Data Analysis – Test Evaluation 2 (Part 2)

Your name here: Paolo Sean

Wed 24 Apr 2024

Instructions (read carefully):

Use Rstudio and avoid repetitive and unnecessary "global" knitting while working your excercises. Prefer to "Run current chuck" to check and debug. Knit at the end of course to generate your PDF.

Please DO NOT mix the R markdown structure! Just give your answers in the chunks right below each question and provide brief text-answers when required. Do not delete anything.

Load your packages. In general you might need to install more during the exam (but most likely not on this occasion). All the relevant information will be given when necessary.

Make sure your PDF contains the OUTPUT !!!

Good luck!

Exercise 6 (10 points in total)

Use your **RND** value given in the first page of the test (Part 1) and mention it in the relevant chunk!

The **AvsB.txt** data set contains information about the consumption of gallons per mile for light trucks produced in Country A (sample 1) and Country B (sample 2).

(1) Load and read the data and give a simple summary statistics for each sample. No need to type by hand anything. Just make sure that your codes gives the relevant answer. **(1.5 points)**

```
AvsB <- read.table(file = "./AvsB.txt",
                       stringsAsFactors = FALSE)

head(AvsB)
```

```
##   V1 V2
## 1 18 24
## 2 15 27
## 3 18 27
## 4 16 25
## 5 17 31
## 6 15 35
```

```r
summary(AvsB)
```

```
##        V1              V2
##  Min.   : 9.00   Min.   :18.00
##  1st Qu.:13.00   1st Qu.:25.50
##  Median :15.00   Median :32.00
##  Mean   :16.03   Mean   :30.48
##  3rd Qu.:18.00   3rd Qu.:34.00
##  Max.   :28.00   Max.   :47.00
```

(2) For the purposes of this particular study, the size of the samples is considered to be **small**. The 1st column refers to Country A and the 2nd to Country B. Create two variables for the two columns and name them **Ampg** and **Bmpg** respectively. **(1.5 points)**

```r
# Ampg = ...
# Bmpg = ...

names(AvsB)<- c("Ampg","Bmpg")
head(AvsB)
```

```
##   Ampg Bmpg
## 1   18   24
## 2   15   27
## 3   18   27
## 4   16   25
## 5   17   31
## 6   15   35
```

(3) For the Ampg (Country A) variable, replace the 1st line entry (initially 18) to 18+**(RMD/10)** (mind the $\pm$ signs) **using an R command (replace)** (look it up if necessary) and run a summary statistics for the updated variable Ampg (keep the same variable name). Your mean value must be different now (a little bit). **(2 points)**

NOTE: If you get stuck you may change it manually in the file externally and continue with the next questions but you will not get this point.

```r
# Enter your RND value below and print it out.
# RND = 15
mean(AvsB$Ampg)
```

```
## [1] 16.02532
```

```r
#before change the first value

RND <- 15
NewAmpg <- replace(AvsB$Ampg, 1, 18 + (RND / 10))

mean(NewAmpg)
```

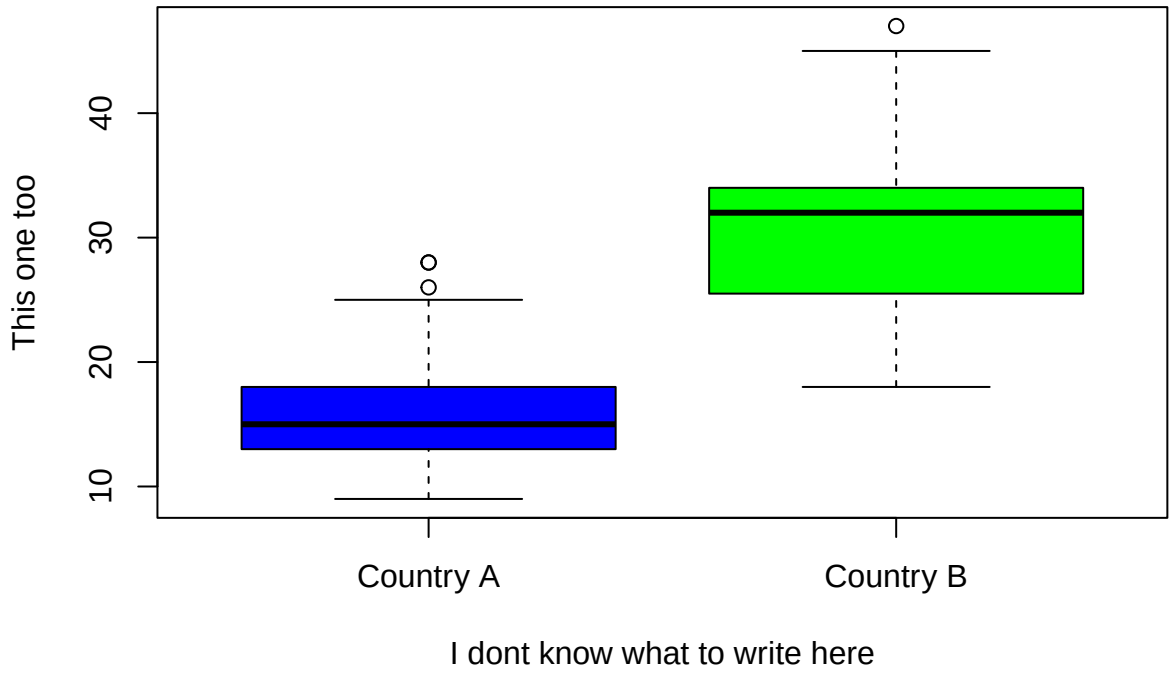
```
## [1] 16.0443
```

```
#
```

(4) Plot one figure with both boxplots in different colors. Edit the labels in the x, y axes and the $x-$tick labels for the each variable (per boxplot). **(2 points)**

NOTE: you will get only **1 point** for a simple boxplot, namely without formatting the labels of axis and $x-$tick.

```
boxplot(AvsB, col = c("blue", "green"),
        xlab = "I dont know what to write here", ylab = "This one too",
        names = c("Country A", "Country B"))
```



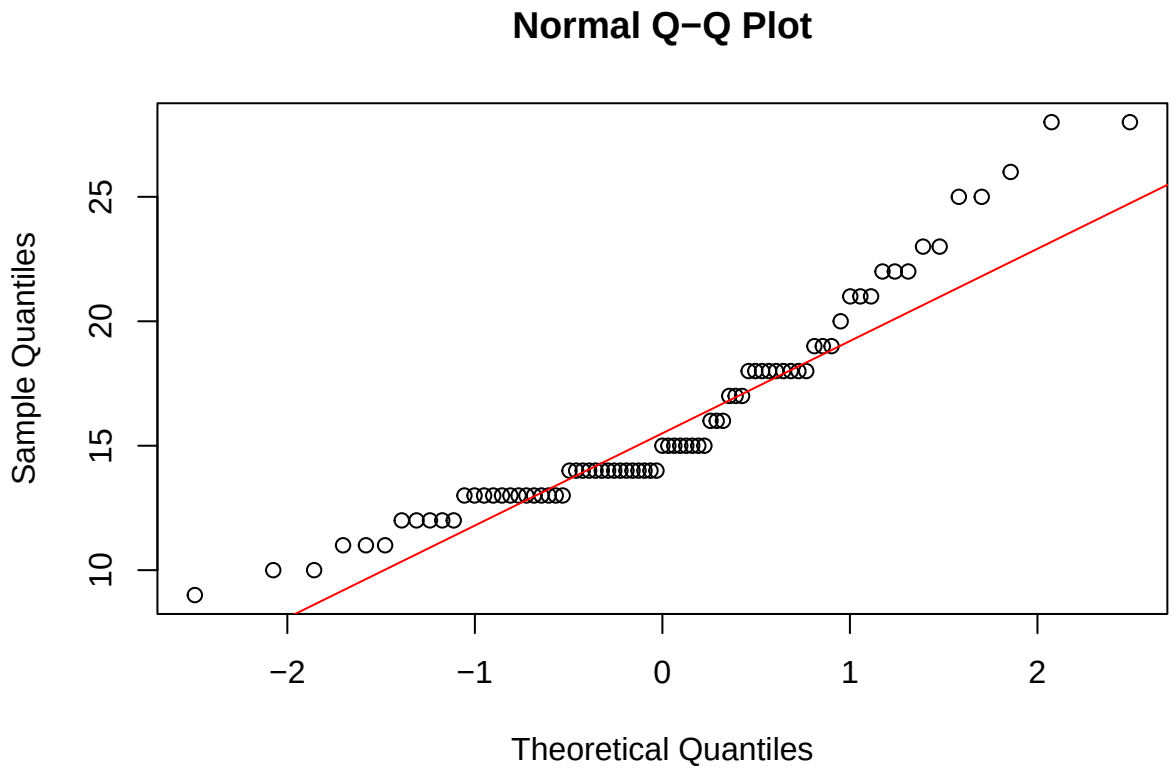
(5) Test the hypothesis that the samples means are equal (assume that their variances are equal and $\alpha = 0.05$). Choose an adequate test-statistic, support with some arguments your choice and check whether the **main assumptions** of the test are fulfilled, for example whether the observations **randomly and normally distributed**. Finally, give a brief interpretation of your result. **(3 points)**

NOTE: You will get **1 point** for choosing the correct statistical test, **1 point** for giving correct answer regarding the assumptions & the relevant code/figure(s) to support it and **1 point** for the correct answer/interpretation of the obtained result.

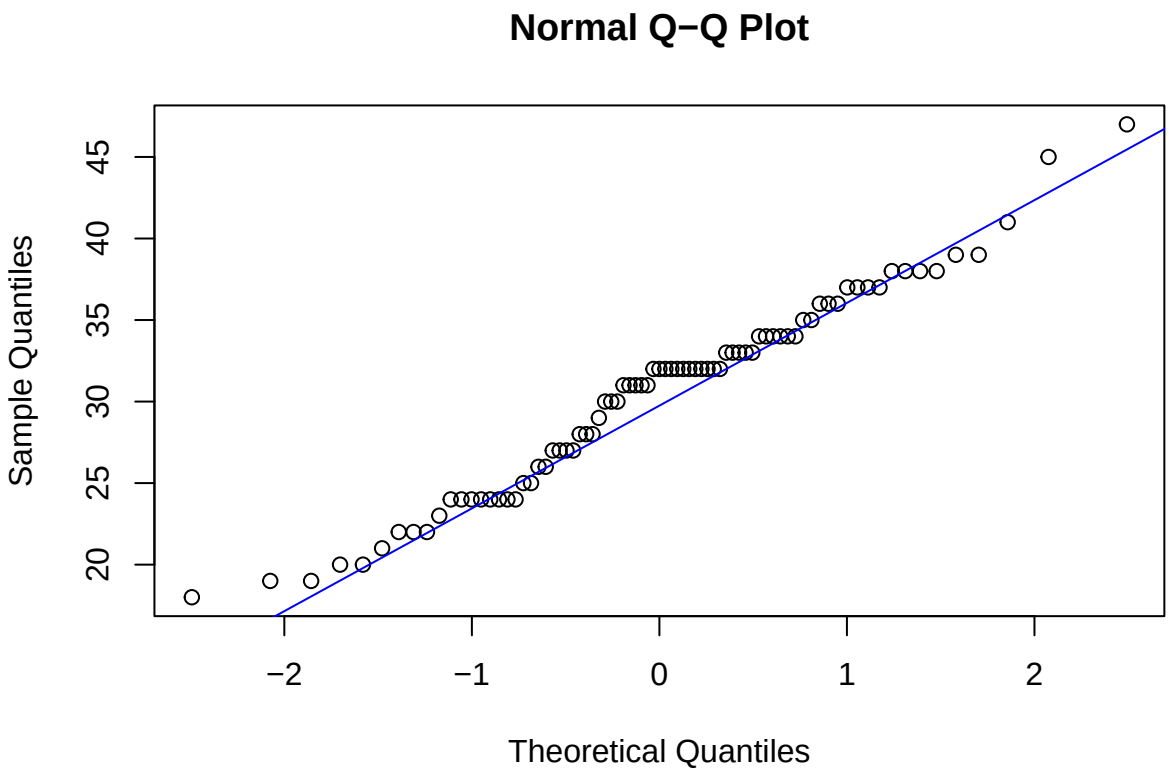
Answer: **We will perform two sample t test because we want to compare the 2 mean of the data**

```
# Check for assumptions
```

```
qqnorm(AvsB$Ampg)
qqline(AvsB$Ampg,col="red")
```




```
qqnorm(AvsB$Bmpg)
qqline(AvsB$Bmpg,col="blue")
```




```
# Perform the statistical test.
# Perform the statistical test.
t.test(AvsB$Ampg, AvsB$Bmpg, var.equal= TRUE)
```

```
##
##  Two Sample t-test
##
## data:  AvsB$Ampg and AvsB$Bmpg
## t = -17.338, df = 156, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.10264 -12.80876
## sample estimates:
## mean of x mean of y
##  16.02532  30.48101
```

Answer: **Interpretation of the obtained result: we fail to reject**