

# Introduction to hypothesis tests

Curs d'Estadística Bàsica per a la Recerca Biomèdica

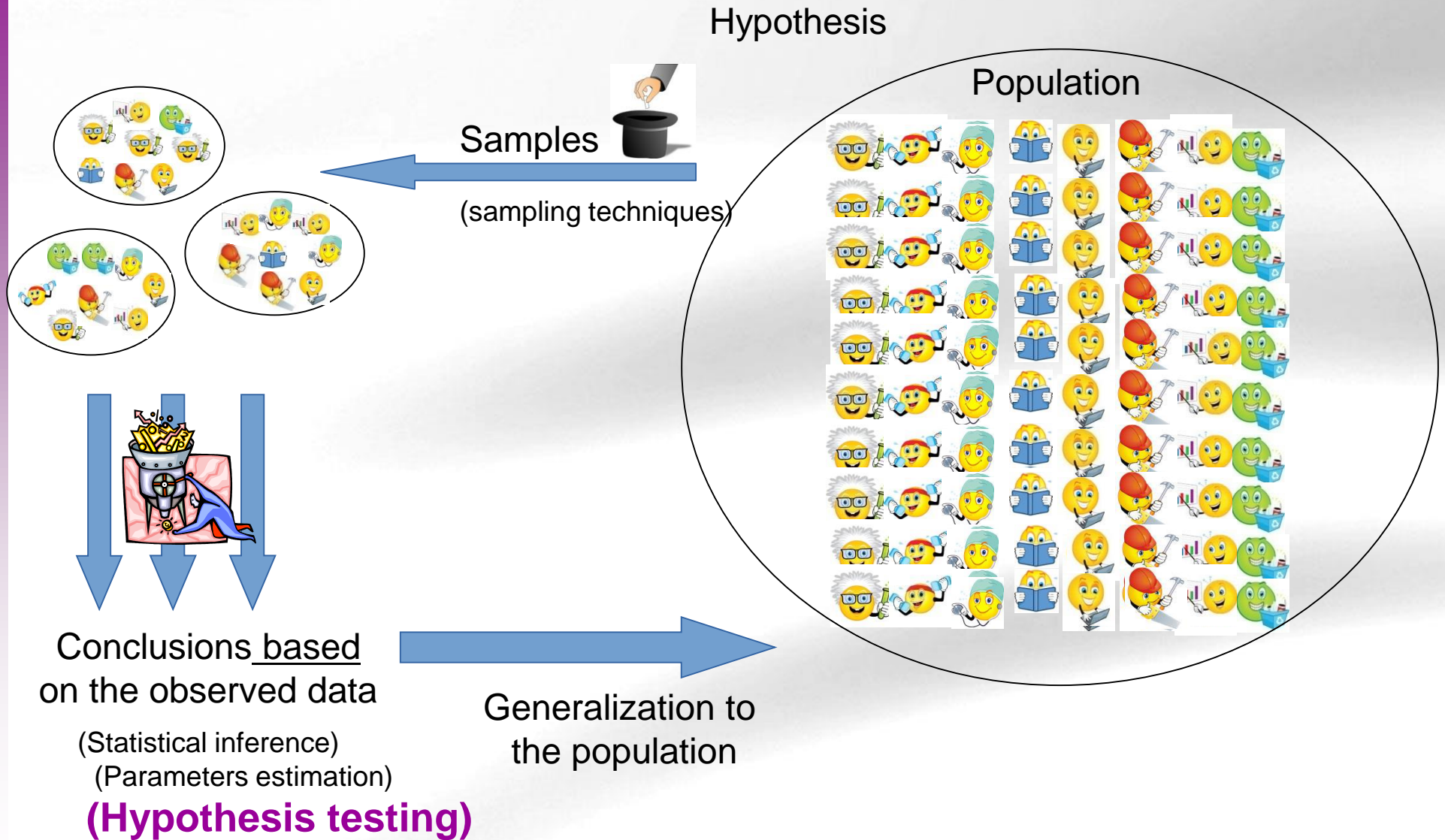
UEB – VHIR

Alex Sánchez-Pla & Santiago Pérez-Hoyos

[{alex.sanchez,santi.perezhoyos}@vhir.org](mailto:{alex.sanchez,santi.perezhoyos}@vhir.org)



# The objectives of statistical inference



# Statistical Inference Questions

## Parameter estimation:

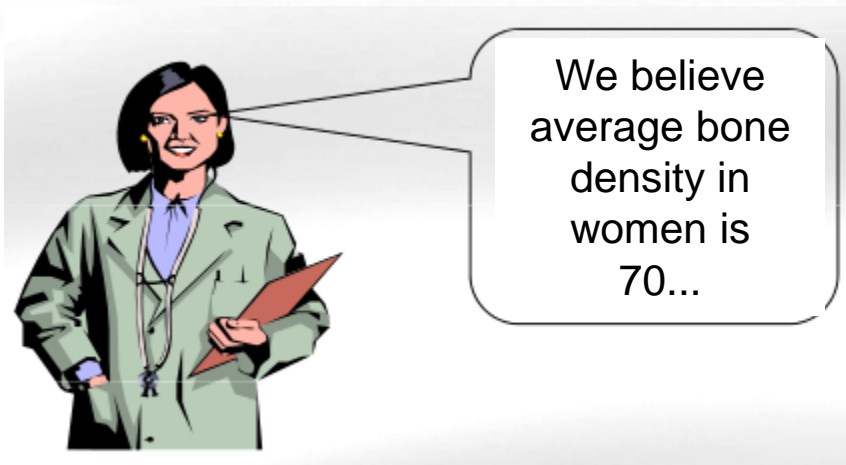
- After assuming population data follow a certain probability distribution (Normal, Binomial, Poisson, etc) goal is find out *which are the value of the parameters that fit best the sample data.*

## Hypothesis testing:

- After making some assumptions about population distribution
- *We wish to check some statement about population parameters*
  - Population mean is equal to 10
  - The mean in population A is equal to the mean in population B
  - The proportion of respondents/non-respondants is associated to treatment
- Hypothesis testing tries to assess if sample data are compatible with the hypothesis assuming that samples differ from populations due to chance.

# Hypothesis testing: Making decisions about populations

---



But... why not to check median, mode or other estimators?

# Case study problem I

---

- A study was performed to study the relation between bone density, age and menopause in women.
- The study had some previous beliefs that researchers wish to check in this study:
  1. The average bone density (measured by variable "BUA") in the population analyzed is 70
  2. BUA mean value is bigger in non-menopausal women than in menopausal or pre-menopausal ones.



## Case study problem II

- NHANES is an epidemiologic study where new lung cancer cases were recorded after 12 years of follow-up (Am J Epidemiol 1997;146:231-243)

		Lung Cancer		
		Yes	No	Total
Fruit Consumptio	High	44	2473	2517
	Low	88	2429	2517
	Total	132	4902	5034

- Given this outcome, *can we conclude that fruit consumption "is a protective factor" for (i.e. is positively associated with less) lung cancer?*
- How can we know this result is not to due chance in sampling?

# Exercise

---

- Explore the osteoporosis data set and try to figure if the claims in case study I *seem to be true* or not.
- Use the data in the table in case study II to figure out if the incidence of lung cancer is associated with fruit consumption.

# The setting of hypothesis testing

---

- We establish two basic hypotheses

## Null Hypothesis ( $H_0$ )

- Observed difference between outcome and population value is due to chance.
- No relationship among exposure and disease.

## Alternative Hypothesis ( $H_1$ )

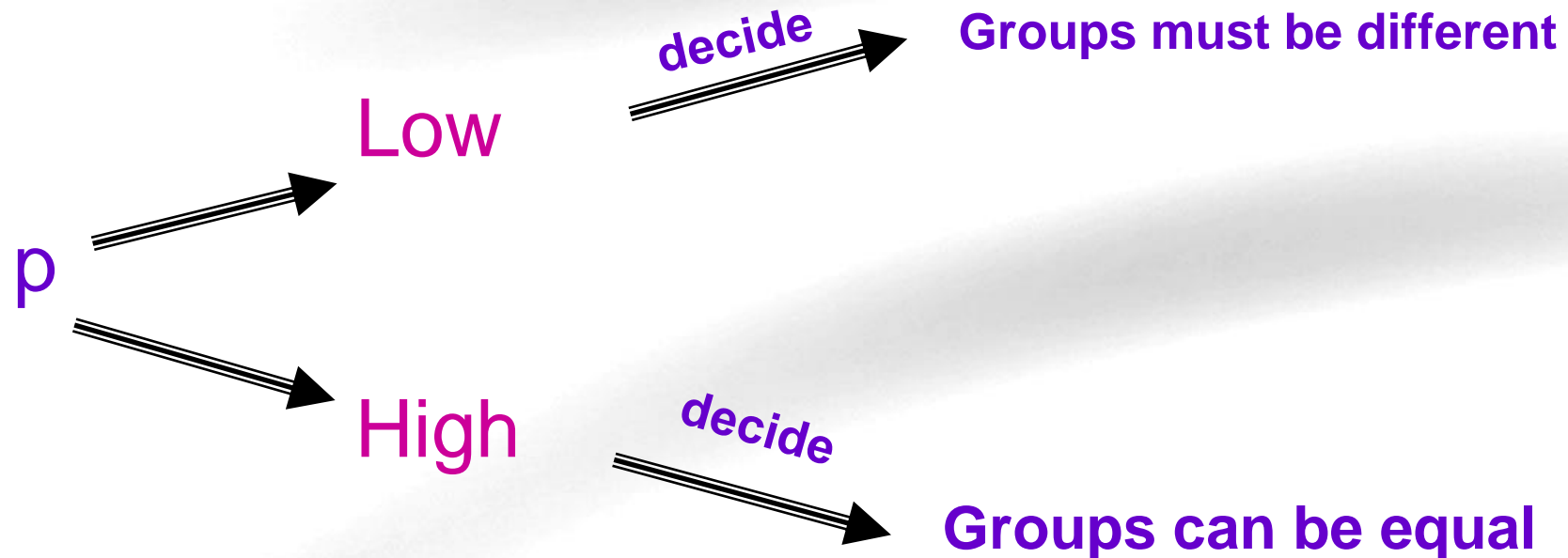
- Observed difference between outcome and population value is due to some "true effect".
- There is relationship among exposure and disease.



# ¿How to decide which hypothesis is more likely?

Calculate probability ( $p$ ) to observe differences in samples between both groups under the null hypothesis of no population differences,

In case study II we would want to *compute the probability that the % of new cancer cases is the same in both groups*





# ¿How to calculate this probability?

---

- This probability ...
  - Depends on the type of study.
  - Depends on type of variable.
  - Depends on the influence of other variables.
- In practice we do not just calculate the probability because it is associated with the error made assuming that the null hypothesis is true
- A step by step process is adopted to take the decision about which hypothesis is considered to be true.

# Hypothesis Testing Steps

---

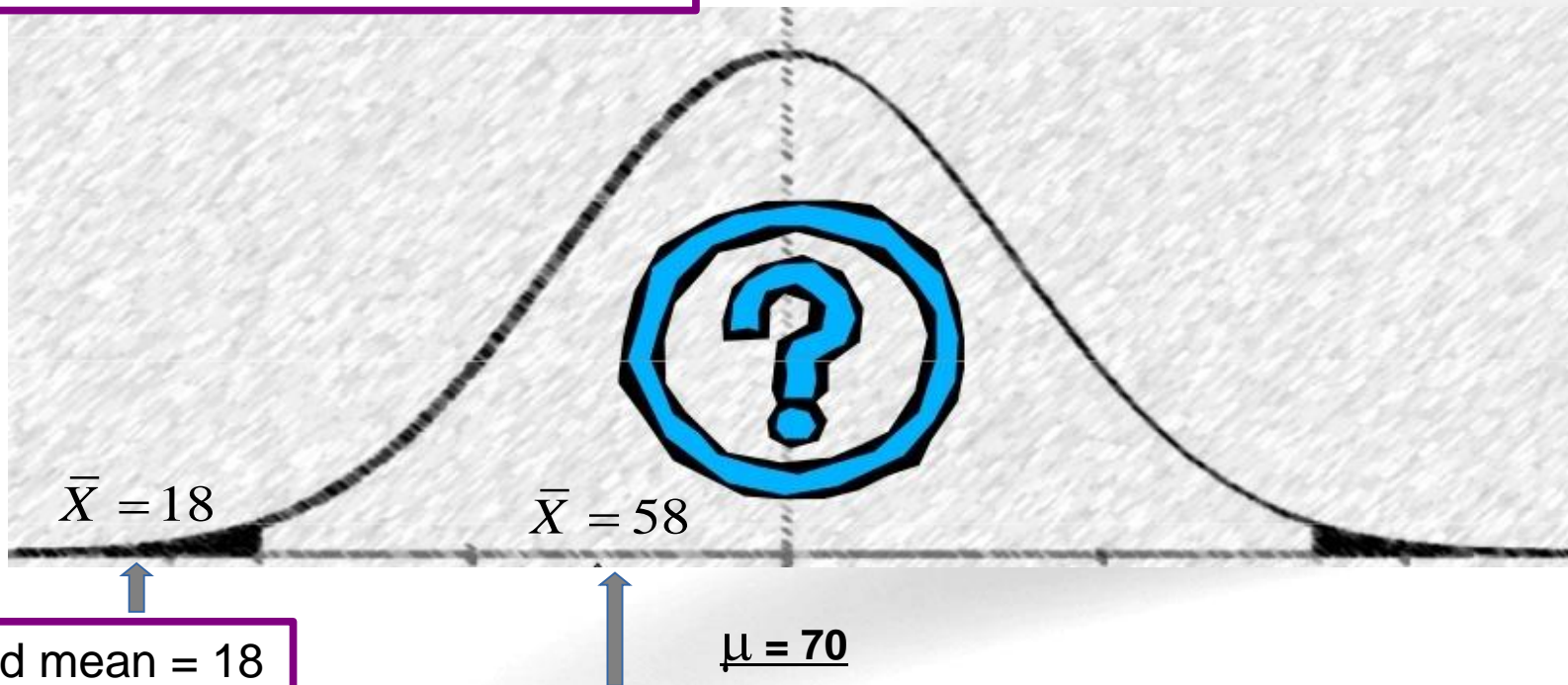
- Establish Null Hypothesis( $H_0$ )
- Establish Alternative Hypothesis ( $H_a$ )
- Select statistical test to calculate probability *under Null Hypothesis*
- Take a sample and compute test value
- Compare test value with a critical value and decide
  - ***If null hypothesis must be rejected***
  - ***Or there is not enough evidence to reject null hypothesis***

# Hypothesis Testing by Example

- Null Hypothesis( $H_0$ ):
  - *Mean bone density is 70*
- Alternative Hypothesis ( $H_a$ )
  - *Mean bone density is not 70*
    - *Might also be " ... greater than 70", "smaller than 70"*
- Select statistical test to calculate probability *under Null Hypothesis*
  - *A reasonable measure to rely on: **sample mean***

# Accepting or rejecting the NULL

$H_1$  only accepted if clear evidence  
that  $H_0$  is not true



If observed mean = 18  
 $H_0$  is rejected

If observed mean = 58  $H_0$  can not be rejected (it  
does not mean  $H_0$  can be accepted!!)

# If the NULL were true ...

---

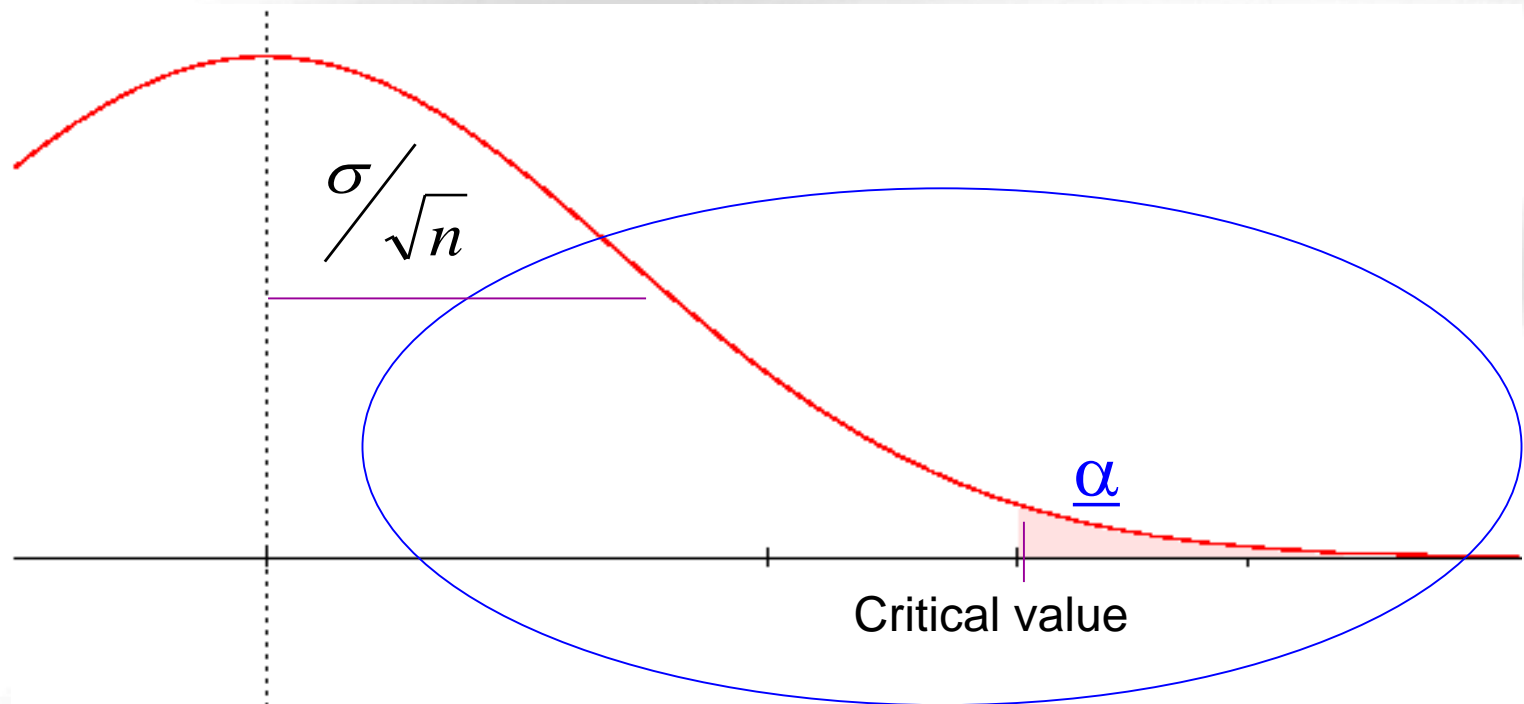
- Most tests require some assumptions to be true (independently of  $H_0$ !)
- Example:
  - The sample mean is approximately normal

$$\bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right), \text{ or: } \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$



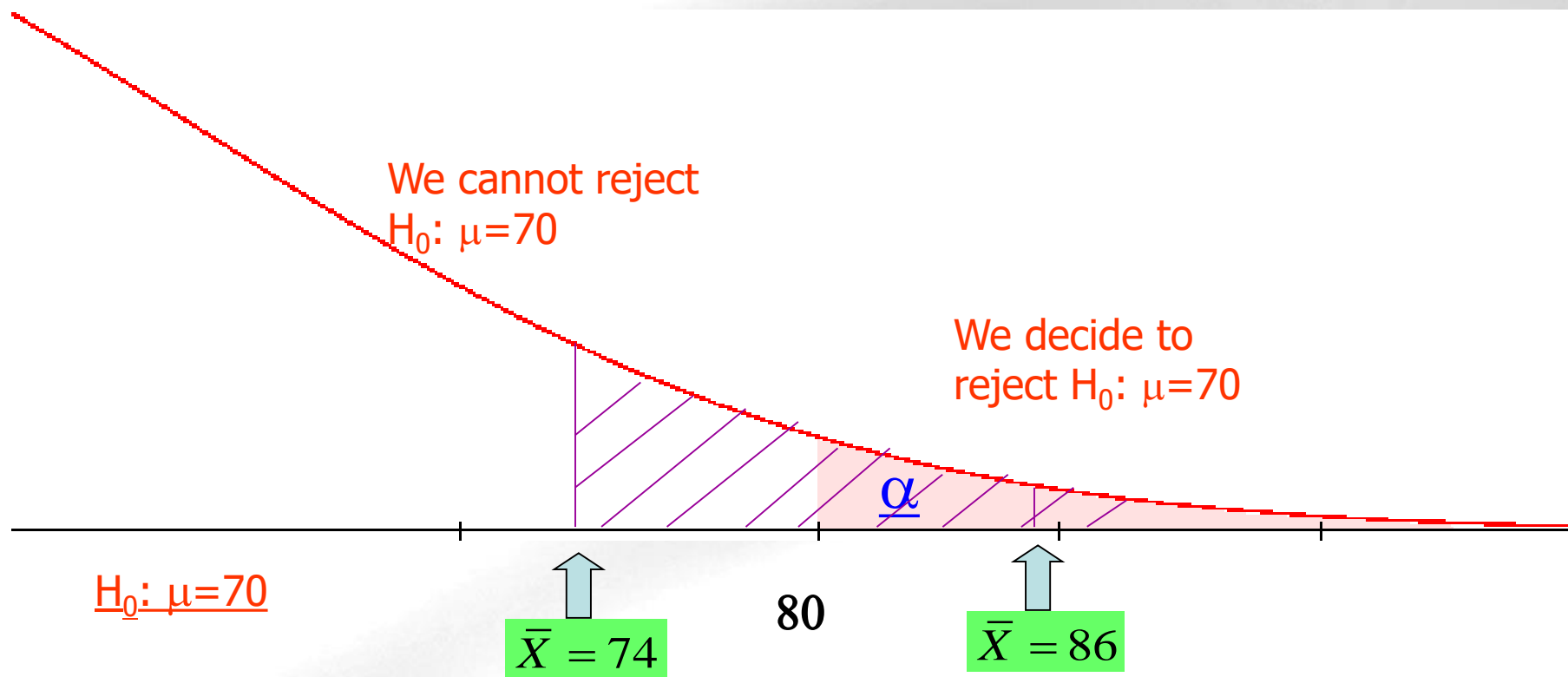
# Critical Value

- At which value of the sample mean does one change from non-rejecting to rejecting the null hypothesis?
  - A value is selected such that the probability that the sample mean exceeds it, if the null hypothesis is true, is “small”, (for example 5%).
  - This value is called “Critical Value” and
  - the probability is called “significance level ( $\alpha$ )”



## Example: Critical value and Sample mean

- ***If  $\sigma=9$ ,  $n=9$  and  $\alpha=0.05$  the critical value will be 80***
  - With a sample mean of 74 we will not reject  $H_0$
  - With a sample mean of 80 we will reject  $H_0$



# Example

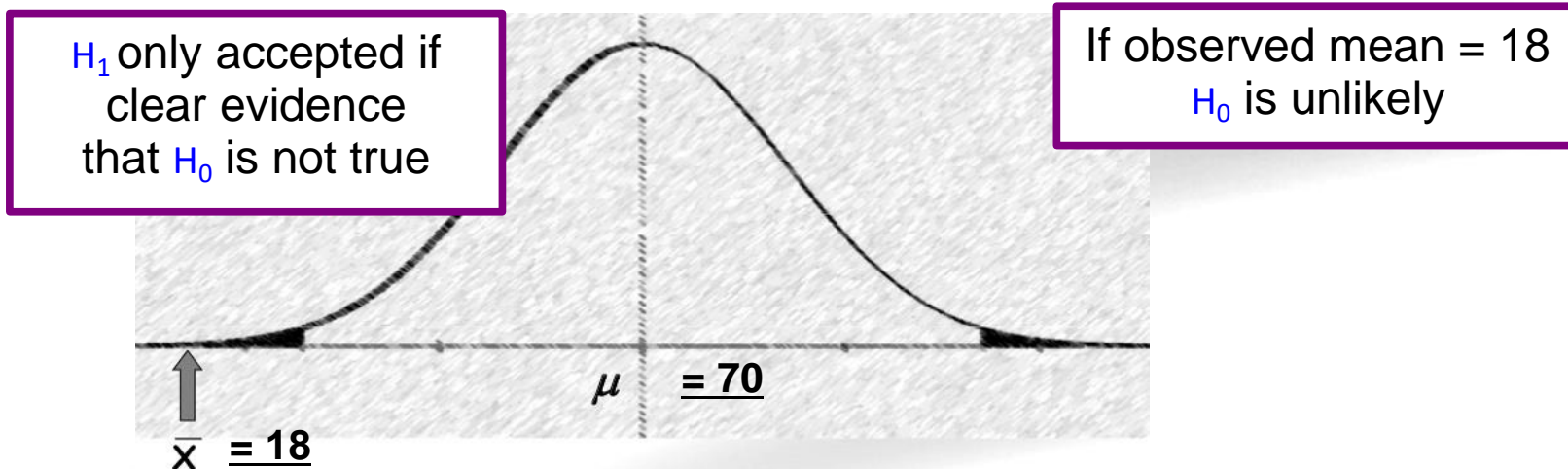
## Null hypothesis ( $H_0$ ):

$H_0$ : The mean of BUA values is 70.0

## Alternative hypothesis ( $H_\alpha = H_a = H_1$ ): the opposite idea

- $H_1$ : The mean of the bua values is not equal to 70.0 (Bilateral)
- $H_1$ : The mean of the bua values is higher(lower) than 70.0 (Unilateral)

Under the null hypothesis if all the samples of one size can be selected the sample distribution is as follows



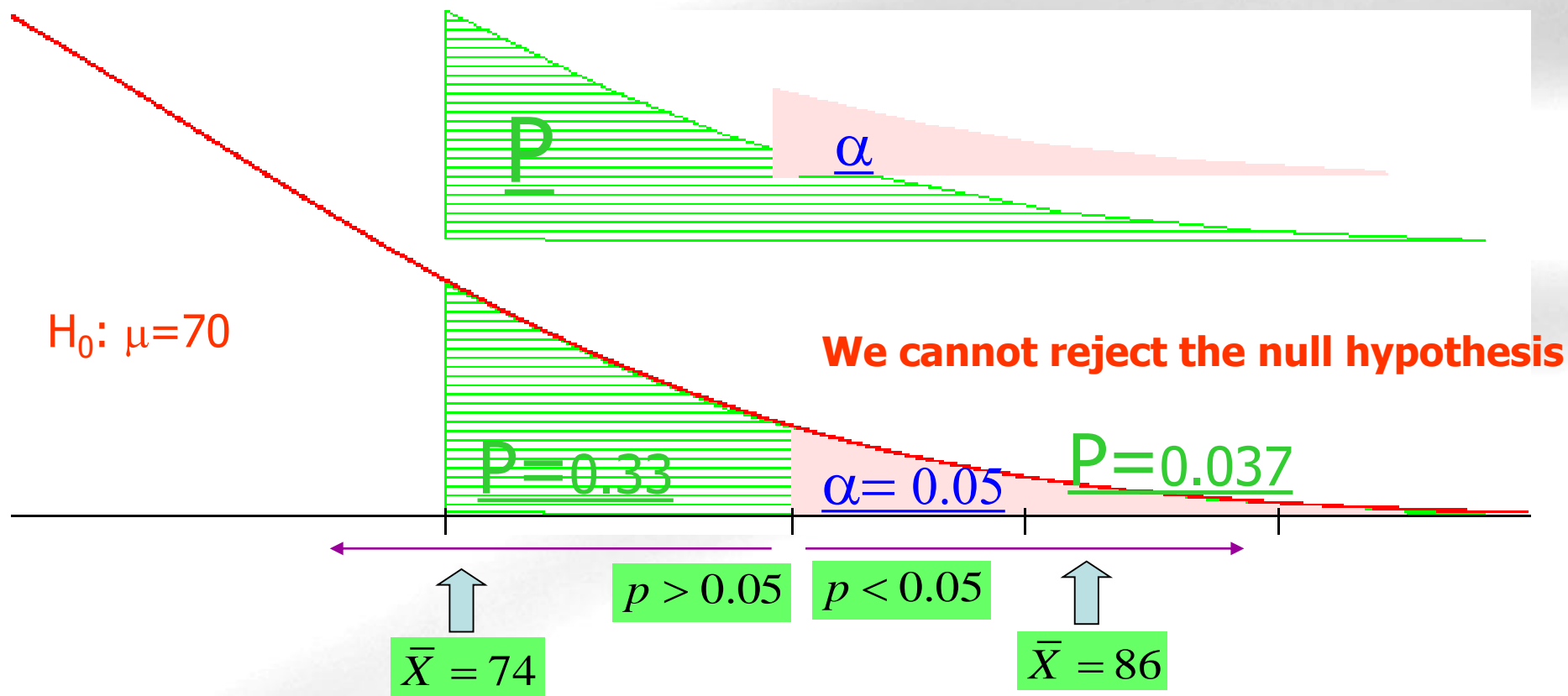
# P values: The alternative

---

- We have based our decision about rejecting  $H_0$  on comparing sample mean (i.e. 74) with the critical value (i.e. 80)
- Instead we can compare the probability of observing at least that sample mean (p value) with the significance level ( $\alpha$ ) (which is the probability of observing at least the critical value),
  - The probability is smaller than  $\alpha$  if (and only if) the sample mean is bigger than the critical value.
    - In such situation we decide to reject  $H_0$
  - The probability is bigger than  $\alpha$  if (and only if) the sample mean is smaller than the critical value.
    - In such situation we cannot reject  $H_0$  so we accept it
- Both criteria (critical value and p-value) are valid for testing hypotheses.

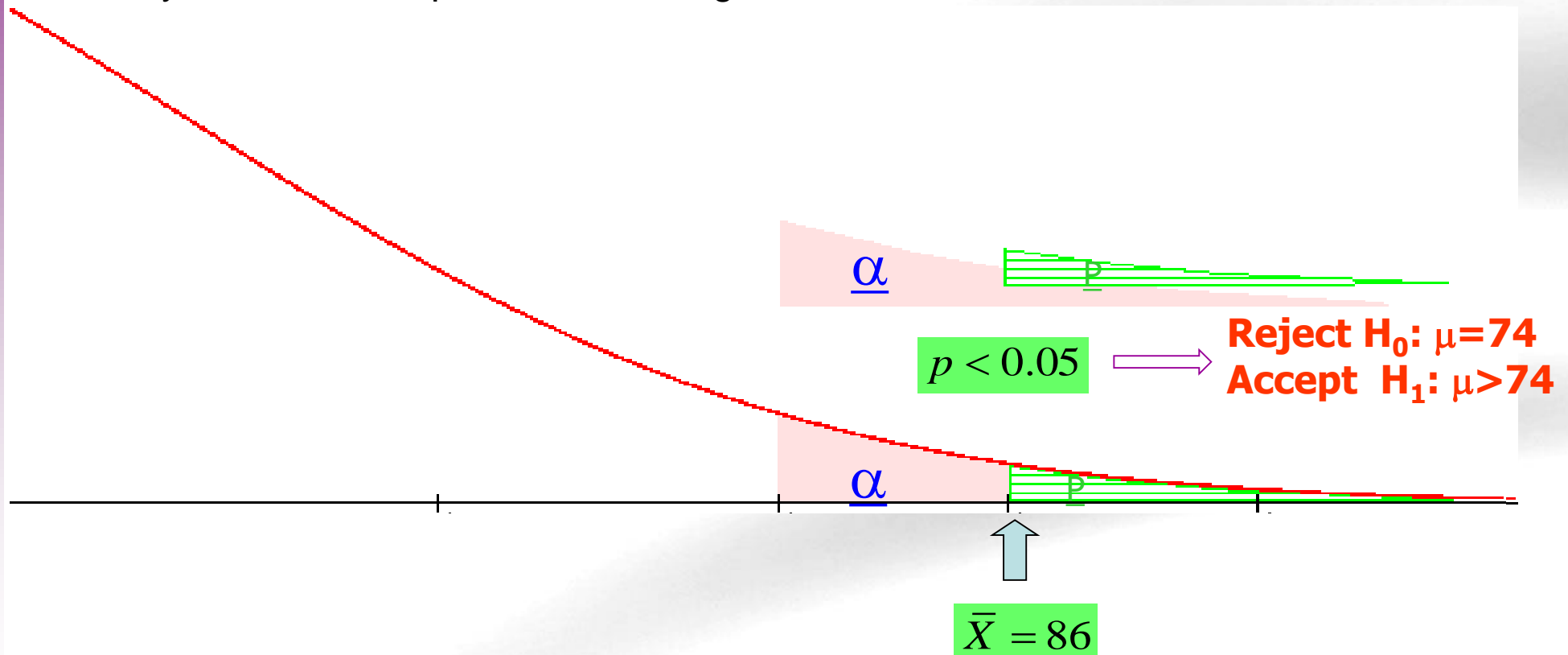
# Example: P-value vs critical value

- *If  $\sigma=27$ ,  $n=9$  and the sample mean is 74 then*
- The probability that assuming that  $H_0$  is true, that is  $\mu=70$ , we can observe by chance ...
  - a sample mean equal or greater than 74 is: 0.328
  - a sample mean equal or greater than 86 is 0.037



# Example: P-value vs critical value

- If  $\sigma=18$ ,  $n=9$  and the sample mean is 86 then
- The probability that assuming that  $H_0$  is true, that is  $\mu=74$ , it can be obtained by chance a sample with a mean greater than 86 is: 0.0378



We usually say the test is statistically significant if  $p < \alpha$



# Summary: $\alpha$ vs $p$

---

$\alpha$  and  $P$  are related but they are not the same ...

- About  $\alpha$ 
  - It is prefixed before experiment
  - Usually low ( 0.05)
  - Linked with critical value (“knowing one, the other is automatically known)
  - Unaffected by the sampling process.
- About  $p$ 
  - It is calculated after the experiment
  - Can take any values in (0,1)
  - After calculation one can know the *achieved significance level*.
  - Depends on the sampling process

# Type of Hypothesis

---

## Confirmation Hypothesis

Aim is to confirm hypothesis about parameters or distributions.

Goodness of fit test to verify hypothesis about the distribution of variable in population

Does arterial pressure in the population follow a normal distribution?

Test to verify values about a parameter.

Is the average "bua" value in our population equal to 70?

Is the proportion of lung cancer cases equal to 2.6%?

# Type of Hypothesis

---

## Independence Hypothesis

Aim is to test hypothesis for relation of variables in a population or no differences of a variable in two or more populations

Is the average "bua" value the same in menopausal and in non-menopausal population?

Is the proportion of lung cancer cases the same in people with high or low fruit consumption?

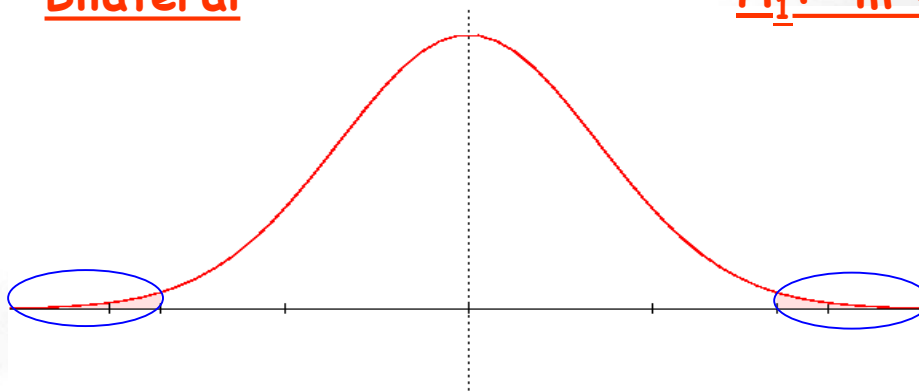
Is CD4 lymphocytes count related with CD8 count in HIV positive?

# Unilateral vs Bilateral

Critical value depends on the type of alternative Hypothesis

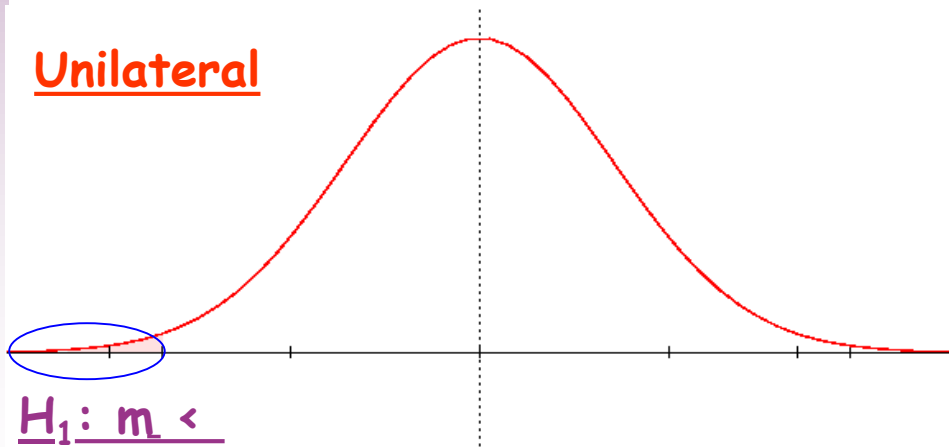
Bilateral

$H_1: m \neq$



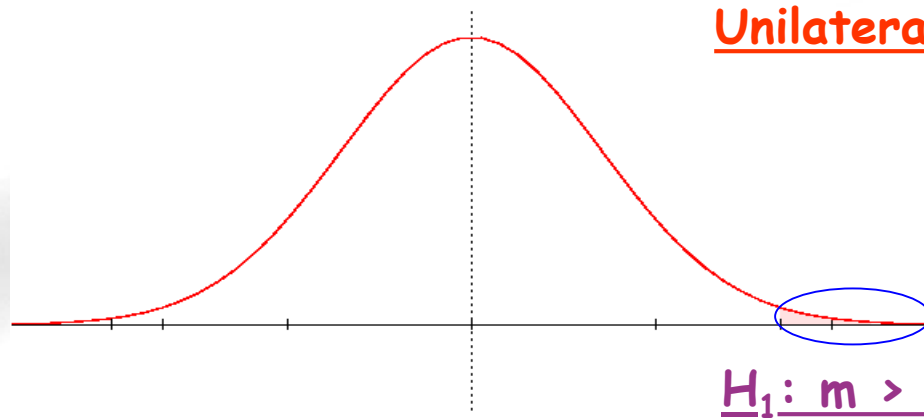
Unilateral

$H_1: m <$



Unilateral

$H_1: m >$

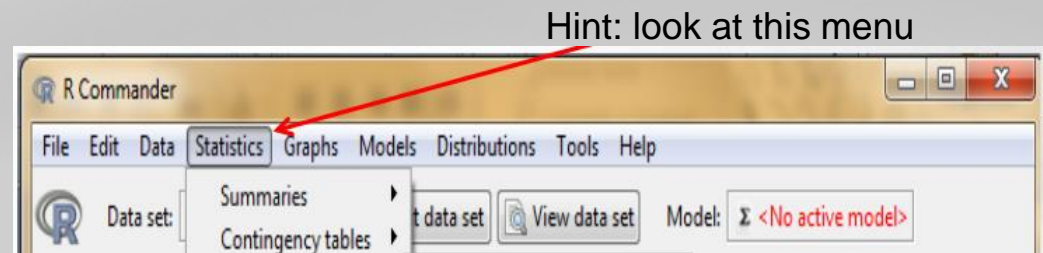


# Example: hypothesis testing with R-commander

Our assumptions:

- The average "bua" value in our population is 70.
- The "bua" mean value in menopausal and non-menopausal women is not the same.

Exercise 2):



- Test if the population mean bone density is 70.0 (Alternative "it is not 70")
- Test if the population mean bone density is equal or not between groups if we separate our observations by "menop" category

```
# osteoAll is the dataset obtained reading osteoporosis.csv
set.seed =123456 # change by your ID to obtain a distinct sample
index100<- sample( 1:nrow(osteoAll), 100)
osteo100 <- osteoAll[index100,]
with(osteo100, t.test (bua, mu=70, alternative="two.sided"))
```

**t.test(bua, mu=70)**

data: bua  $t = 1.8604$ ,  $df = 99$ ,

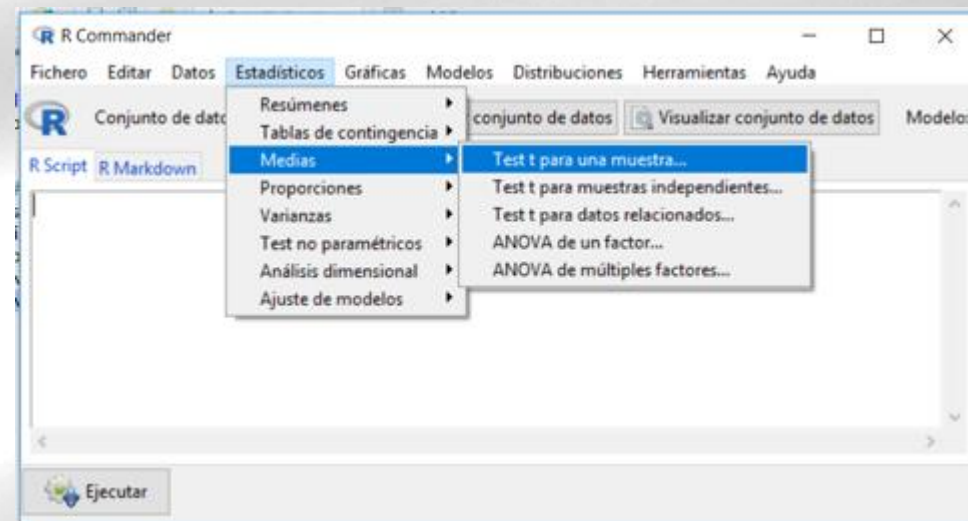
p-value = 0.0658

alternative hypothesis: true mean  
is not equal to 70

95 percent confidence interval:

69.81491 75.74509

sample estimates: mean of x 72.78





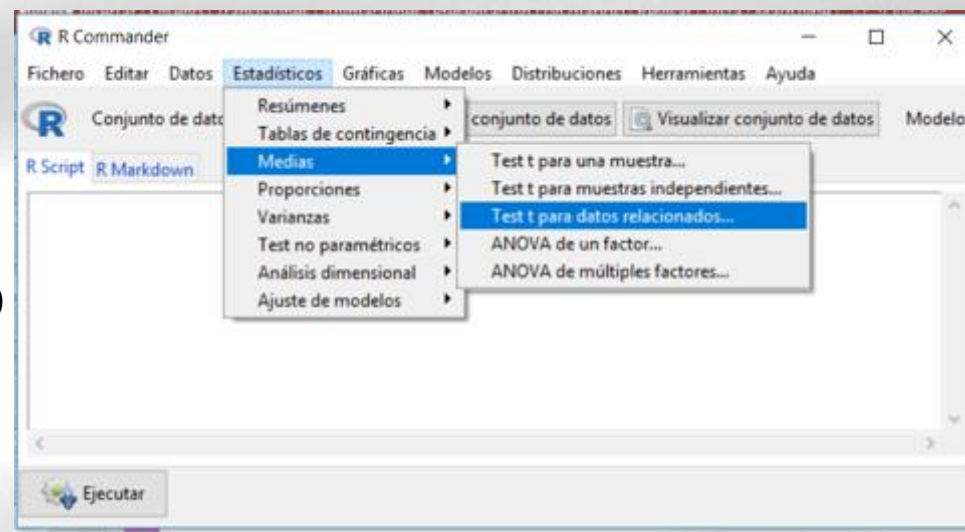
```
# osteoAll is the dataset obtained reading osteoporosis.csv
set.seed =123456 # change by your ID to obtain a distinct sample
index100<- sample( 1:nrow(osteoAll), 100)
osteo100 <- osteoAll[index100,]
with(osteo100, t.test (bua~menop, alternative="two.sided"))
```

## Welch Two Sample t-test

data: bua by menop t = 1.797,  
df = 90.184, p-value = 0.07568  
alternative hypothesis: true  
difference in means is not equal to 0

95 percent confidence interval: -  
0.5399937 10.7761048

sample estimates: mean in group NO mean  
in group SI 76.05556 70.93750



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

R Commander

Fichero Editar Datos Estadísticos Gráficas Modelos Distribuciones Herramientas Ayuda

Conjunto de datos conjunto de datos Visualizar conjunto de datos Modelo: leId=1889"

R Script R Markdown

- Resúmenes
- Tablas de contingencia
- Medias
  - Test t para una muestra...
  - Test t para muestras independientes...
  - Test t para datos relacionados...
- Proporciones
- Varianzas
- Test no paramétricos
- Análisis dimensional
- Ajuste de modelos

Ejecutar

10:1 (Top Level)

Console

```

Loading required package: sandwich
Loading required package: effects
Loading required package: carData

Attaching package: 'carData'

The following objects are masked from 'package:car':

  Guyer, UN, Vocab

lattice theme set by effectsTheme()
See ?effectsTheme for details.
RcmdrMsg: [1] NOTA: Versión de R Commander 2.4-2: Sun Mar 11 21:48:20 2018

Versión del Rcmdr 2.4-2

RcmdrMsg: [2] NOTA: El conjunto de datos osteo100 tiene 100 filas y 15 columnas.
>
  
```

Environment History Connections

Global Environment

Data

oste100	100 obs. of 15 variables
oste101	1000 obs. of 15 variables
tiki_download_f...	1000 obs. of 15 variables

Values

index100	int [1:100] 354 800 204 72 599 970 967 686 ...
set.seed	123456

Files Plots Packages Help Viewer

R: Student's t-Test Find in Topic

t.test [stats] R Documentation

## Student's t-Test

### Description

Performs one and two sample t-tests on vectors of data.

### Usage

```

t.test(x, ...)

## Default S3 method:
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)

## S3 method for class 'formula'
t.test(formula, data, subset, na.action, ...)
  
```

### Arguments

x a (non-empty) numeric vector of data values.

y an optional (non-empty) numeric vector of data values.

alternative a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less". You can specify just the initial letter.

mu a number indicating the true value of the mean (or difference in means if you are

Escribe aquí para buscar

21:50  
11/03/2018

Autoguardado EBB... Alexandre Sanchez Pla

Archivo Inicio Insertar Diseño Transiciones Animaciones Presentación con diaposi Revisar Vista MathType ¿Qué des

R Commander

Fichero Editar Datos Estadísticos Gráficas Modelos Distribuciones Herramientas Ayuda

Conjunto de datos conjunto de datos Visualizar conjunto de datos Modelo:

R Script R Markdown

Ejecutar

Resúmenes  
Tablas de contingencia  
Medias  
Proporciones  
Varianzas  
Test no paramétricos  
Análisis dimensional  
Ajuste de modelos

Test t para una muestra...  
Test t para muestras independientes...  
Test t para datos relacionados...  
ANOVA de un factor...  
ANOVA de múltiples factores...

Diapositiva 28 de 38 Inglés (Reino Unido) 59 %

Autoguardado dia2... - Se h... Alexandre Sanchez Pla

Archivo Inicio Insertar Diseño Transiciones Animaciones Presentación con diaposi Revisar Vista MathType ¿Qué des

Pegar Nueva diapositiva Diapositivas

Fuente Párrafo Dibujos

Formas Organizar Estilos rápidos Edición

Master in Tr Errors and power (in hypothesis testing)

$H_0$   
(innocent)  
(not speculative)

Accepted if data don't show the contrary  
Reject it by mistake (if it is true) has severe consequences

$H_1$   
(guilty)  
(speculative)

Should not be accepted without enough evidence  
Reject it erroneously has less dramatic consequences

Diapositiva 27 de 75 Inglés (Reino Unido) 59 %

## Errors and power (in hypothesis testing)

$H_0$   
(innocent)  
(not speculative)

Data can lead to reject it

Accepted if data don't  
show the contrary

Reject it by mistake (if it is true)  
has severe consequences



$H_1$   
(guilty)  
(speculative)

Should not be accepted without  
enough evidence

Reject it erroneously has less dramatic  
consequences

# Errors after Testing

		True	
		Innocent	Guilty
v e r e d i c t	Innocent	OK	Error
	Guilty	Error	OK

# Types of error

	<b>Null Hypothesis True</b>	<b>Null Hypothesis False</b>
<b>Test does not reject null hypothesis</b>	✓	<b>Type II Error</b> $\beta$
<b>Test rejects null hypothesis</b>	<b>Type I Error</b> $\alpha$	✓ <b>Power (1- <math>\beta</math>)</b>



# Common misunderstandings about the p-value



# Common misunderstandings about the p-value

---

- The p-value is **not** the probability that the null hypothesis is true, nor it is the probability that the alternative hypothesis is false (it is not connected to either of these).
- The p-value **cannot** be used to figure out the probability of a hypothesis being true.
- The p-value is **not** the probability of wrongly rejecting the null hypothesis.
- The p-value is **not** the probability that replicating the experiment would yield the same conclusion.
- The p-value does **not** indicate the size or importance of the observed effect. The two do vary together however: the larger the effect (effect size), the smaller sample size will be required to get a significant p-value.