

# Hypothesis Tests

Bachelor in Computer Science and Engineering

2020/21

## 1. Introduction

A usual way to infer one or more parameters of a population is by formulating a conjecture or hypothesis about their values and to validate it by using the information obtained by measuring a sample dataset. If data predict values for the analyzed parameters that are far from the conjectured one we conclude that our hypothesis is incompatible with the collected data and as such we will reject our hypothesis. However if the sample results compatible with our hypothesis we do not reject it and consider it as valid. The statistical theory about hypothesis testing allows us to quantify the difference that exists between the conjectured values of the parameters and their estimations obtained by the sample data.

### 1.1 Objectives

- To state the null hypothesis about the population parameter that we are investigating (The test could be unilateral and bilateral).
- To select the test statistic according to the type of population we are analyzing (normal or not) and the sample data that we have (large or small).
- To define the critical values that limit the rejection/acceptance regions and to calculate the p-value.
- To take a decision about the null hypothesis by comparing the p-value with the identified critical values.

## 2. Bilateral Test for the population mean, $\mu$ , in the case of large samples

**Example:** A manufacturer of transistor of type BC547B (`Transistor.xlsx`) wants to know if her/his production keeps maintaining its quality standards. In particular, s/he wants to know if on average the current gain of the transistors is still 290. To verify this, s/he takes a sample of size 100 and tests this hypothesis with a significance level 5%.

The hypothesis test implemented in the `t.test` function is valid only under the assumption of normality, and by the Central Limit Theorem also in the case of large samples ( $> 30$  individuals) although in this case the p-value must be recalculated. In the latter case it is not important to know what the exact distribution of the population is. In our example we have 100 observations that is enough to assume normality for the test statistic without previously know the actual distribution of data.

By default `t.test` sets the null hypothesis to  $\mu = 0$ , the alternative hypothesis as bilateral and the significance level to 0.05. If we want to modify these settings we can modify the arguments of this function: `mu`, `alternative` and `conf.level`, respectively. In particular, for the argument `alternative`, there are three possible options:

- To make a Bilateral Test: `"two.sided"`
- To make a Right Unilateral Test: `"greater"`
- To make a Left Unilateral Test: `"less"`

For this example select the default option, "two.sided", since the transistor's manufacturer want to test the following hypotheses:

$$H_0 : \mu = 290$$

$$H_1 : \mu \neq 290$$

```
t.test(Transistor$BC547B, mu = 290, alternative = "two.sided")
```

```
##
## One Sample t-test
##
## data: Transistor$BC547B
## t = -2.7842, df = 99, p-value = 0.006429
## alternative hypothesis: true mean is not equal to 290
## 95 percent confidence interval:
## 276.7953 287.7847
## sample estimates:
## mean of x
## 282.29
```

The above results shows the characteristic measures used to make the test, value of the test statistic,  $t = -2.7842$  and the  $p\text{-value} = 0.006429$ . It should be noted that this  $p\text{-value}$  (also the confidence interval) was calculated using a Student's  $t$  distribution with  $df = 99$ , that is, it is implicitly assumed that the distribution of the variable is normal. Since this assumption has not been verified and is not necessary because the sample size is large enough, we can calculate the  $p\text{-value}$  using the asymptotic distribution

```
2*pnorm(2.7842, lower.tail = FALSE)
```

```
## [1] 0.005365994
```

Since the  $p\text{-value} < 0.05$ , we reject the Null hypothesis (with significance level 5%). That means that the computed difference between the hypothesis ( $\mu = 290$ ) and the parameter estimation (sample mean,  $\bar{x} = 282.29$ ) is too large to be simply explained by the randomness of the sample. This could mean that the average of the current gain changes, perhaps due to a failure in the production process, or that the measuring device is not well calibrated.

By what we have seen in the lectures, the decision taken about the Null Hypothesis by a bilateral test is equivalent to verifying if the hypothesized value of the mean is contained in the confidence interval with the same degree of significance. For our example, we calculate the interval using the function **z-test** (why?)

```
suppressWarnings(library(BSDA))
z.test(Transistor$BC547B, sigma.x = sd(Transistor$BC547B))$conf.int
```

```
## [1] 276.8624 287.7176
## attr(,"conf.level")
## [1] 0.95
```

We observe that the confidence interval for the mean does NOT include the value  $\mu = 290$ , that implies rejecting the Null Hypothesis as well.

### 3. Unilateral Test for the population mean, $\mu$ , in the case of large samples

**Example:** The file `Estaturas.xlsx` contains the heights of 50 male and 50 female students that were 18 and 25 years old. These students belonged to university of Madrid. According to some anthropometric studies, young Spanish people of the same age as the selected sample have an average height of 164 cm and 177 cm, respectively. We want to test if, on average, the graduate students in Madrid are taller than the Spanish average.

First we need to state the Null and the Alternative Hypotheses:

$$H_0 : \mu(\text{female}) = 164 \text{ versus } H_1 : \mu(\text{female}) > 164$$

$$H_0 : \mu(\text{male}) = 177 \text{ versus } H_1 : \mu(\text{male}) > 177$$

We start by testing the data about the female students.

```
t.test(Estaturas$mujeres, mu = 164, alternative = "greater")
```

```
##
## One Sample t-test
##
## data: Estaturas$mujeres
## t = 3.281, df = 49, p-value = 0.0009549
## alternative hypothesis: true mean is greater than 164
## 95 percent confidence interval:
## 165.055      Inf
## sample estimates:
## mean of x
## 166.1574
```

Next we do the same for the male students.

```
t.test(Estaturas$hombres, mu = 177, alternative = "greater")
```

```
##
## One Sample t-test
##
## data: Estaturas$hombres
## t = -1.3041, df = 49, p-value = 0.9009
## alternative hypothesis: true mean is greater than 177
## 95 percent confidence interval:
## 174.5002      Inf
## sample estimates:
## mean of x
## 175.9063
```

We must recalculate the p-values

```
pnorm(2.281, lower.tail = FALSE)
```

```
## [1] 0.01127422
```

```
pnorm(-1.3041, lower.tail = FALSE)
```

```
## [1] 0.9039003
```

According to the obtained p-values we can deduce the following:

- For the case of the female students, the **p-value** is 0.01127422 that means that for the usual values of  $\alpha$  we would have rejected the null hypothesis. Being the value very low it looks clear in this case that the hypothesis should be rejected, and we can conclude that there is high evidence in supporting the fact that the female graduate students in Madrid are on average higher than the corresponding national average.
- For the case of the male students, the **p-value** is large (0.9039003) that means that we cannot reject the null hypothesis with the usual significance levels. Therefore there is not enough evidence about the fact that the male graduate students are taller higher than the corresponding national average.

## 4. Unilateral Test for the population variance of normal distributed samples

Using again the example of the type transistors BC547B, in which we want to know if the production continues to maintain the same level of quality. More specifically, we want to know if the variance of the production has not increased compared to the historical value of 760. The contrast to be made is whether the variance has increased or not, and we can express it as the following hypothesis test

$$H_0 : \sigma^2 = 760$$

$$H_0 : \sigma^2 > 760$$

This test for variances is valid only if the population is normal. We then test for normality before proceeding with the test (see the practice about Goodness-of-fit test for distributions).

```
library(nortest)
pearson.test(Transistor$BC547B)
```

```
##
## Pearson chi-square normality test
##
## data: Transistor$BC547B
## P = 10.24, p-value = 0.4197
```

We can then assume normality (to be precise, we cannot reject the null hypothesis of normal). Therefore we can perform a variance test by

```
suppressWarnings(library(EnvStats))
varTest(Transistor$BC547B, sigma.squared = 760, alternative = "greater")
```

```
##
## Chi-Squared Test on Variance
##
## data: Transistor$BC547B
## Chi-Squared = 99.893, df = 99, p-value = 0.456
## alternative hypothesis: true variance is greater than 760
## 95 percent confidence interval:
## 616.0962      Inf
## sample estimates:
## variance
## 766.8544
```

The **p-value** of the test is considerably bigger than 0.05, therefore we cannot reject the null Hypothesis stating that the variance has not been changed.