Course 02402 Introduction to Statistics Lecture 5:

# One-sample hypothesis test and model control

Per Bruun Brockhoff

DTU Compute
Danish Technical University
2800 Lyngby – Denmark
e-mail: perbb@dtu.dk



## Agenda

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





## Oversigt

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





## Motivating example - sleeping medicine

### Difference of sleeping medicines?

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

Sample, $n = 10$:

| person | $x =$ Beffect - Aeffect |
|---|---|
| 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 - sleeping medicine

The hypothesis of no difference:

$$H_0: \ \mu = 0$$

Sample mean and standard deviation:

$$\bar{x} = 1.670 = \hat{\mu}$$
$$s = 1.13 = \hat{\sigma}$$

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

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 unlike far away from $H_0$, we **reject** $H_0$ - we have found a **significant effect** of sleep medicine B as compared to A.





## Oversigt







## Method 3.22: 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 the *evidence* against the null hypothesis – Table **??**:

| | |
|---|---|
| $p < 0.001$ | 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 3.21 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 - sleeping medicine

The hypothesis of no difference:

$$H_0: \ \mu = 0$$

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, 9))
```

Interpretation of the $p$-value in light of Table **??**:

There is strong evidence agains the null hypothesis.





# Example - sleeping 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)
## Compute the tobs - the observed test statistic:
tobs <- (mean(x) - 0) / (sd(x) / sqrt(n))
## Compute the p-value as a tail-probability
## in the t-distribution:
pvalue <- 2 * (1-pt(abs(tobs), df=n-1))
pvalue

## [1] 0.0012
```





# Example - sleeping medicine - in R - with inbuilt function

```
t.test(x)

##
##  One Sample t-test
##
## data:  x
## t = 5, df = 9, p-value = 0.001
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.86 2.48
## sample estimates:
## mean of x
##       1.7
```





# The definition of hypothesis test and significance (generally)

Definition 3.23. 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-specifed (so-called) *significance level*. And if not, then the null hypothesis is said to be *accepted*.

Definition 3.28. 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 - sleeping medicine

**With $\alpha = 0.05$ we can conclude:**

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

**And hence:**

We have found a **significant effect** af medicine B as compared to A. (And hence that B works better than A)





## Oversigt







## Critical value

**Definition 3.30 - 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:

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

**Metode 3.31: One-sample $t$-test by critical value:**

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

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

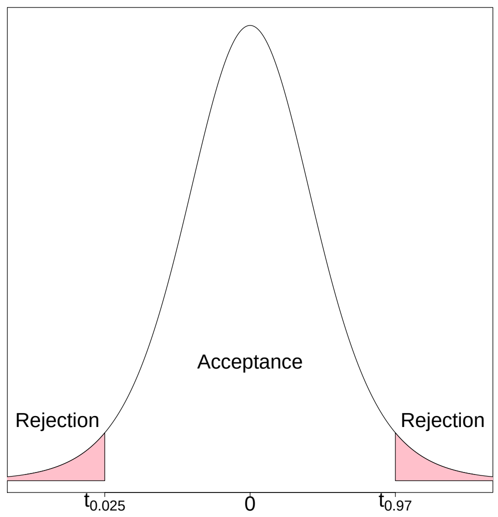
otherwise *accept*.





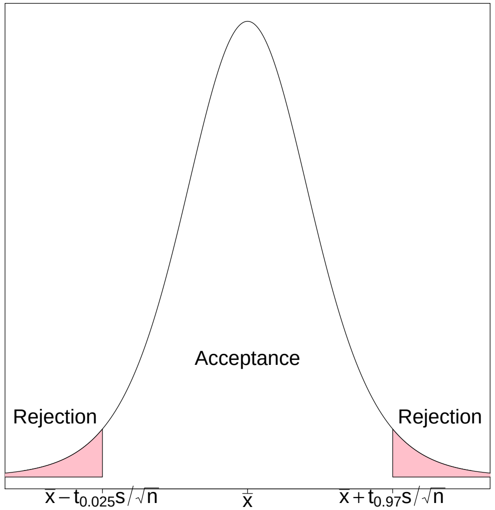
## Critical value and hypothesis test

The acceptance region are the values for $\mu$ not too far away from the data - here on the standardized scale:



## Critical value and hypothesis test

The acceptance region are the values for $\mu$ not too far away from the data - now on the original scale:








## Critical value, confidence interval and hypothesis test

**Theorem ??: 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

$$H_0: \quad \mu = \mu_0$$

**(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 3.33**

A $\mu_0$ inside the confidence interval will fullfill 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.





## Oversigt

## 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 : $H_1: \ \mu < \mu_0$

But we stick to the "non-directional" in this course





## Steps by hypothesis tests - an overview

Generelly a hypothesis test consists of the foloowing steps:

1. Formulate the hypotheses 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 sampling distribution, and compare the $p$-value and 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 3.35 The level $\alpha$ test is:

1. Compute $t_{\text{obs}}$ as before
2. Compute the 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 $p$–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 or errors:

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

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

## 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 the alternative hypotheses are:

$H_0$ : The man is not guilty

$H_1$ : The man is guilty

**That you cannot be proved guilty is not the same as being proved innocent**

Or differently put:
*Accepting* a null hypothesis is NOT a statistical proof of the null hypothesis being true!





## Errors in hypothesis testing

**Theorem 3.38: 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 ($\beta$) |





## Oversigt

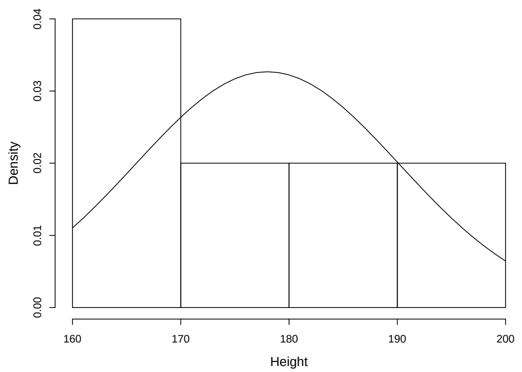






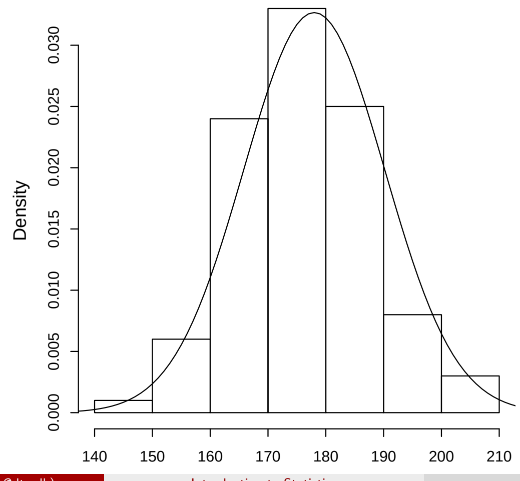
## Example - student heights - are they normally distributed?

```r
x <- c(168,161,167,179,184,166,198,187,191,179)
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:

```r
xr <- rnorm(100, mean(x), sd(x))
hist(xr, xlab="Height", main="", freq = FALSE)
lines(seq(130, 230, 1), dnorm(seq(130, 230, 1), mean(x), sd(x)))
```

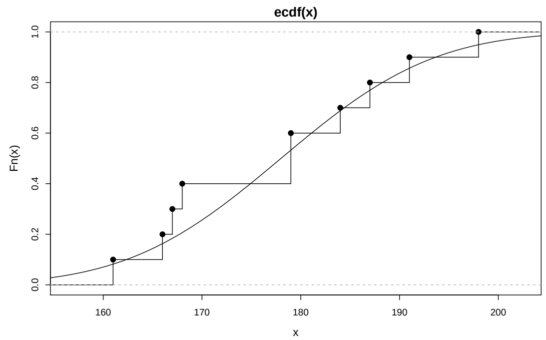







## Example - student heights - ecdf

```r
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)))
```

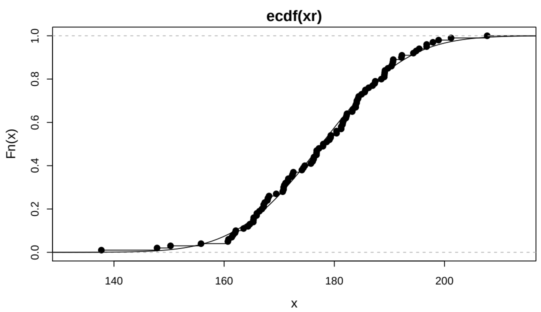







## Example - 100 observations from a normal distribution, ecdf:

```r
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)))
```

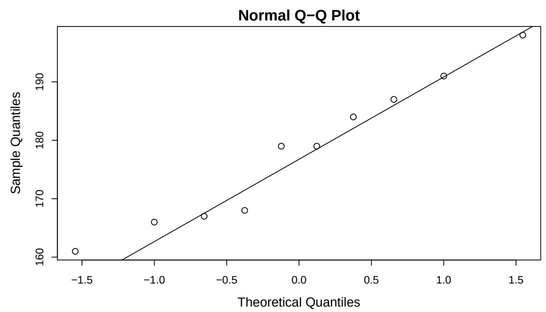




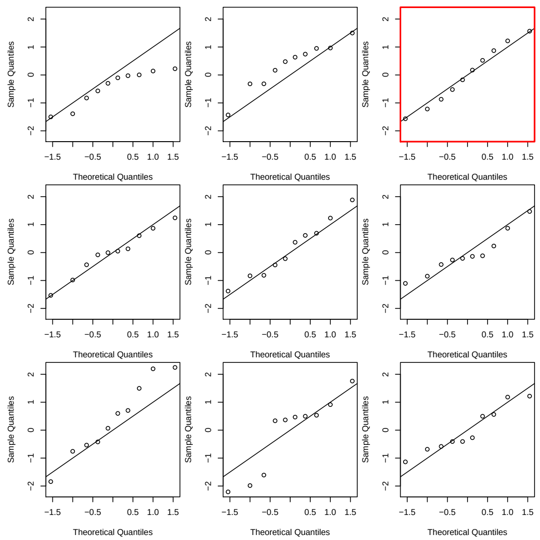



## Example - student heights - Normal Q-Q plot

```r
qqnorm(x)
qqline(x)
```



# Example - student heights - Normal Q-Q plot - compare with other simulated normally distributed data








# Normal Q-Q plot

**Metode 3.41- 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 $n > 10$:

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

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

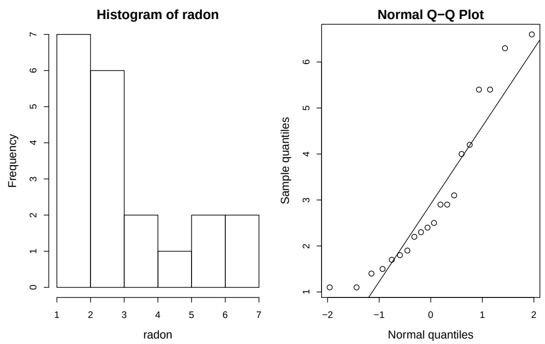
$$p_i = \frac{i-3/8}{n+1/4},\ i=1,\dots,n$$





# 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)
##A HISTOGRAM AND A QQ-PLOT
par(mfrow=c(1,2))
hist(radon)
qqnorm(radon,ylab = 'Sample quantiles',xlab = "Normal quantiles")
qqline(radon)
```



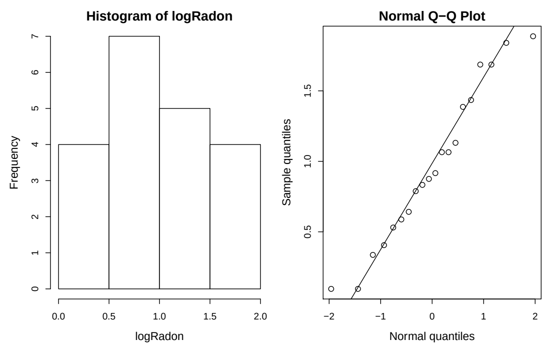





# Example - Radon data - log-transformed are closer to a normal distribution

```
##TRANSFORM USING NATURAL LOGARITHM
logRadon<-log(radon)

hist(logRadon)
qqnorm(logRadon,ylab = 'Sample quantiles',xlab = "Normal quantiles")
qqline(logRadon)
```



# Agenda

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