HEADS
PhD PROGRAMME IN HEALTH DATA SCIENCE

Hypothesis testing

1. The coin example
2. Hypothesis testing for the mean
3. Types of error in hypothesis testing

Basics of Health Intelligent Data Analysis

PhD Programme in Health Data Science

Cláudia Camila Dias

Pedro Pereira Rodrigues

U.PORTO
FMUP FACULDADE DE MEDICINA
UNIVERSIDADE DO PORTO

MEDCIDS
DEPARTAMENTO DE MEDICINA
DA COMUNIDADE, INFORMAÇÃO
E DECISÃO EM SAÚDE

CINTESIS
Health. Research.

Hypothesis test

- With confidence intervals we can infer about a parameter in the population based on an estimate of that parameter.

- **Hypothesis tests** are based on another different (but related) approach.

- The idea now is to measure how much the results observed in the sample are compatible with a **hypothesis** about the population.

Hypothesis test

1. Define the hypothesis

H0 = null hypothesis - no effect on population

2. Set the significance level (alpha) - usually 0.05

3. Get the test statistic with the sample data and obtain the p-value: probability of obtaining the result we obtained or even more extreme, assuming H0 is true.

5. Interpret the p-value:

if p < alpha, we have enough evidence to reject H0

if p ≥ alpha, we do not have enough evidence to reject H0

Hypothesis test

A Randomized Trial of a Low-Carbohydrate Diet for Obesity

N Engl J Med 2003; 349:1000-1002.

Results: Subjects on the low-carbohydrate diet had lost more weight than subjects on the conventional diet at 3 months (mean [±SD], –6.8±5.0 vs. –2.7±3.7 percent of body weight; P=0.001) and 6 months (–7.0±6.5 vs. –3.2±5.6 percent of body weight, P=0.02), but the difference at 12 months was not significant (–4.4±6.7 vs. –2.5±6.3 percent of body weight, P=0.26).

Hypothesis test

A Randomized Trial of a Low-Carbohydrate Diet for Obesity

N Engl J Med 2003; 349:1000-1002.

Results: Subjects on the low-carbohydrate diet had **lost more weight** than subjects on the conventional diet at **3 months** (mean [±SD], –6.8±5.0 vs. –2.7±3.7 percent of body weight; **P=0.001**) and 6 months (–7.0±6.5 vs. –3.2±5.6 percent of body weight, **P=0.02**), but the difference **at 12 months was not significant** (–4.4±6.7 vs. –2.5±6.3 percent of body weight, **P=0.26**).

Hypothesis test

The classic example: Toss a coin 10 times in the air; observe heads or tails.

Suppose we get 5 heads (5 tails). Is the coin balanced?

Suppose we get 7 heads (3 tails). Is the coin balanced?

Suppose we get 9 heads (1 tails). Is the coin balanced?

Suppose we get 10 heads (0 tails). Is the coin balanced?

Hypothesis test

Being the coin balanced, we expected 5 heads and 5 tails (null hypothesis).

Could the results we found (6 heads and 4 tails; 7 heads and 3 tails;…; 10 heads and 0 tails) be due to chance?

- Yes, it can always be due to chance

But to what extent do we accept chance as justification for the obtained result?

Hypothesis test

The classic example: We toss a coin 10 times in the air and observe heads or tails.

Suppose we get 5 heads (5 tails). Is the coin balanced? Probability 0.25

Suppose we get 7 heads (3 tails). Is the coin balanced? Probability 0.12

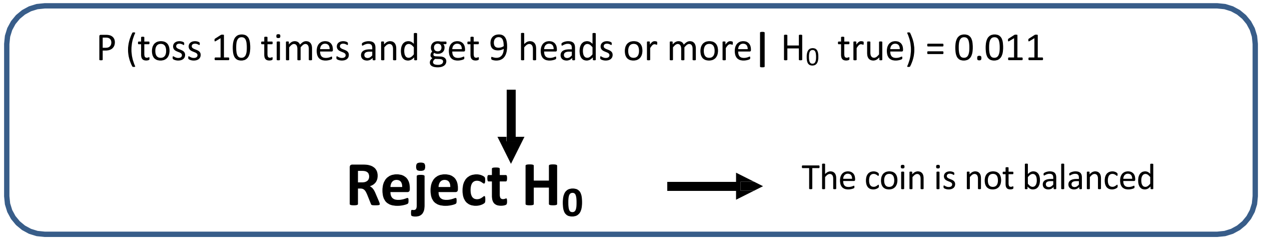
Suppose we get 9 heads (1 tails). Is the coin balanced? Probability 0.01

Suppose we get 10 heads (0 tails). Is the coin balanced? Probability 0.001

Hypothesis test

The null hypothesis is typically the opposite of the hypothesis we want to investigate (no effect)

H0: coin is balanced or heads ratio = 0.5



P (toss 10 times and get 9 heads or more | H_0 true) = 0.011

Reject H_0 → The coin is not balanced

Hypothesis test

In the coin example it is more or less intuitive to assess the extent to which we accept chance as justification for the result obtained.

But in other cases it becomes more difficult.

Hypothesis test

- In the clinical trial comparing the two diets, it was observed that the average weight reduction in one of them was 6.8kg and in the other 2.7kg.

- **Is this result very or little compatible with the null hypothesis that diets are identical?**

> A Randomized Trial of a Low-Carbohydrate Diet for Obesity
>
> *N Engl J Med 2003; 349:1000-1002.*
>
> *Results:* Subjects on the low-carbohydrate diet had lost more weight than subjects on the conventional diet at 3 months (mean [±SD], **–6.8±5.0 vs. –2.7±3.7 percent of body weight; P=0.001**) and 6 months (–7.0±6.5 vs. –3.2±5.6 percent of body weight, P=0.02), but the difference at 12 months was not significant (–4.4±6.7 vs. –2.5±6.3 percent of body weight, P=0.26).

Hypothesis test

H_0: $\mu_1=\mu_0$

or

H_0: $\mu_1-\mu_0=0$

$X_1=-6.8$ e $X_2=-2.7$

Is there evidence against the null hypothesis?

Hypothesis test

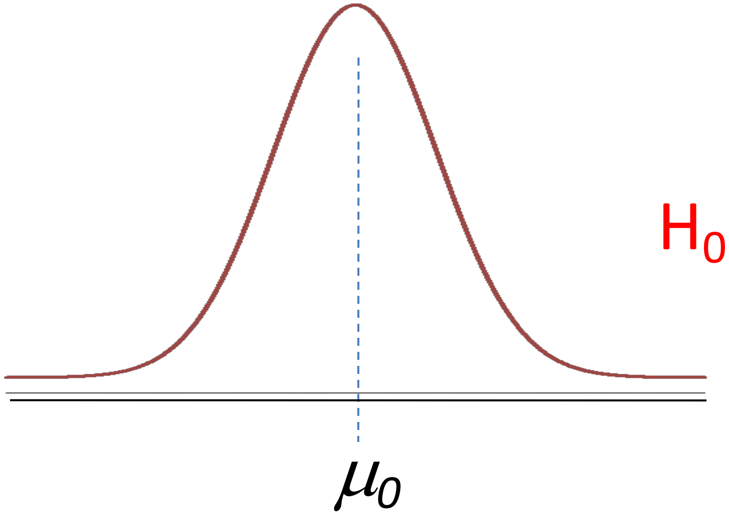
By the **Central Limit Theorem**,

the distribution of the sample means of size n tends to a normal distribution

with mean equal to population mean and standard deviation equal to standard error

Hypothesis test

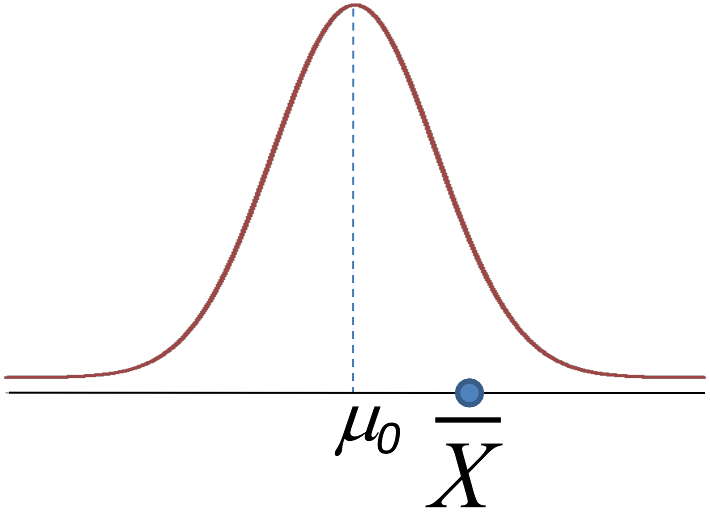
- If null hypothesis is true



H_0: $\mu = \mu_0$

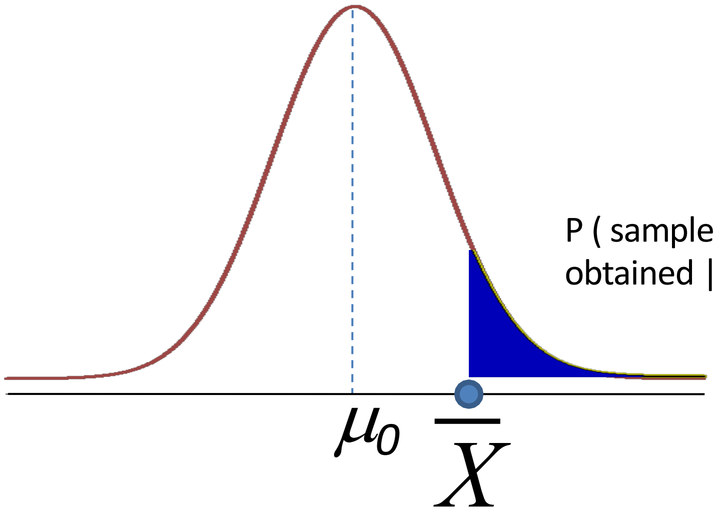
Hypothesis test

Then we can see if the sample we have may have been taken from this population.



Hypothesis test

Then we can see if the sample we have may have been taken from this population.



P (sample with an average as or more extreme than that obtained | H_0 true)

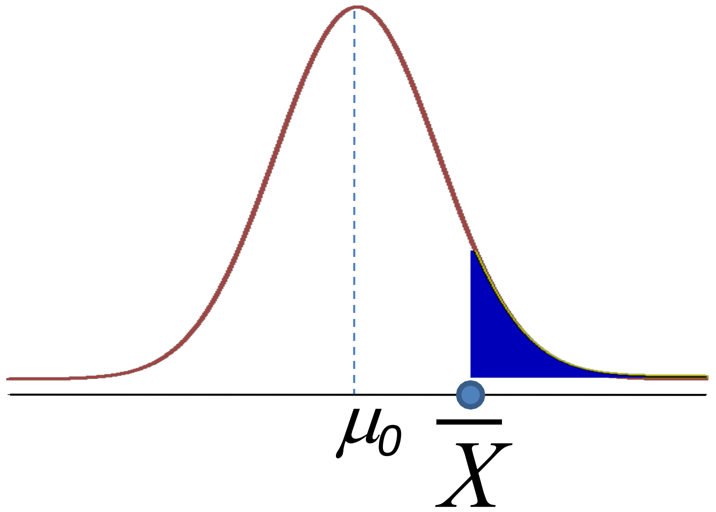
p-value

Hypothesis test

The p- value is the probability of obtaining a result as or more extreme than that observed in the sample, since (assuming) the null hypothesis is true.

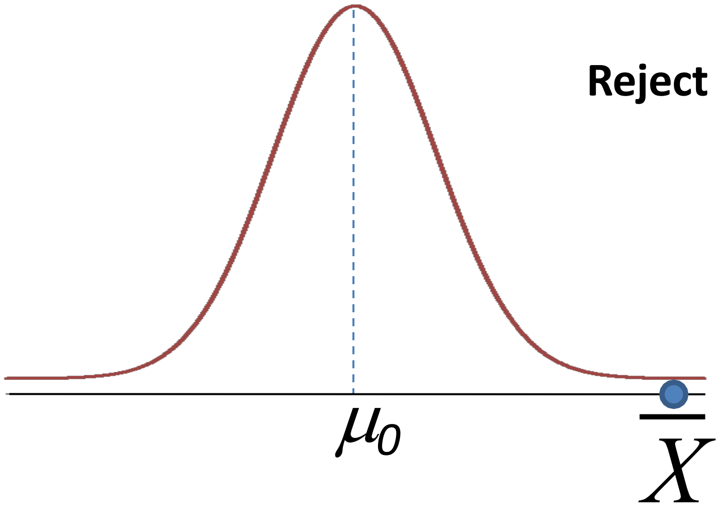
Hypothesis test

The observed sample mean seems compatible with the null hypothesis.



Hypothesis test

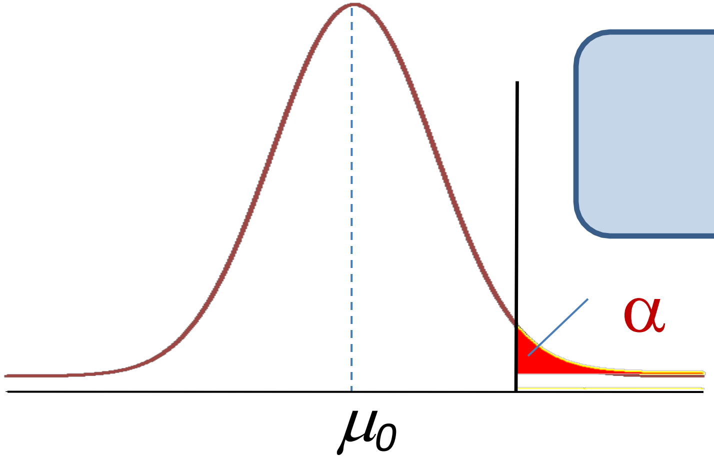
A sample with this mean was hardly obtained from a population where the null hypothesis is true.



Reject null Hypothesis

Hypothesis test

But from what value do we decide we should reject the null hypothesis?



P(reject H_0 | H_0true) = α

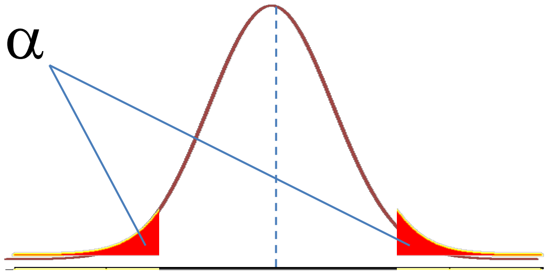
Usually , $\alpha = 0.05$

Hypothesis test

The previous example corresponds to “**one-tail” or “one-side**” tests where the alternative hypothesis (H1) corresponds to

$$H_1: \mu_1 > \mu_0$$

It is, however, more common for the alternative hypothesis (H_1) to be H_1: $\mu_1 \neq \mu_0$ corresponding to a **two-sided test**



Hypothesis test - error

	H_0 true	H_0 false
Accept H_0	Without error	Type II error (β)
Reject H_0	Type I error (α)	Without error(1- β)

test power = 1 – beta = probability of rejecting H_0 when false

that is,

the ability of the test to detect differences when they exist

Hypothesis test - error

α is defined by us when we set the level of significance

To have a higher test power (1-β) we need:

- larger sample size
- smaller variability of observations
- higher effect
- higher α