

Parameter estimation and hypothesis testing

Josep L. Carrasco
Bioestadística. Departament de Fonaments Clínics
Universitat de Barcelona

Working with samples

- Aim: to know the population value of parameters (mean, variance, proportion).
- Procedure: inferential statistics.
- Process:
 - 1) Take a part of the population (sample).
 - 2) Apply a statistical function (estimator) to the sample.
 - 3) Translate the result (estimate) to the population.

Example. A research study aims to know the mean (**parameter**) of the body mass index (BMI) in a population of adult healthy people. A sample of 50 subjects are collected from the population and their BMI is computed.

We compute the arithmetic mean (**estimator**) of the sample values. The resulting mean of the 50 subjects is 23.77. We call this value as the **estimate** of the mean.

- **Sampling error:** difference between the sample estimate and the true value of parameter. Remember that the value of the parameter is unknown, so the sampling error is also unknown.
- **Aim of inferential statistics:** to know, assess and control the sampling error.

A parameter may have different estimators. The best estimator is that with best properties.

Properties of the estimators

Let $\hat{\theta}$ be the estimator of parameter θ .

- The estimators are random variables because their values **randomly change** from one sample to another.

Example. Following the BMI example, the researchers took a different sample of 50 subjects. The mean BMI of this second sample was 24.25.

Given that estimators are random variables they have expectation and variance.

- **Bias.** An estimator is **unbiased** if $E(\hat{\theta}) = \theta$

In the case of the sample mean, assuming that all the x_i (data values) come from the same population with mean μ , i.e $E(x_i) = \mu$

$$E(\bar{x}) = E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} n\mu = \mu$$

- **Standard error.** The standard deviation of an estimator is known as standard error.

Following the case of the sample mean, let's determine its variance assuming that all the x_i are independent and come from the same population with variance σ^2 .

$$Var(\bar{x}) = Var\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n Var(x_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

The standard error of the sample mean is

$$SE(\bar{x}) = \sqrt{Var(\bar{x})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

The standard error also expresses the **magnitude** of the sampling error.

- Sampling error: $\epsilon = \hat{\theta} - \theta$

$$Var(\epsilon) = Var(\hat{\theta} - \theta) = Var(\hat{\theta})$$

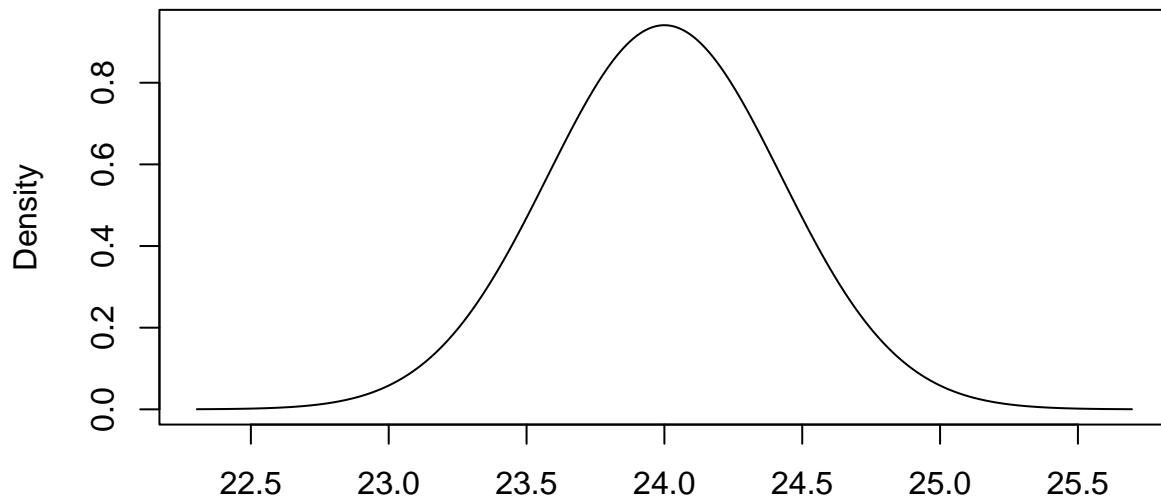
- **Sample distribution.** The probability distribution model of the estimator.

Example. Let's suppose that the BMI values in this population are distributed as a Normal model with a mean of $\mu = 24$ and a standard deviation of $\sigma = 3$.

Following central limit theorem, the sample distribution of the mean in samples of size $n = 50$ will be a Normal distribution with mean $\mu = 24$ and standard deviation (standard error):

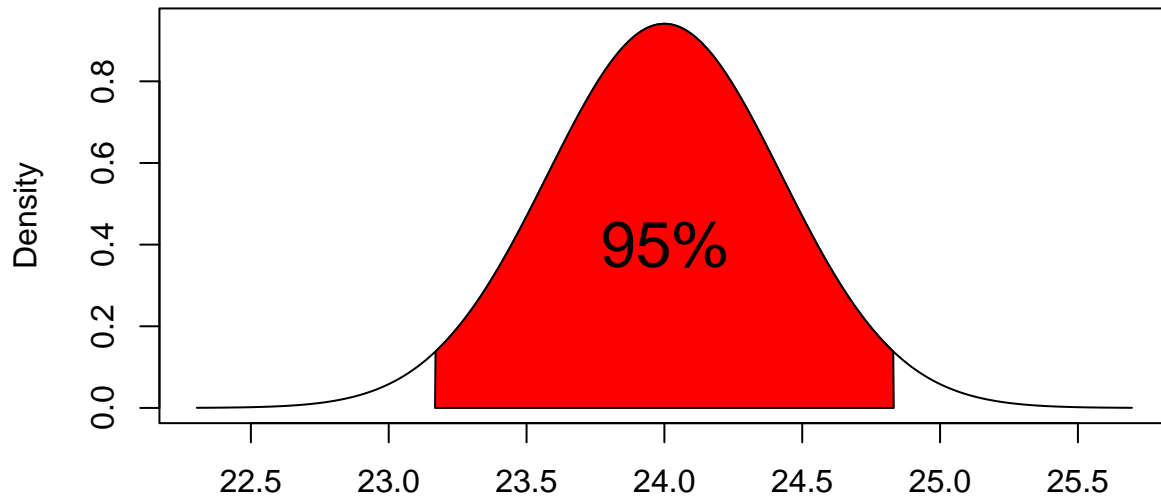
$$\frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{50}} = 0.424$$

Density Function of the sample mean



By knowing the sample distribution we can assess the sampling error. For example, with a probability of 95% the sample means will take values between 23.168 and 24.832.

Density Function of the sample mean



Confidence intervals

The sample distribution is determined by the parameters. However, the parameters are unknown. The way to proceed is assign to parameters the estimates, so that the sample distribution is *guessed* rather than known.

Additionally, the goal of the inference is to predict the parameter value. We could use the estimate and affirm “the mean of BMI is 23.77”. However, the probability that this assertion is true is 0 (remember that the probability of a particular value in a continuous variable is 0).

A better approach is giving a range of values that may include the parameter with a certain probability.

Confidence interval: probability interval applied on the *guessed* sample distribution.

A $(1 - \alpha)$ **confidence interval** is a range of values that has a probability $(1 - \alpha)$ of including the parameter value.

Example. Let’s suppose that we know the standard deviation is $\sigma = 3$ but the mean is unknown. From the results of the sample we obtain an estimate of $\bar{x} = 23.77$. The *guessed* sample distribution is a Normal distribution with mean 23.77 and standard error 0.424.

- 95% confidence interval

```
m<-mean(BMI)
sem<-3/sqrt(50)
l195<-qnorm(0.025,m,sem)
u195<-qnorm(0.975,m,sem)
c(l195,u195)
```

```
[1] 22.94246 24.60554
```

Our guess is that the mean (parameter) is between 22.942 and 24.606 with a 95% of confidence.

Sample size

A confidence interval must be informative, i.e. precise. The precision of an interval is expressed by its length.

For example, let $[23, 27]$ and $[20, 30]$ be two confidence intervals about the mean of the BMI. The first is narrower therefore more precise and informative.

The **length of the confidence intervals** depends on:

- The **level of confidence**. The higher is the confidence the wider is the interval.
- The **sample size**. The larger is the sample size the narrower is the interval.
- The **standard error**. The larger is the standard error the wider is the interval.

When the research is focused on estimating a parameter, the success is assessed by the degree of precision of the confidence intervals.

The common way to control the confidence interval width is to compute the sample size before collecting the sample. The computation approach depends on the parameter to estimate.

Hypothesis testing

- Aim: take a decision about a research hypothesis.
- Hypothetical-deductive method:
 - 1) Define the hypothesis.
 - 2) Apply the falsifiability principle: try to demonstrate the hypothesis is false.
 - 3) Not refusing a hypothesis does NOT mean the hypothesis is true.

Example. Let's suppose we want to know if Biomedical Engineering students at UB are all born in Barcelona province.

- 1) Define the hypothesis.

The Biomedical Engineering students at UB are all born in Barcelona

- 2) Apply the falsifiability principle: try to demonstrate the hypothesis is false.

Ask to students if they are born in Barcelona. If at least one answers "No" the hypothesis will be false.

- 3) Not refusing a hypothesis does NOT mean the hypothesis is true.

If after asking to a number of students all the answers were "Yes", it does not mean the hypothesis is true (unless we asked to the whole population).

- Steps of hypothesis statistical testing procedure:
 - 1) Set the statistical hypotheses
 - 2) Statistical test
 - 3) Expected results
 - 4) Applicability conditions
 - 5) Decision rule

Example. Following with the example of BMI, the researchers wish to test if they can affirm that the mean BMI is greater than 22.5. This is the **research hypothesis**.

In terms of falsifiability principle, they must show that the statement "the mean BMI is lower or equal than 22.5" is false to be able to demonstrate the mean BMI is greater than 22.5.

1. Set the statistical hypotheses.

The statistical hypothesis testing involves expressing the research hypothesis in terms of parameters.

In the example the parameter that is involved in the research hypothesis is the mean.

There are two statistical hypothesis:

- **Null hypothesis.** It is defined as a restriction of the parameter space (equality). It is the hypothesis that we put on a trial to demonstrate its falsifiability.

In the example this hypothesis would be “lower or equal than 22.5”. However, the null hypothesis must set the parameter to a value. The boundary value of 22.5 is the best choice.

- **Alternative hypothesis:** the opposite of the null hypothesis. In the example it should be “greater than 22.5”.

It can be two-sided (difference) or one-sided (greater or lower) depending on the research hypothesis.

So that, the statistical hypotheses in the example are:

$$H_0 : \mu = 22.5$$

$$H_1 : \mu > 22.5$$

2. Statistical test

Choose an appropriate statistical test to assess the falsifiability of the hypothesis.

The procedure usually involves defining a function that assesses the distance between what is observed in data and what the null hypothesis says.

In the example the function could be the difference between the sample mean and the proposed mean in the null hypothesis (μ_0)

$$\theta = \bar{x} - \mu_0$$

3. Expected results

Define the expected results in case the hypothesis was true.

It means to determine the sample distribution of the statistical test if the null hypothesis was true.

From the central limit theorem, we know that if the sample size is large enough the sample mean is distributed as a Normal model.

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

If the null hypothesis was true $\mu = \mu_0$

$$\theta = (\bar{x} - \mu_0) \sim N\left(0, \frac{\sigma}{\sqrt{n}}\right)$$

We could go further and standardize the statistical test, i.e. divide it by its standard error σ/\sqrt{n}

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

The statistical test θ follows a standard Normal model (Normal with mean of 0 and standard deviation of 1).

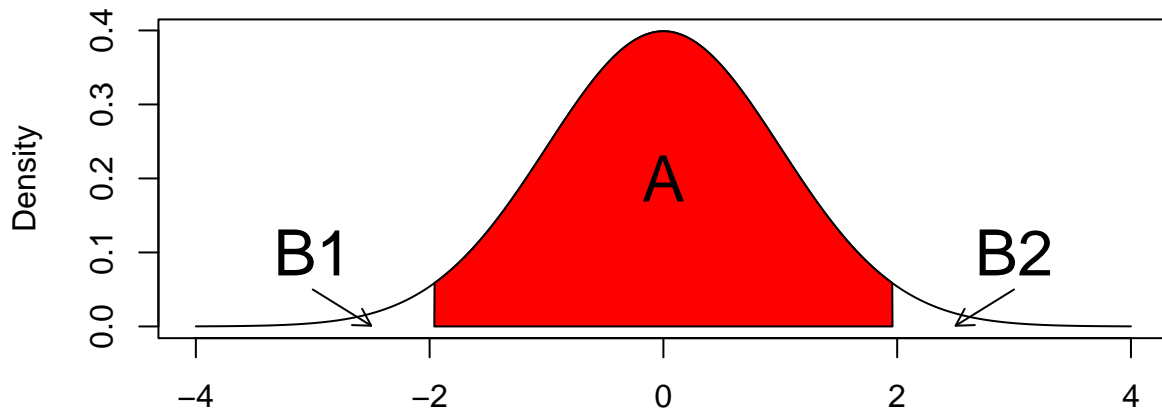
If the null hypothesis is true, we expect θ comes from a standard normal distribution.

- What are therefore the expected values?

We could consider those values with a certain probability, say $(1 - \alpha)$, to appear in case the true hypothesis is true. Commonly $(1 - \alpha)$ is set to be large, most common value is 95%, i.e $\alpha = 5\%$.

In the case of the example, the expected values of θ are drawn in the next figure.

Density Function of θ if H_0 is true

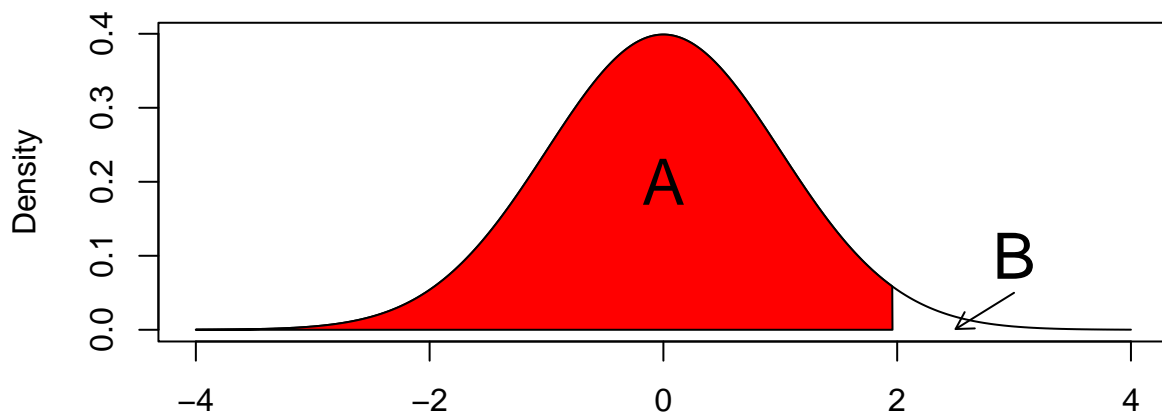


In the figure, zone A would express the expected results. It does not mean that values in areas B1 and B2 are impossible but less expected.

Concerning the example, remember that the null hypothesis was “lower or equal than 22.5” though it had to be set to a particular value. Thus, lower values of θ would be also expected in case the null hypothesis was true. This is a case of a **one-sided** (lower or greater) hypothesis instead of a **two-sided** hypothesis (difference) as it was case of the former figure.

So that, in the example the expected values of θ are:

Density Function of θ if H_0 is true



Where area B includes now the unexpected values.

4. Applicability conditions

Usually to determine the sample distribution of a statistic some conditions must be met (sample size, normality, ...).

Thus, it is necessary to check if these conditions are fulfilled before proceeding with the testing approach.

In the example, in order to assume that the sample mean follows a Normal model (or θ ultimately) one out of two conditions must be met:

- 1) The variable of analysis, BMI in the example, follows a Normal model.
- 2) The sample size is large (commonly 30 is accepted).

5. Decision rule

The decision rule is based on the expected values and it is supported on a probabilistic rule.

- If the statistical test lies in the unexpected values area we are going to **reject** the null hypothesis. We conclude that it is unlikely to observe such a result if the null hypothesis is true.
- If the statistical test does not lie in the unexpected values area we are going to **not reject** the null hypothesis. We conclude that it is quite probable to observe such a result if the null hypothesis is true.

So that, the decision is binary: **reject** or **not reject**.

The reality is also binary: null hypothesis is **true** or **false**.

NOTE. It is never known if the null hypothesis is really TRUE or FALSE. We are just guessing.

Type of decision errors

Based on the decision and reality it is possible to define the four situations that we could deal with.

		Decision	
		NO REJECT H_0	REJECT H_0
Reality	H_0 TRUE	Right Probability = $1 - \alpha$	Type I Error Probability = α
	H_0 FALSE	Type II Error Probability = β	Right Probability = $1 - \beta$

- What is the value of α ?

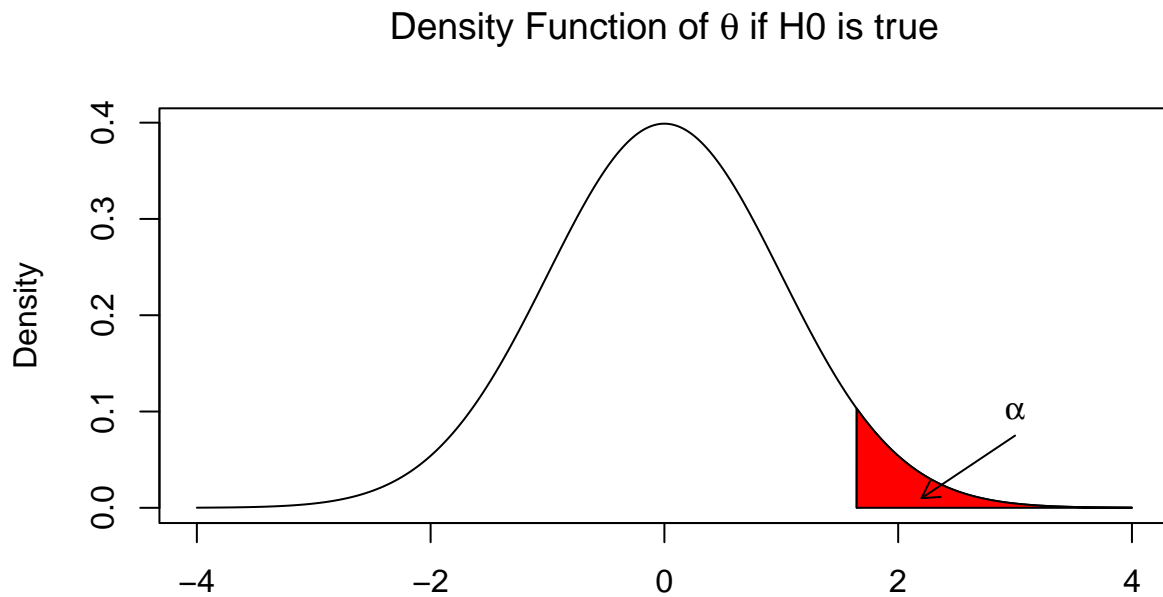
$$\alpha = P(\text{Reject } H_0 | H_0 \text{ is true})$$

It is also known as **significance level**.

Remember that “ H_0 is true” ultimately means θ follows a specific probability model that assumes the null hypothesis is true.

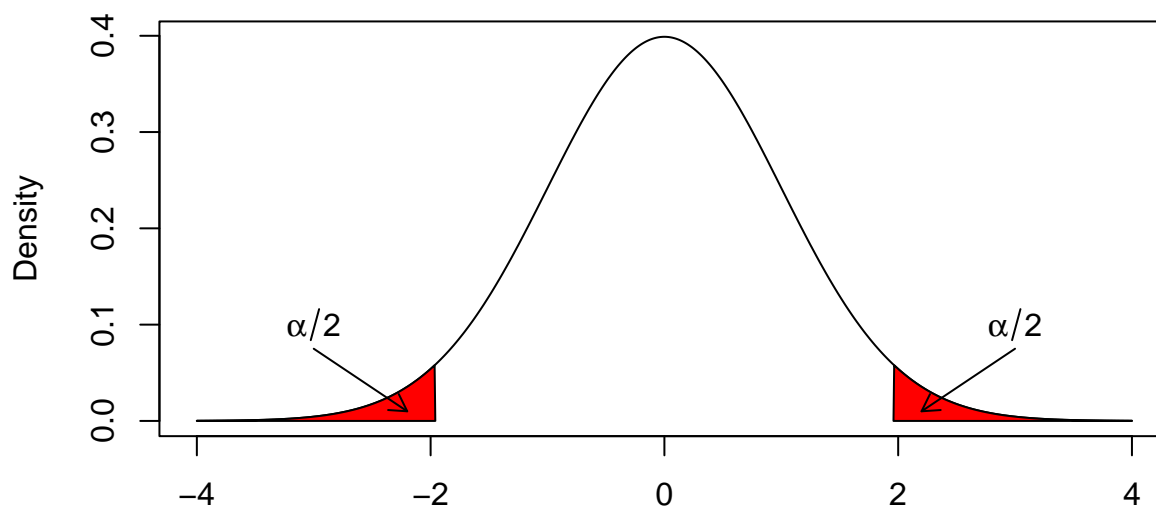
The null hypothesis is rejected in front an unexpected result, hence the probability of rejecting is the area B (or B1 plus B2 in the two-sided case).

Therefore, the rejection area is defined when α is set. Most widely accepted value is $\alpha = 5\%$. That means the risk of rejecting a true null hypothesis is 5%.



In case the alternative hypothesis was two-sided (difference) the rejection area has to be split in two subareas with probability $\alpha/2$.

Density Function of θ if H_0 is true



Power of the test

The probability of type-II error is:

$$\beta = P(\text{No Reject } H_0 | H_0 \text{ is false})$$

The opposite probability is known as **power** of the test:

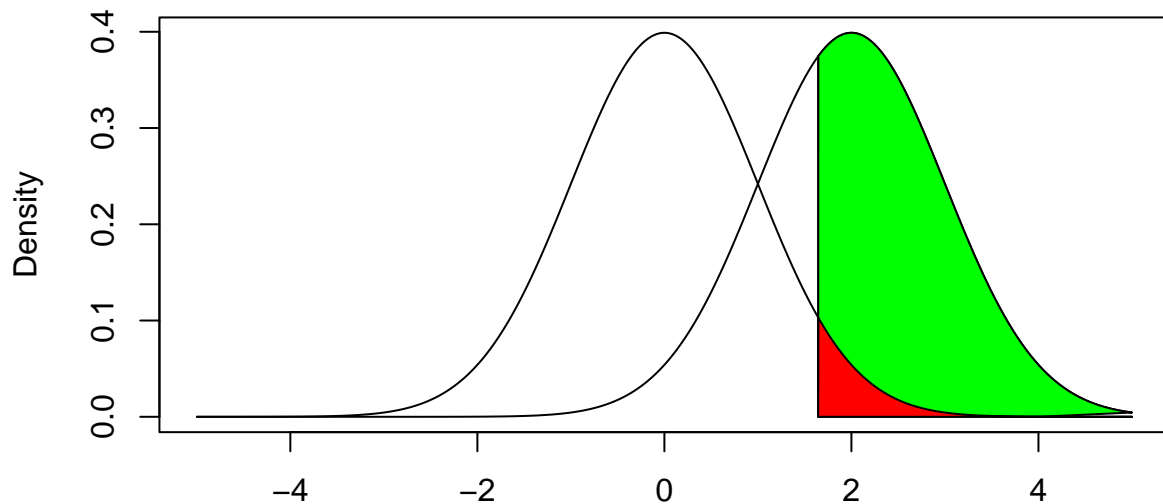
$$1 - \beta = P(\text{Reject } H_0 | H_0 \text{ is false})$$

The power is understood as the ability of the test to find false null hypothesis.

To compute the power the alternative hypothesis must be set to a value.

Let's suppose that under the alternative hypothesis $\theta \sim N(2, 1)$

Density Function of θ under H_0 and H_A



Red area: type-I error probability.

Green area: power of the test.

The power of the test depends on:

- The significance level.
- The difference to detect.
- The sample size.

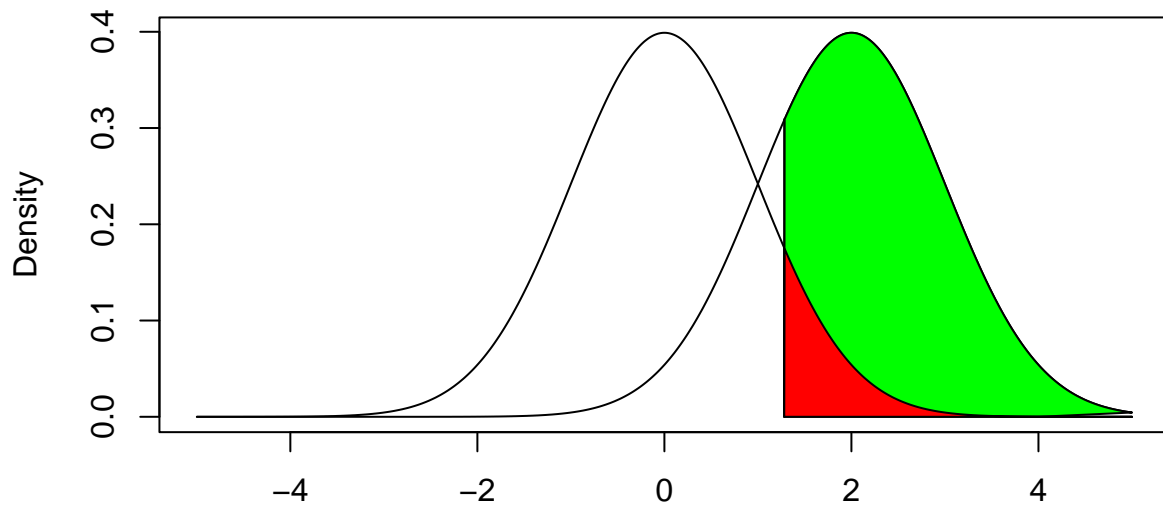
The significance level

The greater is the significance level the greater is the power.

It means that increasing the risk of one error diminishes the other error risk.

Let's draw the density plots again but now we are going to set $\alpha = 10\%$.

Density Function of θ under H_0 and H_A

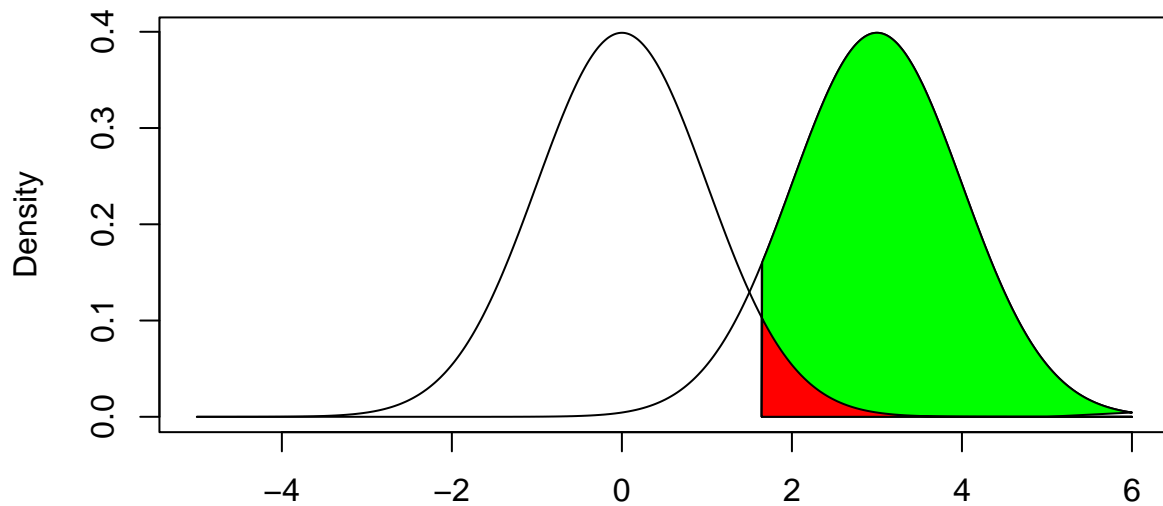


The difference to detect

It means the minimum difference we want to detect as significant, i.e the difference that the test has to detect as significant.

In the first densities chart, the difference was 2 because we had set the alternative hypothesis to 2. Let's see what happens if the alternative hypothesis is set to 3.

Density Function of θ under H_0 and H_A



Notice that:

Red area remains the same. Remember that α is pre-set.

Green area is larger. It means that it is easier to detect a bigger difference.

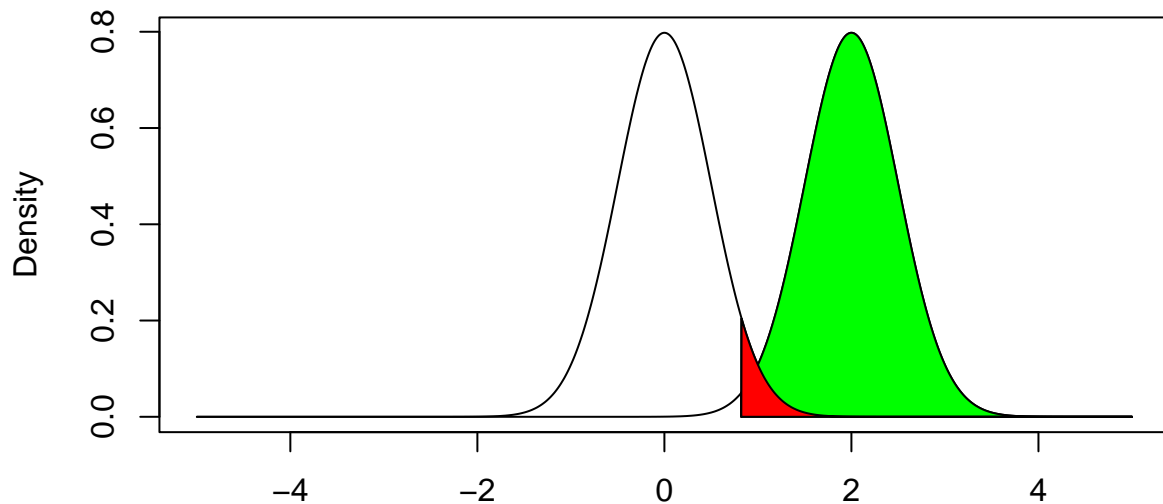
The sample size

The density function amplitude depends on the standard error.

The standard error depends on the sample size.

Let's suppose that the sample size is increased in such a way that the standard error decreases by half.

Density Function of θ under H0 and HA



When planning a research to carry out a hypothesis testing, the sample size is computed to reach a lower bound of power. Common values for the desired power are between 70% and 80%.

P-value

The p-value is also known as **degree of significance**.

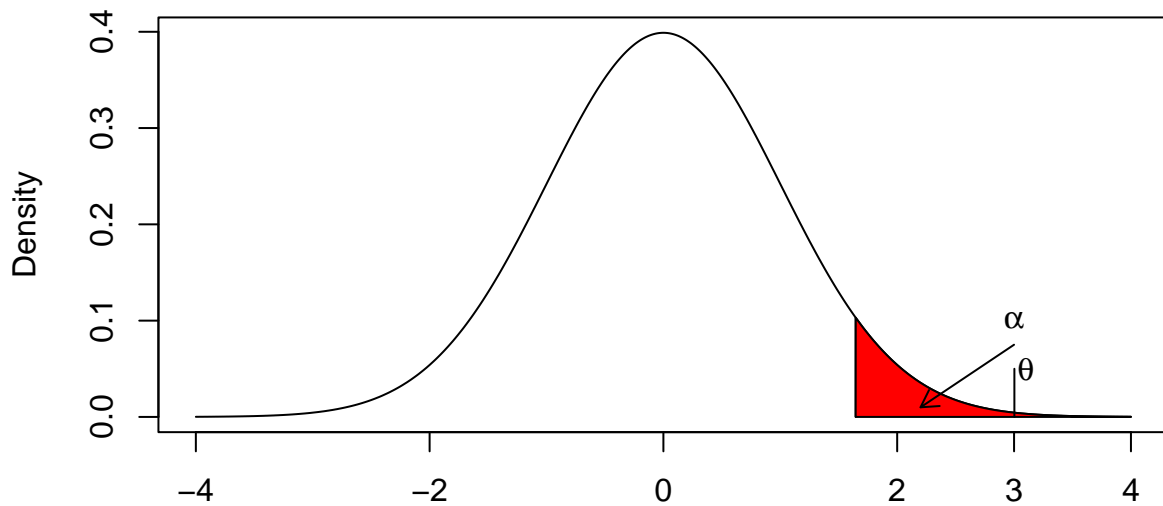
It is ultimately used as a decision rule.

In the example of BMI:

- Sample mean: 23.77
- Standard deviation: 3
- Sample size: 50
- Statistical test

$$\theta = \frac{(\bar{x} - \mu_0)}{\sigma/\sqrt{n}} = \frac{(23.77 - 22.5)}{3/\sqrt{50}} = 2.99$$

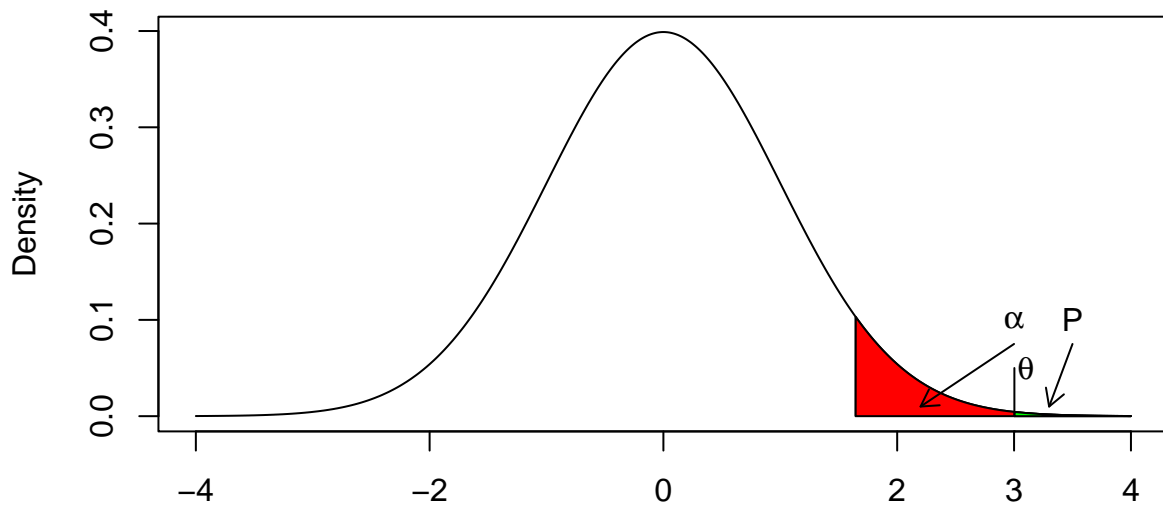
Density Function of θ if H_0 is true



The value of the statistical test (θ) lies in the rejection area.

The P-value is the probability of a result as that observed (θ) or even farther from the null hypothesis IF the null hypothesis is true.

Density Function of θ if H_0 is true



Green area: P-value.

The green area is lower than the red area \rightarrow the P-value is lower than α

Rule of thumb:

- P-value $< \alpha \rightarrow$ Reject the null hypothesis
- P-value $\geq \alpha \rightarrow$ Do not reject the null hypothesis

The P-value is also interpreted as the probability that the difference between the observed result and the null hypothesis proposal is due to sampling variability (random difference).