuel efficiency and relative cheaper in prices. We will use the `car_04_regular.csv` data set (available in the folder) to perform some preliminary analysis.

### 2.1 Question 1: Read data

Read the data first. To make sure you get the data into R we have included the following R-chuck. Notice this chunk works only if you have stored the data `car_04_regular.csv` in a sub-folder called `data`.

```
data1 <- read.csv("data/car_04_regular.csv", header=TRUE)
str(data1)
```

```
## 'data.frame':    226 obs. of  13 variables:
## $ Make.Model   : Factor w/ 226 levels "Acura_MDX","Acura_NSX",...: 3 5 6 4 2 1 7 8 9 10 ...
## $ Continent    : Factor w/ 3 levels "Am","As","E": 2 2 2 2 2 2 3 3 3 3 ...
## $ MPG_City      : int   18 20 23 25 17 17 20 18 17 16 ...
## $ MPG_Hwy       : int   24 28 32 34 24 23 28 25 24 22 ...
## $ Horsepower    : int   225 270 200 160 252 265 170 220 330 250 ...
## $ Weight        : num   3.9 3.58 3.32 2.77 3.2 ...
## $ Length        : num   197 189 183 172 174 ...
## $ Width         : num   71.6 72.2 69.4 67.9 71.3 77 76.3 76.1 74.6 76.1 ...
## $ Seating       : int    5 5 5 4 2 7 5 5 5 5 ...
## $ Cylinders      : int    6 6 4 4 6 6 4 6 6 6 ...
## $ Displacement: num    3.5 3.2 2.4 2 3 3.5 1.8 3 4.2 2.7 ...
## $ Make          : Factor w/ 38 levels "Acura","Audi",...: 1 1 1 1 1 1 2 2 2 2 ...
## $ Transmission: Factor w/ 3 levels "automatic","cont_variable",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
names(data1)
```

```
## [1] "Make.Model" "Continent" "MPG_City" "MPG_Hwy"
## [5] "Horsepower" "Weight" "Length" "Width"
## [9] "Seating" "Cylinders" "Displacement" "Make"
## [13] "Transmission"
```

```
summary(data1)
```

```
##      Make.Model  Continent    MPG_City    MPG_Hwy    Horsepower
## Acura_MDX:  1   Am:74      Min.   :11.0   Min.   :14.0   Min.   :103
## Acura_NSX:  1   As:98      1st Qu.:17.0   1st Qu.:23.0   1st Qu.:160
## Acura_RL :  1   E :54      Median :19.0   Median :26.0   Median :200
## Acura_RSX:  1           Mean  :19.3   Mean   :25.9   Mean   :212
## Acura_TL :  1           3rd Qu.:22.0   3rd Qu.:29.0   3rd Qu.:252
## Acura_TSX:  1           Max.   :33.0   Max.   :39.0   Max.   :390
## (Other) :220
##      Weight      Length      Width      Seating      Cylinders
## Min.   :2.10   Min.   :143   Min.   :65.4   Min.   :2.00   Min.   :2.00
## 1st Qu.:3.16   1st Qu.:178   1st Qu.:69.5   1st Qu.:5.00   1st Qu.:4.00
## Median :3.57   Median :187   Median :71.7   Median :5.00   Median :6.00
## Mean   :3.69   Mean   :186   Mean   :72.0   Mean   :5.12   Mean   :5.69
## 3rd Qu.:4.06   3rd Qu.:194   3rd Qu.:74.4   3rd Qu.:5.00   3rd Qu.:6.00
## Max.   :5.82   Max.   :224   Max.   :80.2   Max.   :8.00   Max.   :8.00
##
##      Displacement      Make      Transmission
## Min.   :1.30   Chevrolet : 15   automatic   :211
## 1st Qu.:2.40   Toyota   : 15   cont_variable: 3
## Median :3.20   Mitsubishi: 11   manual      : 12
## Mean   :3.21   Nissan   : 10
## 3rd Qu.:4.00   Volkswagen: 9
## Max.   :6.00   Ford     : 8
##           (Other) :158
```

- i. How many variables are there in this data? And how many cars are there?

**Answers:** There are 13 variables with 226 cars.

- ii. The variable **Continent** is a categorical variable which labels which region a car is made. How many continent are there?

**Answers:** There are three Continents: Am, As and E.

## 2.2 Question 2: EDA

Fuel efficiency can be measures by **MPG\_City**, mileage per gallon in city or **MPG\_highway**. Let us look into the **MPG\_City**

- i. What are the sample mean and sample standard deviation of of **MPG\_City** among all cars?

```
mean(data1$MPG_City)
```

```
## [1] 19.3
```

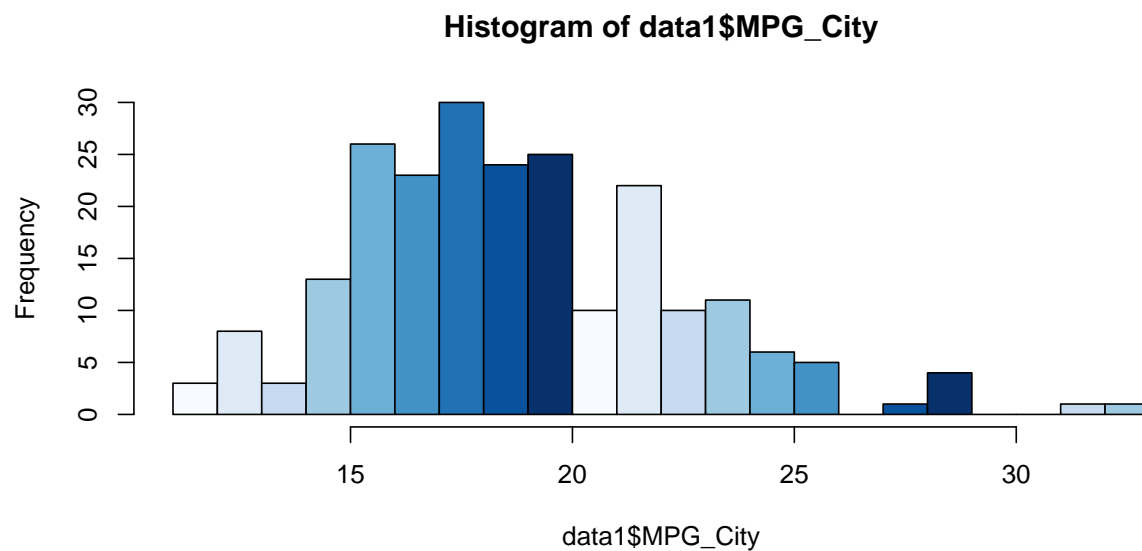
```
sd(data1$MPG_City)
```

```
## [1] 3.7
```

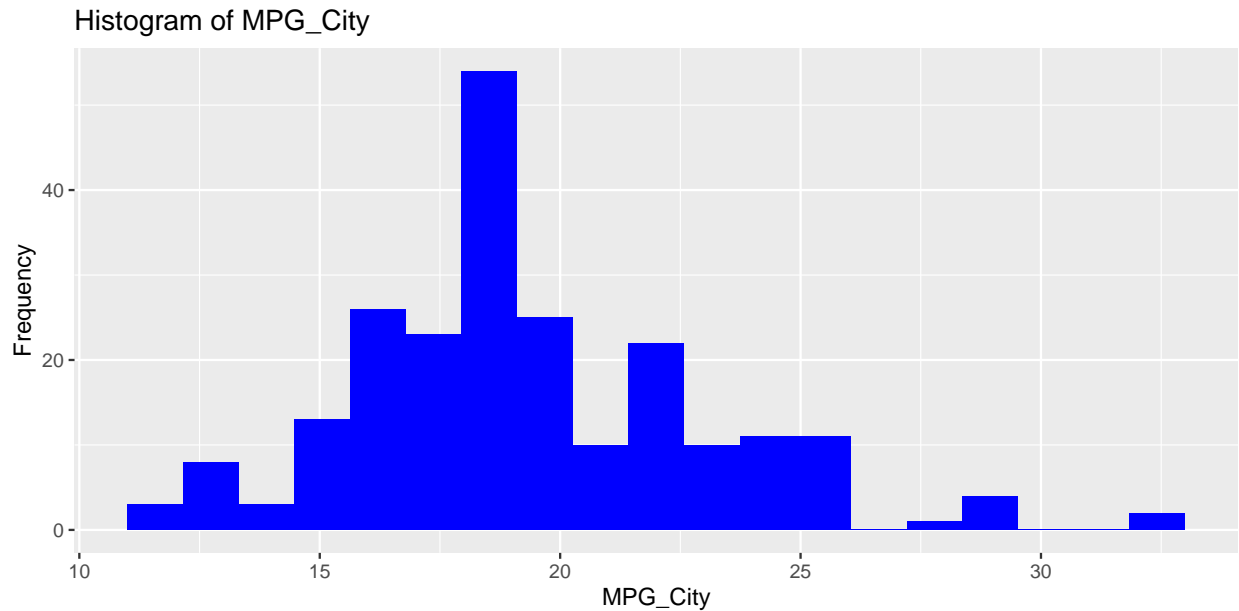
**Answers:** The sample mean of MPG\_City is 19.3 and the sd is 3.7

- ii. Show the histogram of MPG\_City among all cars. Does 'MPG\_City' appear to you a normal variable? Why or why not?

```
# You only need to show hist of MPG_City either use base R or ggplot!!!  
# base R hist()  
hist(data1$MPG_City, breaks = 25, col = blues9)
```



```
# ggplot hist  
ggplot(data1) +  
  geom_histogram(aes(x = MPG_City), bins = 20, fill = "blue") +  
  labs(title = "Histogram of MPG_City",  
        x = "MPG_City",  
        y = "Frequency")
```



**Answers:** The histogram of MPG\_City appears fairly symmetric. I would say the distribution seems to be approximately normal. But if you insist on that the distribution is skewed to right I would agree as well.

## 2.3 Question 3: Comparison

Here is the R-chunk to get summary statistics by Continent.

```
# Break data1 by Continent
data1 %>%
  group_by(Continent) %>%
  summarise(
    mean = mean(MPG_City),
    sd = sd(MPG_City),
    n = n()
  )
```

```
## # A tibble: 3 x 4
##   Continent mean    sd    n
## * <fct>    <dbl> <dbl> <int>
## 1 Am       18.7  2.99   74
## 2 As       20.2  4.20   98
## 3 E       18.3  3.22   54
```

- i. What is the sample mean of MPG\_City among all Asian cars?

**Answers:** For Asian cars, the sample mean of MPG\_City is 20.2

- ii. What is the sample mean of MPG\_City among all European cars?

**Answers:** For European cars, the sample mean of MPG\_City is 18.3

- iii. By comparing the two sample means above, do we have some evidence that Asian cars are more efficient than that of European cars? (The larger `MPG_City` is the more efficient a car gets.) Comment on the limitations of this simple analysis.

**Answers:** Yes, based on the sample means which are good estimates of the population means it appears that Asian cars are more efficient.

**Limitations:** Sample means have variabilities from the true means. The better way would be either to provide a confidence interval for the population means difference or to test that the two means are different. That will be done later in the multiple regression lecture.

Other concerns you would like to add?

**End of the sample quiz**