

Exercise 10. Introduction to the hypothesis testing, one-sample tests

Michal Béréš, Martina Litschmannová

From interval estimates to hypothesis tests

What is a statistical hypothesis test?

Let's have the following:

- random variable X (for example men's height)
- selection from a random variable (height measurement of 30 men)

Statistical testing of hypotheses decides the validity of statistical statement (hypothesis) based on the data obtained:

- H_0 - null hypotheses
- H_A - alternative hypotheses

For example:

$$H_0: \mu_X = 175$$

$$H_A: \mu_X > 175$$

Since this is a statistical decision, it will always be tied to some level of significance α . We can always reach only 2 different decisions:

- I reject H_0 in favor of H_A
 - this means that I claim that H_0 does not apply
 - this decision is with the maximum error α (significance level, type I error) - this means that we are able to influence the size of this error
- I don't reject H_0
 - this means that I claim that due to the obtained data(selection) it is not possible to reject H_0
 - this decision is with error β (type II error), this error is not directly controllable and depends on the type of test used

How hypothesis tests relate to interval estimates and how the level of significance enters them will be shown in the next section.

Interval estimation and significance level

```
In [1]: data = readxl::read_excel("data/uvod.xlsx")
        head(data)
```

A tibble:

6 × 1

data

<dbl>

75.91

109.00

107.89

102.95

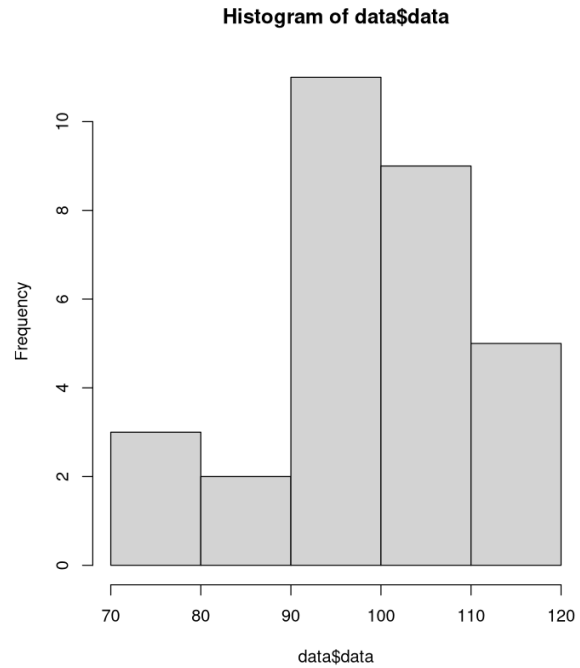
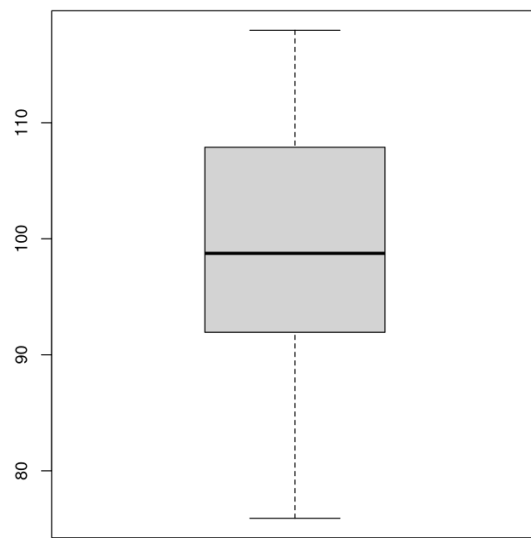
98.93

105.71

```
In [2]: options(repr.plot.width = 12) # width of graphs in Jupyter
```

```
par(mfrow = c(1, 2))           # graph graph matrices 1x2

boxplot(data$data)
hist(data$data)
```



```
In [3]: moments::skewness(data$data)      # oblique
moments::kurtosis(data$data) - 3          # sharpness

shapiro.test(data$data)$p.value           # normality test
```

```
-0.328681869783078
-0.433790248631749
0.537411500830959
```

```
In [4]: length(data$data)
mean(data$data)
sd(data$data)
```

```
30
99.1833333333333
11.3085556056049
```

We make a 95% interval estimate of the mean using a t-test:

```
In [5]: t.test(data$data, alternative = "two.sided", conf.level = 0.95)$conf.int
```

```
94.9606492723303 · 103.406017394336
```

Now imagine that we want to test the hypothesis:

$$H_0: \mu = 100$$

$$H_A: \mu \neq 100$$

What would be the decision with respect to the calculated IO and so the significance level $\alpha = 0.05$?

Well, confidence interval covers the value 100 with maximal error of 5% (confidence 95%). Therefore, we can say we cannot reject this hypothesis at significance value 5%.

Let's further imagine that we want to test the hypothesis:

$$H_0: \mu = 105$$

$$H_A: \mu \neq 105$$

What would be the decision with respect to the calculated IO and so the significance level $\alpha = 0.05$?

Well, now its a different case: confidence interval does not cover the value 105 with maximal error of 5% (confidence 95%). Therefore, we can say we can reject this hypothesis at significance value 5%.

What we just did is called a classic test.

We will show you more classic tests for one-sided alternatives.

$$H_0: \mu = 105$$

$$H_A: \mu > 105$$

```
In [6]: t.test(data$data, alternative = "greater", conf.level = 0.95)$conf.int
```

95.675230152763 · Inf

$$H_0: \mu = 105$$

$$H_A: \mu < 105$$

```
In [7]: t.test(data$data, alternative = "less", conf.level = 0.95)$conf.int
```

-Inf · 102.691436513904

Note that the first of these one-sided alternatives led to a "rejection" of H_0 . This is because of the comparison of the unlikely H_0 with the even less likely H_A .

P-values and connection with CI

An alternative to the classical test(where we create CI) is the so-called pure significance test:

```
In [8]: # H_0: mu=105
# H_A: mu<>105
t.test(data$data, mu = 105, alternative = "two.sided")
```

One Sample t-test

```
data: data$data
t = -2.8173, df = 29, p-value = 0.008629
alternative hypothesis: true mean is not equal to 105
95 percent confidence interval:
 94.96065 103.40602
sample estimates:
mean of x
 99.18333
```

```
In [9]: t.test(data$data, mu = 105, alternative = "two.sided")$p.value
```

0.00862854118186628

The pure significance test results in a p-value. Based on it, we decide whether or not to reject H_0 .

p-value can be understood as the highest possible level of significance, such that our decision is - I do not reject. Thus, the CI/field of acceptance would contain the examined value:

```
In [10]: # H_0: mu=105
# H_A: mu<>105

p.hod = t.test(data$data, mu = 105, alternative = "two.sided")$p.value
p.hod

t.test(data$data, alternative = "two.sided", conf.level = 1 - p.hod)$conf.int
```

0.00862854118186628

93.3666666666667 · 105

```
In [11]: # H_0: mu=105
# H_A: mu>105

p.hod = t.test(data$data, mu = 105, alternative = "greater")$p.value
p.hod

t.test(data$data, alternative = "greater", conf.level = 1 - p.hod)$conf.int
```

0.995685729409067

105 · Inf

In [12]:

```
# H_0: mu=105
# H_A: mu<105

p.hod = t.test(data$data, mu = 105, alternative = "less")$p.value
p.hod

t.test(data$data, alternative = "less", conf.level = 1 - p.hod)$conf.int
```

0.00431427059093314

-Inf · 105

Overview of tests for one sample

Position measures

By position measures we mean the data that determines the position of the data. For data from the normal distribution we can estimate the mean value, for others the median.

a) student's t-test

- we test the mean value
- the data must come from a normal distribution
- exploratory: skewness and sharpness lie in(-2,2)
- exploratory: The QQ graph has points approximately on the line
- exact: using a statistical test, eg Shapiro-Wilk test(shapiro.test(data))

In [13]:

```
# H_0: mu=100
# H_A: mu<>100
t.test(data$data, mu = 100, alternative = 'two.sided')$p.value
```

0.69533398752422

In [14]:

```
# H_0: mu=100
# H_A: mu>100
t.test(data$data, mu = 100, alternative = 'greater')$p.value
```

0.65233300623789

In [15]:

```
# H_0: mu=100
# H_A: mu<100
t.test(data$data, mu = 100, alternative = 'less')$p.value
```

0.34766699376211

b) Wilcoxon test

- we test the median
- the data must come from a symmetric distribution
- exploratory: skewness lies in(-2,2)
- exploratory: histogram looks approximately symmetrical

In [16]:

```
# H_0: X_0.5=100
# H_A: X_0.5<>100
wilcox.test(data$data, mu = 100, alternative = 'two.sided')$p.value
```

0.792159093543887

In [17]:

```
# H_0: X_0.5=100
# H_A: X_0.5>100
wilcox.test(data$data, mu = 100, alternative = 'greater')$p.value
```

0.611715583130717

```
In [18]: # H_0: X_0.5=100
# H_A: X_0.5<100
wilcox.test(data$data, mu = 100, alternative = 'less')$p.value
```

0.396079546771944

c) sign test test

- we test the median
- larger range selection(>10)
- requires "BSDA" library
- as the most robust test, it can also be used for discontinuous data

```
In [19]: # H_0: X_0.5=100
# H_A: X_0.5<>100
BSDA::SIGN.test(data$data, md = 100, alternative = 'two.sided')$p.value
```

0.855535551905632

```
In [20]: # H_0: X_0.5=100
# H_A: X_0.5>100
BSDA::SIGN.test(data$data, md = 100, alternative = 'greater')$p.value
```

0.707667644135654

```
In [21]: # H_0: X_0.5=100
# H_A: X_0.5<100
BSDA::SIGN.test(data$data, md = 100, alternative = 'less')$p.value
```

0.427767775952816

Variability measures

By measures of variability we mean the data determining the dispersion/variability of the data. For data from the normal distribution, we can estimate the standard deviation.

standard deviation test

- we test the standard deviation
- the data must come from a normal distribution
- exploratory: skewness and kurtosis lie in(-2,2)
- Explosive: The QQ graph has points approximately on the line
- exact: using a statistical test, eg Shapiro-Wilk test(shapiro.test(data))
- requires "EnvStats" package
- function in R, compares variance !!!

```
In [22]: # H_0: sigma=10
# H_A: sigma<>10
EnvStats::varTest(data$data, sigma.squared = 10*10,
                  alternative = 'two.sided')$p.value
```

0.288064602223072

```
In [23]: # H_0: sigma=10
# H_A: sigma>10
EnvStats::varTest(data$data, sigma.squared = 10*10,
                  alternative = 'greater')$p.value
```

0.144032301111536

```
In [24]: # H_0: sigma=10
# H_A: sigma<10
EnvStats::varTest(data$data, sigma.squared = 10*10,
                  alternative = 'less')$p.value
```

0.855967698888464

Probability

Test of probability

- We test the probability
- We require sufficient data: $n > \frac{9}{p(1-p)}$
- Clopper's - Pearson's estimate(binom.test)
- does not take data as a parameter, but the number of successes and the number of observation

```
In [25]: pi = 0.3
data_bin = runif(n = 100, min = 0, max = 1) < pi

n = length(data_bin)
x = sum(data_bin)
n
x
```

100

23

```
In [26]: # H_0: pi=0.2
# H_A: pi<>0.2
binom.test(x = x, n = n, p = 0.2, alternative = 'two.sided')$p.value
```

0.453404832266696

```
In [27]: # H_0: pi=0.2
# H_A: pi>0.2
binom.test(x = x, n = n, p = 0.2, alternative = 'greater')$p.value
```

0.261067247669095

```
In [28]: # H_0: pi=0.2
# H_A: pi<0.2
binom.test(x = x, n = n, p = 0.2, alternative = 'less')$p.value
```

0.810912792286201

Examples

```
In [29]: library(dplyr)
library(rstatix)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

Attaching package: 'rstatix'

The following object is masked from 'package:stats':

filter

Example 1.

We have a selection of 216 patients and we measured their protein serum(file testy_jednovyberove.xlsx sheet bilk_serum). Verify that the average protein serum(Albumin) of all patients of this type(population average μ) differs statistically significantly from 35 g/l.

```
In [30]: # Reading data from xlsx file(using readxl package)
albumin = readxl::read_excel("data/testy_jednovyberove.xlsx",
                             sheet = "bilk_serum")
head(albumin)
```

A tibble: 6 × 1

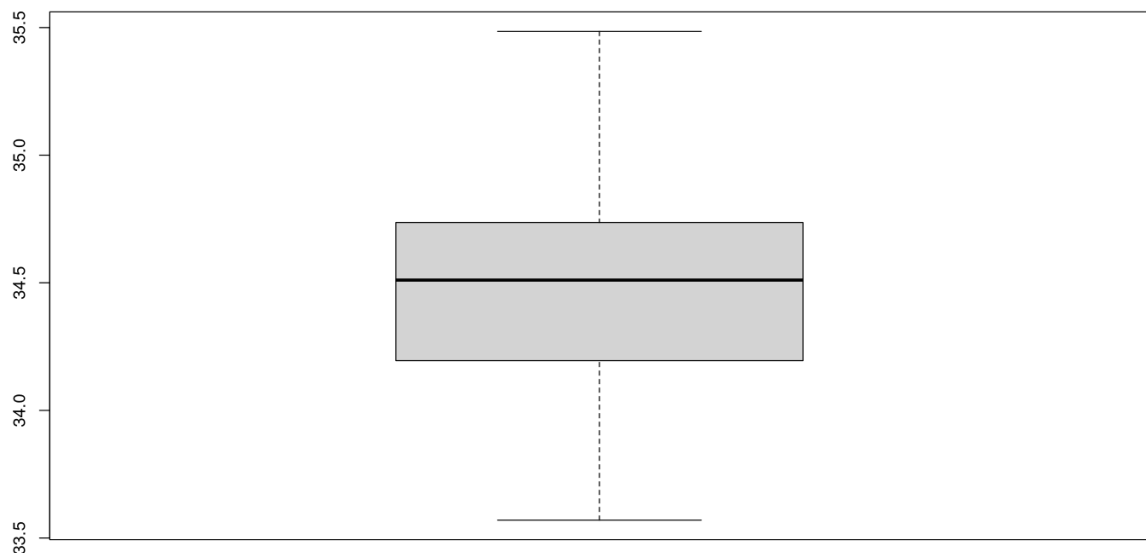
Bílkovinné sérum (g/l)

<dbl>
33.874
34.718
34.957
34.758
35.280
34.621

```
In [31]: colnames(albumin)="value"
```

```
In [32]: # Exploratory analysis
boxplot(albumin$value)
summary(albumin$value)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
33.57	34.20	34.51	34.49	34.74	35.48



```
In [33]: length(albumin$value) # sd is rounded to 3 valid digits
sd(albumin$value)             # sd and position measures are rounded to the nearest thousandth
```

218

0.393682994116148

Position measurement test

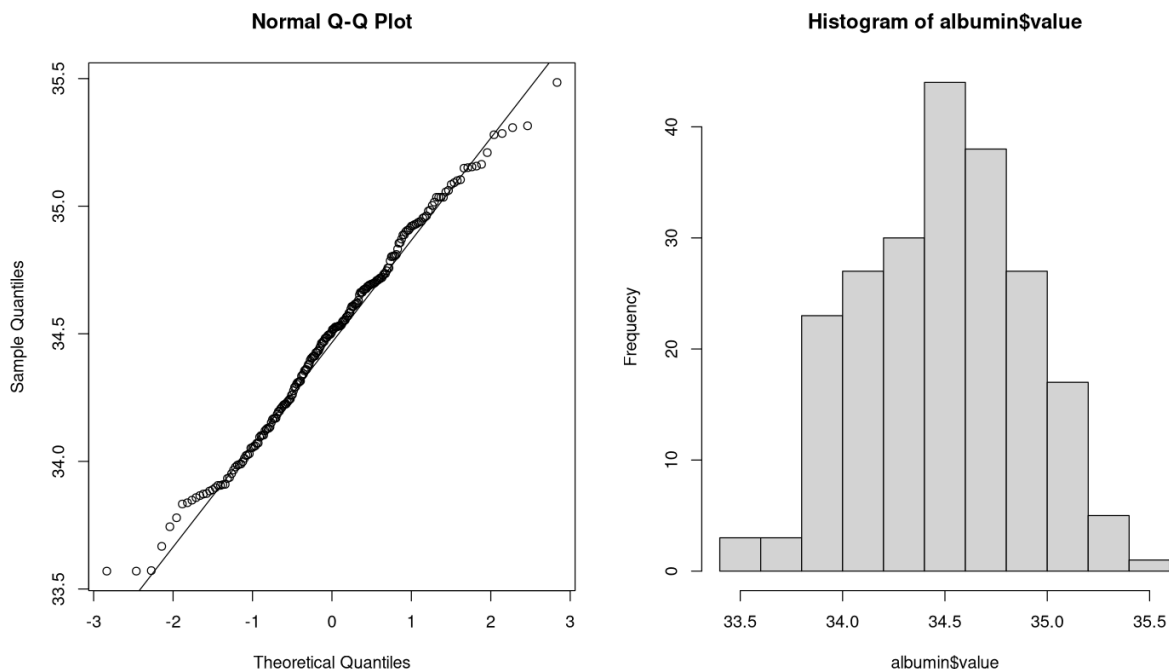
```
In [34]: # Verification of normality - exploratory
moments::skewness(albumin$value) # skew
moments::kurtosis(albumin$value)-3 # sharpness

options(repr.plot.width = 12) # width of graphs in Jupyter
par(mfrow = c(1, 2)) # matrix of 1x2 graphs

qqnorm(albumin$value)
qqline(albumin$value)
hist(albumin$value)
```

-0.0281600650055974

-0.572023254270734



```
In [35]: # We will use the normality test for the final decision on data normality.

# The presumption of normality is verified by the Shapir - Wilk test.
# H0: Data is a selection from the normal distribution.
# Ha: Data is not a selection from the normal distribution.
shapiro.test(albumin$value)
# p-value>0.05 ->Na h1. significance of 0.05, the assumption of normality cannot be rejected.
```

Shapiro-Wilk normality test

```
data: albumin$value
W = 0.99151, p-value = 0.2358
```

```
In [36]: # normal OK ->t.test

# H0: mu=35 g/l
# Ha: mu<>35 g/l

t.test(albumin$value, mu=35, alternative = "two.sided")

# p-value<0.05 ->at significance level of 0.05 we reject the null hypothesis
# in favor of the alternative hypothesis
# The mean albumin value differs statistically significantly from 35 g/l.
```

One Sample t-test

```
data: albumin$value
t = -19.249, df = 217, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 35
95 percent confidence interval:
 34.43421 34.53932
sample estimates:
mean of x
 34.48677
```


Example 2.

Survival times for 100 lung cancer patients treated with the new drug are listed in the tests_jednovyberove.xlsx sheet "preziti". It is known from previous studies that the average survival of such patients without the administration of a new drug is 22.2 months. Can these data suggest that the new drug prolongs survival?

```
In [37]: # Reading data from xlsx file(using readxl package)
preziti = readxl::read_excel("data/testy_jednovyberove.xlsx",
                             sheet = "preziti")
head(preziti)
```

A tibble: 6 × 1

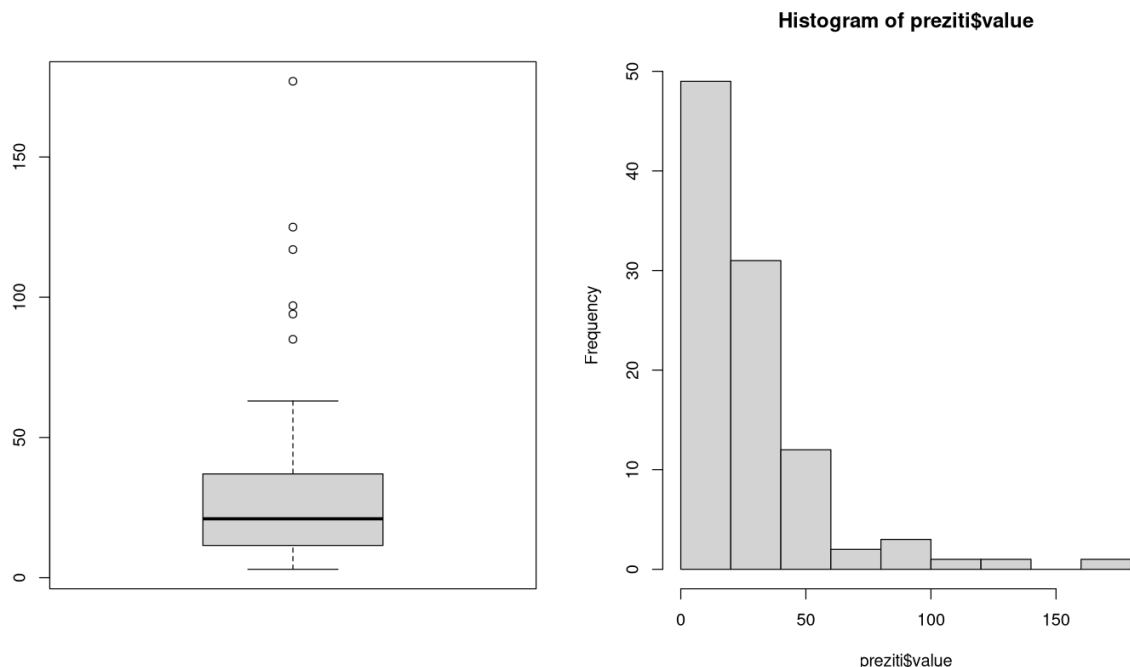
Doba přežití (měsíc)

<dbl>
27
6
9
62
10
10

```
In [38]: colnames(preziti)="value"
```

```
In [39]: ## Exploratory analysis
par(mfrow = c(1, 2)) # graph matrix 1x2

boxplot(preziti$value)
hist(preziti$value)
```



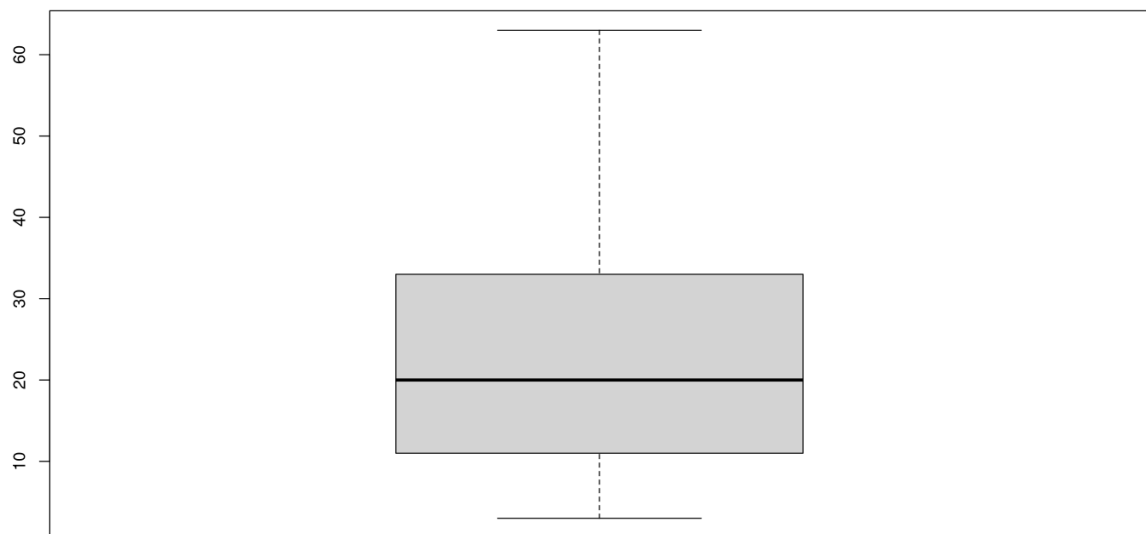
Data contains outliers -> we can delete them. Or note that this is probably an exponential distribution and the outliers are not actually there(the distribution simply behaves this way.)

```
In [40]: # Data contains outliers. We can list them with the help of f-cv boxplot.
preziti$ID = seq(1,length(preziti$value))
outliers = preziti %>% identify_outliers(value)
outliers
# if we decided to remove outliers, then
preziti$value_no_outliers = ifelse(preziti$ID %in% outliers$ID,NA,preziti$value)
```

A tibble: 6 × 4

value	ID	is.outlier	is.extreme
<dbl>	<int>	<lgl>	<lgl>
177	12	TRUE	TRUE
125	70	TRUE	TRUE
85	77	TRUE	FALSE
94	82	TRUE	FALSE
117	86	TRUE	TRUE
97	89	TRUE	FALSE

```
In [41]: # Exploratory analysis for data without remote observations
boxplot(preziti$value_no_outliers)
```



```
In [42]: length(na.omit(preziti$value_no_outliers)) # sd is rounded to 3 valid digits
sd(preziti$value_no_outliers, na.rm=TRUE) # sd and position measurements round. to tenths
```

94

15.3439629732825

Position measure(mean/median) test

```
In [43]: # Verification of normality - exploratory
moments::skewness(preziti$value_no_outliers, na.rm=TRUE)
moments::kurtosis(preziti$value_no_outliers, na.rm=TRUE)-3

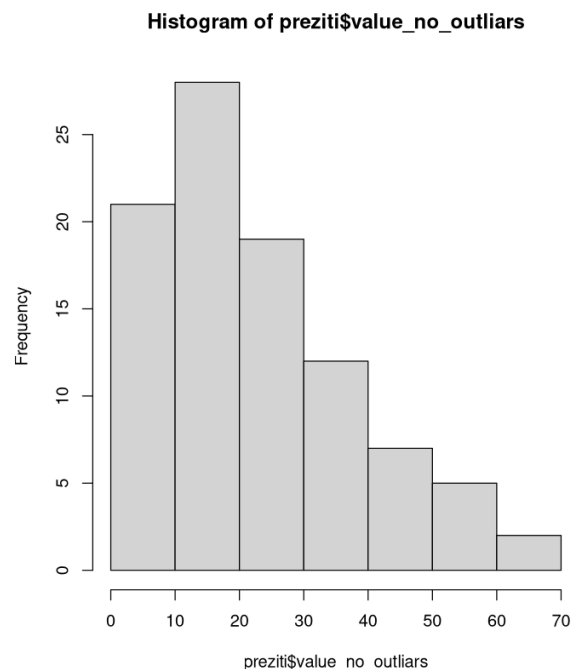
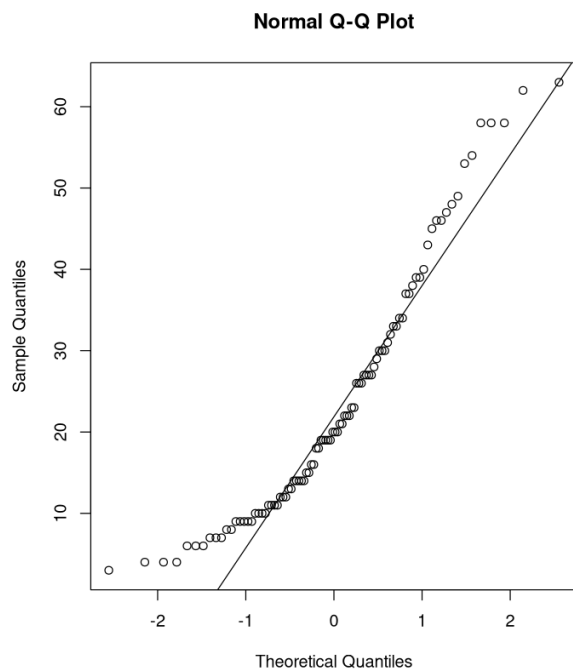
par(mfrow = c(1, 2)) # graph matrix 1x2

qqnorm(preziti$value_no_outliers)
qqline(preziti$value_no_outliers)
hist(preziti$value_no_outliers)

# QQ - graph and history show that the choice of truth. is not a choice of standards. distribution.
# Slanting and pointing corresponds to standards. distribution.
# we will use the normality test.
```

0.833022679217481

-0.174802155213883



```
In [44]: # We verify the assumption of normality by the Shapiro-Wilk's test.
shapiro.test(preziti$value_no_outliers)
# p-value < 0.05 -> at significance 0.05, we reject the assumption of normality
```

Shapiro-Wilk normality test

```
data: preziti$value_no_outliers
W = 0.91835, p-value = 2.021e-05
```

```
In [45]: # exploratory assessment of symmetry - exponential distribution - no symmetry
```

```
In [46]: # normality rejected -> symmetry rejected -> Sign. test
# H0: median=22.2 months
# Ha: median > 22.2 months

BSDA::SIGN.test(preziti$value_no_outliers, md=22.2,
                 alternative="greater")

# p-value > 0.05 -> at significance of 0.05, the null hypothesis cannot be rejected
# Median survival time is not statistically significantly greater than 22.2 months.
```

One-sample Sign-Test

```
data: preziti$value_no_outliers
s = 40, p-value = 0.9393
alternative hypothesis: true median is greater than 22.2
95 percent confidence interval:
 16.99182      Inf
sample estimates:
median of x
      20
```

Achieved and Interpolated Confidence Intervals:

	Conf.Level	L.E.pt	U.E.pt
Lower Achieved CI	0.9393	18.0000	Inf
Interpolated CI	0.9500	16.9918	Inf
Upper Achieved CI	0.9605	16.0000	Inf

```
In [47]: median(preziti$value_no_outliers, na.rm = TRUE)
```

The machine produces piston rings of a given diameter. The manufacturer states that the standard deviation of the ring diameter is 0.05 mm. To verify this information, 80 rings were randomly selected and their diameter was measured (file `testy_jednovyberove.xlsx` sheet `krouzky`). Can the results obtained be considered statistically significant in terms of improving the quality of production?

```
In [48]: # Reading data from xlsx file (using readxl package)
krouzky = readxl::read_excel("data/testy_jednovyberove.xlsx",
                             sheet = "krouzky")
head(krouzky)
```

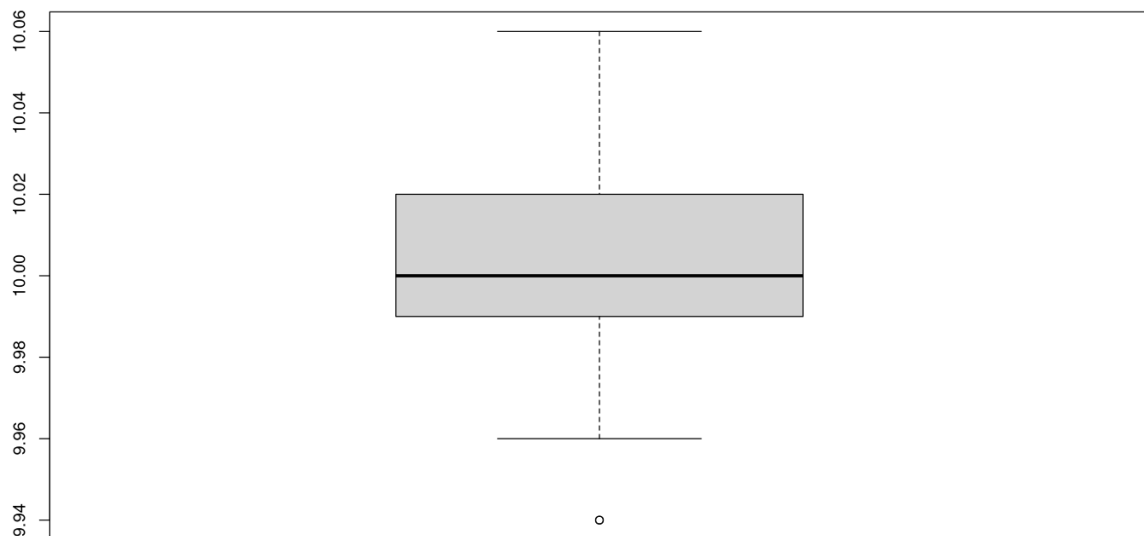
A tibble: 6 × 1

Průměr kroužku (mm)

<dbl>
10.00
10.03
9.96
10.00
9.98
10.02

```
In [49]: colnames(krouzky) = "value"
```

```
In [50]: ## Exploratory analysis
boxplot(krouzky$value)
```



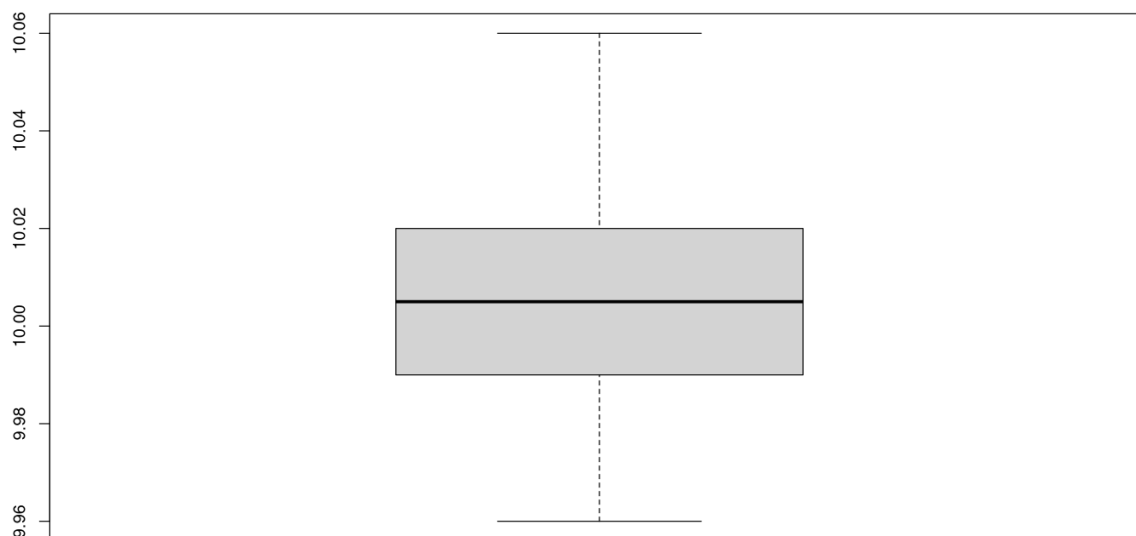
```
In [51]: # Data contains outliers. We can list them with the help of f-ec boxplot.
krouzky$ID = seq(1,length(krouzky$value))
outliers = krouzky %>% identify_outliers(value)
outliers
# if we decided to remove outliers, then
krouzky$value_no_outliers = ifelse(krouzky$ID %in% outliers$ID, NA, krouzky$value)
```

A tibble: 2 × 4

value	ID	is.outlier	is.extreme
<dbl>	<int>	<lgl>	<lgl>
9.94	30	TRUE	FALSE
9.94	41	TRUE	FALSE

```
In [52]: # Exploratory analysis for data without remote observations
summary(krouzky$value_no_outliers, na.rm=TRUE)
boxplot(krouzky$value_no_outliers)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
9.96	9.99	10.01	10.01	10.02	10.06	2



```
In [53]: length(na.omit(krouzky$value_no_outliers))# sd is rounded to 3 valid digits
sd(krouzky$value_no_outliers, na.rm=TRUE) # sd and position measures round. per thousandths
```

78

0.0248461801356396

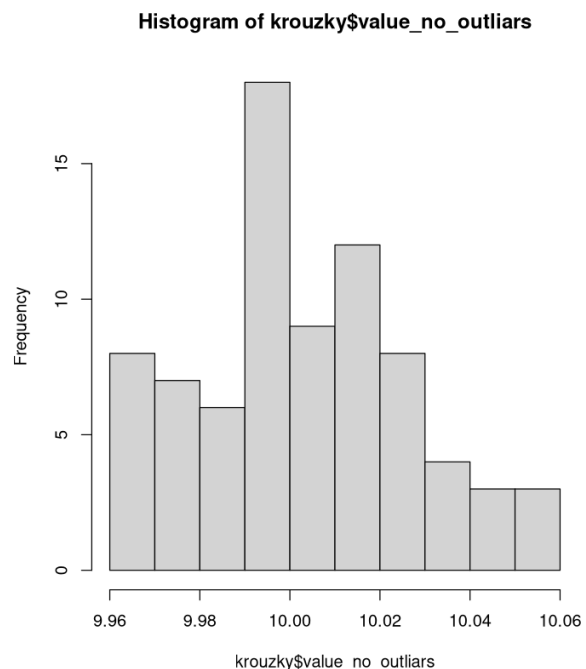
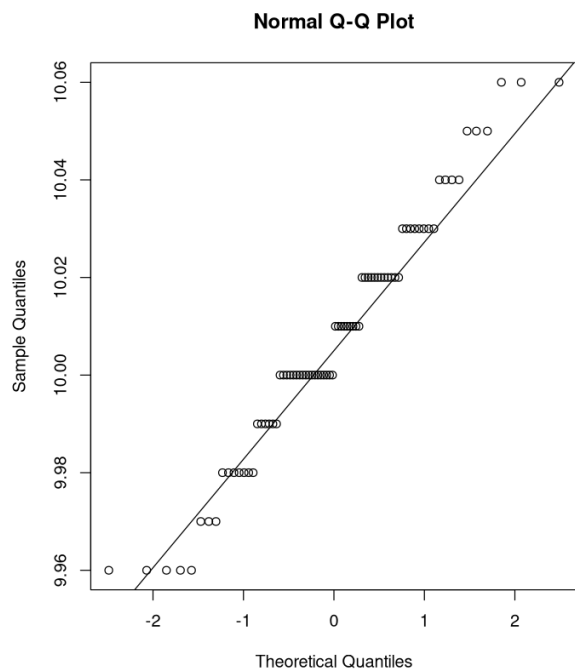
```
In [54]: # Verification of normality - exploratory
moments::skewness(krouzky$value_no_outliers, na.rm=TRUE)
moments::kurtosis(krouzky$value_no_outliers, na.rm=TRUE)-3

par(mfrow = c(1, 2)) # matrix of 1x2 graphs

qqnorm(krouzky$value_no_outliers)
qqline(krouzky$value_no_outliers)
hist(krouzky$value_no_outliers)
# Both skew and sharpness comply with standards. distribution.
# We will use for the final decision on data normality
```

0.0549279162578257

-0.396039943230522



```
In [55]: # normality test.
# We verify the assumption of normality by the Shapirs. Wilkov's test.
shapiro.test(krouzky$value_no_outliers)
# p-value>0.05 ->Na hl. significance of 0.05 cannot be assumed norms. reject
```

Shapiro-Wilk normality test

data: krouzky\$value_no_outliers
W = 0.97005, p-value = 0.0628

```
In [56]: # variability test ->variance test

# H0: sigma=0.05 mm
# Ha: sigma<0.05 mm
EnvStats::varTest(krouzky$value_no_outliers, sigma.squared = 0.05^2,
                  alternative = "less")

# p-value<0.05 ->At the significance level of 0.05 we reject H0 in favor of Ha
```

Warning message in is.not.finite.warning(x):
"There were 2 nonfinite values in x : 2 NA's"
Warning message in EnvStats::varTest(krouzky\$value_no_outliers, sigma.squared = 0.05^2, :
"2 observations with NA/NaN/Inf in 'x' removed."

Chi-Squared Test on Variance

data: krouzky\$value_no_outliers
Chi-Squared = 19.014, df = 77, p-value = 1.354e-12
alternative hypothesis: true variance is less than 0.0025
95 percent confidence interval:
0.0000000000 0.0008225911
sample estimates:
variance
0.0006173327

```
In [57]: # How to find a 95% interval standard deviation estimate?
pom = EnvStats::varTest(krouzky$value_no_outliers,sigma.squared = 0.05^2,
                        alternative = "less", conf.level=0.95)

sqrt(pom$conf.int)
```

Warning message in is.not.finite.warning(x):
"There were 2 nonfinite values in x : 2 NA's"
Warning message in EnvStats::varTest(krouzky\$value_no_outliers, sigma.squared = 0.05^2, :
"2 observations with NA/NaN/Inf in 'x' removed."

LCL: 0 UCL: 0.0286808486207514

Example 4.

TT states that 1% of their resistors do not meet the required criteria. 15 unsuitable resistors were found in the tested delivery of 1000 pieces. Is this result with agreement with TT's assertion or can we reject it?

```
In [58]: n = 1000 # selection range
x = 15 # number of "successes"
p = x/n # relative frequency(probability point estimate)
p
```

0.015

```
In [59]: # Verification of assumptions
9/(p*(1-p))
```

609.137055837563

```
In [60]: # Clopper - Pearson(exact) test
# H0: pi=0.01
# Ha: pi<>0.01

binom.test(x = x, n= n, p = 0.01, alternative="two.sided")
```

Exact binomial test

```
data: x and n
number of successes = 15, number of trials = 1000, p-value = 0.1111
alternative hypothesis: true probability of success is not equal to 0.01
95 percent confidence interval:
 0.00841905 0.02461970
sample estimates:
probability of success
      0.015
```