UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTE OF COMPUTING

SCIENTIFIC METHODOLOGY FOR COMPUTING
MO430

EXERCISE 2

STUDENT: Gian Franco Joel Condori Luna
RA: 234826

UNPAIRED DATA

1.- Teste t:

```
#Read file
data <- read.csv("/exercise_2/ex2.csv", stringsAsFactors = FALSE)

set.seed(10)

#Separate data by people who have diabetes and people who do not have
diabetes
si_diabetes <- data$bp[data$type=="Yes"]
no_diabetes <- data$bp[data$type=="No"]

test <- t.test(si_diabetes, no_diabetes) # test t

print(test)
```

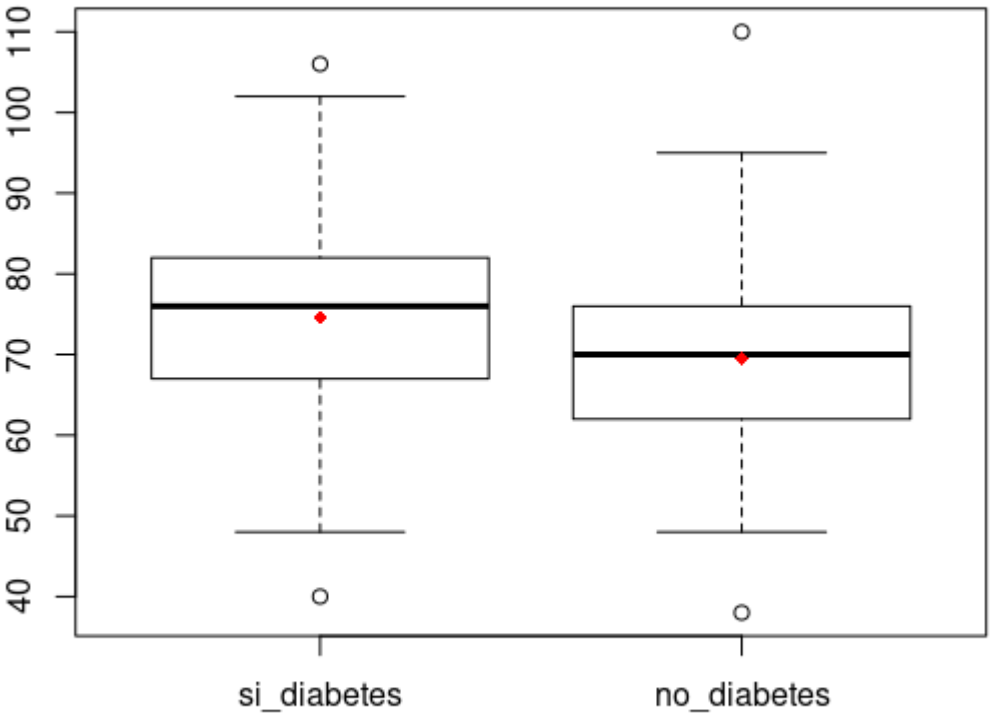
- Result:

```
data:  si_diabetes and no_diabetes
t = 2.9592, df = 130.28, p-value = 0.003665
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  1.671482 8.414080
sample estimates:
  mean of x mean of y
74.58824  69.54545
```

- Diagrams:

```
boxplot(si_diabetes,no_diabetes,names=c("si_diabetes","no_diabetes"))#Sh
ow the diagrams
```

```
medias <- c(mean(si_diabetes),mean(no_diabetes))#Show the mean by a
point
points(medias,pch=18,col="red")#Highlight the mean of a color
```



2.- Wilcoxon rank sum

```
wilcox.test(si_diabetes, no_diabetes, alternative = "two.sided")
```

- Result:

```
Wilcoxon rank sum test with continuity correction

data:  si_diabetes and no_diabetes
W = 5669, p-value = 0.002294
alternative hypothesis: true location shift is not equal to 0
```

3.- Unpaired data report:

- **Which has the highest average pressure? And higher median pressure?**

People with diabetes have higher mean pressure.
The result tells us that the mean for people with diabetes is 74.58824 and people without diabetes is 69.54545.
People with diabetes have a median equal to 76 and people without diabetes 70.

- **Is the difference statistically significant?**

  Yes, since the p-value is less than 0.05.

- **Discuss the difference in p-values.**

  In the test carried out with "t test" it gives us "p-value = 0.003665", while with "wilcox" it gives us "p-value = 0.002294", in this case "wilcox" is being more precise in giving his results.

- **Which of the 2 tests do you think is more appropriate in this case?**

  In this case “wilcox” is giving a lower value, so we could conclude that for this case “wilcox” is more appropriate.

PAIRED DATA

1.- Teste t:

```
#Read file
data <- read.csv("/exercise_2/ex2-paired.csv", stringsAsFactors = FALSE)

set.seed(10)

#'-------------------------
#'teste t
#'-------------------------
test <- t.test(data$November, data$August, paired = TRUE) # teste
```

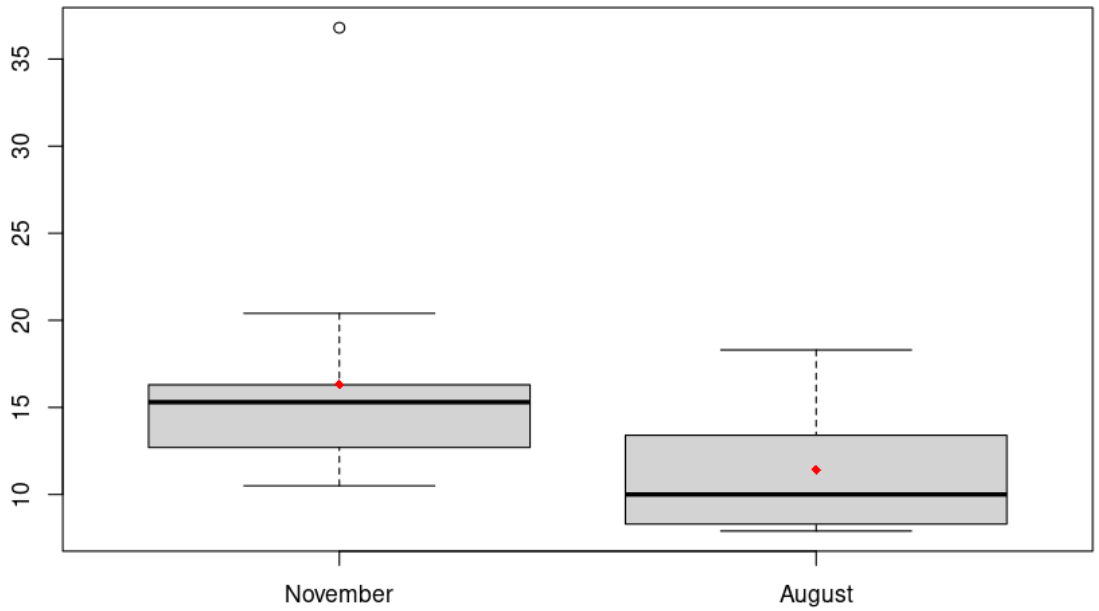
- Result:

```
Paired t-test

data:  data$November and data$August
t = 2.3089, df = 12, p-value = 0.03956
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  0.2760652 9.5239348
sample estimates:
  mean of the differences
4.9
```

- Diagrams:

```
boxplot(data$November,data$August,names=c("November","August"))#Show the diagrams
medias <- c(mean(data$November),mean(data$August))#Show the mean by a point
points(medias,pch=18,col="red")#Highlight the mean of a color
```



2.- Wilcoxon rank sum

```
wilcox.test(data$November, data$August, paired = TRUE)
```

- Result:

```
Wilcoxon signed rank exact test

data:  data$November and data$August
V = 75, p-value = 0.03979
alternative hypothesis: true location shift is not equal to 0
```

3.- Paired data report:

- **Which month has the highest mean and highest median?**

The mean for November is 16.32308 and the mean for August is 11.42308.
The median for November is 15.3 and the mean for August is 10.
Therefore, the month of November has the largest mean and median.

- **Is the difference statistically significant?**

  Yes, since the p-value is less than 0.05.

- **Discuss the difference in p-values.**

  In the test carried out with "t test" it gives us "p-value = 0.03956", while with "wilcox" it gives us "p-value = 0.03979", in this case "t.test" is being more precise in giving his results.

- **Which of the 2 tests do you think is more appropriate in this case?**

  In this case “t.test” is giving a lower value, so we could conclude that for this case “t.test” is more appropriate.