Statistical Computing
Master Data Science
Winter Semester 2019/20

BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Prof. Tim Downie

R Test Example version December 2019

(Time allowed: 85 minutes)

Name, Forename:	Matriculation no.:

Please write your answers in the given boxes. If your answer is too long to fit in the box, please continue your answer on your own separate A4 paper, indicating clearly the question number and part.

Preliminaries 1 mark:

You need to download a file which contains the following R-objects. `Reciepts`, `datf` and `datf2`

- Download the file noting which directory the file is stored in.
- Start RStudio.
- If required set your working directory.
- Start with a clean workspace: `> rm(list=objects())`
- load the Data: `> load("TestOneExample.Rda")`
- As a check type: `> loaded()`
  You should get the text `"The data for the test have loaded properly."`

Exercise 1 (6 marks)

The R object `Receipts` is a sample of till receipt totals (in Euros) from a café-Bar.
Enter the R Commands to obtain the relevant statistics from these data and write down the corresponding output.

(a) The sample size

```
> length(Receipts)          (Example solution)          25
```

(b) The median

```
>
```

(c) The mean

```
>
```

(d) The $p = 0.4$-quantile

```
>
```

(e) Give the command to obtain a histogram of the receipts. To get maximum points your histogram should include the following properties:

- The interval breaks should be at €0, 5, 10 , . . . , 50 located.
- The histogram columns should be coloured grey
- A sensible title and axis labels.

```
>
```

Exercise 2 (6 marks)

You have downloaded and installed two data frames called `datf` and `datf2`. Give the *R*-Code and Output for the following:

(a) One command to give the number of rows and variables in `datf`

>

(b) A command which outputs the variable names in `datf`

>

(c) A frequency table for the variable in `datf` called `ddd`

>

(d) A box-plot for `aaa` with a box for each level in `ddd`

>

(e) The mean value of `aaa` for each level of `ddd`

>

(f) A command which appends the columns of `datf2` to the dataframe `datf`. The output is not required.

```
>
```

(g) What property of `datf` and `datf2` is required for the command in part (f)

Exercise 3 (6 marks)

(a) Enter these commands and write the output:

```
> set.seed(100)
> rnorm(1)
```

(b) Let $X_1, \ldots X_{30}$ be a random sample of 30 noramlly distributed $N(10, 5^2)$ random variables. Enter a command to find the probability $P(X_1 \leqslant 8)$. Write the command and the probability.

```
>
```

(c) Enter a command to find the probability $P(\overline{X} \leqslant 8)$. Write the command and the the probability.

```
>
```

(d) Generate a sample of 30 random numbers with a Normal distribution with mean 10 and standard deviation 5. Assign the output to an object called `randsamp`.

```
>
```

(e) Enter a command to find the proportion of elements in `randsamp` less than or equal to 8

```
>
```

Finishing off (1 mark)

- Add a comment to the *top* of your script file with your name and matriculation number.
- Save your script file.
- Go to the Moodle section *Test script files* and upload your file. If you have any problems with uploading your results, please notify the lecturer.