Exercises for the lecture

Mathematics III

Summer Semester 2023

University of Bremen

Institute for Statistics

Daniel Ochieng
MZH, Room 7240
E-Mail: dochieng@uni-bremen.de

Solutions to Exercise 10

Tasks

37. **Scale types**

(a) Ordinal scale. An ordinal scale is a measurement scale where data is ordered or ranked in a meaningful way, but the differences between values may not be equal or quantifiable. In the case of the number of semesters, the data is ordered in a sequential manner, indicating the progression of time and educational milestones. For example, the order might be: 1st semester, 2nd semester, 3rd semester, and so on.

(b) Nominal scale. In nominal scales, numbers, such as driver's license numbers and product serial numbers, are used to name or identify people, objects, or events.

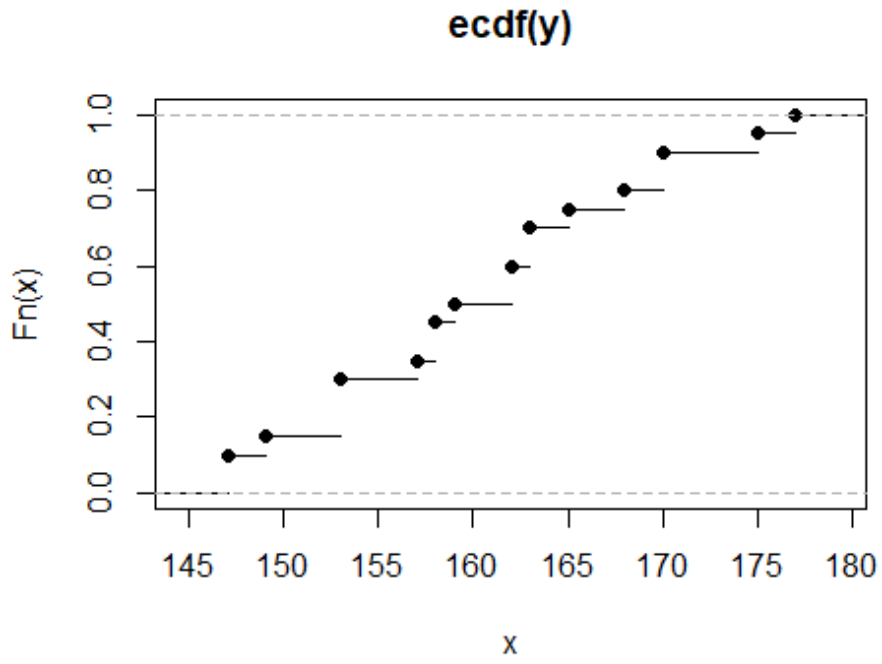
(c) Interval or ordinal. If the grades are on a numerical scale (e.g., 0-100), then the average grade would also be on a numerical scale. This would be an interval scale, as the differences between grades have meaningful quantitative interpretations. For example, the difference between a grade of 80 and 90 represents a higher level of achievement. However, if the grades are on an ordinal scale (e.g., A, B, C, D, F), where the grades represent ordered categories without specific numerical values, the average grade would be on an ordinal scale as well. In this case, the average represents the midpoint of the ordered categories and does not have a precise numerical interpretation.

(d) Ratio scale. It possesses all the properties of an interval scale, such as having meaningful intervals between values. Additionally, a ratio scale includes an absolute zero point, which represents the absence of the measured attribute. In the case of disposable monthly net income, a value of zero euros and cents would indicate no income. On a ratio scale, you can perform various arithmetic operations, such as addition, subtraction, multiplication, and division, and these operations hold meaningful interpretations. For example, you can accurately compare the income levels of individuals, calculate percentage increases or decreases, and compute ratios between incomes.

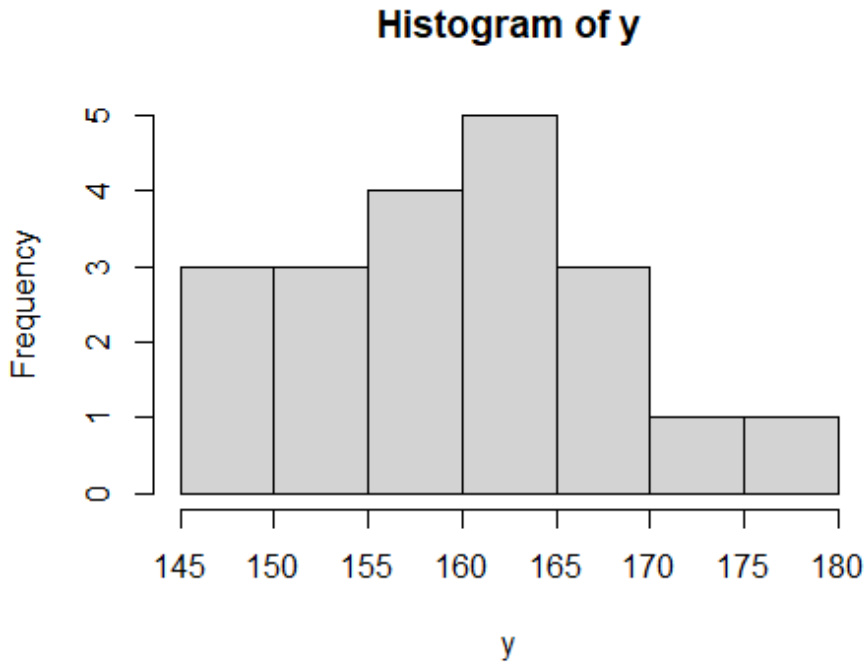
38. **Descriptive Statistics.**

Task 38

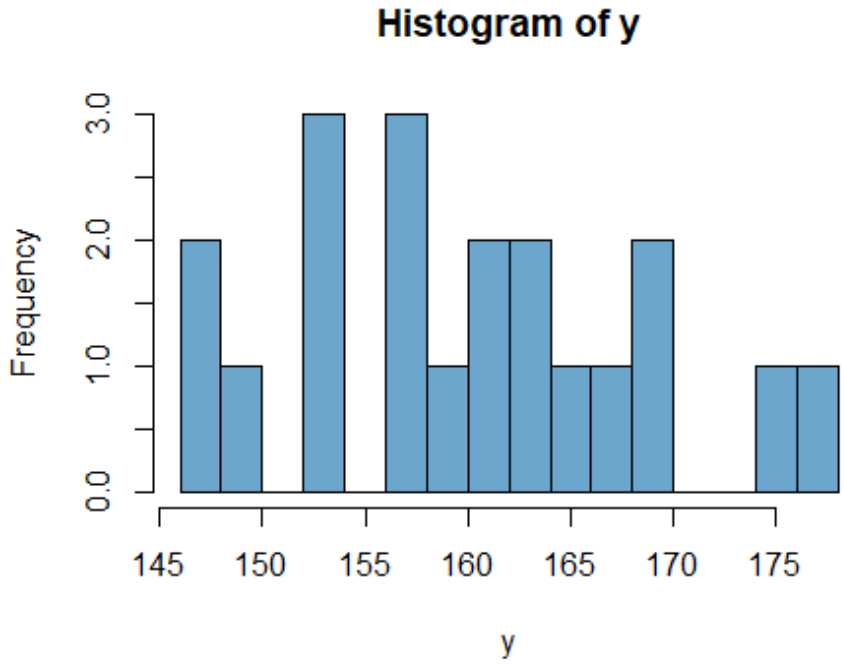
```r
y=c(149, 147, 158, 165, 153, 153, 168, 158, 163, 159, 177, 175, 163,170, 162,
162, 170, 153, 147, 157)

# (a)

plot(ecdf(y))
```




```r
# (b)

hist(y)
```




```r
# (c)

hist(y, col = 'skyblue3', breaks=14)
```



```r
# (d)
#install.packages("moments")
library(moments)

## Warning: package 'moments' was built under R version 4.1.3

skewness(y)

## [1] 0.1815547

#the data is positively skewed.
```

39. **Programming task: descriptive Statistics.**

Task 39

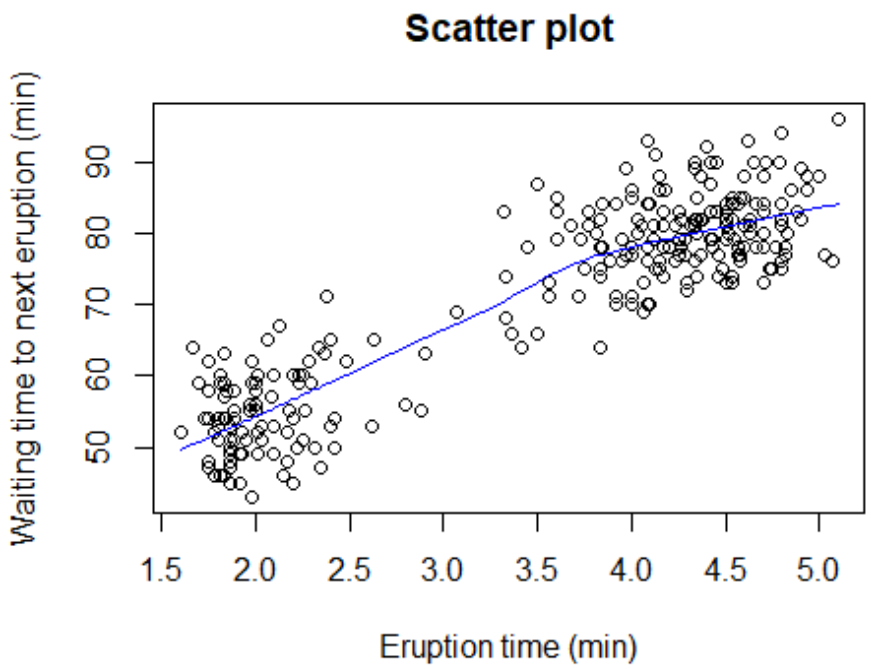
Daniel Ochieng

2023-07-06

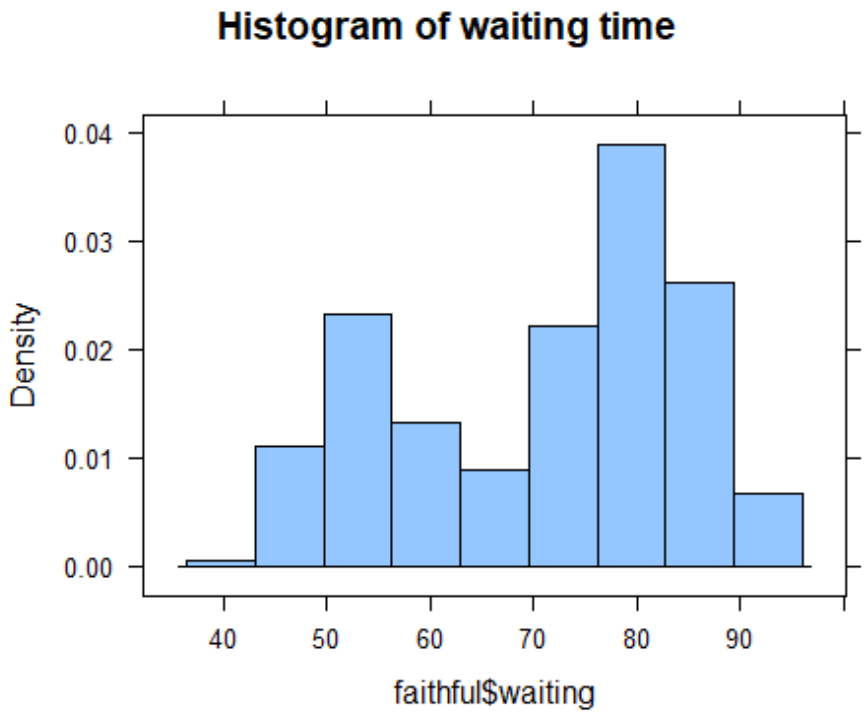
```
## Load required packages

library(stats)
library(graphics)
library(mosaic)

## Warning: package 'mosaic' was built under R version 4.1.3

## Registered S3 method overwritten by 'mosaic':
##   method                           from
##   fortify.SpatialPolygonsDataFrame ggplot2

##
## The 'mosaic' package masks several functions from core packages in order
to add
## additional features.  The original behavior of these functions should not
be affected by this.

##
## Attaching package: 'mosaic'

## The following objects are masked from 'package:dplyr':
##
##     count, do, tally

## The following object is masked from 'package:Matrix':
##
##     mean

## The following object is masked from 'package:ggplot2':
##
##     stat

## The following objects are masked from 'package:stats':
##
##     binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
##     quantile, sd, t.test, var

## The following objects are masked from 'package:base':
##
##     max, mean, min, prod, range, sample, sum

library(psych)
```

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:mosaic':
##
##     logit, rescale

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

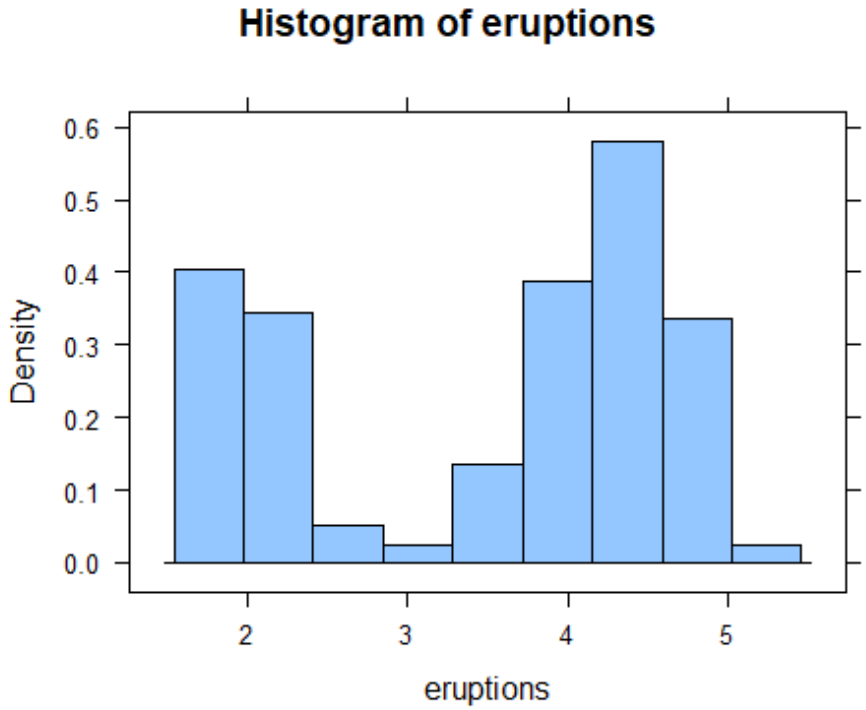
```
## scatter plots

plot(faithful[, -3], main ="Scatter plot",
     xlab = "Eruption time (min)",
    ylab = "Waiting time to next eruption (min)")
lines(lowess(faithful$eruptions, faithful$waiting, f = 2/3, iter = 3),
     col = "blue")
```



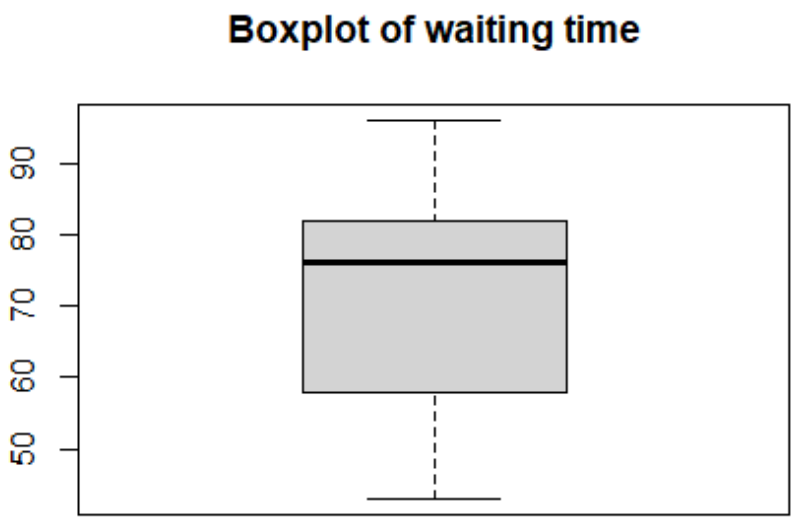

```
## Histograms

histogram(faithful$waiting,main="Histogram of waiting time")
```



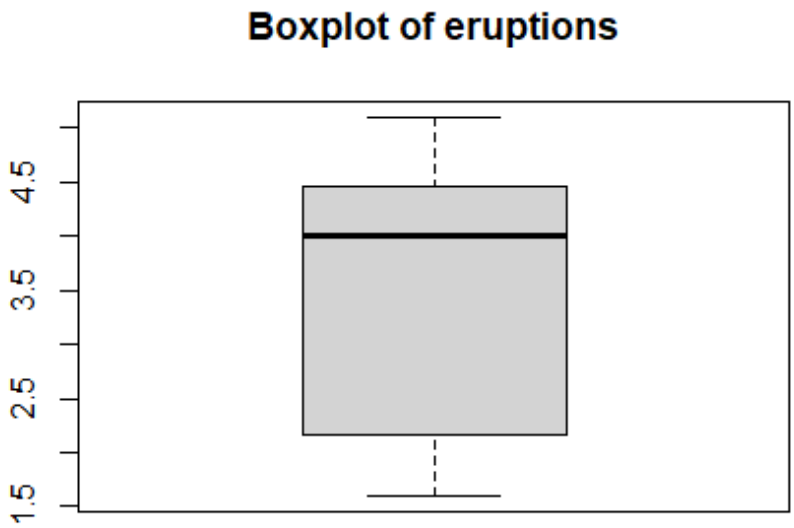

```r
histogram(~eruptions, data = faithful,main="Histogram of eruptions")
```



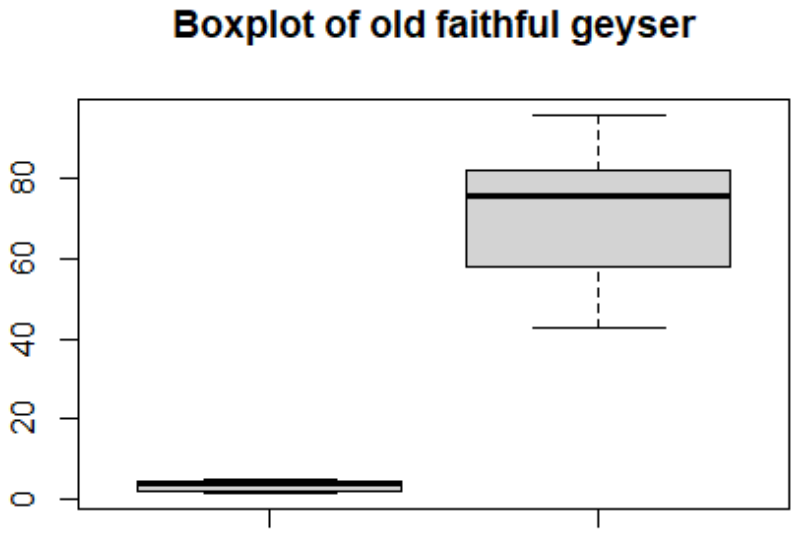
```r
## boxplots
boxplot(faithful$waiting,main="Boxplot of waiting time")
```




```r
boxplot(faithful$eruptions,main="Boxplot of eruptions")
```




```r
boxplot(faithful$eruptions,faithful$waiting,main="Boxplot of old faithful geyser")
```



```r
## Stem and leaf plot
stem(faithful$eruptions)
```

```
##
##   The decimal point is 1 digit(s) to the left of the |
##
##   16 | 070355555588
##   18 | 000022233333335577777777888822335777888
##   20 | 00002223378800035778
##   22 | 0002335578023578
##   24 | 00228
##   26 | 23
##   28 | 080
##   30 | 7
##   32 | 2337
##   34 | 250077
##   36 | 0000823577
##   38 | 2333335582225577
##   40 | 000000335778888800223355557777 8
##   42 | 033355557788002333335555777 8
##   44 | 022223355577800000000233333577788 8
##   46 | 00002333577000000235 78
##   48 | 00000022335800333
##   50 | 0370
```

```r
stem(faithful$waiting)
```

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   4 | 3
##   4 | 55566666777788899999
##   5 | 000001111112222233333334444444 4
##   5 | 55555566667778888999999
##   6 | 00000022223334444
##   6 | 555667899
##   7 | 000011111233333344444 4
##   7 | 55555555666666666777777777777888888888888888999999999
##   8 | 000000001111111111111222222222222333333333333334444444444
##   8 | 5555556666667888888999
##   9 | 00000012334
##   9 | 6
```

```r
## Summary statistics
summary(faithful$eruptions)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.600   2.163   4.000   3.488   4.454   5.100
```

```r
summary(faithful$waiting)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    43.0    58.0    76.0    70.9    82.0    96.0
```

```
describe(faithful$eruptions)
```

```
##    vars   n mean   sd median trimmed  mad min max range  skew kurtosis
se
## X1    1 272 3.49 1.14      4    3.53 0.95 1.6 5.1   3.5 -0.41    -1.51
0.07
```

```
describe(faithful$waiting)
```

```
##    vars   n mean    sd median trimmed   mad min max range  skew kurtosis
se
## X1    1 272 70.9 13.59     76    71.5 11.86  43  96    53 -0.41    -1.16
0.82
```

40. **Multiple select task**

(a) TRUE. The accuracy of the empirical distribution function relies on the assumption of independent and identically distributed (i.i.d.) observations. When observations are dependent, the empirical distribution function may not accurately capture the underlying distribution.

(b) TRUE.It implies that these variables have categorical or nominal values, where the numbers assigned to them do not possess any inherent quantitative meaning. For example, if Y_i represents different categories such as $\{\text{red}, \text{blue}, \text{green}\}$, assigning numerical values like $\{1, 2, 3\}$ would not imply any quantitative relationship between the categories.

(c) FALSE. For dichotomous variables, the empirical distribution function F_n represents the proportion of observations that have a value less than or equal to a given threshold. It can be interpreted as the cumulative probability distribution function for the variable.

(d) TRUE. To construct a histogram from F_n, you can divide the range of the interval-scaled variable into bins or intervals and calculate the proportion of observations falling within each bin. This can be done by using the empirical distribution function F_n to estimate the probabilities of the intervals.