

# Practical Session2a

Juan Vanegas

2023-12-16

## General Information

- The practical session consists of 5 questions in which you are asked to conduct an analysis of a dataset or to use R for simple calculations. Datasets are available in R as data frames.
- Your output should consist of a PDF file which contains the results and R code.
- Solutions will be available online in BB in a later stage (you will receive an email about this via BB).

## Question 1

In this question we focus on the `airquality` data which is available as a data frame in R. Use `help(airquality)` to get information about the data. 1. Define an R object which contains the information about the wind speed. 2. Calculate the mean, median and variance for the wind speed. 3. Produce the figure below.

```
windspeed <- airquality$Wind
sprintf("mean: %f", mean(windspeed))

## [1] "mean: 9.957516"

sprintf("var: %f", var(windspeed))

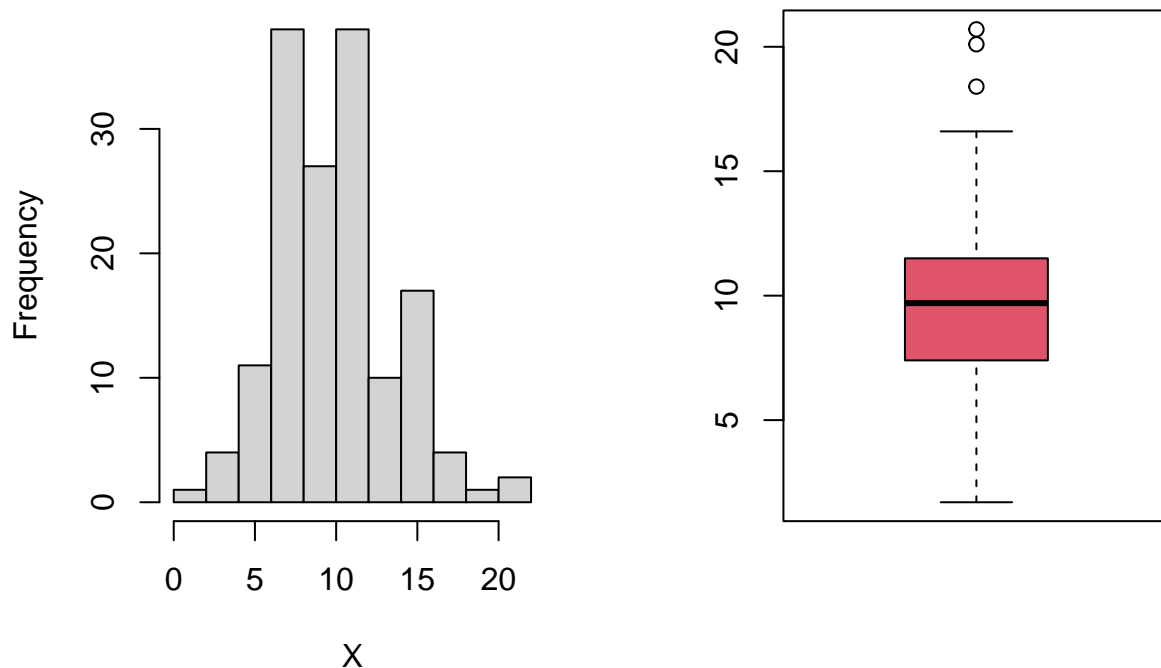
## [1] "var: 12.411539"

sprintf("median: %f", median(windspeed))

## [1] "median: 9.700000"

par(mfrow=c(1,2))
hist(windspeed, xlab = "X", main = "Histogram of wind speed")
boxplot(windspeed, col=2)
```

## Histogram of wind speed



## Question 2 In this question we focus on the ToothGrowth data which available as data frame in R. Use `help(ToothGrowth)` to get information about the data. 1. The response variable is the Tooth length. 2. Test if the Supplement type has an effect on the tooth length.

```
head(ToothGrowth)
```

```
##      len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

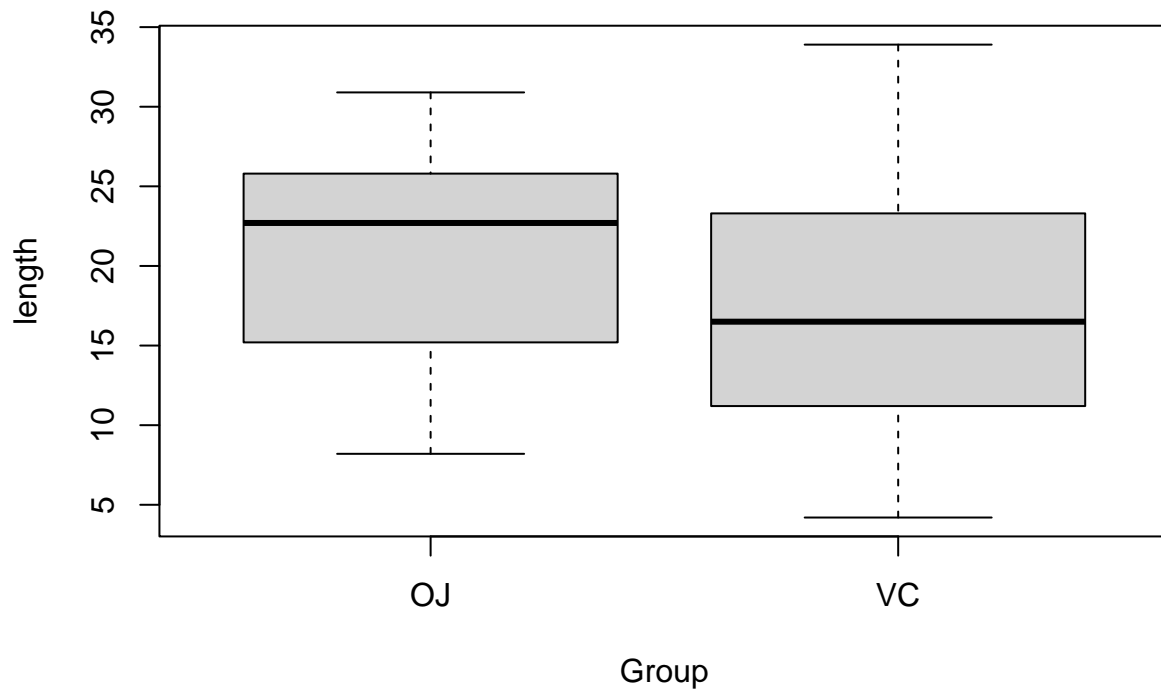
```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
t.test(len~supp, var.equal=TRUE, data=ToothGrowth)
```

```
##
## Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
## -0.1670064  7.5670064
## sample estimates:
```

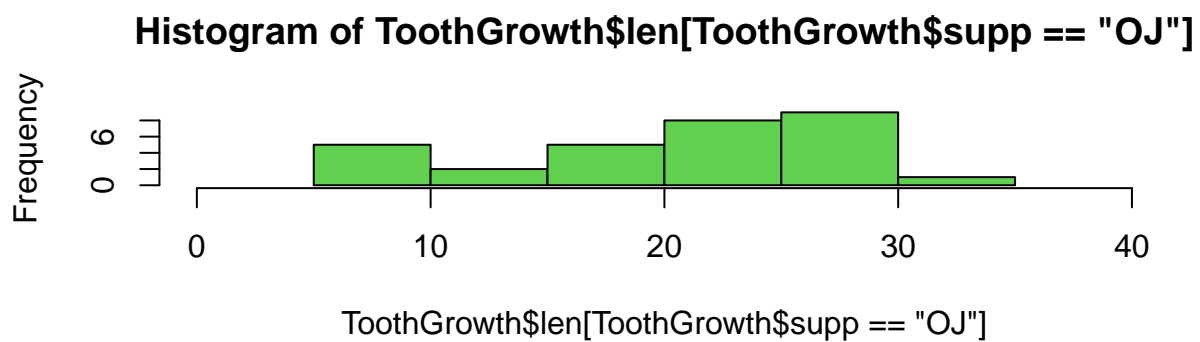
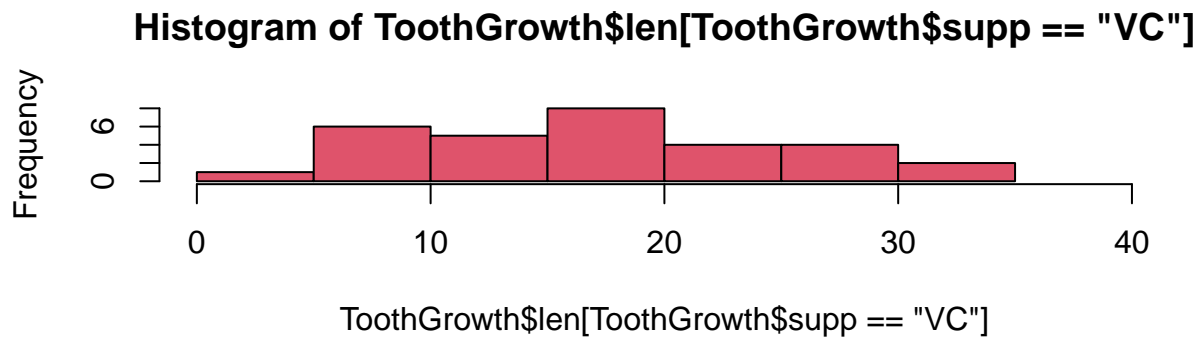
```
## mean in group OJ mean in group VC
##      20.66333      16.96333
boxplot(ToothGrowth$len~ToothGrowth$supp, xlab="Group", ylab = "length")
```



## Question 3

For the ToothGrowth data, produce an histogram for the tooth length (by group) with the following structure.

```
par(mfrow=c(2,1))
hist(ToothGrowth$len[ToothGrowth$supp == "VC"], col=2, xlim=c(0, 40))
hist(ToothGrowth$len[ToothGrowth$supp == "OJ"], col=3, xlim=c(0, 40))
```



## Question 4 Create the following data frame

```
Name <- c("A", "B")
Gender <- c("M", "M")
Height <- c(180, 170)
data.frame(Name, Gender, Height)
```

```
##   Name Gender Height
## 1    A      M    180
## 2    B      M    170
```

## Question 5

In this question we focus on the chickwts data which available as data frame in R. Use `help(chickwts)` to get information about the data. 1. Create a new data frame which contains data only of the horsebean and the soybean groups. How many observations are included in the new dataset. 2. calculate the mean and standard deviation of the weight by group.

```
newdata <- chickwts[chickwts$feed %in% c("horsebean", "soybean"), ]
dim(newdata)
```

```
## [1] 24  2
```

```
tapply(newdata$weight, newdata$feed, mean)
```

```
##   casein horsebean  linseed  meatmeal  soybean sunflower
##      NA   160.2000      NA      NA   246.4286      NA
```

```
tapply(newdata$weight, newdata$feed, sd)
```

##	casein	horsebean	linseed	meatmeal	soybean	sunflower
##	NA	38.62584	NA	NA	54.12907	NA