HYPOTHESIS TESTING

INTRODUCTION

FROM THE DATA TO THE POPULATION MODEL

In previous lesson we knew the ground truth (all population data) and we derived statistics of the sampling distribution

In real life, we have to work the other way around:

- We do not know true model or population values.
- We can only sample some data.
- We want to infer value from data to population.
- We also want to know how *accurate* our prediction can be.

HYPOTHESIS TESTING

In hypothesis testing, we want to confront two complementary ideas (or hypothesis) about our population, and we want to use the available data to decide which of the hypothesis is more plausible.

The two ideas we want to confront are referred to as:

- Null hypothesis H_0
- Alternative hypothesis H_1

The objective is to test the null hypothesis (try to prove it wrong) with the available data. If we can't reject the hypothesis, our confidence in it increases (although we can never prove it right 100%).

SOME DEFINITIONS

- **Null Hypothesis (H_0)**: The hypothesis we considered 'true' unless evidence to the contrary is obtained.
- **Alternative hypothesis** (H_1): A hypothesis against which the null hypothesis is tested and that will be held 'true' if the null hypothesis is declared 'false'.
- **One-sided test**: We test if the alternative is either greater or smaller than the null hypothesis. We decide that in advance.
- **Two-sided test**: We test if the alternative is different (on any direction) that the null hypothesis.
- **Type I error**: The error committed by rejecting a true null hypothesis.
- **Type II error**: The error committed by not rejecting a false null hypothesis.
- **Significance level (α):** The probability of rejecting a null hypothesis that is true.
- **Power of test**: The probability of rejecting a null hypothesis that is false.

EXAMPLE 1

You work at a marketing firm, preparing a web-based campaign for a client.

You have to imagine a new campaign, and you want to test it against the previous campaign that produces a Click-Through-Rate (CTR) of 0,35%.

You decide to test the new campaign on a sample of users, and every day you show the new campaign to 1000 randomly selected users. You want to test if the new campaign is better than the previous one. **What is the null hypothesis and what is the alternative hypothesis?**

The current situation is that the CTR is at least 0,35%, but we will only implement the new strategy if we show that it is strictly higher than this; in this case, this is a **one-sided test** with:

- H0 : $\mu \leq 0{,}35\%$
- H1 : $\mu > 0{,}35\%$

EXAMPLE 2

Your company produce titanium rings that are essential for the construction of space rockets and that must all measure exactly 5 cm in diameter.

The company is considering implementing a new production process, but your manager is worried about the consistency of your product and asks you to do a test with a small sample.

What is the null hypothesis and what is the alternative hypothesis?

The current situation is that the diameter is 5 cm; we will keep the old machines if the new process makes bigger or smaller rings. Then we have a **two-sided test** with hypothesis:

- H0: $\mu = 5cm$
- H1: $\mu \neq 5cm$

WHAT IS THE PROCEDURE ?

Suppose that we want to test if the mean of a certain population is equal to a particular value. To do this test, we need to do the following:

Step 1. Decide on the Null and Alternative Hypothesis (and this includes, deciding if we want to do a one-sided or a two-sided test.

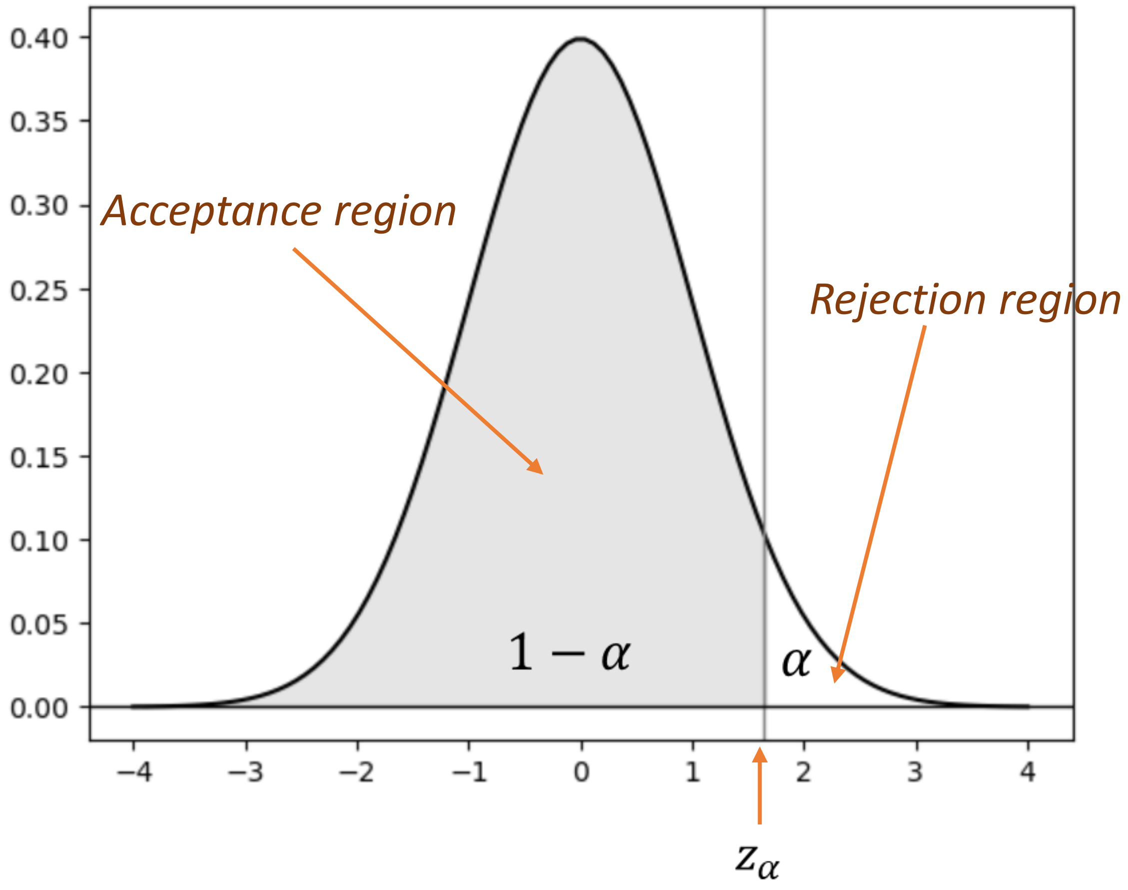
Step 2. Decide on a level of significance α. The smaller the value, the harder will be to reject the null hypothesis. Typical values are 5% or 1% (The confidence level will be then 1- α).

Step 3: Verify that the data and the population satisfy certain conditions to perform the appropriate test (we will see right now which conditions are these, but as an advance, we hope as usual that the population is normally distributed, or that the sample size is large enough).

Step 4: Perform the test and deduce the right conclusion (we will also see later how to do the test).

CRITICAL VALUE (RIGHT SIDE)

Consider a standard normal distribution (mean=0, standard deviation = 1)



For any α between 0 and 1, there is value z_α called critical value , that separates the normal curve in two parts. One with area = $1-\alpha$ and the other with area α

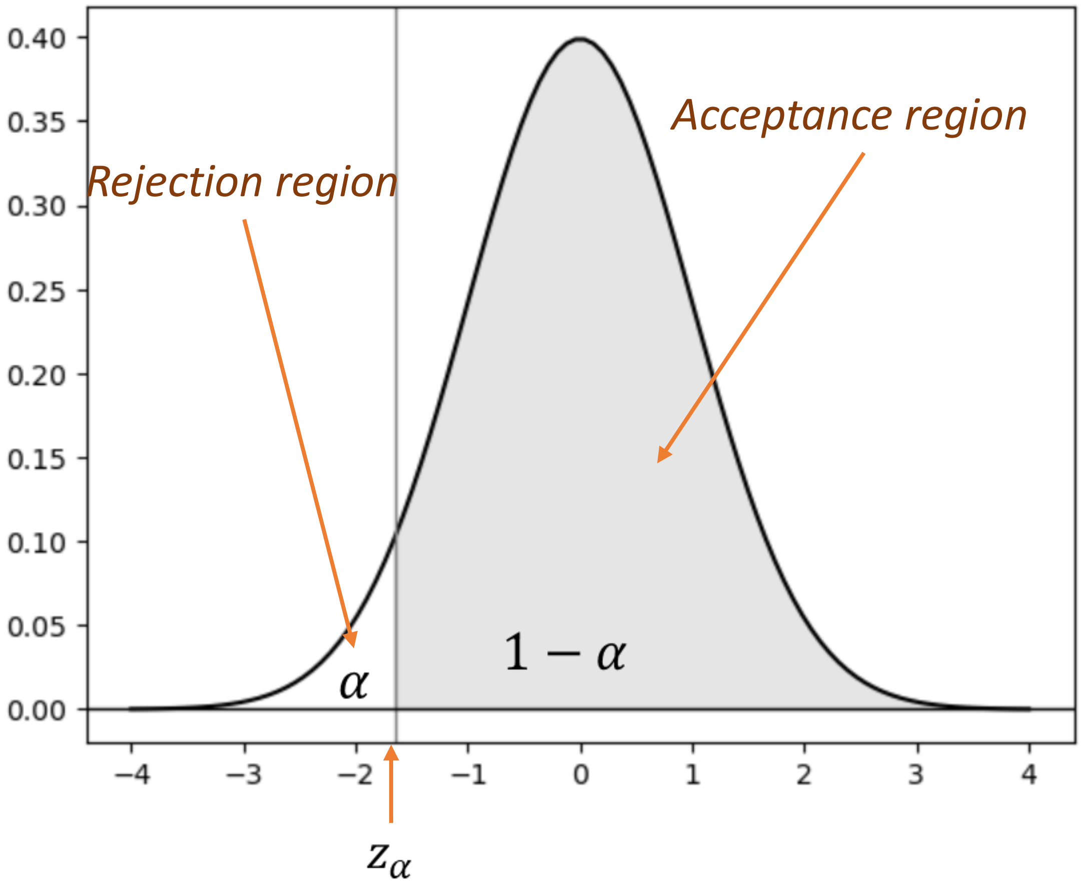
In Python is computed like this, for the right side:

```
import scipy.stats
alpha=.05
z = scipy.stats.norm.ppf(1-alpha)
print(z)
```

```
1.6448536269514722
```

CRITICAL VALUE (LEFT SIDE)

Consider a standard normal distribution (mean=0, standard deviation = 1)

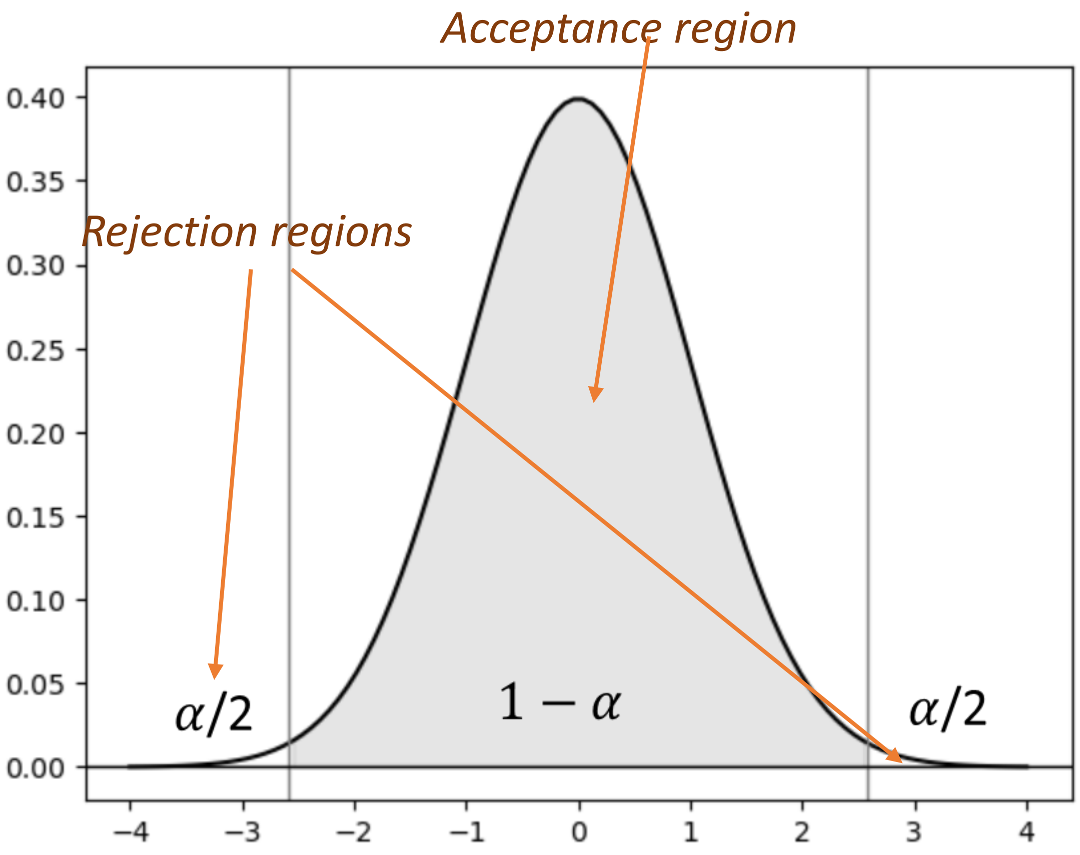


For the left side, the code is similar, but without the `1-` inside the ppf function:

```
import scipy.stats
alpha=.05
z = scipy.stats.norm.ppf(alpha)
print(z)
```

```
-1.6448536269514729
```

CRITICAL VALUE (TWO-SIDED)



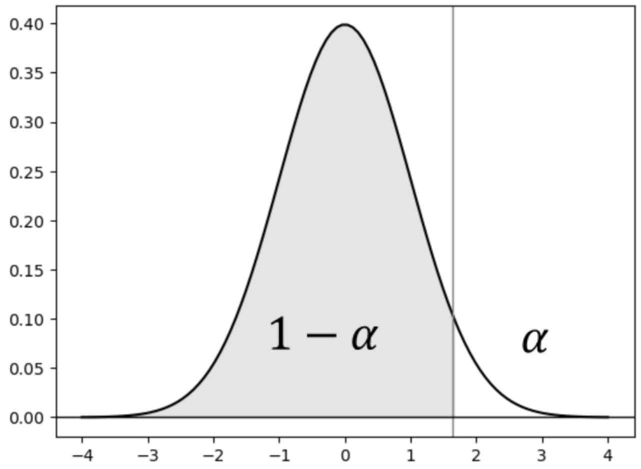
For two-sided, we allocate one half of rejection region to each side, so there are two (symmetrical) critical values, computed as follows:

```
import scipy.stats
alpha=.05
z_right = scipy.stats.norm.ppf(1-alpha/2)
z_left = scipy.stats.norm.ppf(alpha/2)
print(z_left, z_right)
```

```
-1.959963984540054 1.959963984540054
```

ONE-SIDED (RIGHT) TEST, NORMAL DISTRIBUTION AND KNOWN VARIANCE

1-sided test



H0 : $\mu \leq \mu_0$
H1 : $\mu > \mu_0$
α is chosen by us
Known σ, n

1. Compute the sample mean, $\bar{X}$
2. Compute the critical value z_α
3. Compute the Z-value:

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

4. Reject if $Z > z_\alpha$

- Z measures how different is the sample mean from the hypothesized population mean

- z_α represents our tolerance limit. If the measured sample is more different than this, we reject.