Introduction to hypothesis tests

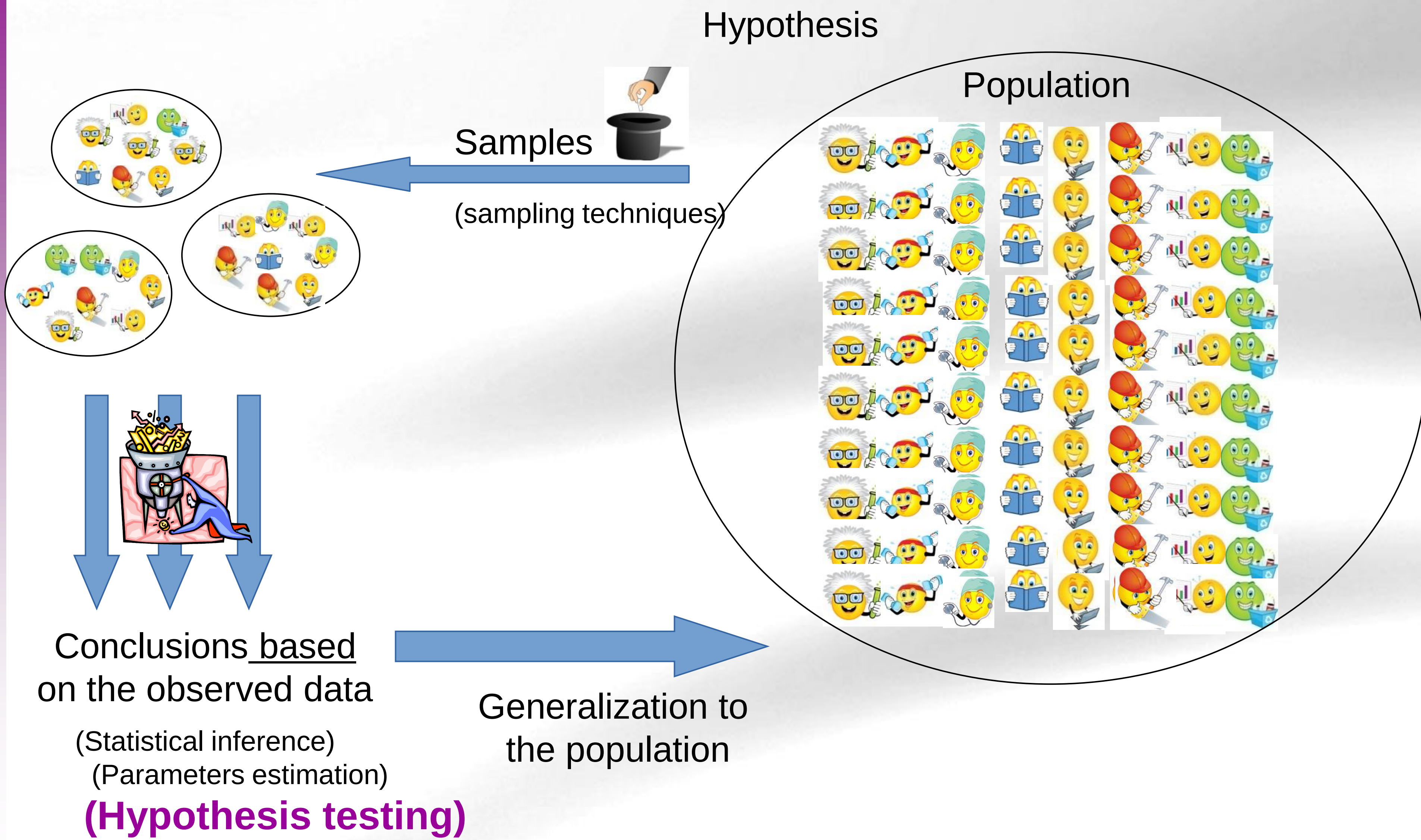
Curs d'Estadística Bàsica per a la Recerca Biomèdica

UEB – VHIR

Alex Sánchez-Pla & Santiago Pérez-Hoyos

santi.perezhoyos@vhir.org

The objectives of statistical inference



Statistical Inference Questions

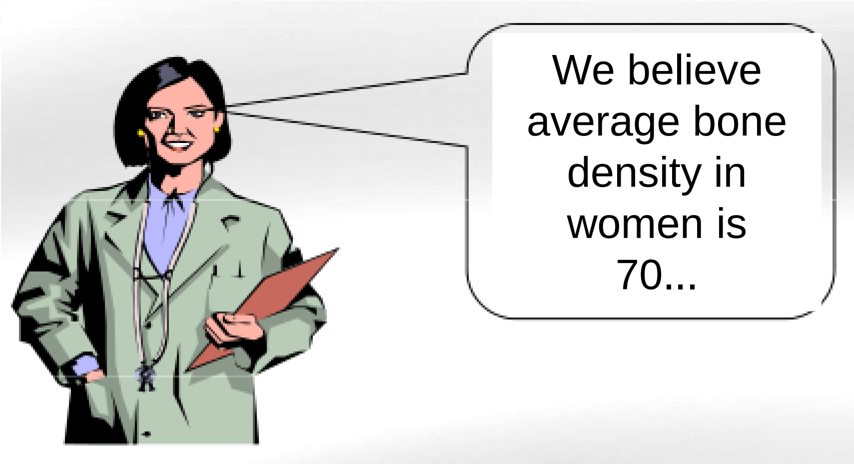
Parameter estimation:

- 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 respondants/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

Our guess:

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

Exercise 1):

- Explore osteoporosis data in order to get an idea about our first guess
- Do de same for the second question

- What other things you can figure out about the bone density in our population?

Case study problem II

- Cohort study with new lung cancer cases after 12 years of follow up (The NHANES Epidemiologic Follow-up Study. Am J Epidemiol 1997;146:231-243)

		Lung Cancer			
		Yes	No	Total	% Disease
Fruit Consumption	High	44	2473	2517	**1.8%**
	Low	88	2429	2517	**3.5%**

Total lung cancer rate =132/ 5034=2.6%

- 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 chance in sampling?

The setting of hypothesis testing

- We establish two basic hypothesis

Null Hypothesis (H_0)

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

Alternative Hypothesis (H_1)

- 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, that is *if lung cancer % is the same for both groups*

p → Low —decide→ **Groups must be different**

p → High —decide→ **Groups can be equal**

¿How to calculate this probability?

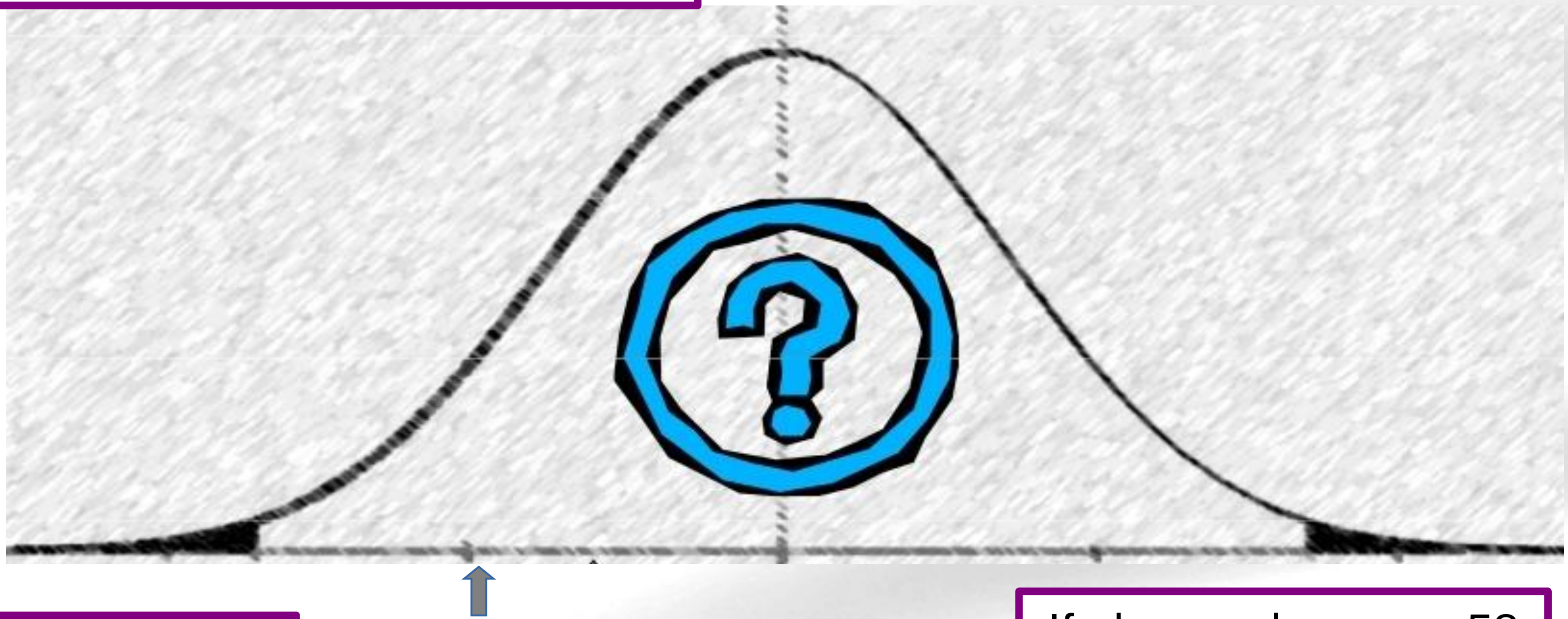
- Depends on type of study.
- Depends on type of variable.
- Depends on the influence of other variables.
- This probability is associated with the error made assuming that the null hypothesis is true when it is not.
- Needs to be carefully interpreted for example…
  - Showing there is statistical association does not mean there is causal relations, moreover if data are obtained from cross-sectional observational studies.

Hypothesis Testing Steps

- Stablish Null Hypothesis(H_0)
- Stablish 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
  - We cannot reject null hypothesis

Accepting or rejecting the NULL

H_1 only accepted if clear evidence that H_0 is not true



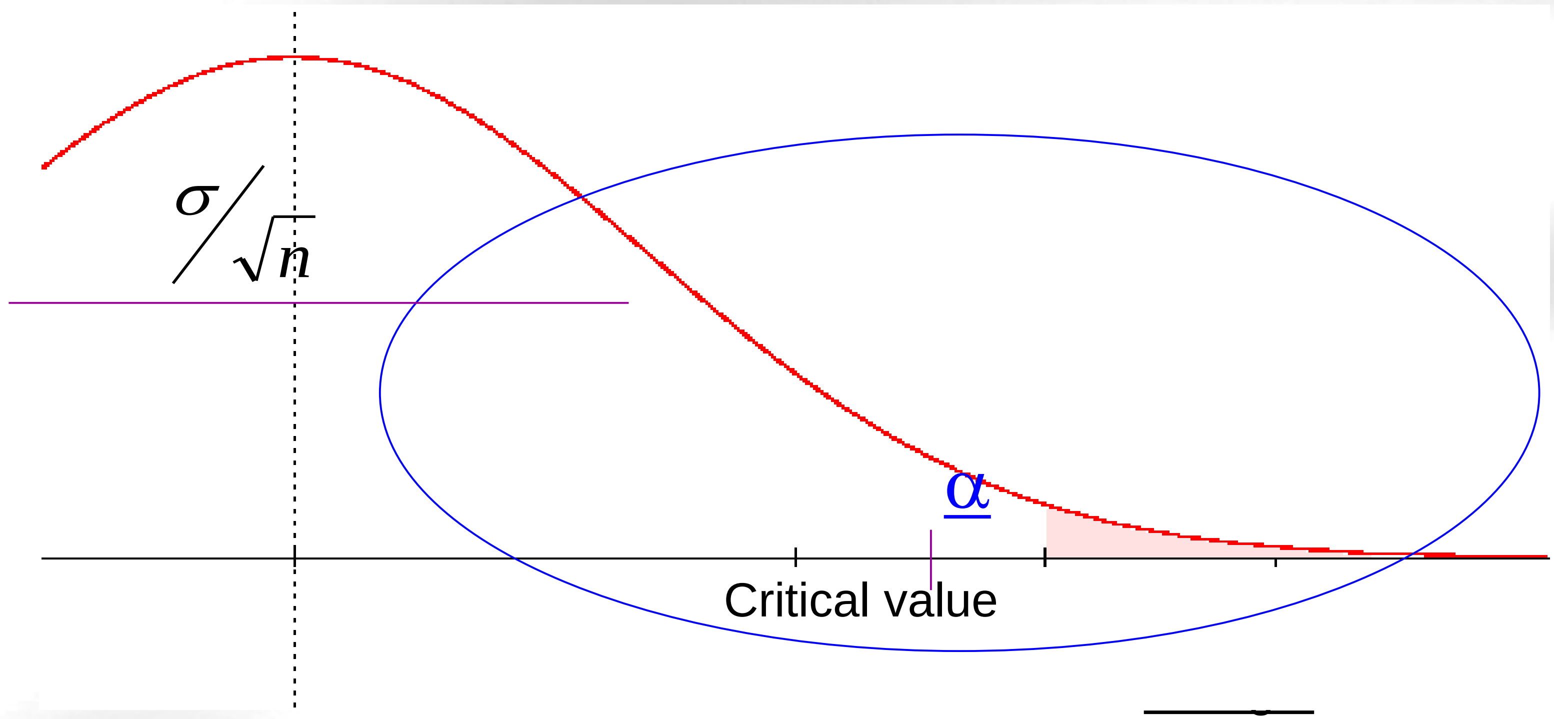
μ = 70

If observed mean = 18
H_0 is rejected

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

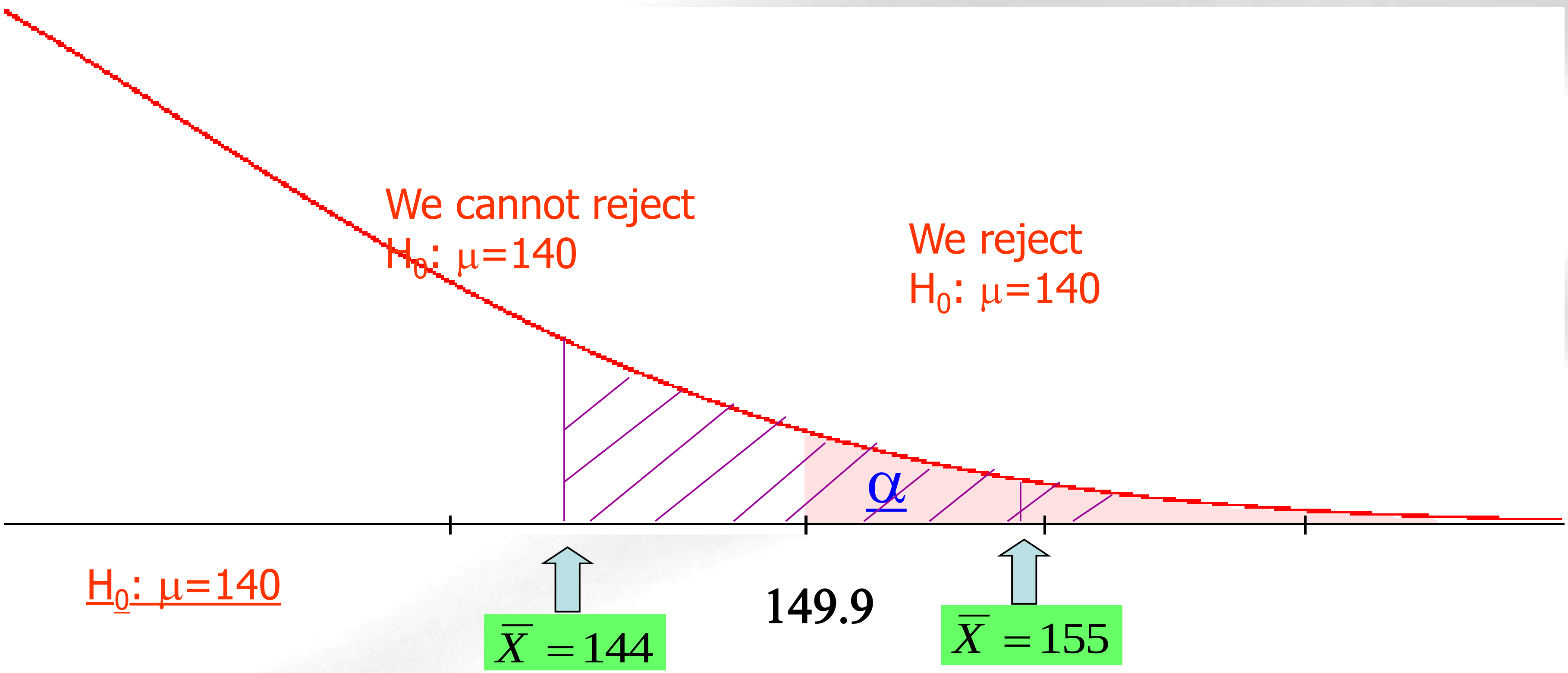
Critical Value

- 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 (α)”



Example: Critical value and Sample mean

- ***If σ =18, n =9 and α=0.05 the critical value will be 149,9***
  - With a sample mean of 144 we will not reject H_0
  - With a sample mean of 155 we will t reject H_0



Example

Null hypothesis (H_0):

$H_{0:}$ The mean of BUA values is 70.0

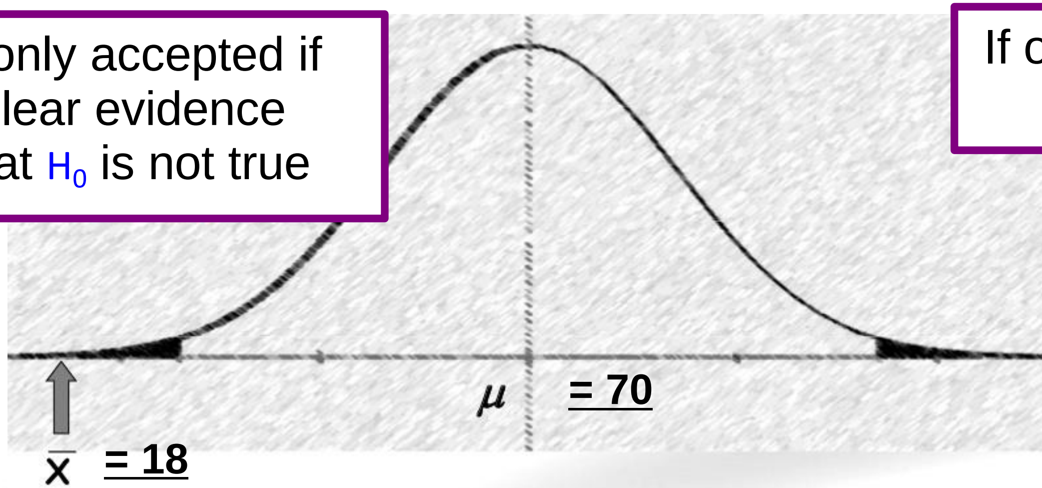
Alternative hypothesis (H_α =H_a =H_1): the opposite idea

- H1: The mean of the bua values is not equal to 70.0 (Bilateral)
- H1: 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 distribition is as follows

H_1 only accepted if clear evidence that H_0 is not true

If observed mean = 18
H_0 is unlikely



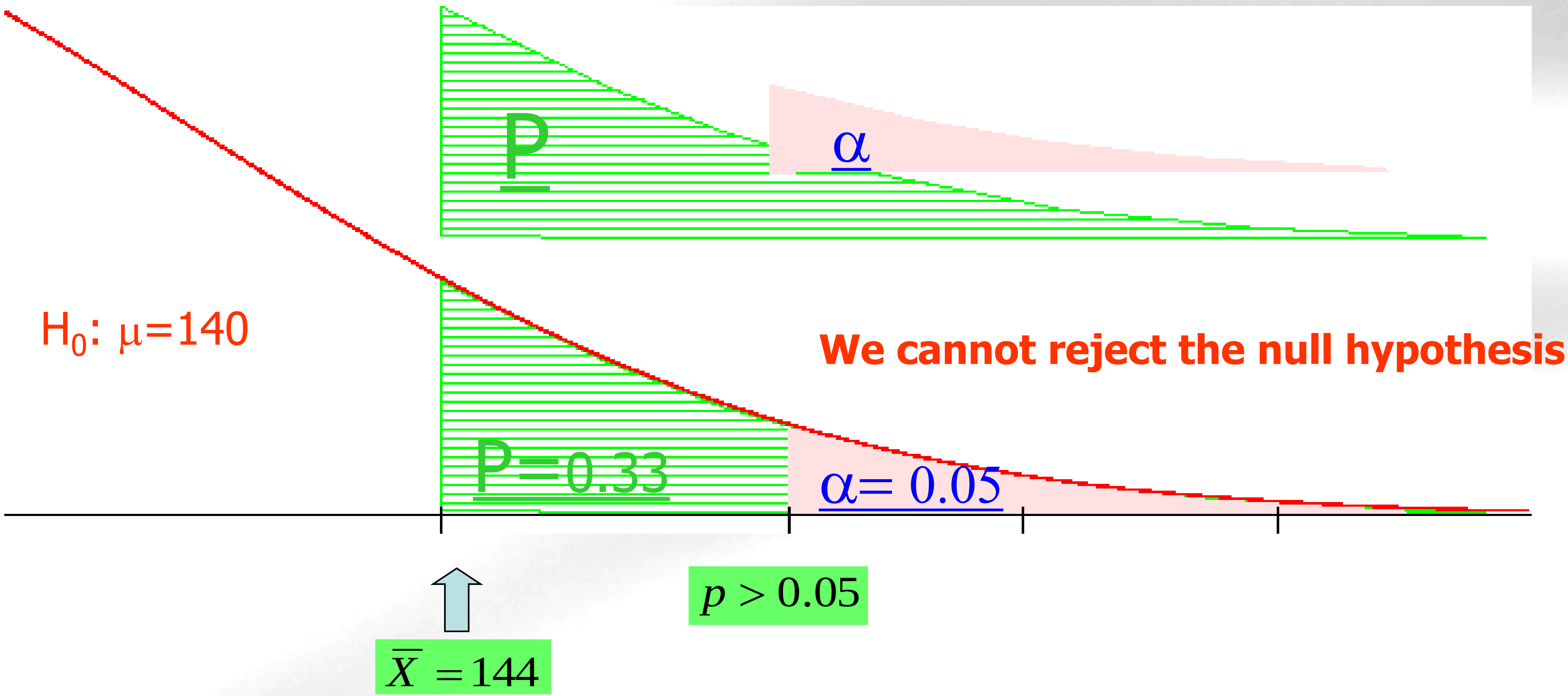
P values: The alternative

- We have based our decision about rejecting H_0 on comparing sample mean (i.e. 144) with the critical value (i.e. 149.9)
- Instead we can compare the probability of observing at least that sample mean (p value) with the significance level (α) (which is the probability of observing at least the critical value),
  - The probabiliy is smaller than alfa if (and only if) the sample mean is bigger than the critical value.
    - In such situation we decide to reject H0
  - The probabiliy is bigger than alfa if (and only if) the sample mean is smaller than the critical value.
    - In such situation we cannot reject H0 so we accept it
- Both criteria (critical value and p-value) are valid for testing hypotheses.

Example: P-value vs critical value

- ***If σ=18, n =9 and the sample mean is 144 then***
- The probability that assuming that H0 is true, that is m=140, we can observe by chance a sample greater than 144 is: 0.328

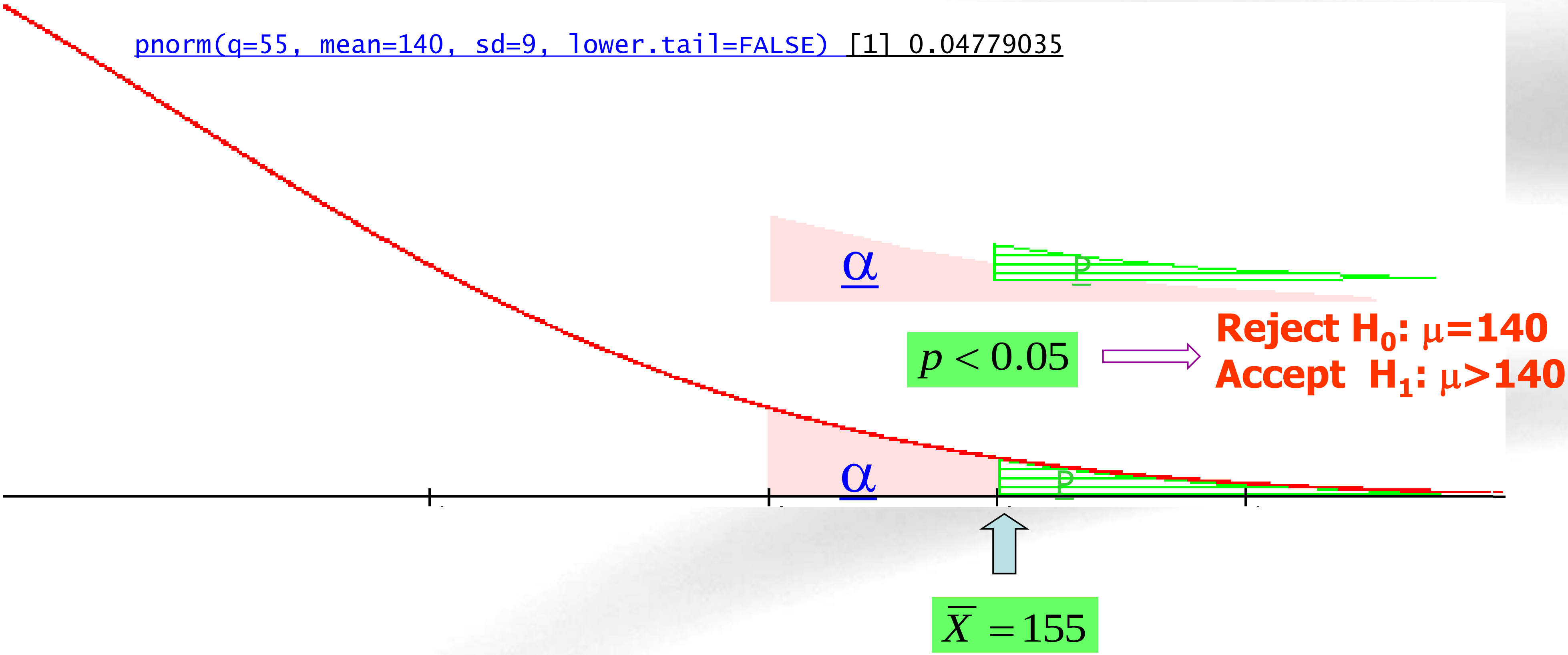
`pnorm(q=144, mean=140, sd=9, lower.tail=FALSE) [1] 0.3283606`



Example: P-value vs critical value

- ***If σ=18, n =9 and the sample mean is 155 then***
- The probability that assuming that H_0 is true, that is μ=140, it can be obtained by chance a sample with a mean greater than 155 is: 0.0478

`pnorm(q=55, mean=140, sd=9, lower.tail=FALSE) [1] 0.04779035`



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

Summary: α vs p

α and P are related but they are not the same …

- About α
  - 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 hipothesis about the distribution of variable in population

- Does arterial pression 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

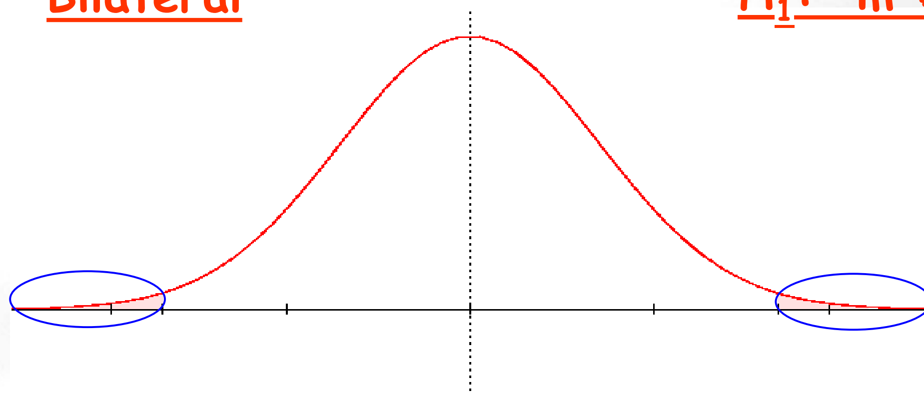
- Is the average "bua" value the same in menopausic and in no menopausic population?
- Is the proportion of lung cancer cases the same in people with high or low fruit consumption?
- Is CD4 lymphocytes count related with CD8 count in HIV positive?

Unilateral vs Bilateral

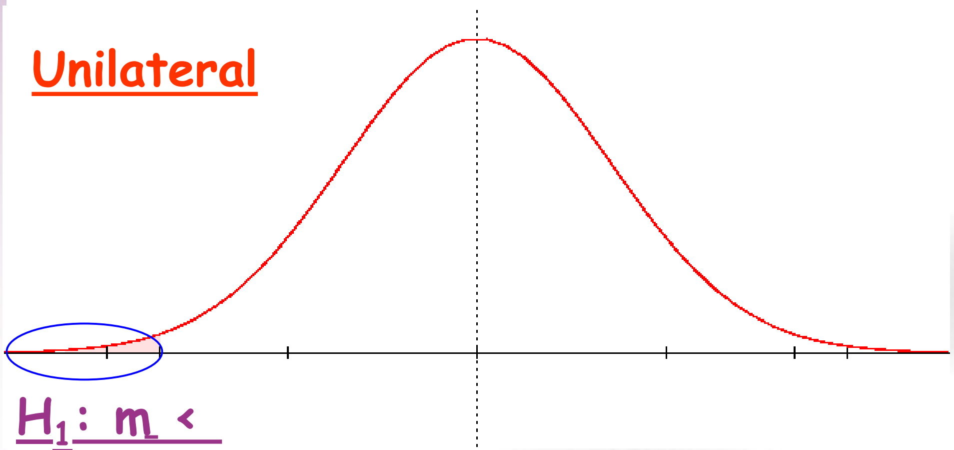
Critical value depends on the type of alternative Hypothesis

Bilateral

H_1: m ≠

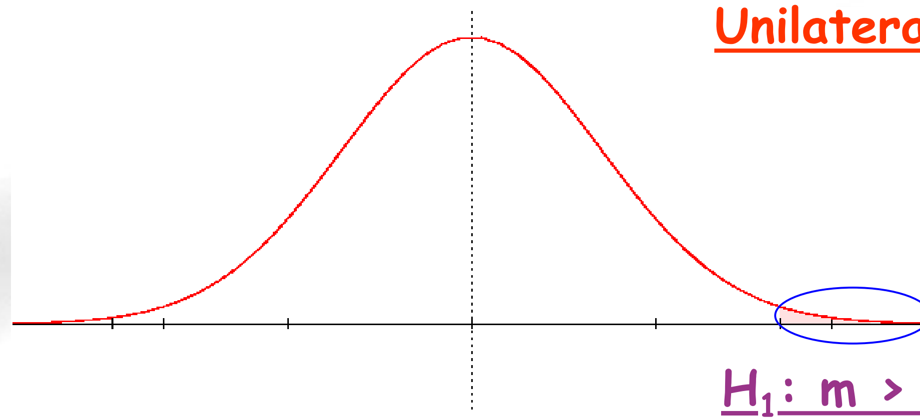


Unilateral



H_1: m <

Unilateral



H_1: m >

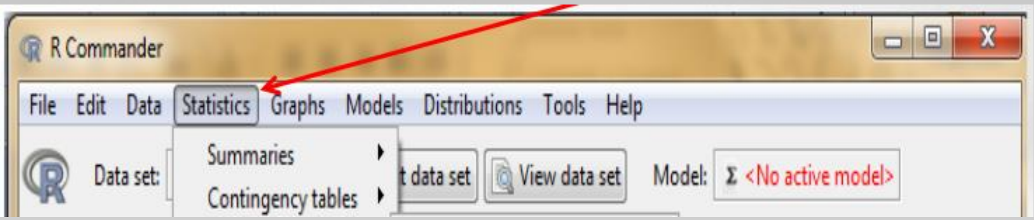
Example: hypothesis testing with R-commander

Our assumptions:

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

Exercise 2):

Hint: look at this menu

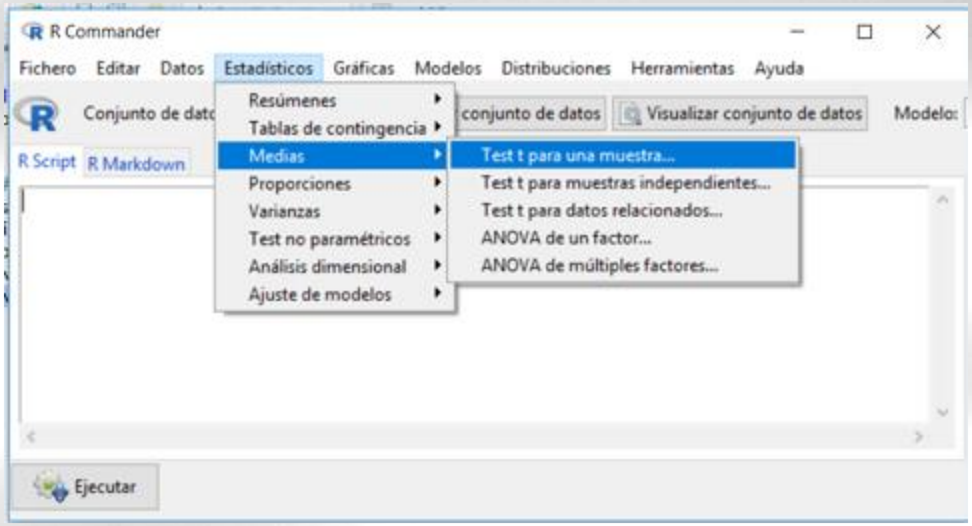


- Test if the population mean bone density is 70.0 (Alternative “it is not 70”)

- Test if the population mean bone density is equal or not between groups if we separate our observations by "menop" category

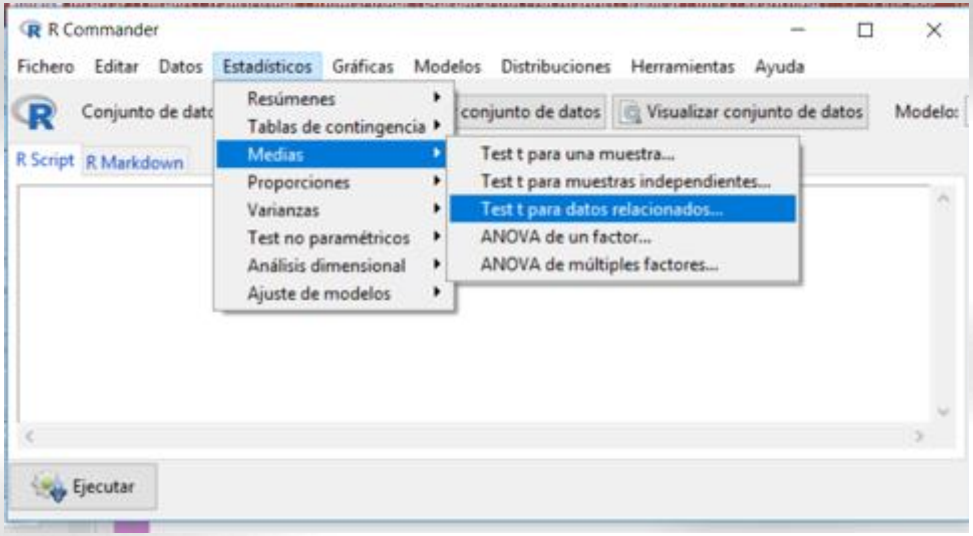
```
# osteoAll is the dataset obtained reading osteoporosis.csv
set.seed =123456 # change by your ID to obtain a distinct sample
index100<- sample( 1:nrow(osteoAll), 100)
osteo100 <- osteoAll[index100,]
with(osteo100, t.test (bua, mu=70, alternative="two.sided"))
```

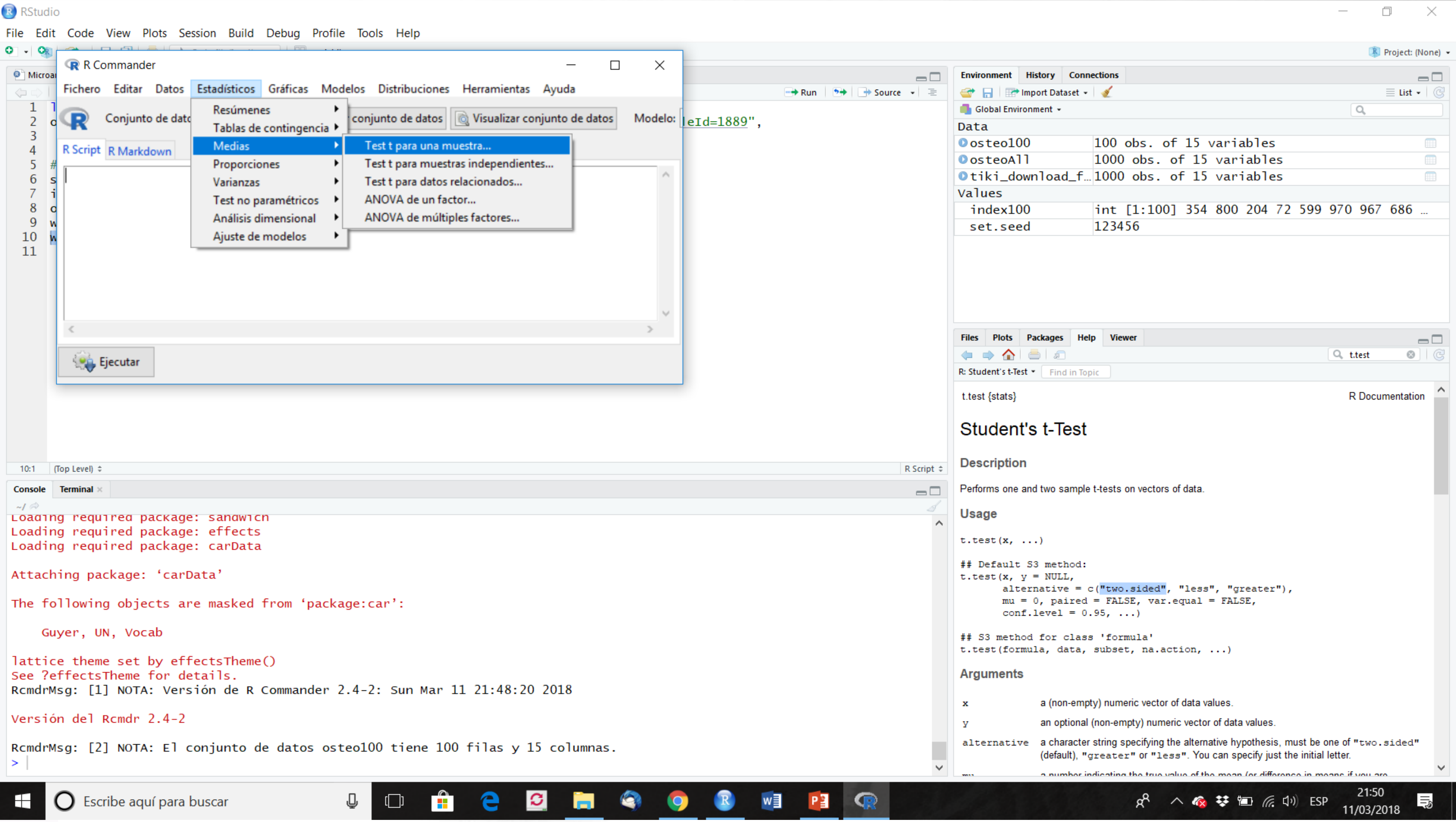
```
t.test(bua, mu=70)

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"))
```

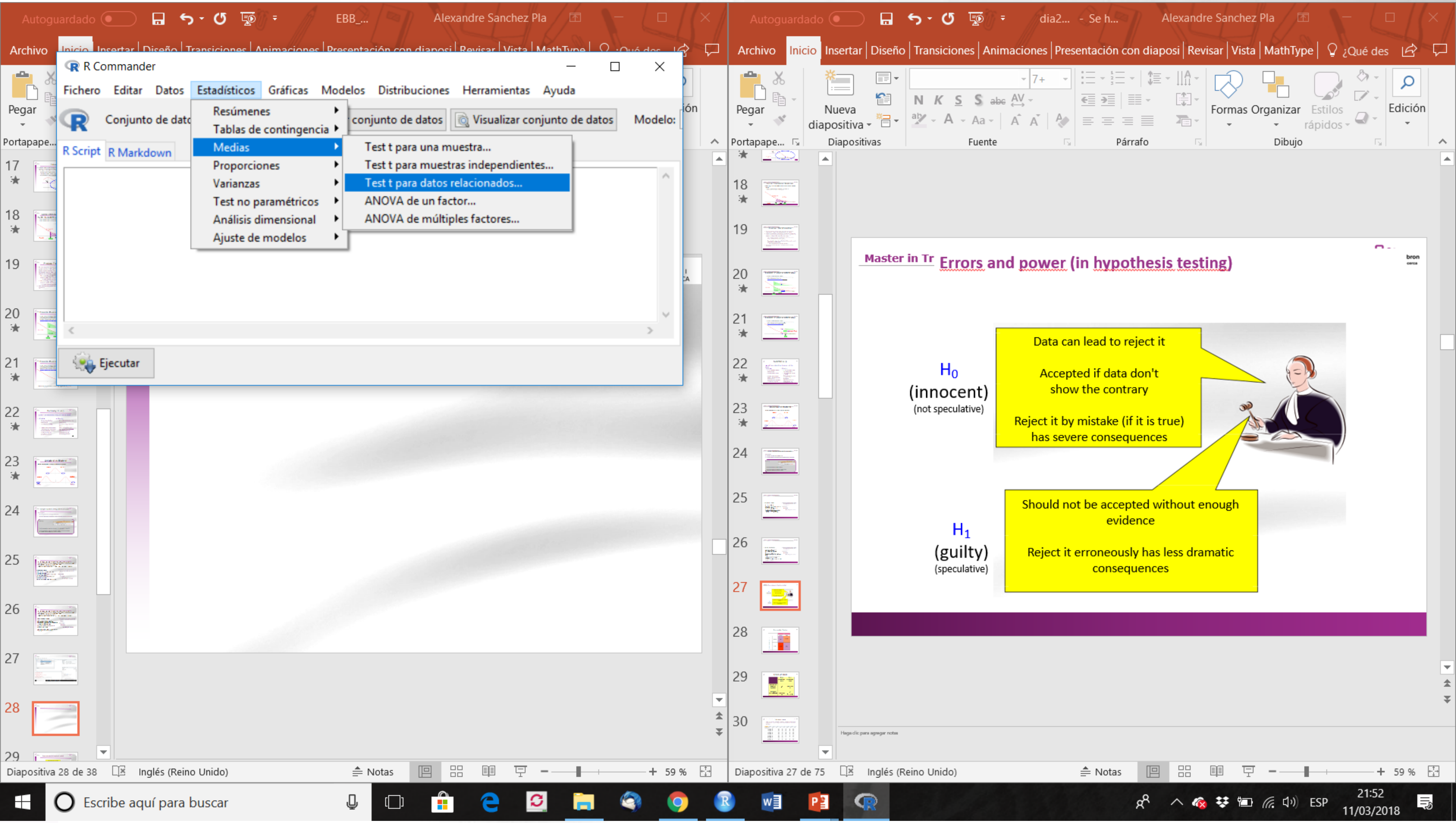
```
Welch Two Sample t-test

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
Visualizar conjunto de datos
Modelo:
R Script R Markdown
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...
Ejecutar
Environment History Connections
Import Dataset
Global Environment
Data
osteo100 100 obs. of 15 variables
osteoAll 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
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.
Console Terminal
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.
Escribe aquí para buscar
21:50
11/03/2018

Vall d'Hebron Institut de Recerca

UNITAT D'ESTADÍSTICA I BIOINFORMÀTICA



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
verdict	Innocent	OK	Error
	Guilty	Error	OK

Types of error

	Null Hypothesis True	Null Hypothesis False
Test does not reject null hypothesis	✓	Type II Error β
Test rejects null hypothesis	Type I Error α	✓ Power (1- β)

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.