University of St.Gallen

EFMD EQUIS ACCREDITED

Methods: Statistics (4,120)

9. Test Theory and Hypothesis Tests

Spring 2022

Prof. Dr. Roland Füss

Swiss Institute of Banking and Finance (s/bf)

Contents

Learning Objectives

After this lecture, you know how:

- to transform a verbal statement into appropriate **null** and **alternative hypotheses**, including the determination of whether a **two-tailed test** or a **one-tailed test** is appropriate.
- to describe what is meant by **Type I** and **Type II** errors and explain how their probabilities can be reduced in hypothesis testing.
- to carry out a **hypothesis test for a population mean** or a **population proportion** and interpret the results of the test.
- to explain the p**-value** for a hypothesis test and distinguish it from the **level of significance** α.

Literature

Levine, D.M., K. A. Szabat, and D.F. Stephan. (2016). *Business Statistics: A First Course*, 7th ed. United States: Pearson, **Chapter 9**.*

Stinerock, R. (2018). *Statistics with R*. United Kingdom: Sage. **Chapters 9-10**.*

Shira, Joseph (2012). *Statistische Methoden der VWL und BWL*, 4th ed. Munich et al.: Pearson Studium, **Chapter 15**.

Weiers, R. M. (2011). *Introductory Business Statistics*,7th ed., Canada: Thomson South-Western, **Chapters 10 and 11**.

*Mandatory literature

I. Introduction Example

Cola Zero vs. Original:

Two students do not agree whether *Cola Zero* tastes as good as the *Original* (assumption). For this reason, they carry out a **blind test** (random sample) where 20 fellow students are asked to taste both versions. If *Cola Zero* tastes as good as the *Original*, *Cola Zero* should be chosen in approximately 50% of the cases (chance level).

However, the test shows that in total 8 students (8/20 = 40%) prefer *Cola Zero*.

- **Can you infer that *Cola Zero* does not taste as good as the original?**
- **Does this value still lie in a range of statistical chance?**

I. Basic Idea

Facts or questions such as the previous introductory example can be answered by using the concepts of test theory. The statement about the population is precisely formulated by the **null hypotheses** and the contrary statement is given by the **alternative hypotheses.**

Cola Zero vs. Original:

Null hypothesis (H_0):

In 50% of the cases *Cola Zero* is preferred as it tastes equally good.

Alternative hypothesis (H_1):

Cola Zero is not preferred in 50% of cases as it does not tastes equally good.

I. Basic Idea

The **hypothesis test** (significance test) assesses the validity of these hypotheses by drawing a conclusion from the sample to the population. The following **assumptions** are important:

- The null hypothesis is valid ("the benefit of the doubt ").
- The data are collected by means of a random sample.

There is always the risk of a false conclusion as the decision of the hypothesis test is based on only one sample. This risk is denoted by the **significance level** α, which has to be determined *subjectively* and before the test is conducted.

As part of the hypothesis test, a **test statistic** (z or t) is assigned to the drawn sample. The test statistic is subsequently evaluated in order to conclude whether or not H_0 should be rejected.

I. Basic Idea

The evaluation of the test statistic is done by consulting a suitable (and already known) **test distribution**. In case the test statistic is located at one of the two tails of the test distribution (i.e., if it exceeds a critical value) H_0 can be rejected.

The entirety of all possible test values is thus divided into an area where H_0 is rejected and one where H_0 is not rejected. This partitioning is determined by the selected level of significance.

- If the test statistic **falls into the acceptance region** of H_0, H_0 is *not* rejected
- If the test statistic **does not fall into the acceptance region** of H_0, it is regarded as being statistically significant and H_0 is rejected.

II. Null and Alternative Hypothesis

A hypothesis can either be formulated **two-tailed** (non-directional) or **one-tailed** (directional), with one-tailed hypotheses being either left-tailed or right-tailed.

Cola Zero vs. Original:

Two-tailed hypothesis:

Cola Zero is preferred in 50% of cases.

H_0: $\pi = 0.50$, H_1: $\pi \neq 0.50$

One-tailed, left-tailed hypothesis:

Cola Zero is preferred in *at least* 50% of cases.

H_0: $\pi \geq 0.50$, H_1: $\pi < 0.50$

One-tailed, right-tailed hypothesis:

Cola Zero is preferred in a *maximum of* 50% of cases.

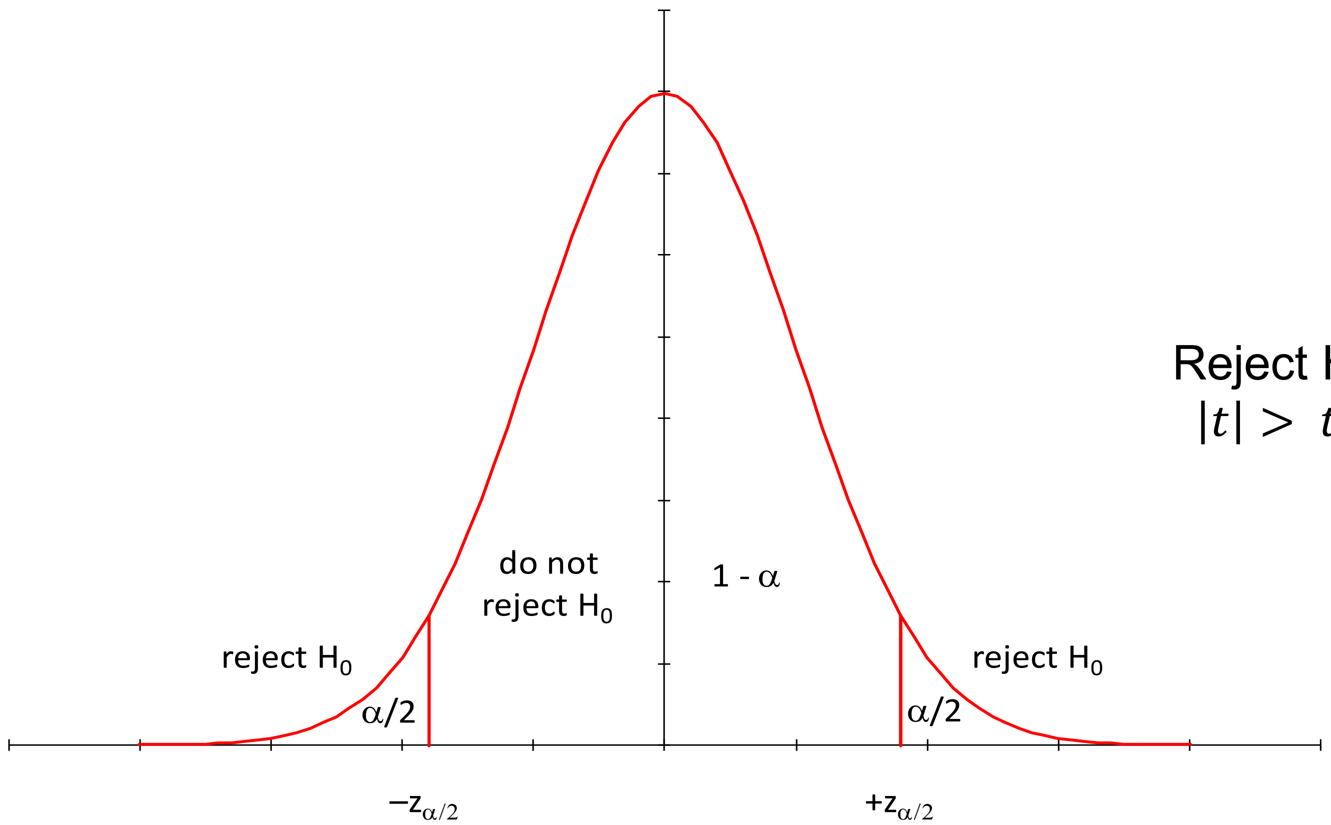
H_0: $\pi \leq 0.50$, H_1: $\pi > 0.50$

III. Two-tailed Tests

Non-directional hypothesis:

Cola Zero is preferred in 50% of cases.

H_0: $\pi = 0.50$, H_1: $\pi \neq 0.50$



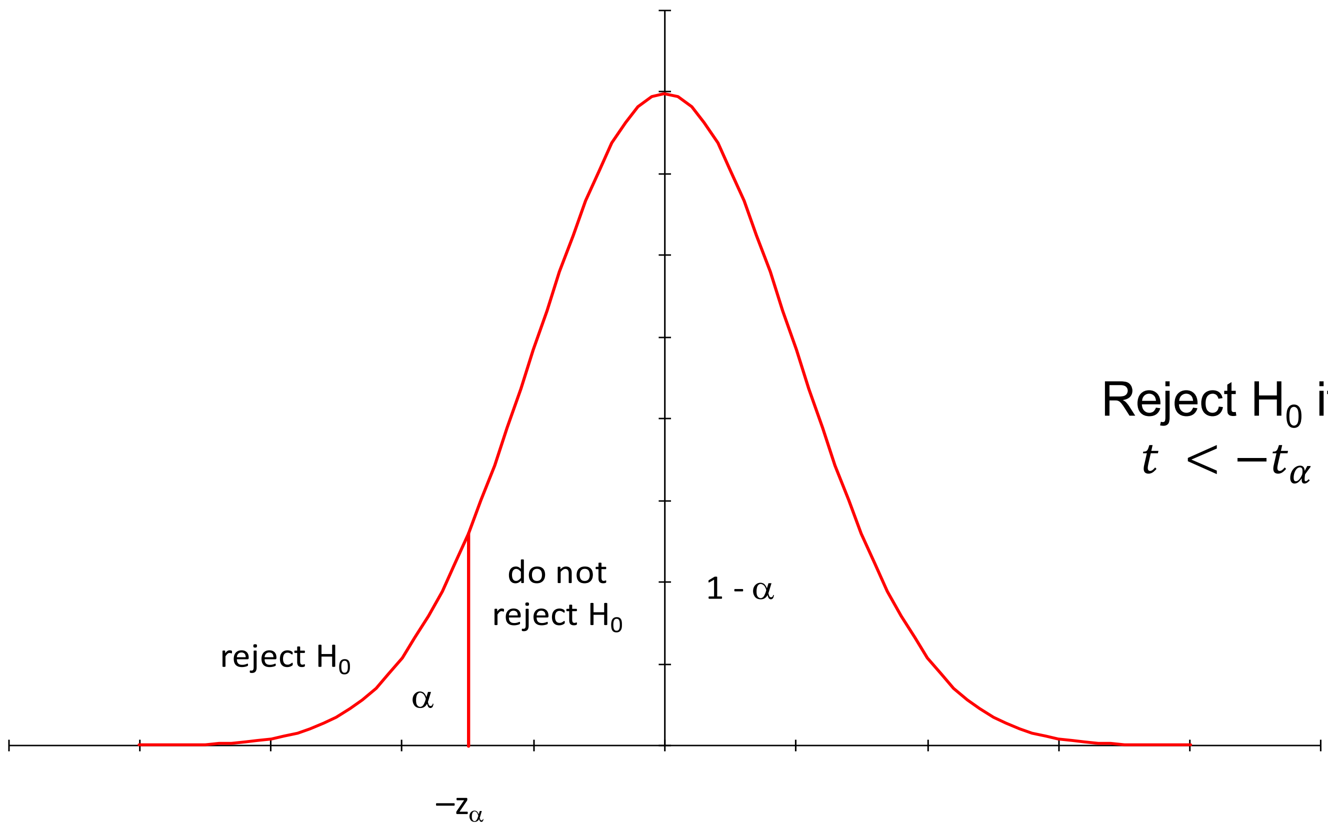
Reject H_0 if:
$|t| > t_{\alpha/2}$

III. One-tailed Tests (left-tailed)

Directional hypothesis:

Cola Zero is preferred in at least 50% of cases.

H_0: $\pi \geq 0.50$, H_1: $\pi < 0.50$



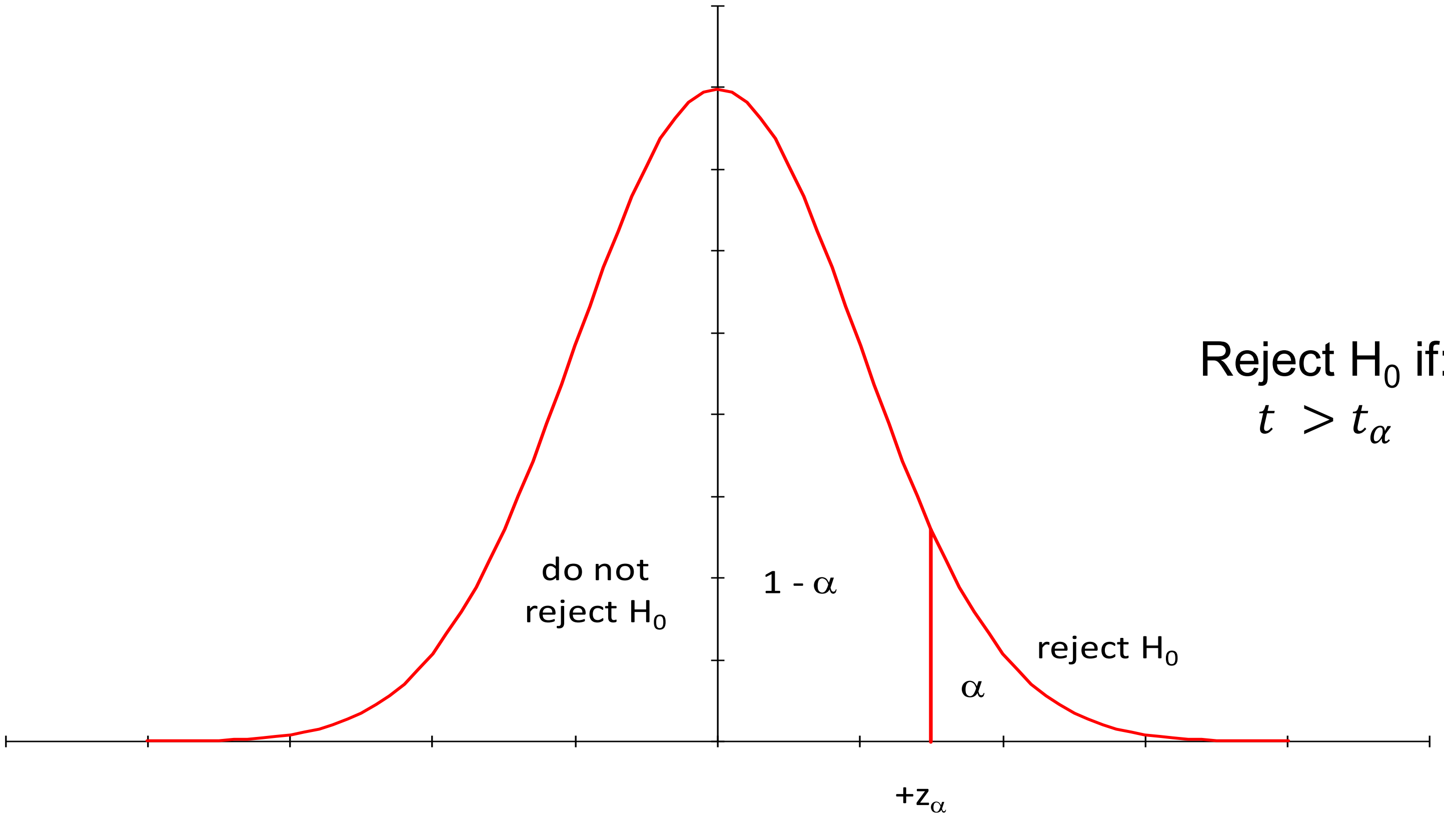
Reject H_0 if:
$t < -t_\alpha$

III. One-tailed Tests (right-tailed)

Directional hypothesis:

Cola Zero is preferred in at most 50% of cases.

H_0: $\pi \leq 0.50$, H_1: $\pi > 0.50$



Reject H_0 if:
$t > t_\alpha$

III. Two-tailed vs. One-tailed Tests (R-Example)

Open the file "L9-Example_1.R" in R-Studio and follow the R-Code for the reproduction of the introductory example.

```r
# introductory example: cola zero
#---------------------------------------
# Results of the tasting (0 = preferably original, 1 = preferably cola zero)
results<-c(0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1)
# 40% prefer cola zero to the original
mean(results)

# two-tailed hypothesis:
# cola zero is preferred in 50% of cases. H0: p = 0.5, H1: p =! 0.5
t.test(results,conf.level=0.95,mu=0.5)

# one-tailed hypothesis (left):
# cola zero is preferred in at least 50% of cases. H0: p >= 0.5, H1: p < 0.5
t.test(results,conf.level=0.95,mu=0.5,alternative = "less")

# one-tailed hypothesis (right):
# cola zero is preferred in  50% of cases at maximum. H0: p <= 0.5, H1: p > 0.5
t.test(results,conf.level=0.95,mu=0.5,alternative = "greater")
```

IV. Procedure of Hypothesis Testing

In hypothesis testing there are **five steps** to follow:

1. Formulation of the **null** and **alternative hypothesis** as well as selection of the **significance level**
2. Selection of an appropriate **test statistic** and **test distribution**, assuming validity of the null hypothesis
3. Determination of the **critical value(s)**
4. **Calculation** of the test statistic
5. **Decision** and interpretation

V. Test Statistic (for μ, σ known)

Assumptions:

The population follows a normal distribution and σ is **known**.

Test statistic:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \quad : \quad N(0,1)$$

with $\bar{x}$ mean sample value

μ_0 hypothesized population mean

σ population standard deviation

n sample size

V. Test Statistic (for μ, σ known: Example)

Welding robot:

A properly adjusted welding robot needs on average 1.3250 minutes for a particular weld seam. Previous experiences show that the standard deviation of one cycle is 0.0396 minutes. A randomly drawn sample of 80 cycles exhibits a sample mean of 1.3229 minutes. Should the robot be readjusted based on a hypothesis test (significance level α = 0.05)?

1. **Hypotheses and significance levels:**

   H_0: $\mu = 1.3250$; H_1: $\mu \neq 1.3250$; $\alpha = 0.05$

2. **Test statistic and test distribution:**

   σ is known → normal distribution

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

V. Test Statistic (for μ, σ known: Example)

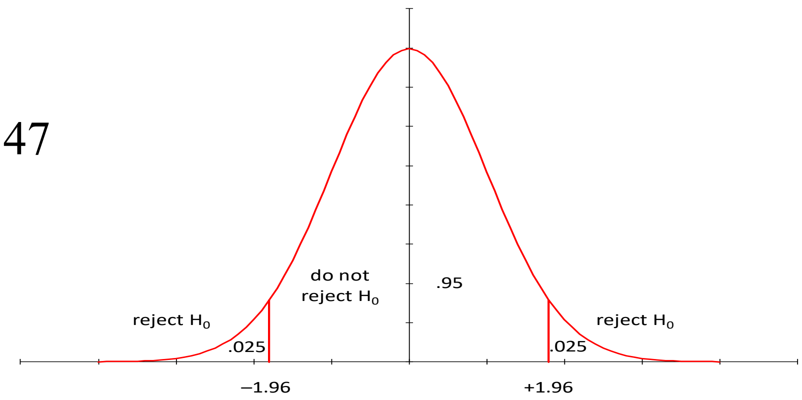
3. Critical value(s):

Two-tailed hypothesis test with α = 0.05 → $z_{\alpha/2}$ = ± 1.96

```
> qnorm(0.025)
[1] -1.959964
> qnorm(0.975)
[1] 1.959964
```

4. Calculation of the test statistic:

$$z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{1.3229 - 1.3250}{\frac{0.0396}{\sqrt{80}}} = \frac{-0.0021}{0.00443} = -0.47$$



5. Decision and interpretation:

Since the test variable z = -0.47 lies within the range of $z_{\alpha/2}$ = ± 1.96, H_0 cannot be **rejected** at a significance level of 5%. Hence, it cannot be concluded from the test result that the robot should be readjusted.

V. Test Statistic (for μ, σ known: Example)

Lightbulbs:

A factory's light bulbs have an average life of 1,030 hours with a standard deviation of 90 hours. An engineer designs a new feature to prolong the life of lightbulbs. However, the production manager fears that this is not the case. Therefore, he tests 40 randomly drawn lightbulbs with the new feature which has an average life of 1,061.6 hours. Should the procuction manager use the new feature in the future (significance level α = 0.05)?

1. **Hypotheses and significance levels:**

   H_0: $\mu \leq$ 1'030; H_1: $\mu >$ 1'030; α = 0.05

2. **Test statistic and test distribution:**

   σ is known → normal distribution

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

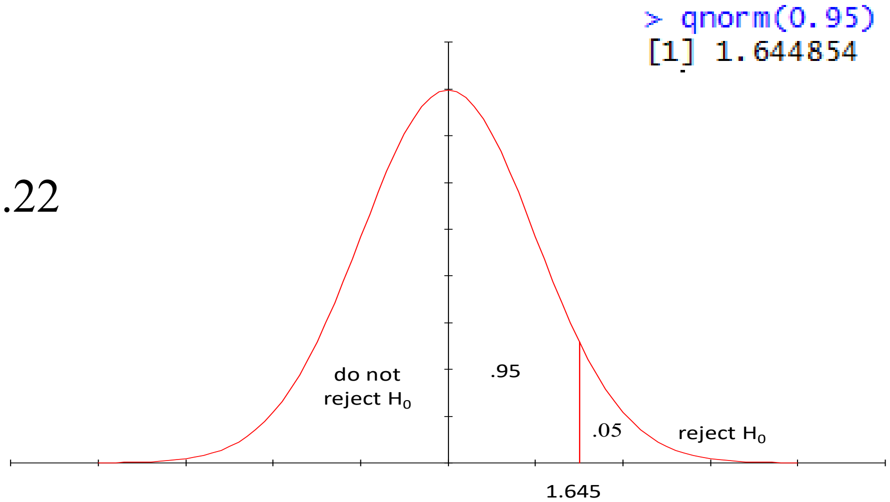
V. Test Statistic (for μ, σ known: Example)

3. Critical value(s):

One-tailed, right-tailed hypothesis test with $\alpha = 0.05 \rightarrow z_\alpha = 1.645$

4. Calculation of the test statistic:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{1061.6 - 1030}{\frac{90}{\sqrt{40}}} = \frac{31.6}{14.23} = 2.22$$



5. Decision and interpretation:

As the test statistic $z = 2.22$ is above the critical value $z_\alpha = 1.645$, H_0 can be **rejected** at a significance level of 5%. It follows, that the device actually extends the life of light bulbs (based on the sample).

V. Test Statistic (for μ, σ unknown)

Assumptions:

The population follows a normal distribution and σ is **unknown**.

Test statistic: $$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \quad t - distribution,\ df = n - 1$$

with $\bar{x}$ sample mean

μ_0 hypothesized population mean

s sample standard deviation

n sample size

df number of degrees of freedom

V. Test Statistic (for μ, σ unknown: Example)

Construction company:

A firm produces panels with a target thickness of 0.25 cm on average. The panel thickness is normally distributed. A sample of 10 panels exhibits a mean of 0.253 cm and a standard deviation of 0.003 cm. Test the hypothesis that the machine is working accurately at a significance level of $\alpha = 0.05$. Note that the population standard deviation is not known!

1. **Hypotheses and significance levels:**

   H_0: $\mu = 0.25$; H_1: $\mu \neq 0.25$; $\alpha = 0.05$

2. **Test statistic and test distribution:**

   σ is unknown → *t-distribution* with n - 1 = 9 degrees of freedom $t = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

V. Test Statistics (for μ, σ unknown: Example)

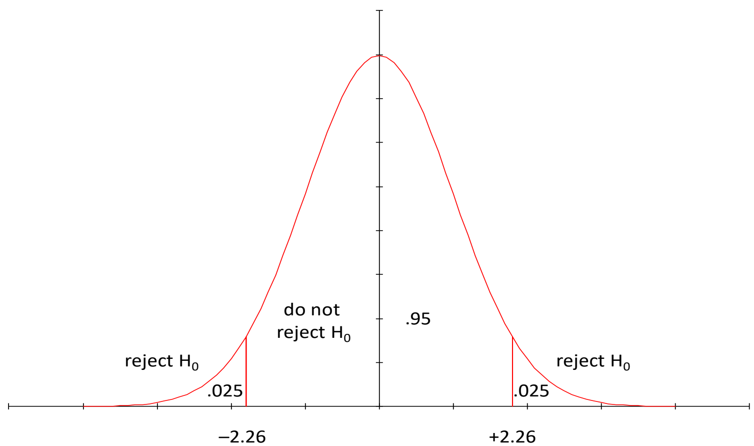
3. Critical value(s):

Two-tailed hypothesis test with α = 0.05 → $t_{\alpha/2}$ = ± 2.262

```
> qt(0.025,9)
[1] -2.262157
> qt(0.975,9)
[1] 2.262157
```

4. Calculation of the test statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.253 - 0.25}{0.003/\sqrt{10}} = 3.162$$



5. Decision and interpretation:

The null hypothesis H_0 can be **rejected** at a significance level of 5% as the test statistic t = 3,162 is above the critical value $t_{\alpha/2}$ = 2,262. This is the proof that the machine is **not** working accurately.

V. Overview: Test Statistic for the Mean

(1) σ known, population normally distributed

↓

z-Test

↓

z-statistic

↓

$$z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$$

(2) σ unknown, population normally distributed and n<30

↓

t-Test

↓

t-statistic

↓

$$t = \frac{\overline{x} - \mu_0}{s / \sqrt{n}} \sim t_{n-1}$$

(3) σ known/unknown and n ≥ 30

↓

z-Test

↓

z-statistic

↓

$$z = \frac{\overline{x} - \mu_0}{\hat{\sigma} / \sqrt{n}} \sim N(0,1)$$

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}}$$

$$\hat{\sigma} = \begin{cases} \sigma & \text{if variance of population known} \\ s, & \text{otherwise } (s = \text{sample std. dev.}) \end{cases}$$

VI. Test Statistic for the Proportion

Prerequisites:

The proportion *p of* a sample of a dichotomous population (π_0: hypothesized proportion of the population) is **normally distributed** with a sufficiently large sample size *n, i.e*. if $n \times \pi_0 \geq 5$ and $n \times (1 - \pi_0) \geq 5$:

Test statistic: $$z = \frac{p - \pi_0}{\sigma_p} \quad : \ N(0,1)$$

with p sample proportion

π_0 hypothesized population proportion

σ_p standard error of the sample proportion = $\sigma_p = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$

n sample size

VI. Test Statistic for the Proportion (Example)

Graduate survey:

A university claims that 70% of graduates get their first employment in a position directly related to their undergraduate field of study. A survey among 200 recent graduates shows that 66% get an employment that is directly related to their undergraduate field of study. Can the university's statement be confirmed by using a hypothesis test with a significance level of 0.05?

1. **Hypotheses and significance levels:**

   H_0: $\pi = 0.70$; H_1: $\pi \neq 0.70$; $\alpha = 0.05$

2. **Test statistic and test distribution:**

   $200 \times 0.70 \geq 5$ and $200 \times 0.30 \geq 5 \rightarrow$ normal distribution

$$z = \frac{p - \pi_0}{\sigma_p} \qquad \sigma_p = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

VI. Test Statistic for the Proportion (Example)

3. Critical value(s):

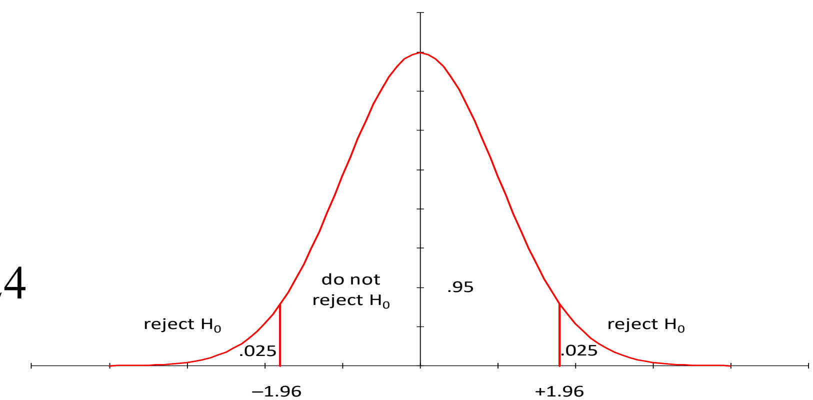
Two-tailed hypothesis test with = 0.05 → $z_{\alpha/2}$ = ± 1.96

```
> qnorm(0.025)
[1] -1.959964
> qnorm(0.975)
[1] 1.959964
```

4. Calculation of the test statistic:

$$z = \frac{p - \pi_0}{\sigma_p} = \frac{0.66 - 0.70}{0.0324} = -1.23$$

$$\sigma_p = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{0.70(1-0.70)}{200}} = 0.0324$$



5. Decision and interpretation:

As z = -1.23 lies within the range $z_{\alpha/2}$ = ± 1.96, the statement of the university (H_0) cannot be **rejected.** Therefore, it can be assumed that approximately 70% of graduates find a job within 3 months after graduation.

VI. Overview: Test Statistic for the Proportion

Approximation of the underlying binomial distribution if the **two** prerequisites

$n \times \pi_0 \geq 5$ and $n \times (1 - \pi_0) \geq 5$ are both given.

↓

z-Test

↓

z-Statistic

↓

$$z = \frac{p - \pi_0}{\sigma_p} : N(0,1) \longleftarrow \sigma_p = \sqrt{\frac{\pi_0 (1 - \pi_0)}{n}}$$

VII. Type I and Type II Error

A hypothesis tests draws the **right conclusion** about the population in the following two cases:

- H_0 is **wrong** and is **rejected**.
- H_0 is **true** and is **not rejected**.

However, there is always the danger of error. A wrong conclusion about the population results if H_0 is **true** and is **rejected** (Type I error) and if H_0 is **wrong** and is **not rejected** (Type II error):

		Reality	
		H_0 is false	*H_0 is true*
Test	*Reject H_0*	**Correct**	**Type I error** (False Positive)
	Accept H_0	**Type II error** (False Negative)	**Correct**

The **significance level** α indicates the probability of a Type I error, while β reflects the probability of a Type II error. In hypothesis testing, there is a necessary **trade-off** between Type I and Type II errors.

VIII. *p*-Value

Besides the described procedure of hypothesis testing, in which the test statistic is compared with the critical value(s), one can test a hypothesis also by means of the ***p*-value** (empirical significance level).

The *p*-value is calculated from the sample (usually via statistical software) and assumes values between 0 and 1. The *p*-value indicates **the probability of obtaining a result that is equal to (or more extreme than) what was observed, even though the null hypothesis is true.** Thus, the *p*-value indicates how extreme a result is: the smaller the *p*-value, the stronger the evidence against H_0 (and the stronger the evidence in favor of H_1).

The *p*-value corresponds to the area which is described by the corresponding distribution function and the interval $[x_{test}; \infty]$. H_0 is discarded if the ***p*-value** is smaller than the **significance level.**

VIII. *p*-Value (R-Example)

Open the file "L9-Example_2.R" in R-Studio and follow the R-Code to calculate the *p*-values for the previous examples.

```
# example: welding robot
#-------------------------------------
# test statistic z = -0.47
pnorm(0.47,lower.tail=FALSE) + pnorm(-0.47)
# p-value = 0.6384 > 0.05 therefore, H0 is not rejected.

# example: lightbulbs
#-------------------------------------
# test statistics z = 2.22
pnorm(2.22,lower.tail =FALSE)
# p-value = 0.01321 < 0.05 therefore, H0 is rejected.

# example: construction company
#-------------------------------------
# test statistic t = 3.162
pt(3.162,9,lower.tail=FALSE) + pt(-3.162,9)
# p-value = 0.01151 < 0.05 therefore, H0 is rejected.

# example: graduate survey
#-------------------------------------
# test statistic z = -1.23
pnorm(1.23,lower.tail=FALSE) + pnorm(-1.23)
# p-value = 0.2187 > 0.05 therefore, H0 is not rejected.
```