

# Course 02402 Introduction to Statistics

## Lecture 5: Hypothesis testing

DTU Compute  
Technical University of Denmark  
2800 Lyngby – Denmark

## Overview

- 1 Motivating example - sleep medicine
- 2 One-sample  $t$ -test and  $p$ -value
- 3 Critical value and relation to the confidence interval
- 4 Hypothesis tests in general
  - The alternative hypothesis
  - The general method
  - Errors in hypothesis testing
- 5 Checking the normality assumption
  - The normal q-q plot
  - Transformation towards normality

## Overview

- 1 Motivating example - sleep medicine
- 2 One-sample  $t$ -test and  $p$ -value
- 3 Critical value and relation to the confidence interval
- 4 Hypothesis tests in general
  - The alternative hypothesis
  - The general method
  - Errors in hypothesis testing
- 5 Checking the normality assumption
  - The normal q-q plot
  - Transformation towards normality

## Motivating example - sleep medicine

### Difference between sleep medicines?

In a study, the aim is to compare two kinds of sleep medicine, A and B. 10 test persons tried both kinds of medicine, and the following 10 *differences* between the two types of medicine were measured:  
(For person 1, sleep medicine B was 1.2 sleep hours better than medicine A, etc.):

Sample,  $n = 10$ :

person	$x = \text{B effect} - \text{A effect}$
1	1.2
2	2.4
3	1.3
4	1.3
5	0.9
6	1.0
7	1.8
8	0.8
9	4.6
10	1.4

## Example - sleep medicine

The hypothesis of no difference:

$$H_0: \mu = 0$$

where  $\mu$  represents mean difference in sleep length.

Sample mean and std. deviation:

$$\bar{x} = 1.670 = \hat{\mu}$$

$$s = 1.13 = \hat{\sigma}$$

Is data in accordance with the null hypothesis  $H_0$ ?

$$\text{Data: } \bar{x} = 1.67, H_0: \mu = 0$$

NEW  $p$ -value:

$$p\text{-value} = 0.00117$$

(Computed under the scenario that  $H_0$  is true).

NEW Conclusion:

As the data is far away from  $H_0$  (unlikely under  $H_0$ ), we **reject**  $H_0$ . There is a **significant effect** of medicine B compared to A.

## Overview

- 1 Motivating example - sleep medicine
- 2 One-sample  $t$ -test and  $p$ -value
- 3 Critical value and relation to the confidence interval
- 4 Hypothesis tests in general
  - The alternative hypothesis
  - The general method
  - Errors in hypothesis testing
- 5 Checking the normality assumption
  - The normal q-q plot
  - Transformation towards normality

## Method 3.23: One-sample $t$ -test and $p$ -value

How to compute the  $p$ -value?

For a (quantitative) one sample situation, the (non-directional)  $p$ -value is given by:

$$p\text{-value} = 2 \cdot P(T > |t_{\text{obs}}|)$$

where  $T$  follows a  $t$ -distribution with  $(n - 1)$  degrees of freedom.

The observed value of the test statistics to be computed is

$$t_{\text{obs}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where  $\mu_0$  is the value of  $\mu$  under the null hypothesis:

$$H_0: \mu = \mu_0$$

## The definition and interpretation of the $p$ -value (completely general)

The  $p$ -value expresses *evidence* against the null hypothesis – Table 3.1:

$p < 0.001$	Very strong evidence against $H_0$
$0.001 \leq p < 0.01$	Strong evidence against $H_0$
$0.01 \leq p < 0.05$	Some evidence against $H_0$
$0.05 \leq p < 0.1$	Weak evidence against $H_0$
$p \geq 0.1$	Little or no evidence against $H_0$

Definition 3.22 of the  $p$ -value:

**The  $p$ -value** is the probability of obtaining a test statistic that is at least as extreme as the test statistic that was actually observed. This probability is calculated under the assumption that the null hypothesis is true.

## Example - sleep medicine

The hypothesis of no difference:

$$H_0: \mu = 0$$

where  $\mu$  represents mean difference in sleep length.

Compute the test-statistic:

$$t_{\text{obs}} = \frac{1.67 - 0}{1.13/\sqrt{10}} = 4.67$$

Compute the  $p$ -value:

$$2P(T > 4.67) = 0.00117$$

```
2 * (1 - pt(4.67, df = 9))
```

Interpretation of the  $p$ -value in light of Table 3.1:

There is strong evidence against the null hypothesis.

## Example - sleep medicine - in R, manually

```
# Enter data
x <- c(1.2, 2.4, 1.3, 1.3, 0.9, 1.0, 1.8, 0.8, 4.6, 1.4)
n <- length(x) # sample size

# Compute 'tobs' - the observed test statistic
tobs <- (mean(x) - 0) / (sd(x) / sqrt(n))

# Compute the p-value as a tail-probability
# in the relevant t-distribution:
2 * (1 - pt(abs(tobs), df = n-1))

## [1] 0.001166
```

## Example - sleeping medicine - in R, with built-in function

```
t.test(x)

##
## One Sample t-test
##
## data: x
## t = 4.7, df = 9, p-value = 0.001
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.8613 2.4787
## sample estimates:
## mean of x
##      1.67
```

## Definition of a hypothesis test and significance (general)

**Definition 3.24. Hypothesis test:**

We say that we *carry out a hypothesis test* when we decide against a null hypothesis or not, using the data.

A null hypothesis is *rejected* if the  $p$ -value, calculated after the data has been observed, is less than some  $\alpha$ , that is if the  $p$ -value  $< \alpha$ , where  $\alpha$  is some pre-specified (so-called) *significance level*.

Otherwise, the null hypothesis is said to be '*accepted*'.

**Definition 3.29. Statistical significance:**

An *effect* is said to be (*statistically*) *significant* if the  $p$ -value is less than the significance level  $\alpha$ .

Often, we use  $\alpha = 0.05$ .

## Example - sleep medicine

With  $\alpha = 0.05$ :

Since the  $p$ -value is less than  $\alpha$ , we **reject** the null hypothesis.

In conclusion:

We have found a **significant effect** of medicine B when compared to A (and B works better than A).

## Overview

- 1 Motivating example - sleep medicine
- 2 One-sample  $t$ -test and  $p$ -value
- 3 Critical value and relation to the confidence interval
- 4 Hypothesis tests in general
  - The alternative hypothesis
  - The general method
  - Errors in hypothesis testing
- 5 Checking the normality assumption
  - The normal q-q plot
  - Transformation towards normality

## Critical value

Definition 3.31 - the critical values of the  $t$ -test:

The  $(1 - \alpha)100\%$  critical values for the (non-directional) one-sample  $t$ -test are the  $(\alpha/2)100\%$  and  $(1 - \alpha/2)100\%$  quantiles of the  $t$ -distribution with  $n - 1$  degrees of freedom:

$$t_{\alpha/2} \text{ and } t_{1-\alpha/2}$$

Method 3.32: One-sample  $t$ -test by critical value:

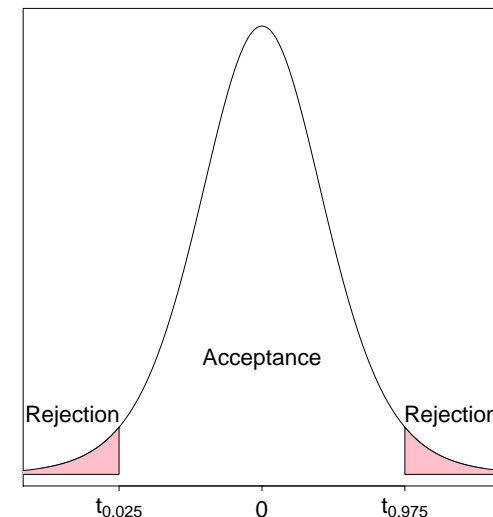
A null hypothesis is *rejected* if the observed test-statistic is more extreme than the critical values:

$$\text{If } |t_{\text{obs}}| > t_{1-\alpha/2} \text{ then reject}$$

otherwise *accept*.

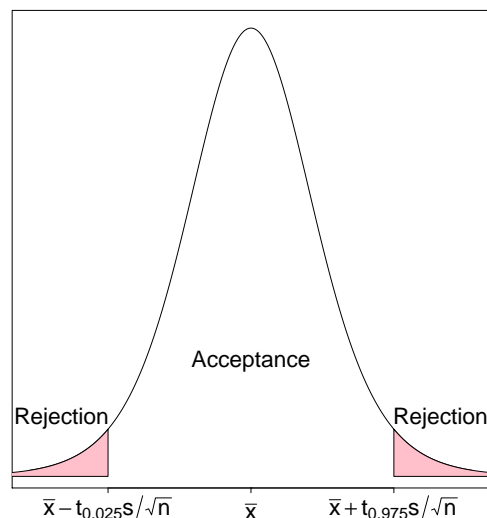
## Critical value and hypothesis test

The acceptance region consists of the values of  $\mu$  which are not too far away from the sample mean - here on the standardized scale:



## Critical value and hypothesis test

The acceptance region consists of the values of  $\mu$  which are not too far away from the sample mean - now on the original scale:



## Critical value, confidence interval and hypothesis test

Theorem 3.33: Critical value method = Confidence interval method

We consider a  $(1 - \alpha) \cdot 100\%$  confidence interval for  $\mu$ :

$$\bar{x} \pm t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

The confidence interval corresponds to the acceptance region for  $H_0$  when testing the (non-directional) hypothesis

$$H_0: \mu = \mu_0$$

(New) interpretation of the confidence interval:

The confidence interval covers those values of the parameter that we believe in given the data.

(Those values that we accept by the corresponding hypothesis test.)

## Proof:

Remark 3.34

A  $\mu_0$  inside the confidence interval satisfies that

$$|\bar{x} - \mu_0| < t_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

which is equivalent to

$$\frac{|\bar{x} - \mu_0|}{\frac{s}{\sqrt{n}}} < t_{1-\alpha/2}$$

and again to

$$|t_{\text{obs}}| < t_{1-\alpha/2}$$

which then exactly states that  $\mu_0$  is accepted, since the  $t_{\text{obs}}$  is within the critical values.

## Overview

- 1 Motivating example - sleep medicine
- 2 One-sample  $t$ -test and  $p$ -value
- 3 Critical value and relation to the confidence interval
- 4 Hypothesis tests in general
  - The alternative hypothesis
  - The general method
  - Errors in hypothesis testing
- 5 Checking the normality assumption
  - The normal q-q plot
  - Transformation towards normality

## The alternative hypothesis

So far - implied: (= non-directional)

The alternative to  $H_0: \mu = \mu_0$  is  $H_1: \mu \neq \mu_0$ .

BUT there are other possible settings, e.g. one-sided (= directional), "less":

The alternative to  $H_0: \mu = \mu_0$  is  $H_1: \mu < \mu_0$ .

We stick to the "non-directional" in this course!

## Steps of a hypothesis test - an overview

Generally, a hypothesis test consists of the following steps:

- 1 Formulate the hypothesis and choose the level of significance  $\alpha$  (choose the "risk-level").
- 2 Calculate, using the data, the value of the test statistic.
- 3 Calculate the  $p$ -value using the test statistic and the relevant distribution. Compare the  $p$ -value to the significance level  $\alpha$  and make a conclusion.

**OR:**

Alternatively, make a conclusion based on the relevant critical value(s).

## The one-sample t-test again

Method 3.36 The level  $\alpha$  one-sample t-test:

- 1 Compute  $t_{\text{obs}}$  as before.
- 2 Compute evidence against the *null hypothesis*  $H_0: \mu = \mu_0$  vs. the *alternative hypothesis*  $H_1: \mu \neq \mu_0$  by the

$$p\text{-value} = 2 \cdot P(T > |t_{\text{obs}}|),$$

where the  $t$ -distribution with  $n - 1$  degrees of freedom is used.

- 3 If  $p\text{-value} < \alpha$ , we reject  $H_0$ . Otherwise, we accept  $H_0$ .

**OR:**

The rejection/acceptance conclusion could alternatively, but equivalently, be made based on the critical value(s)  $\pm t_{1-\alpha/2}$ :

If  $|t_{\text{obs}}| > t_{1-\alpha/2}$  we reject  $H_0$ , otherwise we accept  $H_0$ .

## Errors in hypothesis testing

Two kind of errors can occur (but only one at a time!):

Type I: Rejection of  $H_0$  when  $H_0$  is true.

Type II: Non-rejection (acceptance) of  $H_0$  when  $H_1$  is true.

The risks of the two types of errors are:

$$P(\text{Type I error}) = \alpha$$

$$P(\text{Type II error}) = \beta$$

## Court of law analogy

A man is standing in a court of law:

A man is standing in a court of law, accused of criminal activity.

The null- and the alternative hypotheses are:

$H_0$ : The man is not guilty.

$H_1$ : The man is guilty.

Not being able to prove that the man is guilty is not the same as *proving* that he is innocent.

Or, put differently:

Accepting a null hypothesis is NOT a statistical proof of the null hypothesis being true!

## Errors in hypothesis testing

Theorem 3.39: Significance level = The risk of a Type I error

The significance level  $\alpha$  in hypothesis testing is the overall Type I risk:

$$P(\text{Type I error}) = P(\text{Rejection of } H_0 \text{ when } H_0 \text{ is true}) = \alpha$$

Two possible truths vs. two possible conclusions:

	Reject $H_0$	Fail to reject $H_0$
$H_0$ is true	Type I error ( $\alpha$ )	Correct acceptance of $H_0$
$H_0$ is false	Correct rejection of $H_0$ (Power)	Type II error ( $\beta$ )

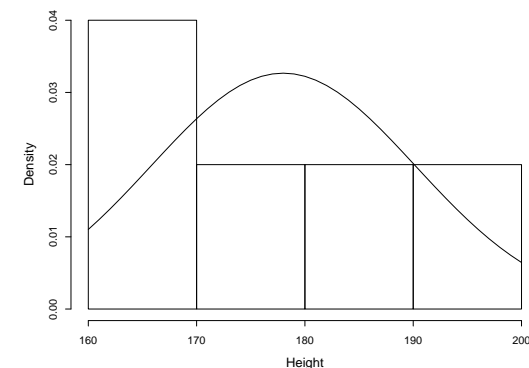
## Overview

- 1 Motivating example - sleep medicine
- 2 One-sample  $t$ -test and  $p$ -value
- 3 Critical value and relation to the confidence interval
- 4 Hypothesis tests in general
  - The alternative hypothesis
  - The general method
  - Errors in hypothesis testing
- 5 Checking the normality assumption
  - The normal q-q plot
  - Transformation towards normality

## Example - student heights

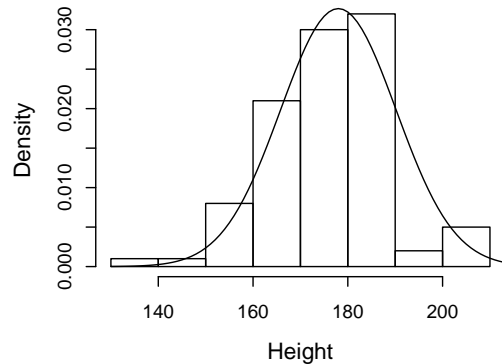
```
# Student heights data
x <- c(168, 161, 167, 179, 184, 166, 198, 187, 191, 179)

# Density histogram of student height data together with normal pdf
hist(x, xlab = "Height", main = "", freq = FALSE)
lines(seq(160, 200, 1), dnorm(seq(160, 200, 1), mean(x), sd(x)))
```



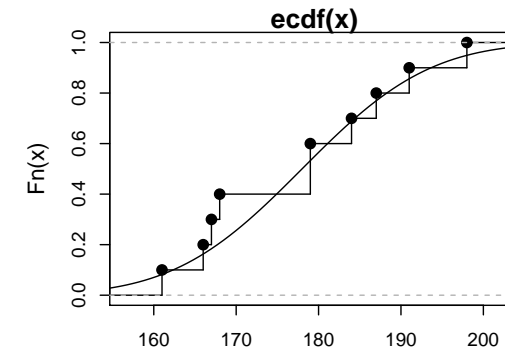
## Example - 100 observations from a normal distribution

```
# Density histogram of simulated data from normal distribution
# (n = 100) together with normal pdf
xr <- rnorm(100, mean(x), sd(x))
hist(xr, xlab = "Height", main = "", freq = FALSE, ylim = c(0, 0.032))
lines(seq(130, 230, 1), dnorm(seq(130, 230, 1), mean(x), sd(x)))
```



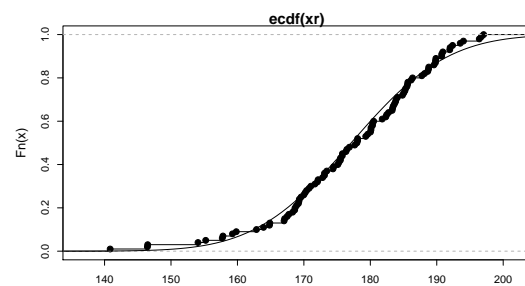
## Example - student heights - ecdf

```
# Empirical cdf for student height data together
# with normal cdf
plot(ecdf(x), verticals = TRUE)
xp <- seq(0.9*min(x), 1.1*max(x), length.out = 100)
lines(xp, pnorm(xp, mean(x), sd(x)))
```



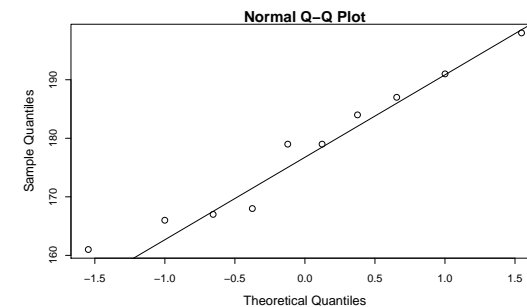
## Example - 100 observations from a normal distribution - ecdf

```
# Empirical cdf of simulated data from normal distribution
# (n = 100) together with normal cdf
xr <- rnorm(100, mean(x), sd(x))
plot(ecdf(xr), verticals = TRUE)
xp <- seq(0.9*min(xr), 1.1*max(xr), length.out = 100)
lines(xp, pnorm(xp, mean(xr), sd(xr)))
```



## Example - student heights - normal q-q plot

```
# Normal q-q plot of student heights
qqnorm(x)
qqline(x)
```





# Normal q-q plot

## Method 3.42- The formal definition

The ordered observations  $x_{(1)}, \dots, x_{(n)}$  are plotted versus a set of expected normal quantiles  $z_{p_1}, \dots, z_{p_n}$ . Different definitions of  $p_1, \dots, p_n$  exist:

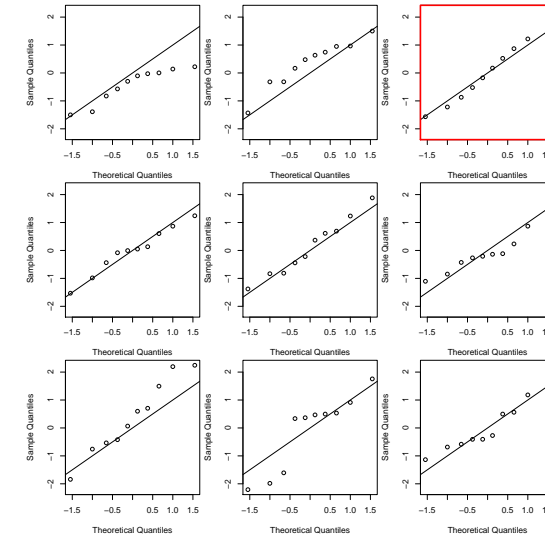
- In R, when  $n > 10$ :

$$p_i = \frac{i - 0.5}{n}, i = 1, \dots, n$$

- In R, when  $n \leq 10$ :

$$p_i = \frac{i - 3/8}{n + 1/4}, i = 1, \dots, n$$

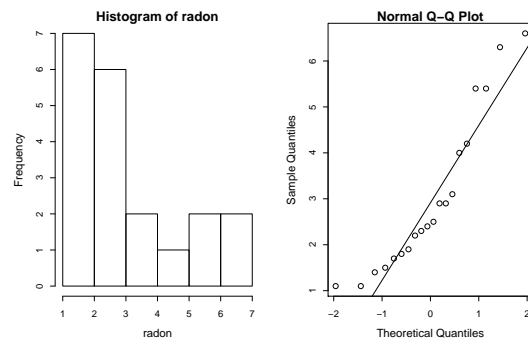
# Example - student heights - compare with other simulated normal distributed data



# Example - Radon data

```
## Reading in the data
radon <- c(2.4, 4.2, 1.8, 2.5, 5.4, 2.2, 4.0, 1.1, 1.5, 5.4, 6.3,
          1.9, 1.7, 1.1, 6.6, 3.1, 2.3, 1.4, 2.9, 2.9)

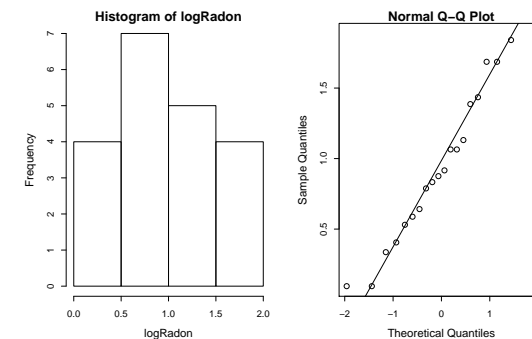
## Histogram and q-q plot of data
par(mfrow = c(1,2))
hist(radon)
qqnorm(radon)
qqline(radon)
```



# Example - Radon data - log-transformed data are closer to a normal distribution

```
# Transform data using the natural logarithm
logRadon <- log(radon)

## Histogram and q-q plot of transformed data
par(mfrow = c(1,2))
hist(logRadon)
qqnorm(logRadon)
qqline(logRadon)
```



## Agenda

- 1 Motivating example - sleep medicine
- 2 One-sample  $t$ -test and  $p$ -value
- 3 Critical value and relation to the confidence interval
- 4 Hypothesis tests in general
  - The alternative hypothesis
  - The general method
  - Errors in hypothesis testing
- 5 Checking the normality assumption
  - The normal q-q plot
  - Transformation towards normality