Statistisk Dataanalyse 2023

Instruktør: Vivek Misra

Exercise Class NR7

Solutions to the Tasks

Task 1 – Description / Solution

- Install R and check the version of the software you have installed. You can do that by typing R. Version () in the console.

- Solution: Please follow the teachers advisory during the lecture for the installment of R. This tool is very useful for daily work regarding statistics, but also (very) benefical during the final exam in this Subject.

Task 2 - Description

- Create the following vector in R:

{8,9,9,14,8,8,10,7,6,9,7,8,10,14,11,8,14,11}

a) For the data assigned to this vector, calculate the following:
   I. Mean
   II. Median
   III. Standard Deviation

b) Construct a histogram for the data.

c) Construct a boxplot for the data.

Task 2A - Solution

- We will write the following commandoes shown below:

```
vector <- c(8, 9, 9, 14, 8, 8, 10, 7, 6, 9, 7, 8, 10, 14, 11, 8, 14, 11)

#Task 2.A (i)
mean(vector)

#Task 2.A (ii)
median(vector)

#Task 2.A (iii)
sd(vector)
```

- The following result is founded:

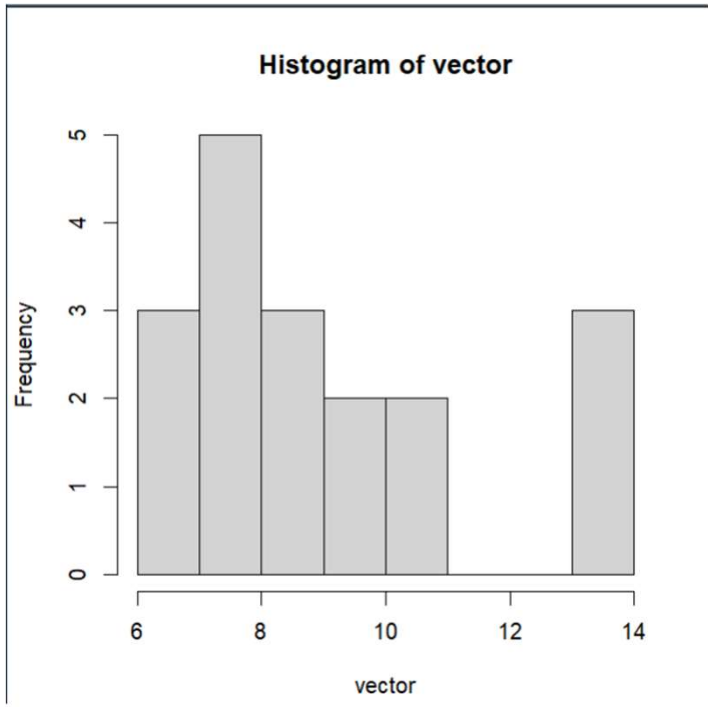
```
> vector <- c(8, 9, 9, 14, 8, 8, 10, 7, 6, 9, 7, 8, 10, 14, 11, 8, 14, 11)
> mean(vector)
[1] 9.5
> median(vector)
[1] 9
> sd(vector)
[1] 2.455486
```

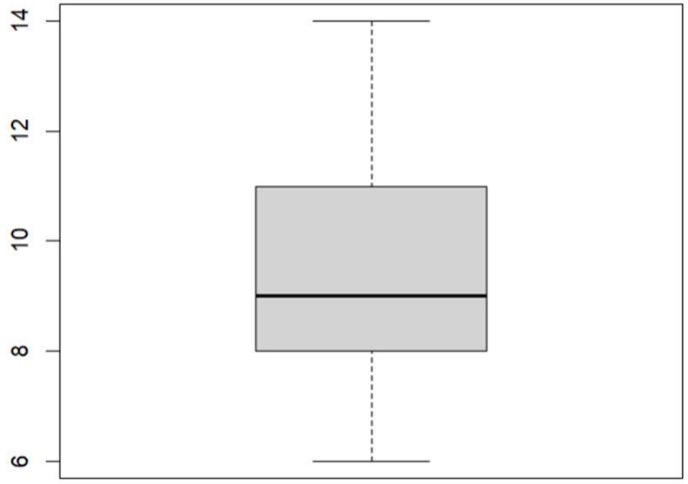
Task 2BC - Solution

- We will write the following command shown below:

```
#Task 2.B
hist(x=vector)

#Task 2.C
boxplot(vector)
```





Task 2 - Description

- Here is the description:

Step 1: Create a data frame in R called ***data.comput*** with data on 5 laptop computers regarding their memory, storage and display size. You know the following:

- Computer 1 has 8 GB RAM of memory, 500 GB storage drive and 13 inches display.
- Computer 2 has 16 GB RAM of memory, 500 GB storage drive and 15 inches display.
- Computer 3 has 16 GB RAM of memory, 1000 GB storage drive and 13 inches display.
- Computer 4 has 8 GB RAM of memory, 240 GB storage drive and 15 inches display.
- Computer 5 has 16 GB RAM of memory, 500 GB storage drive and 17 inches display.

Step 2: Calculate the mean, median and standard deviation for the variables "memory", "storage", and "display".

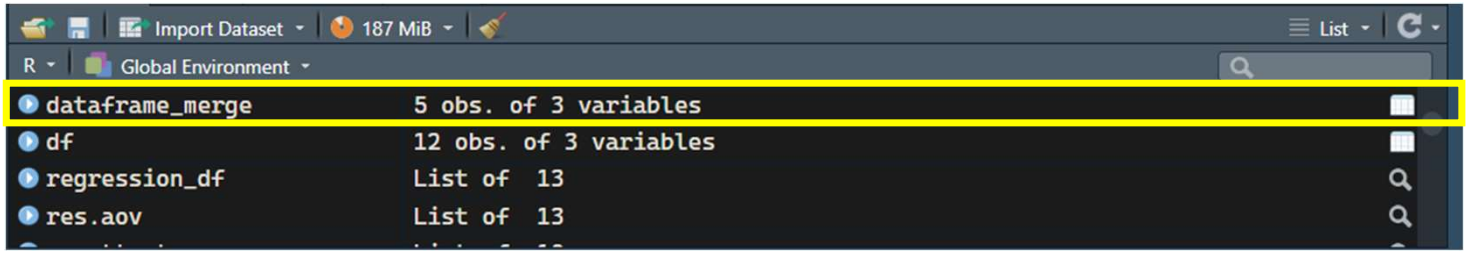
Step 3: Save the workspace (environment) containing the data frame ***data.comput*** in a work directory that is convenient to you. To practice how to open it again, close the R session and open the workspace again and see if you can easily recover the objects (i.e., data, values) of the previous session.

Task 3.1 - Solution

- We have the following commands:

```
#Task 3.1
memory <- c(8,16,16,8,16)
storage_drive <- c(500,500,1000,240,500)
display <- c(13,15,13,15,17)
dataframe_merge <- data.frame(memory,storage_drive,display)
#Print in the commandline terminal
dataframe_merge
```

- Now click on the right window where the global enviroment is.
  - Click on the variable name for data.frame which in this case is dataframe_merge.



	memory	storage_drive	display
1	8	500	13
2	16	500	15
3	16	1000	13
4	8	240	15
5	16	500	17

Task 3.2 & 3.3 - Solution

Index Axis	Memory	Storage_Drive	Display

Commands	`#Memory`
	`mean(dataframe_merge$memory)`
	`median(dataframe_merge$memory)`
	`var(dataframe_merge$memory)`
	`sd(dataframe_merge$memory)`
		`#Storage_Drive`
		`mean(dataframe_merge$storage_drive)`
		`median(dataframe_merge$storage_drive)`
		`var(dataframe_merge$storage_drive)`
		`sd(dataframe_merge$storage_drive)`
			`#Display`
			`mean(dataframe_merge$display)`
			`median(dataframe_merge$display)`
			`var(dataframe_merge$display)`
			`sd(dataframe_merge$display)`

Index Axis	**Memory Results**	**Storage_Drive Results**	**Display Results**
Mean	`> mean(dataframe_merge$memory)` `[1] 12.8`	`> mean(dataframe_merge$storage_drive)` `[1] 548`	`> mean(dataframe_merge$display)` `[1] 14.6`
Median	`> median(dataframe_merge$memory)` `[1] 16`	`> median(dataframe_merge$storage_drive)` `[1] 500`	`> median(dataframe_merge$display)` `[1] 15`
Variance	`> var(dataframe_merge$memory)` `[1] 19.2`	`> var(dataframe_merge$storage_drive)` `[1] 76520`	`> var(dataframe_merge$display)` `[1] 2.8`
Standard Deviation	`> sd(dataframe_merge$memory)` `[1] 4.38178`	`> sd(dataframe_merge$storage_drive)` `[1] 276.6225`	`> sd(dataframe_merge$display)` `[1] 1.67332`

After writing the script, then remember to press Ctrl+S, and save the program. Then open the file again!

Task 4 - Description

Import the dataset "Air_passengers.xlsx", which contains data on the number of passengers that have flew in a specific airplane per month. Now do the following:

a. Summarize the data. What is the minimum and maximum number of passengers who flew in this airplane?
b. Make a histogram using the default *hist()* function. How would you describe the data distribution?
c. Define the number of breaks and choose 5 breaks. The HELP tab can help you here.
d. Change the number of breaks now to 20. Compare this histogram with the one obtained in item b.

Task 4.A - Solution

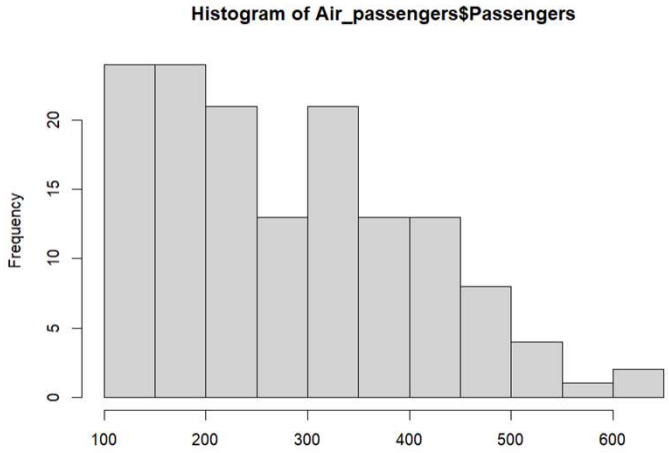
- Solution: I have created a table, where I have shown the commands.

Commands	Results
Minimum `min(Air_passengers$Passengers)`	`> min(Air_passengers$Passengers)` `[1] 104`
Maximum `max(Air_passengers$Passengers)`	`> max(Air_passengers$Passengers)` `[1] 622`

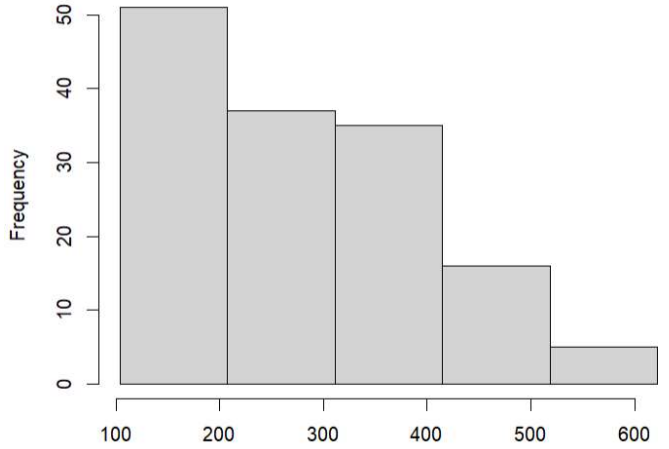
Task 4.B & 4.C - Solution

- Solution: We will use the hist(x=dataframe$column_with_numbers) command for the file.

```
hist(x=Air_passengers$Passengers)
```



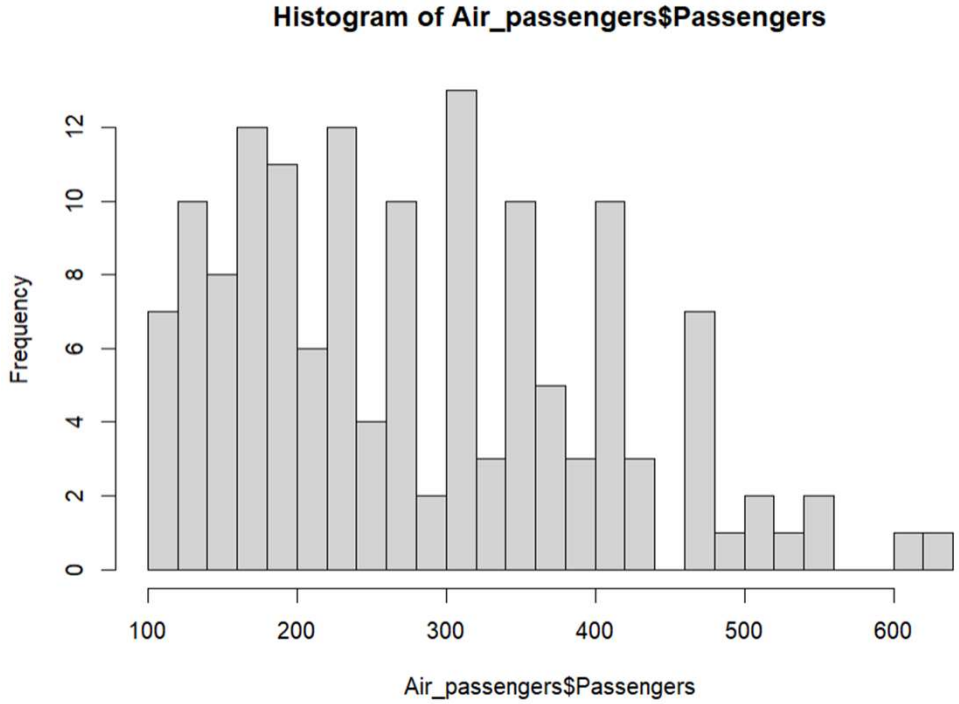

```
Air_passenger_histogram <- hist(Air_passengers$Passengers, breaks=seq(min(Air_passengers$Passengers), max(Air_passengers$Passengers), length.out=6))
```



Task 4.B & 4.C - Solution

- Solution: We will use the hist(x=dataframe$column_with_numbers) command for the file.

```
hist(x=Air_passengers$Passengers, breaks=20)
```

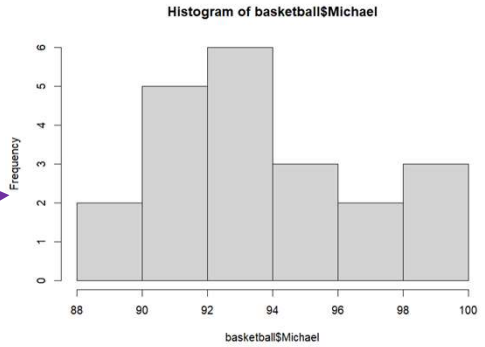


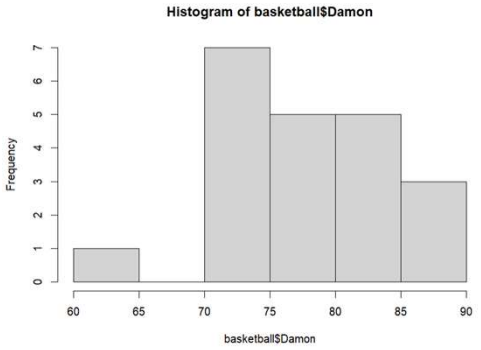
Task 5 - Description

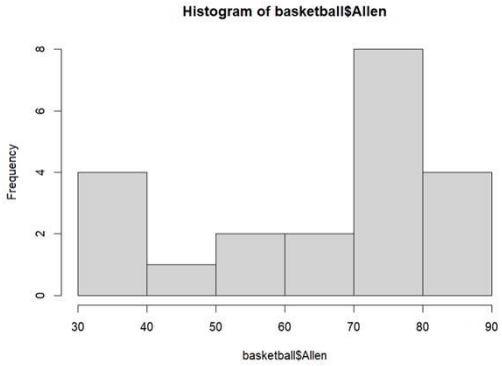
5) Import the dataset "basketball.csv", which contains the scores obtained by three professional basketball players in the pre-season games. Make a boxplot for each of the players. When looking at the boxplots, who seems to be the best player? Can we be sure on this result?

Task 5 - Solution

Persons	Commands
Michael	`hist(x=basketball$Michael)`
Damon	`hist(x=basketball$Damon)`
Allen	`hist(x=basketball$Allen)`







Tak for i dag!

Instruktør: Vivek Misra