Wirtschafts- und Sozialstatistik
Universität Trier

Elements of Statistics

Chapter 9:
Hypothesis testing

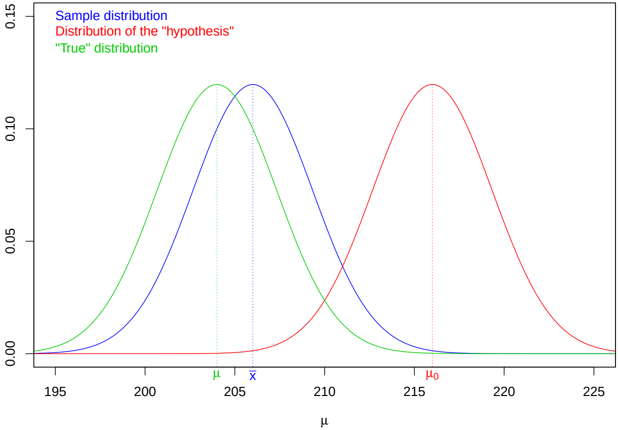
Ralf Münnich, Jan Pablo Burgard and Florian Ertz

University of Trier
Faculty IV
Economic and Social Statistics Department

Winter term 2021/22

Wirtschafts- und Sozialstatistik
Universität Trier

General idea



- Analysis of sample (estimation, e.g. using $\overline{x}$)
- Hypothesis for population (e.g. μ_0 for μ)
- True distribution *unknown* in reality

Hypothesis testing

Wirtschafts- und Sozialstatistik
Universität Trier

- Formulation of a *belief* for the population
  Example: Population is $N(210; 30^2)$-distributed
  $\longrightarrow$ Working hypothesis
- Check if sample results are *compatible* with working hypothesis
- Formulation of a decision rule:
  - right decision regarding working hypothesis
  - wrong decision regarding working hypothesis
- It can only be checked if the observations in the sample are realisations of the distribution postulated in the working hypothesis.
  - Is the sample result very unlikely given the working hypothesis?
  - Is the sample result *plausible* given the working hypothesis?

Example 9.1: Typewriting

Wirtschafts- und Sozialstatistik
Universität Trier

The long-time experience in a two-year typewriting class is that the final 10 minute exam yields test results which are roughly $N(210; 30^2)$-distributed. A new teaching method has been used for the last class and for a sample of $n = 81$ pupils the following values have been recorded: $\bar{x} = 218$ and $s^{*2} = 28^2$.

The working hypothesis is:
The new method does actually improve results.

The null hypothesis is:
The new method does not improve results.

Formally:
$\overline{X} \sim N(210; 30^2)$

We check if the $n = 81$ sample values can be realisations of this distribution.

Example 9.2: Surgery technique

The long-time experience is that in 22.8% of cases complications arise from a severe type of surgery. In a specialised hospital a new technique is studied. Out of the $n = 22$ surgeries already performed, 4 complications resulted $(4/22 = 18.18\%)$.

Can we infer that the new technique is actually to be considered a progress?

Or do we have to consider this result as *random*?

Wirtschafts- und Sozialstatistik
Universität Trier

Example 9.3: Product batches

A certain batch of a product is considered acceptable if the share of scrap is at most 5%. A simple random sample of size $n = 100$ yielded 7 pieces of scrap.

Should we dismiss this batch?

How many elements should be checked to avoid a wrong decision?

One-sample tests

1. Test of localisation
2. Test of deviation
3. Combination of 1. and 2.
4. Test of distribution
   - Distribution law, e.g. normal distribution
   - Distribution including parameters

Example 9.4: Readiness for school

Wirtschafts- und Sozialstatistik
Universität Trier

It is examined if children are ready for school. The $n = 400$ children examined are divided into two subgroups of size $n_1 = 240$ and $n_2 = 160$, respectively, depending on the mother's working status. The following values are observed for a metric test score: $\bar{x}_1 = 66.3$ and $\bar{x}_2 = 78.4$ as well as $s_1^* = 12.8$ and $s_2^* = 13.2$.

Working hypothesis:

a) Working mothers' children reach a different score than non-working mothers' children *on average* (higher/lower).

b) The former group's members' score is *10 points higher* on average.

Two-sample tests

- Comparison of localisation
- Comparison of deviation
- Comparison of distribution

In Example 9.4 we may divide children into k subgroups. We would then speak of a k-sample case from which we abstract here.

Example 9.5: Statistics exam

Wirtschafts- und Sozialstatistik
Universität Trier

$n = 200$ students have to take two statistics exams, a written and an oral one.

In both exams students can reach a maximum of 100 points. The sample results are $\overline{x}_{\mathrm{W}} = 64.3$ with $s_{\mathrm{W}} = 9.8$ for the written exam and $\overline{x}_{\mathrm{O}} = 69.3$ with $s_{\mathrm{O}} = 7.6$ for the oral exam.

It should be checked if the two results are not independent of each other or the results of the oral exam are systematically better than the results of the written exam.

In contrast to Example 9.4 there are not two groups of observations here, but there is rather one group of students with two observations per student which may be presented as a tuple $(x_{i,\mathrm{W}}; x_{i,\mathrm{O}})$. Therefore, we speak of connected samples here, instead of unconnected samples like in Example 9.4.

Types of hypotheses (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Every claim about the distribution of a variable or the relation of variables is called a statistical hypothesis.

We distinguish between working hypotheses and null hypotheses. The user's interest typically lies in the working hypothesis.

In statistical hypothesis testing the null hypothesis takes center stage. A statistical procedure to reach a decision regarding the *compatibility* of sample result and null hypothesis is called statistical test.

Types of hypotheses (2)

Wirtschafts- und Sozialstatistik
Universität Trier

In general, only *statistically noticeable* results of an investigation can indicate a possible incompatibility of sample result and null hypothesis. Therefore, the null hypothesis should (if possible at all) be chosen in such a way that if such an incompatibility occurs and the null hypothesis must consequently be dismissed, the actual *belief* of the investigator (working hypothesis) is substantiated. This can be achieved by reversing the working hypothesis.

If no such incompatibility is found, it does not mean that the null hypothesis is correct. Using the observations of the sample an obvious (significant) contradiction between the sample and the null hypothesis (just) cannot be detected.

Parameter and distribution hypotheses

Wirtschafts- und Sozialstatistik
Universität Trier

Parameter hypothesis

Determination of parameters of a hypothesis (see Example 10.1):

- $H_0 : \mu = \mu_0 = 210$
- $H_0 : \sigma^2 = \sigma_0^2 = 30^2$
- $H_0 : \mu = \mu_0 = 210 \wedge \sigma^2 = \sigma_0^2 = 30^2$

Distribution hypothesis

Determination of the distribution type of a hypothesis (see Example 10.1):

- Fully specified:
  H_0 : The realised variable is $N(210; 30^2)$-distributed.
- Partially specified:
  H_0 : The realised variable is normally distributed.
  $\longrightarrow$ Goodness of fit tests

Point and interval hypotheses

Point hypothesis

Determination of an exact parameter value:
$H_0 : \mu = \mu_0 = 210$

Interval hypothesis

Determination of a value interval:

- One-sided hypothesis (greater than or equal to):
  $H_0 : \mu \geq \mu_0 = 210$
- One-sided hypothesis (less than or equal to):
  $H_0 : \mu \leq \mu_0 = 210$

Alternative hypothesis

Alternative hypothesis

In applications (and because of theoretical considerations) a null hypothesis is always *joined* by an alternative hypothesis. It is specifically called complementary if it is defined as H_0 *does not apply*.

If the null hypothesis is $H_0 : \mu = \mu_0 = 210$ then $H_1 : \mu \geq \mu_0 = 210$ is one of the possible alternative hypotheses and $H_1 : \mu \neq \mu_0 = 210$ is the complementary alternative hypothesis.

The complementary alternative hypothesis to $H_0 : \mu \geq \mu_0 = 210$ is $H_1 : \mu < \mu_0 = 210$.

General approach to hypothesis testing

Wirtschafts- und Sozialstatistik
Universität Trier

Basis: Simple random sample of size n (WR or WOR)

- S is set of all possible sample realisations of the sample vector $(X_1, \ldots, X_n)$
- H_0 assumed to be true
  (at first parameter point hypothesis)
- Decomposition of $S = C \dot{\cup} \overline{C}$ with $P(C)$ small, so that
  - $(x_1, \ldots, x_n) \in C \Rightarrow H_0$ is rejected
  - $(x_1, \ldots, x_n) \in \overline{C} \Rightarrow H_0$ is not rejected

  is used as the basis for the test decision.

$P(C) = \alpha$ is called *probability of error* or *level of significance*.
C is called *critical region*.

The decomposition of S is accomplished using a suitable sample function which is called test statistic. Its distribution under H_0 is called null distribution.

Example 9.6: see Ex. 9.1 (1)

Let X be normally distributed with known variance $\sigma^2 = 30^2$. We want to test

$$H_0 : \mu = \mu_0 = 210$$

using a level of significance of $\alpha = 0.05$.

Under H_0 we have $\overline{X} \sim N\left(210; \left(\frac{30}{9}\right)^2\right)$.

Data input in R:

```
sigma <- 30; mu0 <- 210; alpha <- 0.05; n <- 81
```

Very large or very low realisations of $\overline{X}$ are considered to be unlikely.

Example 9.6: see Ex. 9.1 (2)

Then

$$P\left(\overline{X} < \mu_0 - z\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.025$$

and

$$P\left(\overline{X} > \mu_0 + z\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.025$$

characterise the $\alpha \cdot 100\%$ (here: 5%) of possible values which are *the least plausible* given H_0. Precisely, we have:

$$C = \{(x_1, \ldots, x_n) | \overline{x} < 203.5 \vee \overline{x} > 216.5\} \quad .$$

Calculation of the *borders* of C in R:

```
C <- vector()
C[1] <- mu0 - qnorm(1-alpha/2)*sigma/sqrt(n)
C[2] <- mu0 + qnorm(1-alpha/2)*sigma/sqrt(n)
C
```

```
[1] 203.4668   216.5332
```

Types of error in testing

		Null hypothesis H_0 is	
		not rejected	rejected
H_0 is	true	Right decision	Type 1 error
	false	Type 2 error	Right decision

Type 1 error (α error): A correct hypothesis is wrongly rejected.

Type 2 error (β error): A false hypothesis is wrongly not rejected.

Both errors cannot be simultaneously omitted (their probability *kept low*).

The level of significance α is the supremal probability that the null hypothesis is wrongly rejected – regarding all possible parameter values under H_0.

Example 9.7: see Ex. 9.6 (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Test of $H_0 : \mu = \mu_0 = 210$, with $X \sim N(\mu; 30^2)$, $n = 81$ and $\alpha = 0.05$

a) Probability of a type 1 error:
H_0 is true, but it is rejected.

$$P((x_1, \ldots, x_n) \in C | \mu = 210, \alpha = 0.05) = 0.05$$

Determination of significance level α!

Calculation of the probability for a type 1 error in R:

```
Prob_a <- pnorm(q = C[1], mean = mu0, sd = sigma/sqrt(n)) +
       1 - pnorm(q = C[2], mean = mu0, sd = sigma/sqrt(n))

Prob_a
```

```
[1] 0.05
```

Example 9.7: see Ex. 9.6 (2)

Wirtschafts- und Sozialstatistik
Universität Trier

b) Probability of a type 2 error:
H_0 is false, but it is not rejected.
The answer depends on μ, e.g. $\mu = 216$.

$$P\Big((x_1, \ldots, x_n) \in \overline{C} | \mu = 216, \alpha = 0.05\Big) =$$

$$\Phi\left(216.5|216; \frac{900}{81}\right) - \Phi\left(203.5|216; \frac{900}{81}\right) = 0.56$$

Calculation of the probability for a type 2 error in R:

```
Prob_b <- pnorm(q = C[2], mean = 216, sd = sigma/sqrt(n)) -
           pnorm(q = C[1], mean = 216, sd = sigma/sqrt(n))

round(Prob_b, digits=2)
```

```
[1] 0.56
```

Power function and operating characteristic

Wirtschafts- und Sozialstatistik
Universität Trier

Power function

The power function $\alpha(\pi)$ specifies the probability of rejecting the null, given n, α_0 and H_0, for all possible parameter values π. We have:

$$\alpha(\pi) = \alpha(\pi|n, \alpha_0, H_0) = P\big((X_1, \dots, X_n) \in C|\pi\big)$$

Operating characteristic

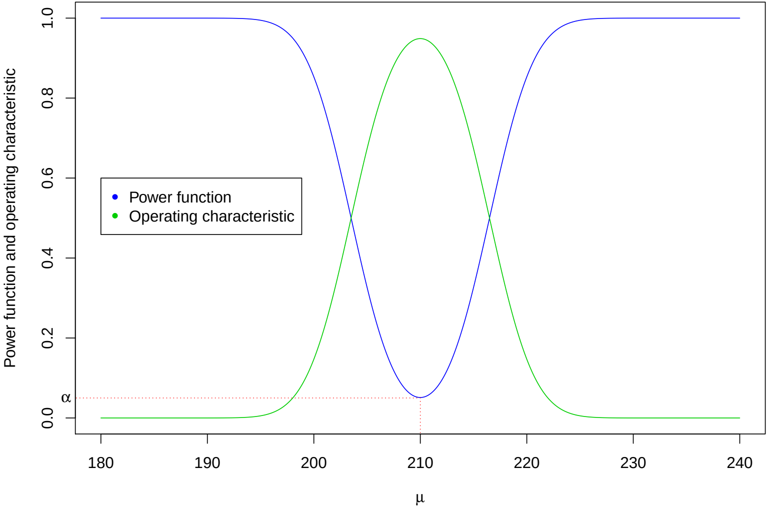
The operating characteristic $\beta(\pi)$ specifies the probability of not rejecting the null, given n, α_0 and H_0, for all possible parameter values π. We have:

$$\beta(\pi) = \beta(\pi|n, \alpha_0, H_0) = P\big((X_1, \dots, X_n) \in \overline{C}|\pi\big)$$

For all possible π the following holds: $\alpha(\pi) + \beta(\pi) = 1$.

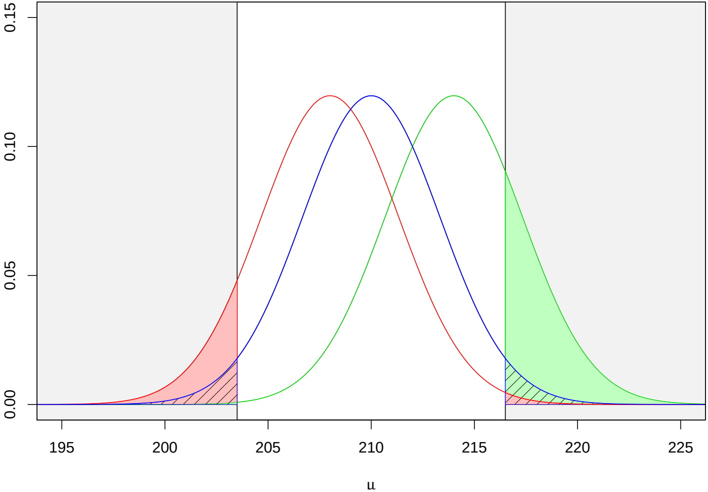
Wirtschafts- und Sozialstatistik
Universität Trier

Power function and operating characteristic $H_0 : \mu = \mu_0 = 210$



Determination of the power function for $H_0 : \mu = \mu_0 = 210$

Wirtschafts- und Sozialstatistik
Universität Trier



Example 9.8: One-sided hypothesis

Contrary to Example 10.7, now the null is $H_0 : \mu \geq \mu_0 = 210$ (one-sided hypothesis; greater than or equal to). The critical region of the sample distribution $\overline{X} \sim N\left(\mu; \frac{100}{9}\right)$ for $\mu \geq 210$ is to be determined. The significance level is at its maximum exactly at $\mu = 210$ and should not surpass the given value of $\alpha = 0.05$. Using

$$
\begin{aligned}
&P\left(\overline{X} < \mu_0 - z(1-\alpha) \cdot \frac{\sigma}{\sqrt{n}}\right) \\
&\quad = P\left(\overline{X} < 210 - 1.64 \cdot \frac{10}{3}\right) = P\left(\overline{X} < 204.5\right) = 0.05
\end{aligned}
$$

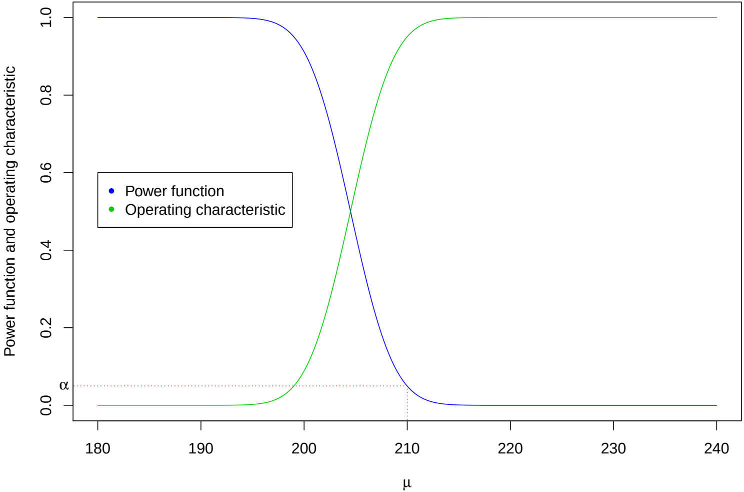
we get the critical region

$$C = \left\{(x_1, \ldots, x_n) \middle| \overline{x} < 204.5\right\}$$

The probability for a type 1 error is *at most* 5%.
Analoguously, for the complementary null hypothesis $H_0 : \mu \leq \mu_0 = 210$ we have: $C = \left\{(x_1, \ldots, x_n) \middle| \overline{x} > 215.5\right\}$.

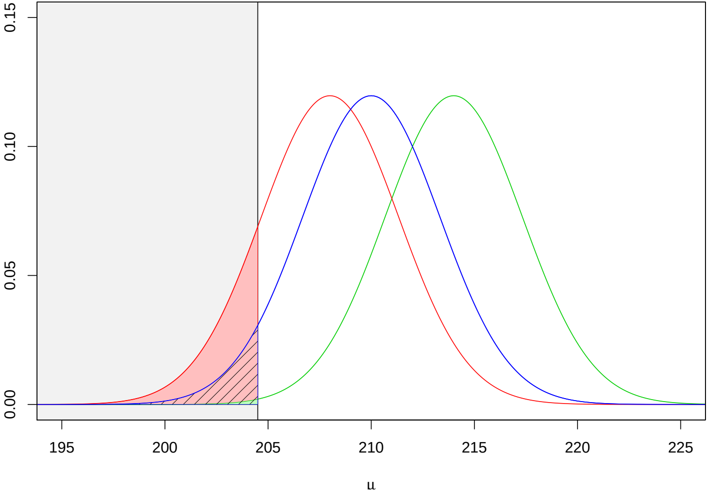
Wirtschafts- und Sozialstatistik
Universität Trier

Power function and operating characteristic $H_0 : \mu \geq \mu_0 = 210$



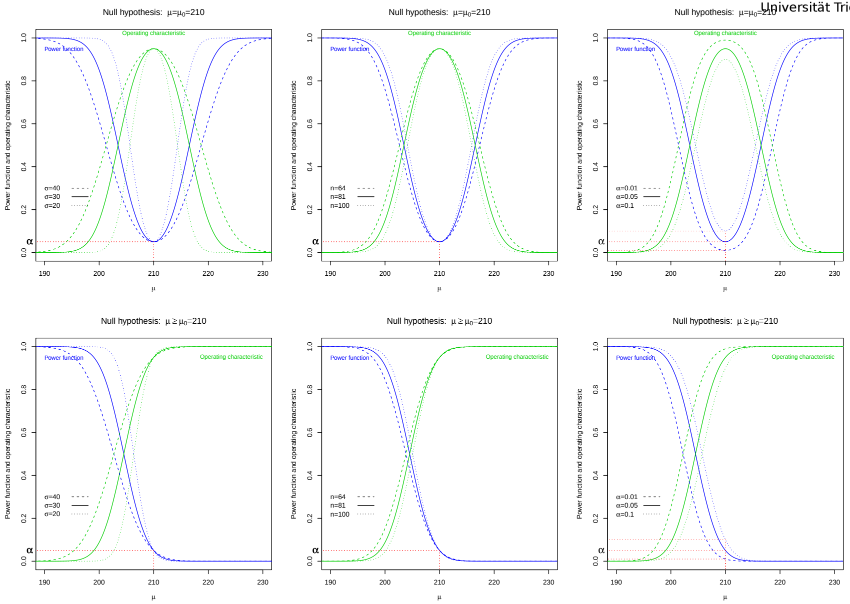
Determination of the power function for $H_0 : \mu \geq \mu_0 = 210$

Wirtschafts- und Sozialstatistik
Universität Trier



Influence of n, σ^2 and α on PF and OC

Wirtschafts- und Sozialstatistik
Universität Trier



General approach of hypothesis testing

Wirtschafts- und Sozialstatistik
Universität Trier

1. Postulation of working hypothesis and, consequently, postulation of
   - null hypothesis H_0 and of
   - alternative hypothesis H_1
2. Setting of significance level α
3. Statement/Calculation of test statistic
4. Setting of critical region C
5. Implementation of test and interpretation of results

In practice, the null hypothesis is derived by negation of the working hypothesis. Therefore, a rejection of the null hypothesis lends support the the working hypothesis. Such a negation is only possible for one-sided hypotheses.

Example 9.9: Clinical trial (1)

Wirtschafts- und Sozialstatistik
Universität Trier

In a clinical trial after oral intake of an antibiotic a concentration of approx. 8 $\mu g/l$ of the agent was measured at the centre of inflammation. A doctor assumes that under *normal* conditions at most 5 $\mu g/l$ are plausible. A control study with $n = 100$ patients yielded $\overline{x} = 3$ $\mu g/l$ and $s^2 = 25$ $(\mu g/l)^2$.
Test the doctor's assumption using $\alpha = 0.05$.

- $H_0 : \mu \geq \mu_0 = 5$ and $H_1 : \mu < \mu_0 = 5$ (negation of working h.)
- $\alpha = 0.05$

Data input in R:

```
SpMean <- 3; SpVar <- 25
mu0 <- 5    ; alpha <- 0.05; n <- 100
```

Example 9.9: Clinical trial (2)

- Test statistic: $\frac{\overline{X} - \mu_0}{S} \cdot \sqrt{n} \sim N(0;1)$ $(n = 100 > 30!)$
- Critical region: $C = \{(x_1, \ldots, x_n) | \frac{\overline{x} - \mu_0}{s} \cdot \sqrt{n} < z(\alpha)\}$
- $\frac{3-5}{5} \cdot \sqrt{100} = -4 < -1.645$, so that $-4 \in C$
  H_0 is rejected and the working hypothesis is statistically supported (at $\alpha = 0.05$).

Test decision in R:

```
Teststat <- (SpMean - mu0)/sqrt(SpVar) * sqrt(n)
c_stat <- qnorm(p = alpha)
```

```
Teststat
```

```
[1] -4
```

```
c_stat
```

```
[1] -1.644854
```

```
Teststat < c_stat
```

```
[1] TRUE
```

p value or *exceeding probability*

Software packages typically report the so called p value (*exceeding probability*) of a test.

The p value indicates the probability that, given H_0, the observed value of the test statistic or a value which is *even more unfavourable* (for the null hypothesis) results.

H_0 is rejected if the p value is smaller than the significance level chosen **in advance**.

Example 9.9: Clinical trial (3)

In Example 9.9 a value of the test statistic of $z = -4$ was calculated. In terms of the null hypothesis all values of the test statistic z' with $z' < -4$ are *worse*. Therefore, the respective p value (not tabulated) is

$$P(Z < -4) = \Phi(-4) = 3.167 \cdot 10^{-5} \quad .$$

As $3.167 \cdot 10^{-5} \ll 0.05$ the null hypothesis has to be rejected.

Determination of the p value in R:

```
p_value <- pnorm(q = Teststat)

p_value
```

```
[1] 3.167124e-05
```

Attention:

Two-sided null hypotheses have a second interval of *more unfavourable* values. Accordingly, the calculated value might have to be doubled.

Wirtschafts- und Sozialstatistik
Universität Trier

Test for μ

We are looking for the null distribution of the standardised random variable $\overline{X}$. Analoguously to estimation procedures we get:

- POP normally distributed, σ^2 known

$$\frac{\overline{X}-\mu_0}{\sigma}\cdot\sqrt{n}\sim N(0;1)$$

- POP normally distributed, σ^2 unknown

$$\frac{\overline{X}-\mu_0}{S}\cdot\sqrt{n}\sim t(n-1)$$

- POP distribution unknown, $\mathrm{Var}\,(X)$ known, $n>30$

$$\frac{\overline{X}-\mu_0}{\sqrt{\mathrm{Var}\,(X)}}\cdot\sqrt{n}\sim N(0;1)$$

- POP distribution unknown, $\mathrm{Var}\,(X)$ unknown, $n>30$, WR or WOR

$$\frac{\overline{X}-\mu_0}{S}\cdot\sqrt{n}\sim N(0;1)\qquad\text{or}\qquad\frac{\overline{X}-\mu_0}{S\cdot\sqrt{\frac{N-n}{N-1}}}\cdot\sqrt{n}\sim N(0;1)$$

Different cases of hypothesis testing for μ

Wirtschafts- und Sozialstatistik
Universität Trier

1. Two-sided problem ($H_0 : \mu = \mu_0$):
   Comparison of test statistic with $\alpha/2$– or $1-\alpha/2$–quantile of null distribution

$$C = \left\{ (x_1, \ldots, x_n) \middle| \frac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{n} < z(\alpha/2) \vee \frac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{n} > z(1-\alpha/2) \right\}$$

2. One-sided problem:
   a) Greater than or equal to ($H_0 : \mu \geq \mu_0$):
      Comparison of test statistic with α–quantile of null distribution

$$C = \left\{ (x_1, \ldots, x_n) \middle| \frac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{n} < z(\alpha) \right\}$$

   b) Less than or equal to ($H_0 : \mu \leq \mu_0$):
      Comparison of test statistic with $1-\alpha$–quantile of null distribution

$$C = \left\{ (x_1, \ldots, x_n) \middle| \frac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{n} > z(1-\alpha) \right\}$$

Example 9.10: Two-sided test (1)

The following $n = 20$ values of a normally distributed random variable are observed: 4.30; 4.30; 3.88; 6.81; 5.17; -0.53; 4.05; 3.34; 6.07; 6.58; 6.94; 0.96; 6.14; 4.20; 8.35; 7.01; 3.39; 3.96; 3.19; 4.21.

The null hypothesis $H_0 : \mu = \mu_0 = 3$ is to be tested using $\alpha = 0.05$. We know that $\sigma^2 = 4$.

- $H_0 : \mu = \mu_0 = 3,\ H_1 : \mu \neq \mu_0 = 3,\ \alpha = 0.05,\ n = 20,\ \sigma^2 = 4$

Data input in R:

```
x9_10 <- c(4.30, 4.30, 3.88, 6.81, 5.17, -0.53, 4.05, 3.34,
           6.07, 6.58, 6.94, 0.96, 6.14, 4.20, 8.35, 7.01,
           3.39, 3.96, 3.19, 4.21)
sigma <- sqrt(4)
mu0 <- 3
alpha <- 0.05
n <- 20
SpMean <- mean(x9_10)
```

Wirtschafts- und Sozialstatistik
Universität Trier

Example 9.10: Two-sided test (2)

- $\dfrac{\overline{X}-\mu_0}{\sigma}\cdot\sqrt{n}\sim N(0;1)$, $z(0.025)=-1.96$, $z(0.975)=1.96$
- $C=\left\{(x_1,\ldots,x_n)\,\middle|\,\dfrac{\overline{x}-\mu_0}{\sigma}\cdot\sqrt{n}<z(\alpha/2)\vee\dfrac{\overline{x}-\mu_0}{\sigma}\cdot\sqrt{n}>z(1-\alpha/2)\right\}$
- As $\overline{x}=4.616$ we have $z=\dfrac{\overline{x}-\mu_0}{\sigma}\cdot\sqrt{n}=\dfrac{4.616-3}{2}\cdot\sqrt{20}=3.613$.
  As $3.613\in C$ $(3.613>1.96)$, H_0 is rejected.

Test decision in R:

```
Teststat <- (SpMean - mu0)/sigma * sqrt(n)
c_stat <- vector()
c_stat[1] <- qnorm(p = alpha/2)
c_stat[2] <- qnorm(p = 1 - alpha/2)

round(Teststat, digits = 3)
```

```
[1] 3.613
```

```
Teststat < c_stat[1] | Teststat > c_stat[2]
```

```
[1] TRUE
```

Example 9.11: see Ex. 9.10 (1)

Now let the distribution of the population and the variance σ^2 be unknown. Once again, we want to test the null hypothesis $H_0 : \mu = \mu_0 = 3$ using a significance level of $\alpha = 0.05$.

- $H_0 : \mu = \mu_0 = 3,\ H_1 : \mu \neq \mu_0 = 3,\ \alpha = 0.05,\ n = 20;$
- $\dfrac{\overline{X} - \mu_0}{S} \cdot \sqrt{n} \sim N(0;1),\ z(0.025) = -1.96,\ z(0.975) = 1.96$
- $C = \left\{ (x_1, \ldots, x_n) \middle| \dfrac{\overline{x} - \mu_0}{s} \cdot \sqrt{n} < z(\alpha/2) \vee \dfrac{\overline{x} - \mu_0}{s} \cdot \sqrt{n} > z(1-\alpha/2) \right\}$
- As $\overline{x} = 4.616$ and $s^2 = 4.4912$ we have
  $z = \dfrac{\overline{x} - \mu_0}{s} \cdot \sqrt{n} = \dfrac{4.616 - 3}{\sqrt{4.4912}} \cdot \sqrt{20} = 3.410.$
  As $3.410 \in C$ $(3.410 > 1.96)$, H_0 is rejected.

Notice that $t(0.975;19) = 2.093 > 1.96 = z(0.975)$.
(t test is used in software packages.)
Attention: $n \not> 30$!
Approximation is still inadmissible according to text book.

Example 9.11 (2): Test of point hypothesis

Application of hypothesis test in R

```
t.test(x9_10,alternative="two.sided",mu=3,conf.level=0.95)
```

```
One Sample t-test

data:  x9_10
t = 3.4102, df = 19, p-value = 0.002936
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 3.624168   5.607832
sample estimates:
mean of x
 4.616
```

The output contains the test statistic $t = 3.4102$ (t-distribution with $df = 19$ degrees of freedom), the p value 0.0029, the point estimate $\overline{x} = 4.616$ and the corresponding confidence interval $[3.624; 5.608]$.

Example 9.11 (3): Test of one-sided hypothesis (greater)

Application of hypothesis test in R

```
t.test(x9_10,alternative="less",mu=3,conf.level=0.95)
```

```
One Sample t-test

data:  x9_10
t = 3.4102, df = 19, p-value = 0.9985
alternative hypothesis: true mean is less than 3
95 percent confidence interval:
 -Inf   5.435393
sample estimates:
mean of x
 4.616
```

The test statistic $t = 3.4102$ has to be compared to $t(0.05; 19) = -1.729$. H_0 is not rejected here, which is illustrated by the p value as well: $0.9985 \gg 0.05$. **Notice: Output states alternative hypothesis!**

Example 9.11 (4): Test of one-sided hypothesis (less)

Wirtschafts- und Sozialstatistik
Universität Trier

Application of hypothesis test in R:

```
t.test(x9_10,alternative="greater",mu=3,conf.level=0.95)
```

```
One Sample t-test

data:  x9_10
t = 3.4102, df = 19, p-value = 0.001468
alternative hypothesis: true mean is greater than 3
95 percent confidence interval:
 3.796607    Inf
sample estimates:
mean of x
 4.616
```

The test statistic $t = 3.4102$ has to be compared to $t(0.95; 19) = 1.729$. H_0 is rejected here, which is illustrated by the p value as well: $0.001468 \ll 0.05$. **Notice: Output states alternative hypothesis!**

Example 9.12: One-sided test (1)

Wirtschafts- und Sozialstatistik
Universität Trier

A sample of size $n = 25$ is drawn from a normally distributed population. The following values have been calculated $\overline{x} = 197.7$ and $s^2 = 42.25$.

Test the null hypothesis $H_0 : \mu \geq \mu_0 = 200$ using $\alpha = 0.05$.

- $H_0 : \mu \geq \mu_0 = 200,\ H_1 : \mu < \mu_0 = 200,\ \alpha = 0.05,\ n = 25;$
- $\dfrac{\overline{X} - \mu_0}{s} \cdot \sqrt{n} \sim t(n-1),\ t(0.05; 24) = -1.711$
- $C = \left\{(x_1, \ldots, x_n) \middle| \dfrac{\overline{x} - \mu_0}{s} \cdot \sqrt{n} < t(\alpha; n-1)\right\}$
- As $\overline{x} = 197.7$ and $s^2 = 42.25$ we have
  $t = \dfrac{\overline{x} - \mu_0}{s} \cdot \sqrt{n} = \dfrac{197.7 - 200}{\sqrt{42.25}} \cdot \sqrt{25} = -1.769.$
  As $-1.769 \in C$ $(-1.769 < -1.711)$, H_0 is rejected.

Data input in R:

```
load("Example9-12.RData")
```

Example 9.12: One-sided test (2)

Test decision in R:

```
SpMean <- mean(x9_12)
SpVar <- var(x9_12)
mu0 <- 200
alpha <- 0.05
n <- length(x9_12)

t.test(x = x9_12, alternative = "less", mu = mu0,
       conf.level = 1 - alpha)
```

```
        One Sample t-test

data:  x9_12
t = -1.7725, df = 24, p-value = 0.0445
alternative hypothesis: true mean is less than 200
95 percent confidence interval:
     -Inf 199.9199
sample estimates:
mean of x
 197.6956
```

Example 9.13: see Ex. 9.12 (1)

Test the null hypothesis $H_0 : \sigma^2 \leq \sigma_0^2 = 30$ using $\alpha = 0.05$.

- $H_0 : \sigma^2 \leq \sigma_0^2 = 30,\ H_1 : \sigma^2 > \sigma_0^2 = 30,\ \alpha = 0.05,\ n = 25;$

Data input in R:

```
sigmaq0 <- 30
```

Example 9.13: see Ex. 9.12 (2)

- $\frac{n-1}{\sigma_0^2} \cdot s^2 \sim \chi^2(n-1)$, $\chi^2(0.95; 24) = 36.415$
- $C = \left\{(x_1, \ldots, x_n) \middle| \frac{n-1}{\sigma_0^2} \cdot s^2 > \chi^2(1-\alpha; n-1)\right\}$
- As $s^2 = 42.25$ we have $\chi^2 = \frac{n-1}{\sigma_0^2} \cdot s^2 = \frac{24}{30} \cdot 42.25 = 33.8$.

  As $33.8 \notin C$ $(33.8 \not> 36.415)$, H_0 is not rejected.

Test decision in R:

```
c_stat <- qchisq(p = 1 - alpha, df = n - 1)
Teststat <- (n-1) / sigmaq0 * SpVar
```

```
c_stat
```

```
[1] 36.41503
```

```
Teststat
```

```
[1] 33.80213
```

```
Teststat > c_stat
```

```
[1] FALSE
```

Wirtschafts- und Sozialstatistik
Universität Trier

Example 9.14: Another urn example

There are red and black balls in an urn. It is postulated that the share of red balls in the urn is $\theta = 0.4$. The following test is performed: $n = 10$ balls are drawn with replacement. H_0 will be rejected if 0, 1, 9, or 10 red balls are drawn. Then

$$\alpha = P(X = 0, 1, 9, 10 | \theta = \theta_0 = 0.4) = \binom{10}{0} \cdot 0.4^0 \cdot 0.6^{10} + \binom{10}{1} \cdot 0.4^1 \cdot 0.6^9 + \binom{10}{9} \cdot 0.4^9 \cdot 0.6^1 + \binom{10}{10} \cdot 0.4^{10} \cdot 0.6^0 = 0.0481$$

is the significance level and

$$\beta(0.5) = 1 - 2 \cdot \left(\binom{10}{0} \cdot 0.5^{10} \cdot 0.5^0 + \binom{10}{1} \cdot 0.5^9 \cdot 0.5^1 \right) = 0.9785$$

is the probability of a type 2 error if $\theta = 0.5$.

Example 9.15: Percentage test (1)

Wirtschafts- und Sozialstatistik
Universität Trier

In a study on poverty in Germany $n = 100$ individuals' income is recorded. A person is considered poor if her income is below the poverty threshold. It is postulated that exactly 25% of the population have to be considered poor. This null hypothesis is tested using $\alpha = 0.05$.

$H_0 : \theta = \theta_0 = 0.25,\ H_1 : \theta \neq \theta_0,\ \alpha = 0.05,\ n = 100$

Data input in R:

```
theta0 <- 0.25; alpha <- 0.05; n <- 100
```

According to de Moivre and Laplace's theorem we have:
$\frac{X - n\cdot\theta_0}{\sqrt{n\cdot\theta_0\cdot(1-\theta_0)}} \sim N(0;1)$.
X is the number of poor people $(n \cdot p)$ in the sample.
We should consider a continuity correction as well.
The approximation conditions hold, as
$100 \cdot 0.25 \cdot 0.75 = 18.75 > 9$ and $0.1 \leq 0.25 \leq 0.9$.

Example 9.15: Percentage test (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Test of approximation conditions in R:

```
n * theta0 * (1 - theta0) > 9
[1] TRUE
0.1 <= theta0 & theta0 <= 0.9
[1] TRUE
```

- $$C = \left\{(x_1, \ldots, x_n) \middle| \frac{n \cdot p + 0.5 - n \cdot \theta_0}{\sqrt{n \cdot \theta_0 \cdot (1 - \theta_0)}} < z(\alpha/2) \right.$$
$$\left. \vee \quad \frac{n \cdot p - 0.5 - n \cdot \theta_0}{\sqrt{n \cdot \theta_0 \cdot (1 - \theta_0)}} > z(1 - \alpha/2) \right\}$$
- The sample yields 30 poor people:

$$z = \frac{30 + 0{,}5 - 25}{\sqrt{100 \cdot 0.25 \cdot 0.75}} = 1.270 < 1.96$$

H_0 is not rejected and the continuity correction is redundant.

Example 9.15: Percentage test (3)

Wirtschafts- und Sozialstatistik
Universität Trier

Hypothesis test in R:

```
p <- 0.3

c_stat <- vector()
c_stat[1] <- qnorm(p = alpha/2)
c_stat[2] <- qnorm(p = 1 - alpha/2)

Teststat <- vector()
Teststat[1] <- (n*p+0.5-n*theta0) / sqrt(n*theta0*(1-theta0))
Teststat[2] <- (n*p-0.5-n*theta0) / sqrt(n*theta0*(1-theta0))

c_stat
```

```
[1] -1.959964  1.959964
```

```
Teststat
```

```
[1] 1.270171 1.039230
```

```
Teststat[1] < c_stat[1] | Teststat[2] > c_stat[2]
```

```
[1] FALSE
```

Wirtschafts- und Sozialstatistik
Universität Trier

Example 9.15: Percentage test (4)

- The sample yields 40 poor people:

$$z = \frac{40 - 0.5 - 25}{\sqrt{100 \cdot 0.25 \cdot 0.75}} = 3.349 > 1.96$$

H_0 is rejected and the continuity correction is redundant.

- The sample yields 33 poor people:

$$z = \frac{33 + 0.5 - 25}{\sqrt{100 \cdot 0.25 \cdot 0.75}} = 1.963 > 1.96$$

$$z = \frac{33 - 0.5 - 25}{\sqrt{100 \cdot 0.25 \cdot 0.75}} = 1.732 < 1.96$$

Here we would make different decisions. In case of doubt we do not reject the null (conservative testing).

Two connected samples

Wirtschafts- und Sozialstatistik
Universität Trier

Observations are pairs of values: $(X_1, X_2) \rightarrow D = X_1 - X_2$.
We finally get to the one-sample case. We have:

$H_0 : \mu_1 - \mu_2 = \delta_0$

Often $\delta_0 = 0$ is of interest (homogeneity hypothesis).

Notice that:

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - \delta_0}{S_D} \cdot \sqrt{n} \sim N(0; 1)$$

We use the central limit theorem (CLT) here
($n > 30$, if X_1 and X_2 are not already normally distributed).

Furthermore, we have $d = \overline{x}_1 - \overline{x}_2$ and $s_d^2 = s_1^2 + s_2^2 - 2 \cdot s_{12}$.

Example 9.16: Weddings (1)

For $n = 50$ weddings, the age of the groom and the age of the bride are recorded (x_{1i}, x_{2i}). The descriptive statistics for the sample are: $\overline{x}_1 = 31.5$; $\overline{x}_2 = 28.5$; $s_1^2 = 27.3$; $s_2^2 = 23.8$ and $s_{12} = 15.6$. We want to test the null that the *typical* bride is at least 5 years younger than the *typical* groom ($\alpha = 0.05$).

- $H_0: \mu_1 - \mu_2 \geq \delta_0 = 5$ vs. $H_1: \mu_1 - \mu_2 < 5$, $\alpha = 0.05$ and $n = 50 > 30$!

Data input in R:

```
SpMean_X1 <- 31.5; SpMean_X2 <- 28.5
SpVar_X1 <- 27.3  ; SpVar_X2 <- 23.8
Cov_X1_X2 <- 15.6

delta0 <- 5; alpha <- 0.05; n <- 50
```

Checking approximation conditions in R:

```
n > 30
```

```
[1] TRUE
```

Example 9.16: Weddings (2)

- $\dfrac{D-\delta_0}{S_d}\cdot\sqrt{n}\sim N(0;1)$, $z(0.05)=-1.645$, $s_d^2=s_1^2+s_2^2-2s_{12}=19.9$
- $C=\left\{(x_1,\ldots,x_n)\middle|\dfrac{d-\delta_0}{s_d}\cdot\sqrt{n}<z(\alpha)\right\}$
- We have $z=\dfrac{d-\delta_0}{s_d}\cdot\sqrt{n}=\dfrac{31.5-28.5-5}{\sqrt{19.9}}\cdot\sqrt{50}=-3.17$.

  As $-3.17\in C$ $(-3.17<-1.645)$, H_0 is rejected.

Test decision in R:

```
S_d2 <- SpVar_X1 + SpVar_X2 - 2 * Cov_X1_X2
Teststat <- (SpMean_X1 - SpMean_X2 - delta0) /
            sqrt(S_d2) * sqrt(n)
c_stat <- qnorm(p = alpha)
```

```
S_d2
```

```
[1] 19.9
```

```
c_stat
```

```
[1] -1.644854
```

```
Teststat
```

```
[1] -3.170213
```

```
Teststat < c_stat
```

```
[1] TRUE
```

Comparison of expected values of two independent samples

Wirtschafts- und Sozialstatistik
Universität Trier

$H_0: \mu_1 - \mu_2 = \delta_0$ vs. $H_1: \mu_1 - \mu_2 \neq \delta_0$

$\overline{X}_1 \sim N(\mu_1; \frac{\sigma_1^2}{n_1})$ and $\overline{X}_2 \sim N(\mu_2; \frac{\sigma_2^2}{n_2})$ are stochastically independent.

ND, variances known

Test statistic: $Z = \dfrac{\overline{X}_1 - \overline{X}_2 - \delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Test distribution: Standard normal distribution

Arbitrary distribution, variances known

Test statistic: $Z = \dfrac{\overline{X}_1 - \overline{X}_2 - \delta_0}{\sqrt{\dfrac{\mathrm{Var}(X_1)}{n_1} + \dfrac{\mathrm{Var}(X_2)}{n_2}}}$

Test distribution: Standard normal distribution (CLT, $n_1, n_2 > 30$)

Wirtschafts- und Sozialstatistik
Universität Trier

ND, variances unknown but identical

We use the weighted average of the sample variances

$$S^2 := \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2} \,.$$

Test statistic: $T = \dfrac{\overline{X}_1 - \overline{X}_2 - \delta_0}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

Test distribution: t-distribution with $n_1 + n_2 - 2$ degrees of freedom

Arbitrary distribution, variances unknown and potentially not identical

Test statistic: $Z = \dfrac{\overline{X}_1 - \overline{X}_2 - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$

Test distribution: Standard normal distribution (CLT, $n_1, n_2 > 30$)

Wirtschafts- und Sozialstatistik
Universität Trier

ND, variances unknown and not identical

Fisher-Behrens problem (approximative method)

Test statistic: $T = \dfrac{\overline{X}_1 - \overline{X}_2 - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$

Test distribution: t-distribution with $[k]$ degrees of freedom, using

$$k = \frac{1}{\dfrac{c^2}{n_1 - 1} + \dfrac{(1-c)^2}{n_2 - 1}} \qquad \text{with} \qquad c = \frac{\dfrac{S_1^2}{n_1}}{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}\,.$$

Example 9.17: Independent samples (1)

Two independent samples from normally distributed populations yielded in R:

```
x9_17 <- c(3,5,8,7,9,12,17,10,5,9,14,22)
y9_17 <- c(17,25,9,21,12,14,10,12,14,10,16,12,19,14,28)
```

A test for homogeneity of expected values ($\alpha = 0{,}01$) yielded in R:

```
t.test(x = x9_17, y = y9_17, alternative="two.sided",
       var.equal=FALSE, paired=FALSE, mu=0, conf.level=0.99)
```

```
       Welch Two Sample t-test
data:  x9_17 and y9_17
t = -2.5519, df = 23.968, p-value = 0.01751
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -11.4240159   0.5240159
sample estimates:
mean of x mean of y
 10.08333  15.53333
```

With $y_{15} = 8$ instead of 28, the p value would have been 0.05066.

Calculate the test statistic and other relevant values *by hand*.

Example 9.17: Independent samples (2)

An *expert* wants to be sure that the mean of the first population is smaller than the mean of the second population. Accordingly, he postulates $H_0 : \mu_1 - \mu_2 \geq 0$ (the negation of his working hypothesis).
Hypothesis test in R:

```
t.test(x=x9_17,y=y9_17,alternative="less",var.equal=FALSE,
        paired=FALSE, mu=0, conf.level=0.99)
```

```
Welch Two Sample t-test

data:  x9_17 and y9_17
t = -2.5519, df = 23.968, p-value = 0.008756
alternative hypothesis: true difference in means is less than 0
99 percent confidence interval:
 -Inf   -0.1270684
sample estimates:
mean of x mean of y
 10.08333  15.53333
```

The null has to be rejected this time!

Wirtschafts- und Sozialstatistik
Universität Trier

Comparison of proportions of two independent samples

To be tested: $H_0 : \theta_1 - \theta_2 = \delta_0$ ($\geq$ or $\leq$ possible as well)

$\delta \neq \delta_0 = 0$

Test statistic: $$Z = \frac{P_1 - P_2 - \delta_0}{\sqrt{\dfrac{P_1 \cdot (1 - P_1)}{n_1} + \dfrac{P_2 \cdot (1 - P_2)}{n_2}}}$$

Test distribution: Standard normal distribution (CLT)

$\delta = \delta_0 = 0$ (Homogeneity hypothesis)

We have $P = \dfrac{n_1 \cdot P_1 + n_2 \cdot P_2}{n_1 + n_2}$.

Test statistic: $$Z = \frac{P_1 - P_2}{\sqrt{P \cdot (1 - P) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Test distribution: Standard normal distribution (CLT)

Example 9.18: see Ex. 9.4 (1)

In case of non-working mothers there were $n_1 \cdot p_1 = 60$ children ready for school, $n_2 \cdot p_2 = 24$ for working mothers. The null hypothesis that the share of children ready for school is 5% higher for non-working mothers is to be tested ($\alpha = 0.05$).

Data input in R:

```
n1 <- 240
n2 <- 160
p_X1 <- 60 / n1
p_X2 <- 24 / n2
delta0 <- 0.05
alpha <- 0.05
```

Example 9.18: see Ex. 9.4 (2)

Checking approximation conditions in R:

```
(n1 * p_X1 * (1 - p_X1)) > 9 & (n2 * p_X2 * (1 - p_X2)) > 9
[1] TRUE
0.1 <= p_X1 & p_X1 <= 0.9
[1] TRUE
0.1 <= p_X2 & p_X2 <= 0.9
[1] TRUE
```

Test decision in R:

```
c_stat <- vector()
c_stat[1] <- qnorm(p = alpha/2)
c_stat[2] <- qnorm(p = 1 - alpha/2)
Teststat <- (p_X1-p_X2-delta0) / (sqrt((p_X1 * (1-p_X1)) / n1
            + (p_X2 * (1-p_X2)) / n2))
c_stat
[1] -1.959964  1.959964
Teststat
[1] 1.258634
Teststat < c_stat[1] | Teststat > c_stat[2]
[1] FALSE
```

Example 9.18: see Ex. 9.4 (3)

Now we would like to test the homogeneity hypothesis. The conditions remain the same.

We get $p = 84/400 = 0.21$ and therefore:

$$z = \frac{0.25 - 0.15 - 0}{\sqrt{0.21 \cdot 0.79 \cdot \left(\frac{1}{240} + \frac{1}{160}\right)}} = 2.4055\,.$$

As the approximation conditions are met, H_0 is rejected.

Hypothesis test in R:

```
p <- (n1 * p_X1 + n2 * p_X2) / (n1 + n2)
Teststat_new <- (p_X1-p_X2) / sqrt(p*(1-p)*(1/n1+1/n2))
```

```
p
```

```
[1] 0.21
```

```
Teststat_new
```

```
[1] 2.405539
```

```
Teststat_new < c_stat[1] | Teststat_new > c_stat[2]
```

```
[1] TRUE
```

Comparison of variances of two independent samples

Wirtschafts- und Sozialstatistik
Universität Trier

We want to test $H_0 : \sigma_1^2 = \sigma_2^2$. Then, $\frac{n_1 - 1}{\sigma_1^2} \cdot S_1^2 \sim \chi^2_{n_1-1}$ and $\frac{n_2 - 1}{\sigma_2^2} \cdot S_2^2 \sim \chi^2_{n_2-1}$ are stochastically independent. Given H_0 we have

$$\frac{S_1^2}{S_2^2} \sim F(n_1 - 1; n_2 - 1) \quad .$$

Notice that $f(p; k_1; k_2) = \frac{1}{f(1-p; k_2; k_1)}$.

Example 9.19: Fill quantities (1)

Wirtschafts- und Sozialstatistik
Universität Trier

A producer of lemonades wants to buy a new bottling plant and has to choose one of two alternatives. A test revealed no difference in the average fill quantities but an employee thinks that he observed less variation in fill quantities for the first machine. Another test of 101 fillings per plant yielded the following results: $\overline{x} = 0.99971$, $s_x = 0.01039$, $\overline{y} = 0.99853$ and $s_y = 0.02078$. Let the fill quantities be normally distributed for both machines.

We use the null $H_0 : \sigma_1^2 \geq \sigma_2^2$. Then, H_0 has to be rejected if the test statistic is too small. We have:

$$\frac{s_x^2}{s_y^2} = 0.24966 < f(0.05; 100; 100) = \frac{1}{f(0.95; 100; 100)} = 0.7185 \quad .$$

H_0 is rejected.

Example 9.19: Fill quantities (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Hypothesis test in R:

```
load("Beispiel11-19.RData")

var.test(x = x9_19,y = y9_19,ratio = 1,alternative="less",
         conf.level = 0.95)
```

```

        F test to compare two variances

data:  x9_19 and y9_19
F = 0.24966, num df = 100, denom df = 100, p-value = 1.275e-11
alternative hypothesis: true ratio of variances is less than 1
95 percent confidence interval:
 0.0000000 0.3474606
sample estimates:
ratio of variances
         0.2496628
```

χ^2 tests

Wirtschafts- und Sozialstatistik
Universität Trier

The χ^2 tests build a class of tests comprising

- hypotheses regarding the distribution,
- hypotheses regarding independence,
- homogeneity of two or more distributions

and many other variations from which we abstract here.

X may be nominal, ordinal or metric. We are looking at a countable number of categories, values or classes $1, \ldots, m$.

We postulate a hypothesis regarding the distribution over the m categories, values or classes. θ_j^0 $(j = 1, \ldots, m)$ are the corresponding shares. A null hypothesis regarding the distribution of X is transformed into a postulation regarding the θ_j^0.

Pearson's χ^2 test

$H_0 : \theta_1 = \theta_1^0 \wedge \cdots \wedge \theta_m = \theta_m^0$ vs. $H_1 : \theta_1 \neq \theta_1^0 \wedge \cdots \wedge \theta_m \neq \theta_m^0$

Test statistic: $\chi^2 = \sum_{j=1}^{m} \frac{\left(n \cdot p_j - n \cdot \theta_j^0\right)^2}{n \cdot \theta_j^0} = \sum_{j=1}^{m} \frac{\left(n_j - n \cdot \theta_j^0\right)^2}{n \cdot \theta_j^0}$

If H_0 holds, χ^2 is asymptotically χ^2_{m-1}-distributed for $n \to \infty$; therefore, n should be *large*.

Approximation conditions:

$m = 2$: $n \geq 30 \wedge n \cdot \theta_j^0 \geq 5$ for all j

$m > 2$: $n \geq 30 \wedge n \cdot \theta_j^0 \geq 1$ for all j and $n \cdot \theta_j^0 < 5$ for not more than 20% of categories, values or classes

The critical region is always one-sided:

$$C = \left\{(x_1, \ldots, x_n) \,|\, \chi^2 > \chi^2_{m-1}(1-\alpha)\right\}$$

- The quality of the test is determined by the number and localisation of the classes.
- A coarsening of the classification using a combination of classes might be needed.

Wirtschafts- und Sozialstatistik
Universität Trier

Example 9.20: Testing a dice (1)

A dice is suspicious and therefore suspected to be *crooked*. To test this hypothesis, the dice is rolled $n = 300$ times. The recorded outcomes are $n_1 = 40$; $n_2 = 45$; $n_3 = 80$; $n_4 = 55$; $n_5 = 45$; $n_6 = 35$. The null that the dice *is not crooked* is to be tested using $\alpha = 0.05$.

- $H_0 : \theta_j = \theta_j^0 = \frac{1}{6}$ vs. $H_1 : \theta_j \neq \theta_j^0 = \frac{1}{6}$, $\alpha = 0.05$, $n = 300$

Data input in R:

```
Sp9_20 <- c(40,45,80,55,45,35)
n <- 300
m <- 6
theta0 <- rep(x = 1/6, times = m)
alpha <- 0.05
```

Checking approximation conditions in R:

```
m > 2
```

```
[1] TRUE
```

```
all(n * theta0 >= 1)
```

```
[1] TRUE
```

```
n >= 30
```

```
[1] TRUE
```

```
sum(n * theta0 < 5) / m <= 0.2
```

```
[1] TRUE
```

Example 9.20: Testing a dice (2)

- $\chi^2 = \sum_{j=1}^{m} \frac{\left(n_j - n \cdot \theta_j^0\right)^2}{n \cdot \theta_j^0}$; $\chi_5^2(0.95) = 11.070$
- $C = \left\{(x_1, \ldots, x_n) \,|\, \chi^2 > 11.0705\right\}$
- $\chi^2 = \frac{(40-50)^2}{50} + \frac{(45-50)^2}{50} + \frac{(80-50)^2}{50} + \frac{(55-50)^2}{50} + \frac{(45-50)^2}{50} + \frac{(35-50)^2}{50} = 26$

  As $26 \in C$ $(26 > 11.0705)$, H_0 is rejected.

Example 9.20: Testing a dice (3)

Test decision in R:

```
c_stat <- qchisq(p = 1-alpha, df = m-1)
```

```
[1] 11.0705
```

```
chisq.test(x = Sp9_20, p = theta0 * n, rescale.p = TRUE,
           correct = FALSE) # ATTENTION: argument rescale.p
```

```
        Chi-squared test for given probabilities

data:  Sp9_20
X-squared = 26, df = 5, p-value = 8.924e-05
```

Example 9.21: Test for standard normal distribution (1)

Wirtschafts- und Sozialstatistik
Universität Trier

The frequency distribution of a variable y in the sample looks as follows ($n = 1000$):

Class	$y \leq -1$	$-1 < y \leq 0$	$0 < y \leq 1$	$y > 1$
Frequency	165	360	314	161

Test the null that *variable y follows a standard normal distribution* using $\alpha = 0.05$.

- First, the theoretical frequencies $n \cdot \theta_j^0$ have to be determined.
  $1000 \cdot \theta_1^0 = 1000 \cdot \Phi(-1) = 1000 \cdot (1 - 0.8413) = 159$. As the SND is symmetric, we have $\theta_4^0 = 159$ and from $\theta_2^0 = \theta_3^0$ finally follows that $n \cdot \theta_2^0 = n \cdot \theta_3^0 = 341$.
  $H_0 : \theta_1 = \theta_1^0 = 0.159 \wedge \theta_2 = \theta_2^0 = 0.341 \wedge \theta_3 = \theta_3^0 = 0.341 \wedge$
  $\theta_4 = \theta_4^0 = 0.159$, H_1 complementary, $\alpha = 0.05$ and $n = 1000$

Example 9.21: Test for standard normal distribution (2)

Data input in R:

```
load("Example9-21.RData")

y9_21_kl <- table(y9_21)
n <- 1000
m <- 4
theta0 <- c(159, 341, 341, 159) / n
alpha <- 0.05
```

Checking approximation conditions in R:

```
m > 2
```

```
[1] TRUE
```

```
n >= 30
```

```
[1] TRUE
```

```
all(n * theta0 >= 1)
```

```
[1] TRUE
```

```
sum(n * theta0 < 5) / m <= 0.2
```

```
[1] TRUE
```

Example 9.21: Test for standard normal distribution (3)

Wirtschafts- und Sozialstatistik
Universität Trier

- $\chi^2 = \sum_{j=1}^{m} \frac{\left(n_j - n \cdot \theta_j^0\right)^2}{n \cdot \theta_j^0}$; $\chi_3^2(0.95) = 7.815$
- $C = \left\{(x_1, \ldots, x_n) \,|\, \chi^2 > 7{,}815\right\}$
- $\chi^2 = \frac{(165-159)^2}{159} + \frac{(360-341)^2}{341} + \frac{(314-341)^2}{341} + \frac{(161-159)^2}{159} = 3.448$

  As $3.448 \notin C$ ($3.448 \not> 7.815$), H_0 is not rejected.

Hypothesis test in R:

```
chisq.test(x = y9_21_kl, p = theta0 * n, rescale.p = TRUE,
           correct = FALSE)
```

```
        Chi-squared test for given probabilities

data:  y9_21_kl
X-squared = 3.4481, df = 3, p-value = 0.3276
```

Example 9.21: Test for standard normal distribution (4)

In this case the hypothesis has been fully specified as the distribution with all its parameters has been stated (SND). If we still need to estimate ω parameters from the sample, we need to use the following test distribution: $\chi^2_{m-\omega-1}$.

In the present case we might also have tested for a normal distribution. μ and σ^2 would have to be estimated using the known estimators $\overline{X}$ and S^2 using sample data. The theoretical frequencies would then have to be calculated again. The test distribution would then be χ^2_1.

Please test again, this time using $\overline{x} = -0.1$ and $s^2 = 1.047$.
(Hint: H_0 is rejected)

Wirtschafts- und Sozialstatistik
Universität Trier

χ^2 test for independence

The variables of interest have m and r values, categories or classes, respectively. n_{jk} is the joint frequency and θ_{jk} is the joint probability with the corresponding marginal probabilities $\theta_{j\cdot}$ and $\theta_{\cdot k}$.
We use

$$H_0 : \theta_{jk} = \theta_{j\cdot} \cdot \theta_{\cdot k} \qquad \text{for all } j = 1, \ldots, m \text{ and } k = 1, \ldots, r$$

as the null. The theoretical frequencies on the RHS of H_0 are built using $\theta_{j\cdot} = n_{j\cdot}/n$ and $\theta_{\cdot k} = n_{\cdot k}/n$, respectively.
Then

$$\chi^2 = \sum_{j=1}^{m} \sum_{k=1}^{r} \frac{\left(n_{jk} - n \cdot \theta_{j\cdot} \cdot \theta_{\cdot k}\right)^2}{n \cdot \theta_{j\cdot} \cdot \theta_{\cdot k}} = \sum_{j=1}^{m} \sum_{k=1}^{r} \frac{\left(n_{jk} - \frac{n_{j\cdot} \cdot n_{\cdot k}}{n}\right)^2}{\frac{n_{j\cdot} \cdot n_{\cdot k}}{n}}$$

is the test statistic. As the test distribution we use $\chi^2_{(m-1)\cdot(r-1)}$.

Two-dimensional frequencies

Wirtschafts- und Sozialstatistik
Universität Trier

Cat. of 1st variable \ Cat. of 2nd variable	1	$\cdots$	k	$\cdots$	r	$\sum$
1	n_{11}	$\cdots$	n_{1k}	$\cdots$	n_{1r}	$n_{1\cdot}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$	$\vdots$
j	n_{j1}	$\cdots$	n_{jk}	$\cdots$	n_{jr}	$n_{j\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$
m	n_{m1}	$\cdots$	n_{mk}	$\cdots$	n_{mr}	$n_{m\cdot}$
$\sum$	$n_{\cdot 1}$	$\cdots$	$n_{\cdot k}$	$\cdots$	$n_{\cdot r}$	n

Wirtschafts- und Sozialstatistik
Universität Trier

Example 9.22: exam (see Schaich, 1998) (1)

For $n = 627$ graduated students at a faculty of a university, the number of semesters and the final grade have been recorded (see table below). We want to test for independence of number of semesters and final grade using $\alpha = 0.05$.

Load data in R:

```
load("Example9-22.RData")

alpha <- 0.05
n <- sum(Sp9_22)
m <- dim(Sp9_22)[1]
r <- dim(Sp9_22)[2]

theta0 <- margin.table(x = Sp9_22, margin = 1) %*%
          t(margin.table(x = Sp9_22, margin = 2)) /
          margin.table(x = Sp9_22)
```

Example 9.22: exam (see Schaich, 1998) (2)

Semesters / Grade	8	9	10	11	12-13	14-15	over 15	$\sum$
1	9 (3.1)	19 (10.5)	23 (23.0)	5 (10.8)	4 (8.2)	3 (6.0)	4 (5.4)	67
2	10 (7.3)	34 (24.7)	68 (54.2)	11 (25.5)	11 (19.4)	21 (14.1)	3 (12.9)	158
3	8 (8.6)	34 (29.2)	67 (64.1)	32 (30.1)	20 (23.0)	13 (16.7)	13 (15.2)	187
4	0 (4.9)	10 (16.7)	30 (36.7)	28 (17.2)	18 (13.1)	6 (9.6)	15 (8.7)	107
5	2 (5.0)	1 (16.9)	27 (37.0)	25 (17.4)	24 (13.3)	13 (9.6)	16 (8.8)	108
$\sum$	29	98	215	101	77	56	51	627

Empirical frequencies (theoretical frequencies under H_0)

Example 9.22: exam (see Schaich, 1998) (3)

Checking approximation conditions in R:

```
n >= 30
```

```
[1] TRUE
```

```
sum(theta0 == 0) == 0
```

```
[1] TRUE
```

```
sum(theta0 < 5) / (m*r) <= 0.2
```

```
[1] TRUE
```

- $H_0: \theta_{jk} = \theta_{j\cdot} \cdot \theta_{\cdot k}$ vs. $H_1: \theta_{jk} \neq \theta_{j\cdot} \cdot \theta_{\cdot k}$ for all $j = 1, \ldots, m$ and $k = 1, \ldots, r$, $\alpha = 0.05$, $n = 627$
- $\chi^2 = \sum_{j=1}^{m} \sum_{k=1}^{r} \frac{\left(n_{jk} - \frac{n_{j\cdot} \cdot n_{\cdot k}}{n}\right)^2}{\frac{n_{j\cdot} \cdot n_{\cdot k}}{n}}$, $\chi^2_{24}(0.95) = 36.415$,
- $C = \left\{(x_1, \ldots, x_n) \,|\, \chi^2 > 36.415\right\}$
- $\chi^2 = \frac{(9 - 3.1)^2}{3.1} + \cdots + \frac{(16 - 8.8)^2}{8.8} = 120.54$
  As $120.54 \in C$ $(120.54 > 36.415)$, H_0 is rejected.

Example 9.22: exam (see Schaich, 1998) (4)

Wirtschafts- und Sozialstatistik
Universität Trier

Test decision in R:

```
c_stat <- qchisq(p = 1 - alpha, df = (m-1) * (r-1))
c_stat
```

```
[1] 36.41503
```

```
chisq.test(x = Sp9_22, correct = FALSE, rescale.p = FALSE)
```

```

	Pearson's Chi-squared test

data:  Sp9_22
X-squared = 120.55, df = 24, p-value = 7.741e-15

Warning message:
In chisq.test(x = Sp9_22, correct = FALSE) :
  Chi-squared approximation may be incorrect
```

Wirtschafts- und Sozialstatistik
Universität Trier

χ^2 test for homogeneity

We want to test m distributions for homogeneity. We look at m distributions with r categories, values or classes each. The distributions have to be considered homogeneous if the relative frequencies are identical and do not depend on the respective distribution. Therefore, the test can immediately be derived from the χ^2 test for independence:

Test statistic: $$\chi^2 = \sum_{j=1}^{m} \sum_{k=1}^{r} \frac{\left(h_{jk} - \frac{n_j \cdot h_{\cdot k}}{n}\right)^2}{\frac{n_j \cdot h_{\cdot k}}{n}}.$$

Here, h_{jk} are the observed frequencies in the j-th distribution and the k-th category, value or class. n_j is the absolute frequency of observations of distribution j.

Test distribution: $\chi^2_{(m-1)\cdot(r-1)}$

Wirtschafts- und Sozialstatistik
Universität Trier

Example 9.23: Income distributions (1)

In a study on income distributions (poor, middle, rich) in three regions A, B and C interviews were conducted with $n = 200$ people in each region:

Region	poor	middle	rich	$\sum$
A	51	112	37	200
B	65	121	14	200
C	55	115	30	200
$\sum$	171	348	81	600

Test if the income distributions are homogeneous in the three regions using $\alpha = 0.05$.

Read data in R:

```
load("Example9-23.RData")
alpha <- 0.05
n<-sum(Sp9_23);m<-dim(Sp9_23)[1];r<-dim(Sp9_23)[2]
theta0 <- margin.table(x = Sp9_23, margin = 1) %*%
          t(margin.table(x = Sp9_23, margin = 2)) /
          sum(x = Sp9_23)
```

Example 9.23: Income distributions (2)

Checking approximation conditions in R:

```
n >= 30
```

```
[1] TRUE
```

```
sum(theta0 == 0) == 0
```

```
[1] TRUE
```

```
sum(theta0 < 5) / (m*r) <= 0.2
```

```
[1] TRUE
```

- $H_0 : F_A = F_B = F_C$, $\alpha = 0.05$, $n = 600$
- Test statistic: see above, approximation conditions fulfilled
- $C = \left\{(x_1, \ldots, x_n) \,|\, \chi^2 > 9.488\right\}$
- $\chi^2 = \dfrac{(51 - 200 \cdot 171/600)^2}{200 \cdot 171/600} + \cdots + \dfrac{(30 - 200 \cdot 81/600)^2}{200 \cdot 81/600} = 12.483$
  As $12.483 \in C$ $(12.483 > 9.488)$, H_0 is rejected.

Wirtschafts- und Sozialstatistik
Universität Trier

Example 9.23: Income distributions (3)

Hypothesis test in R:

```
chisq.test(x = Sp9_23, correct = FALSE, rescale.p = FALSE)
```

```
        Pearson's Chi-squared test

data:  Sp9_23
X-squared = 12.483, df = 4, p-value = 0.0141
```

Alternative determination of the theoretical frequencies under H_0 and of the critical value in R:

```
c_stat <- qchisq(p = 1 - alpha, df = (m-1) * (r-1))
theta0_alternative <- chisq.test(x = Sp9_23, correct = FALSE,
                                 rescale.p = FALSE)$expected
```

```
c_stat
```

```
[1] 9.487729
```

```
theta0_alternative
```

```
  poor  middle  rich
A 57    116     27
B 57    116     27
C 57    116     27
```

More tests (1)

Wirtschafts- und Sozialstatistik
Universität Trier

Tests for distribution In praxis, the Kolmogorow-Smirnow test (KS test) is frequently used. Here, one measures the maximum distance between the theoretical distribution function and the relative empirical sum function in order to use it as test statistic.

Tests for correlation The correlation coefficient of Bravais-Pearson has in general no good properties regarding inference. An exception is the bivariate normal distribution! In this case, two tests can be used ($R_{xy} = \widehat{\varrho}_{XY}$ is the emp. corr.):

- $H_0 : \varrho_{XY} = 0$

$$\frac{R_{xy}}{\sqrt{1 - R_{xy}^2}} \cdot \sqrt{n-2} \sim t_{n-2}$$

Since stochastic independence and uncorrelatedness are equal in the case of a bivariate normal distribution, this test is sometimes also called test for independence.

More tests (2)

Wirtschafts- und Sozialstatistik
Universität Trier

Tests for correlation

- $H_0 : \varrho_{XY} = \varrho_0 \ (\varrho_0 \neq 0)$, $n > 25$

$$\frac{1}{2}\left(\ln\frac{1+R_{xy}}{1-R_{xy}} - \ln\frac{1+\varrho_0}{1-\varrho_0}\right)\cdot\sqrt{n-3} \sim N(0;1)$$

Tests for correlation in R:

```
cor.test(x, y,
         alternative = c("two.sided", "less", "greater"),
         method = c("pearson", "kendall", "spearman"),
         conf.level = 0.95)
```

Variance analysis Test for homogeneity of means of at least two distributions. The test is based on a decomposition of the total variance within and between the single distributions. As a test distribution, one uses the F distribution.