

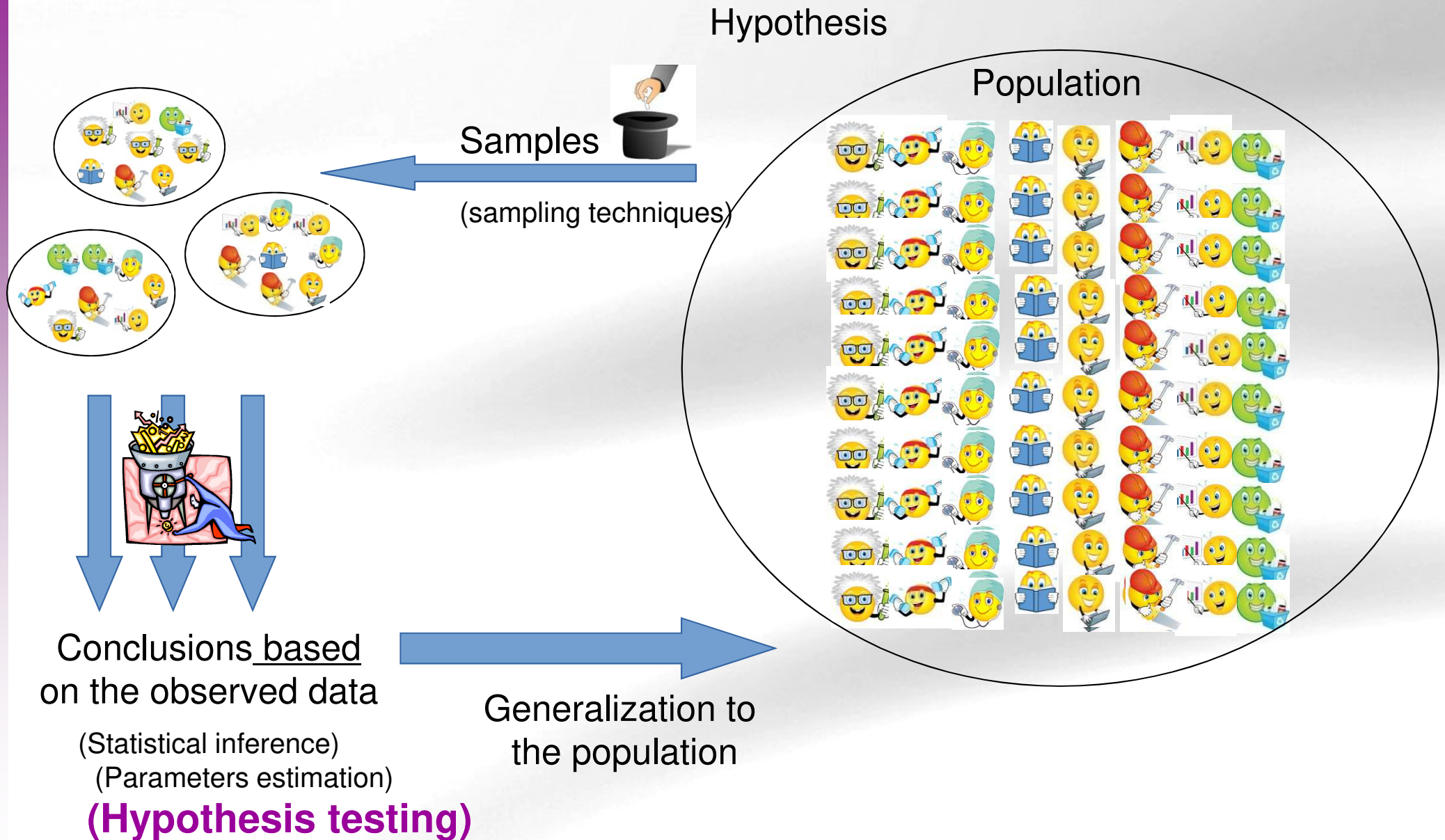
Curs d'Estadística Bàsica per a la Recerca Biomèdica

UEB – VHIR

Santiago Pérez-Hoyos

santi.perezhoyos@vhir.org

The objectives of statistical inference



Statistical Inference Questions

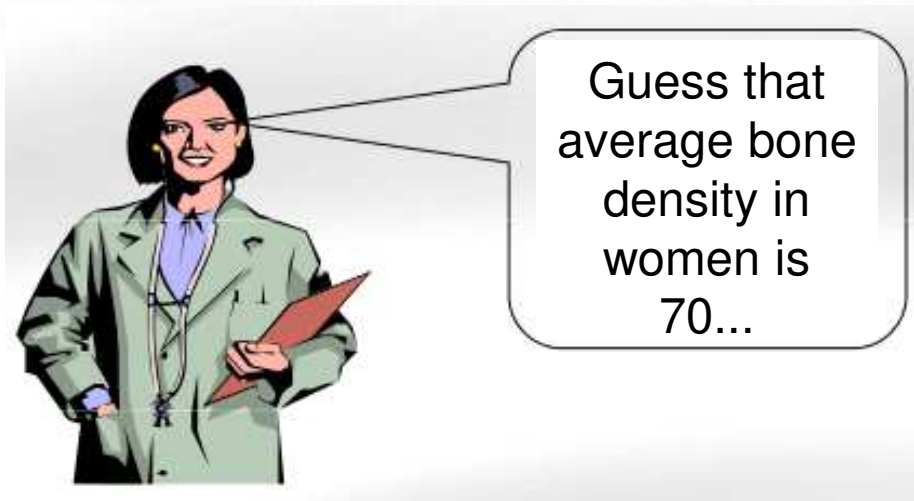
Parameter estimation:

- After assuming population data follow a certain probability distribution (normal, Binomial, Poisson, etc) which are the value of the parameters that better fit the sample data.

Hypothesis testing:

- We have an assumption about the population data parameters
 - Population mean is equal to 10
 - The mean in population A is equal to the mean in population B
- Hypothesis testing tries to verify if sample data are compatible with that hypothesis

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 menopausal and non-menopausal women is not the same

Exercise 1):

- Explore osteoporosis data in order to get an idea about our first guess
- Do the 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 = $132 / 5034 = 2.6\%$

- After this outcome can we say that fruit consumption is a protective factor for lung cancer?
- How can I say this results is not caused by sampling or random?

-
- We can establish two basic hypotheses

Null Hypothesis (H_0)

- Outcome is observed by chance.
- No relationship among exposure and disease.

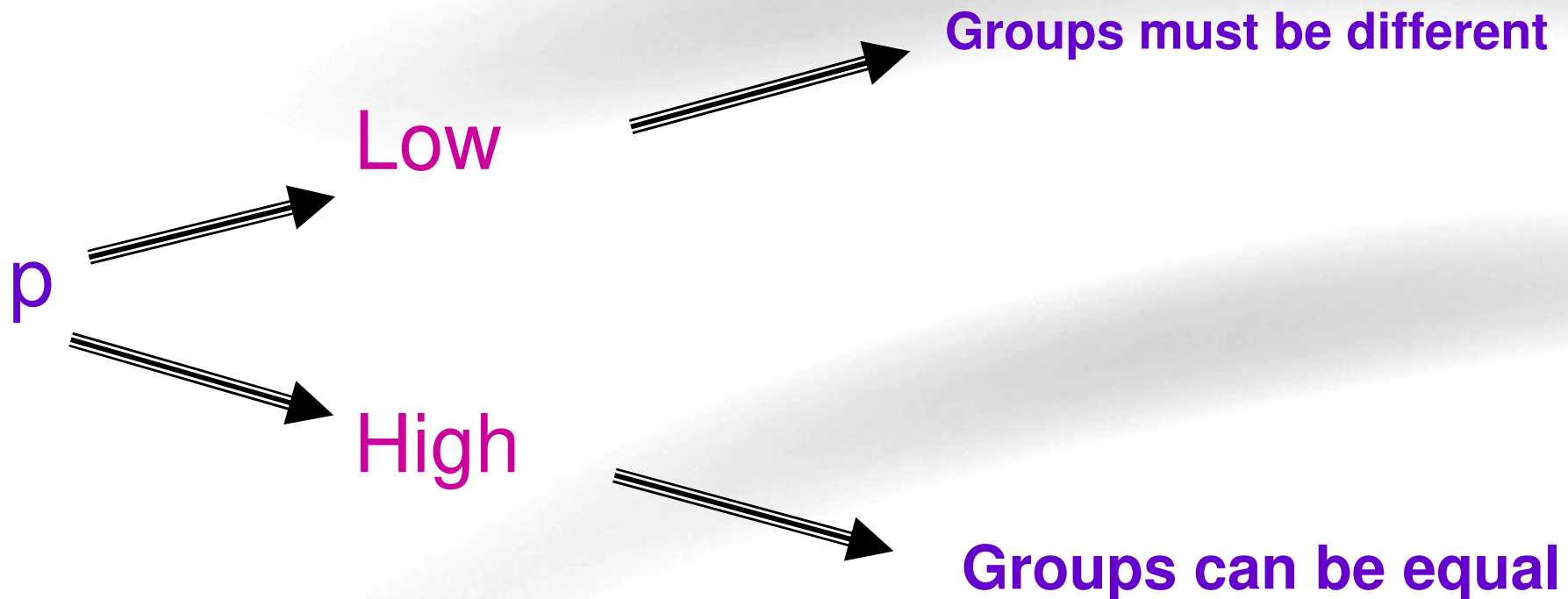
Alternative Hypothesis (H_1)

- There is relationship among exposure and disease.

¿How to decide with hypothesis is more likely?

Calculate probability (p) to observe differences between both groups under the hypothesis of no differences

Lung cancer % is the same for both groups 2.6%



¿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 the error of assume that there is relation when there is NOT.
- Statistical relations do not mean causal relations, moreover if are obtained in cross-sectional studies.

Hypothesis Testing

- Test population hypothesis from samples
 - Establish Null Hypothesis(H_0)
 - Establish Alternative Hypothesis (H_a)
 - Select statistical test to calculate probability ***under Null Hypothesis***
 - Decide after comparing test value with a critical value or probability under null hypothesis.

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 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 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?

Parametric Test

It is assumed that the variable under study follow a particular distribution and values about parameters are tested

- Distribution of proportion of lung cancer is binomial

$$H_0: p = 3\%$$

- BUA is the same in menopausal and non menopausal and variable is normal or symmetric

$$H_0: \mu_{\text{Menopausal}} = \mu_{\text{Non menopausal}}$$

- Distribution is binomial and proportion of lung cancer is the same in high and low fruit consumers

$$H_0: p_{\text{High fruit}} = p_{\text{Low fruit}}$$

Non Parametric Test

No distribution is assumed and test are related to distribution not to values about parameters

- Distribution of bua follow a normal distribution
- Bua is the same in menopausic and non menopausic and variable is normal or symmetric

H_0 : distribution in Menopausic= Distribution in non menopausic

- Lung cancer is not related to fruit consumption. They are independent.

Example

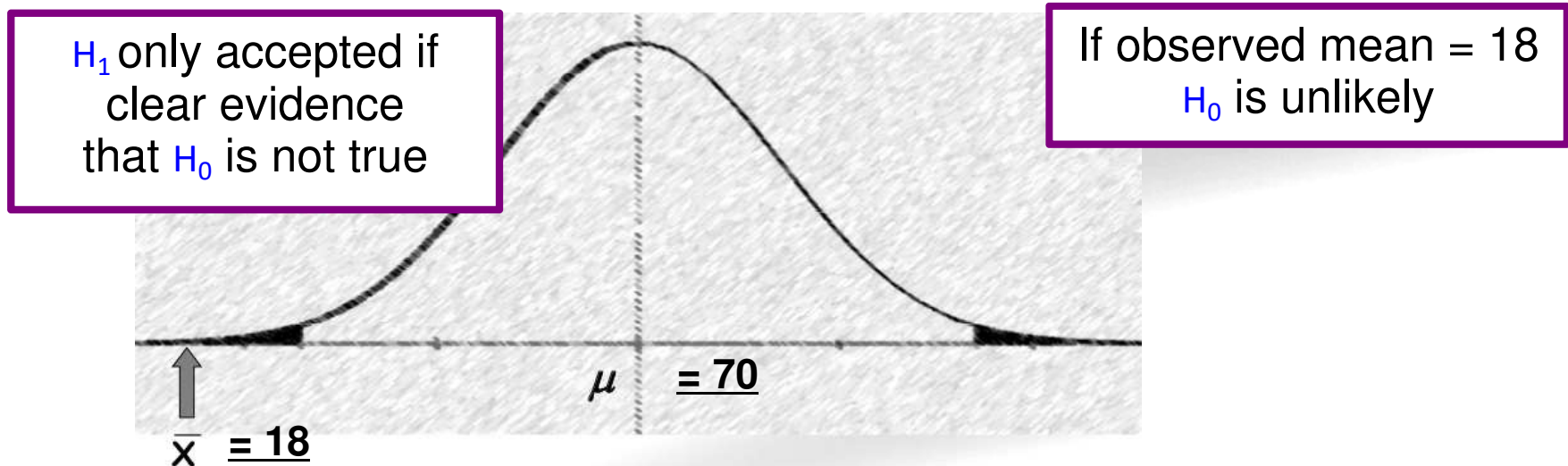
Null hypothesis (H_0):

H_0 : The mean of BUA values is 70.0

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

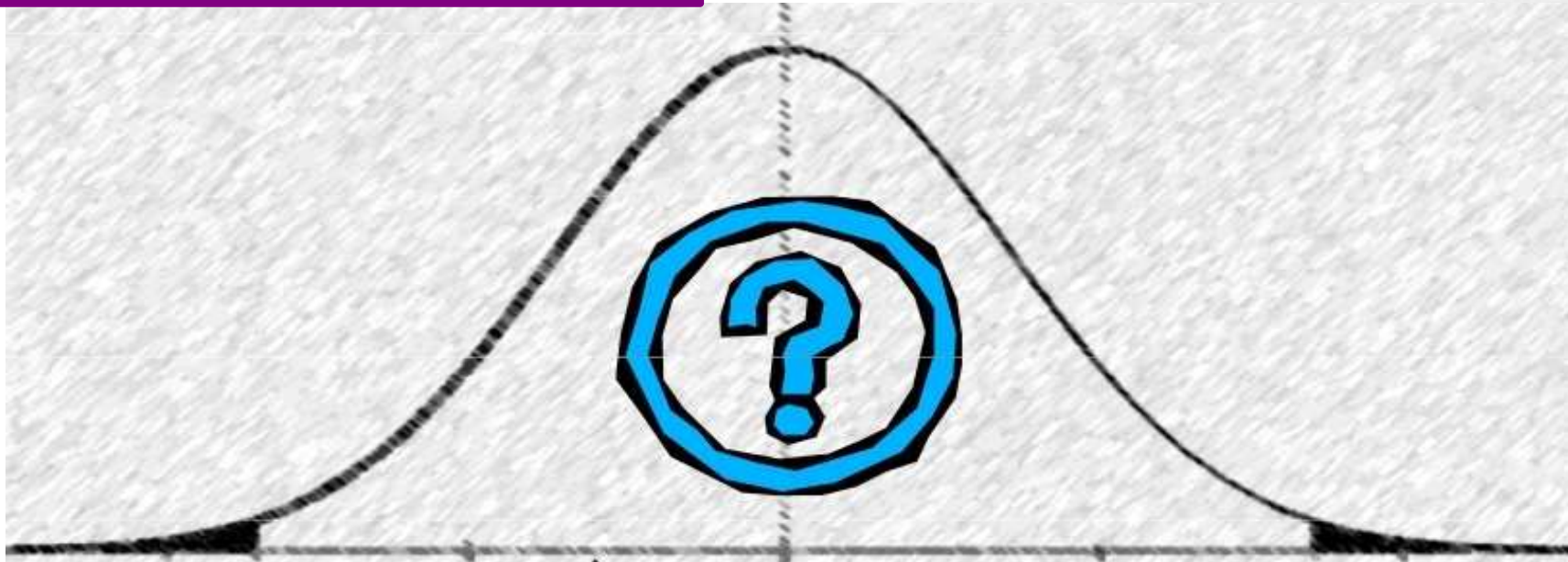
- H_1 : The mean of the bua values is not equal to 70.0 (Bilateral)
- H_1 : The mean of the bua values is higher(lower) than 70.0 (Unilateral)

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



Accepting or rejecting the NULL

H_1 only accepted if clear evidence
that H_0 is not true



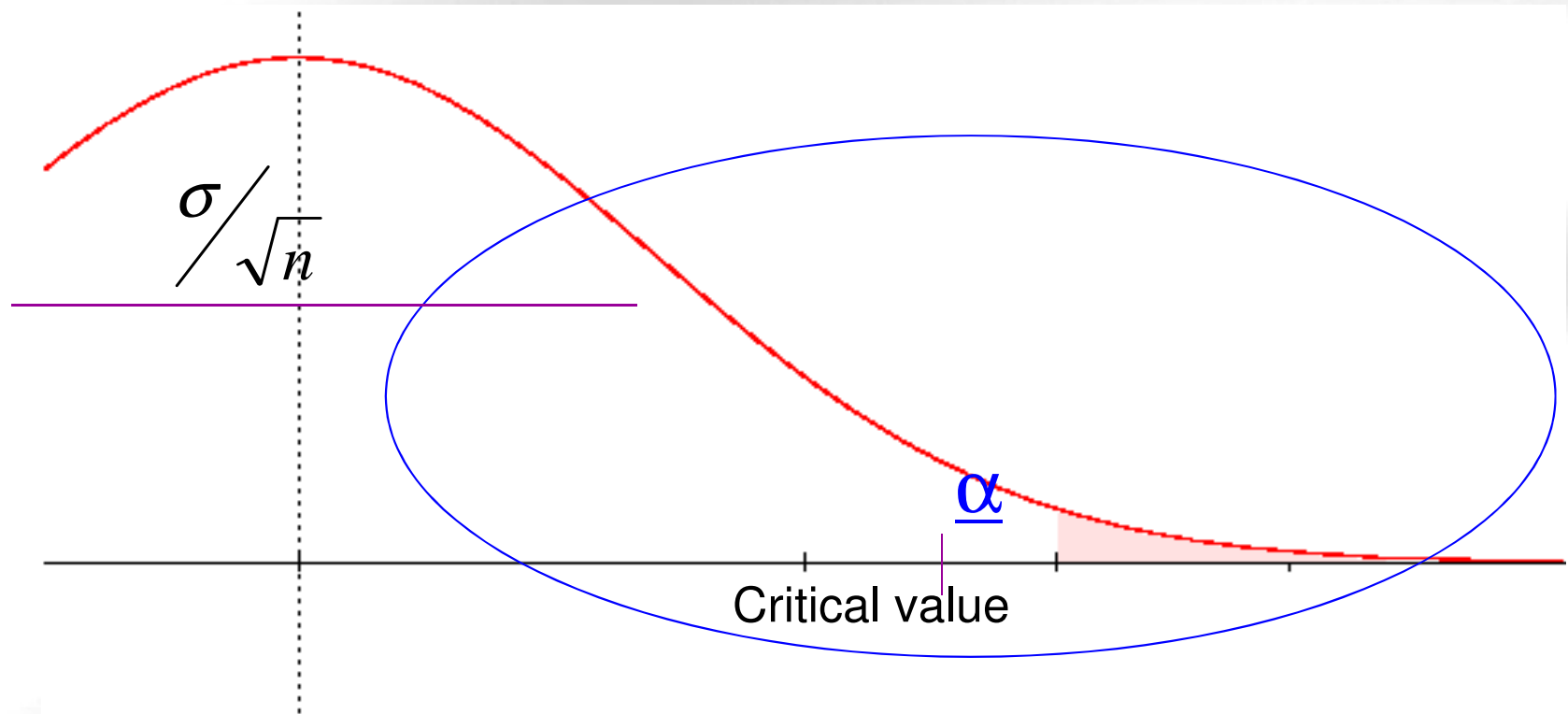
If observed mean = 18
 H_0 is rejected

$\mu = 70$

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

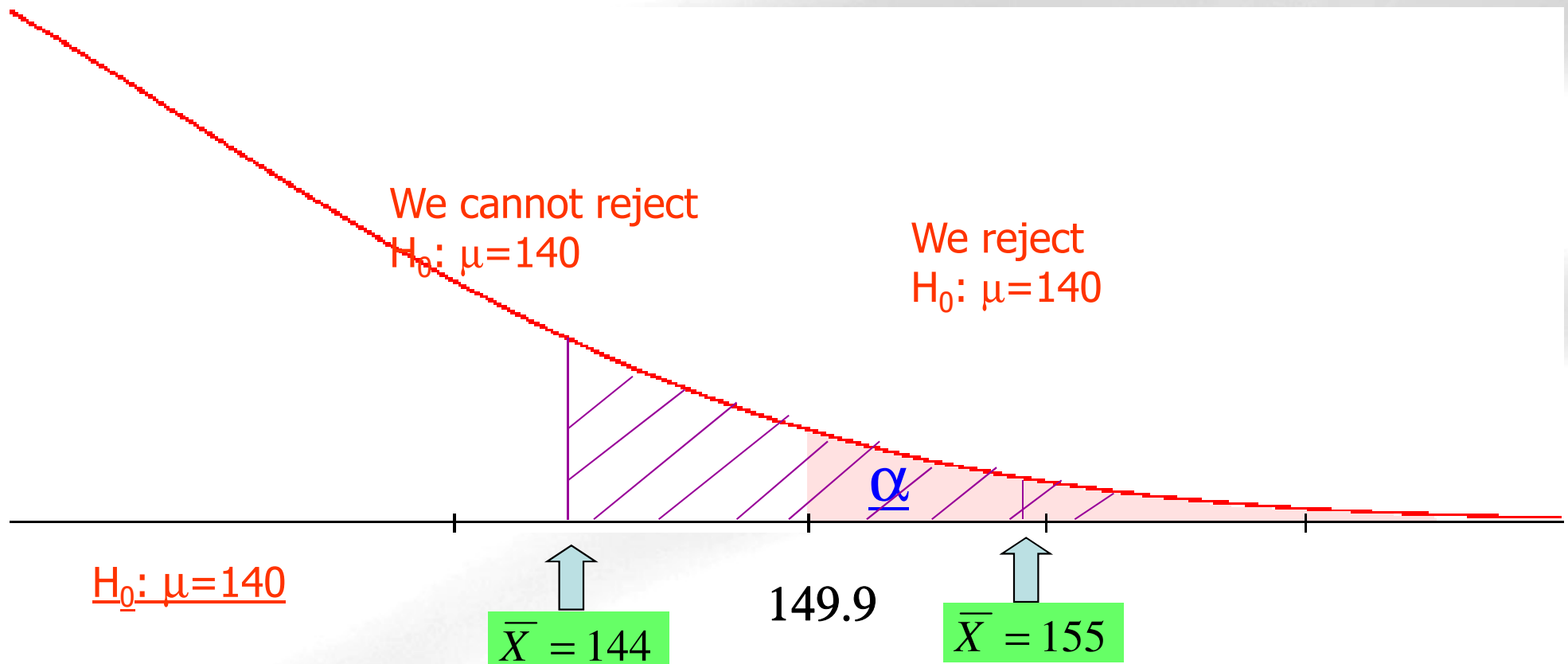
Critical Value

- 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 $\sigma=18$, $n=9$ and $\alpha=0.05$ 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



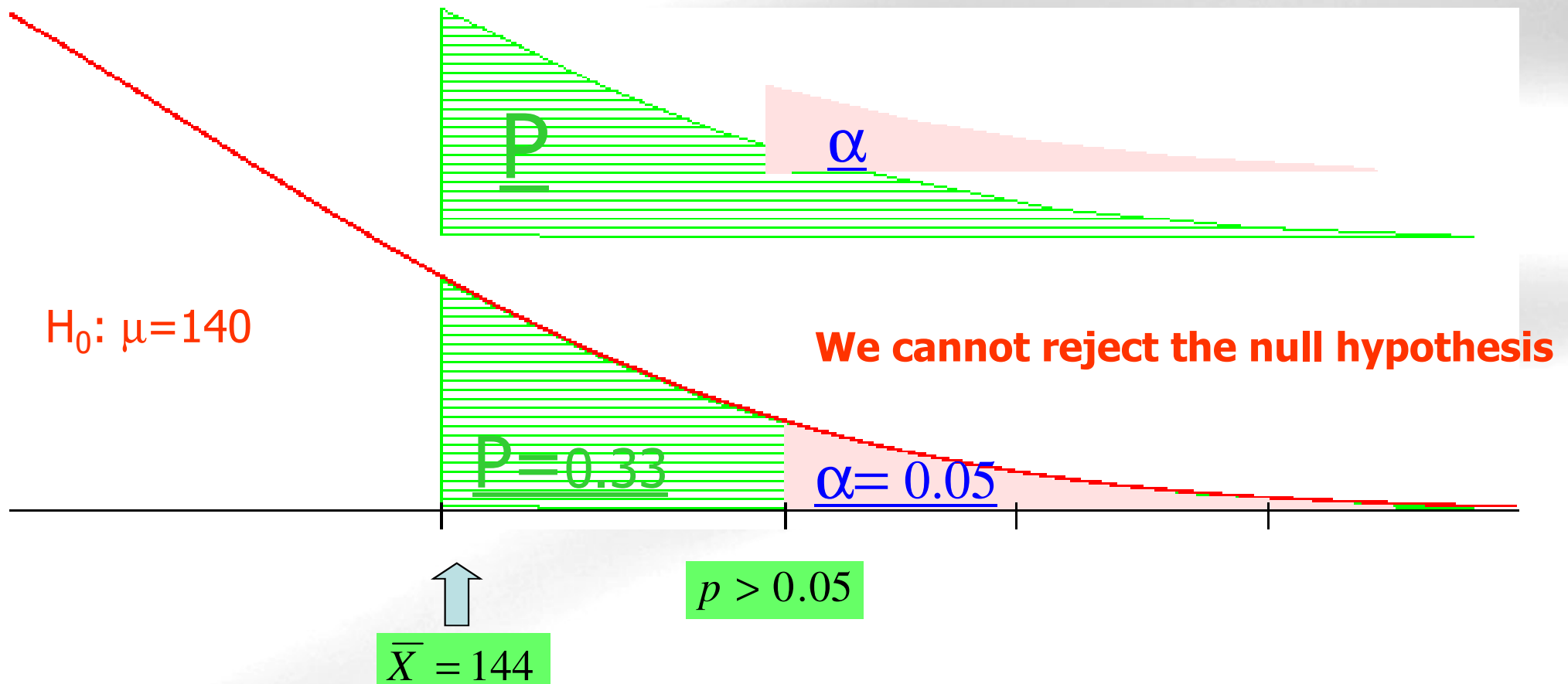
P values: The alternative

- 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 probability is smaller than α 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 α 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 $\sigma=18$, $n=9$ and the sample mean is 144 then*
- The probability that assuming that H_0 is true, that is $\mu=140$, we can observe by chance a sample greater than 144 is: 0.328

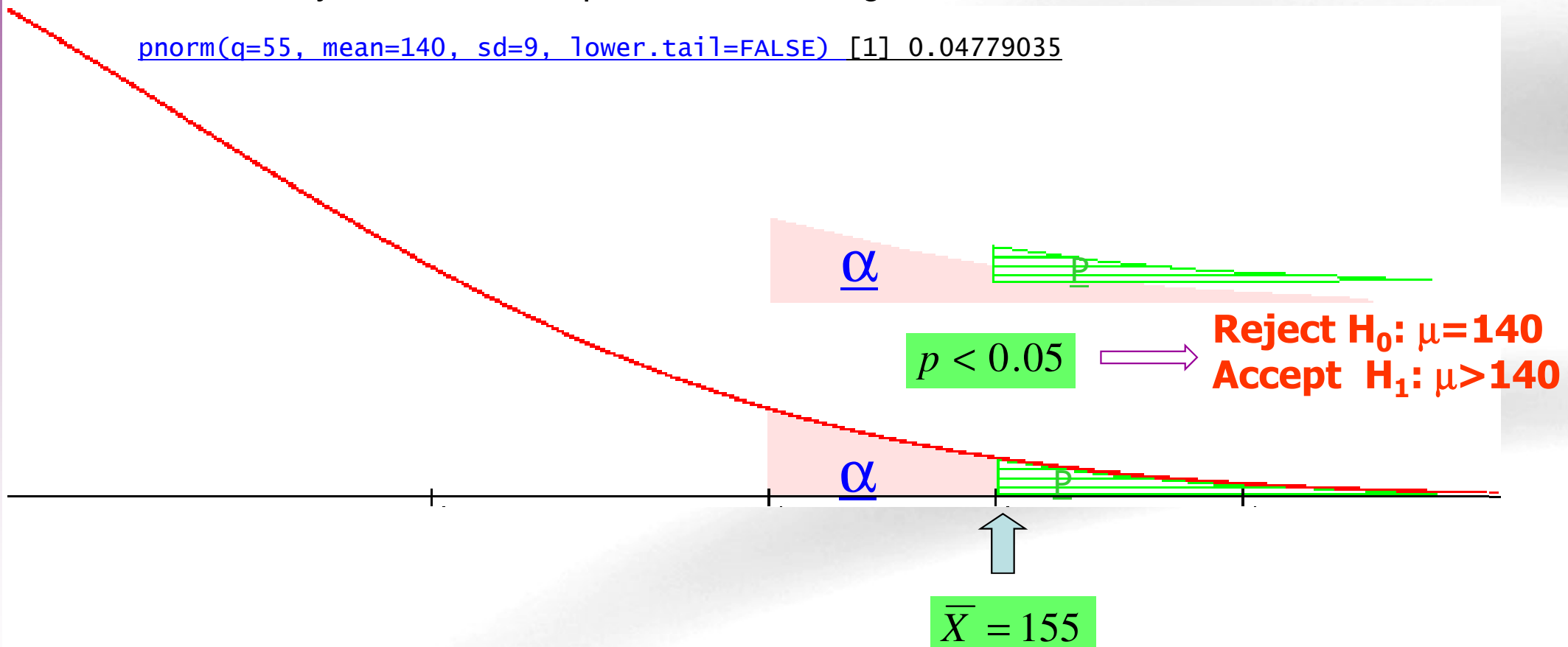
`pnorm(q=144, mean=140, sd=9, lower.tail=FALSE)` [1] 0.3283606



Example: P-value vs critical value

- *If $\sigma=18$, $n=9$ and the sample mean is 155 then*
- The probability that assuming that H_0 is true, that is $\mu=140$, 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 $p < \alpha$

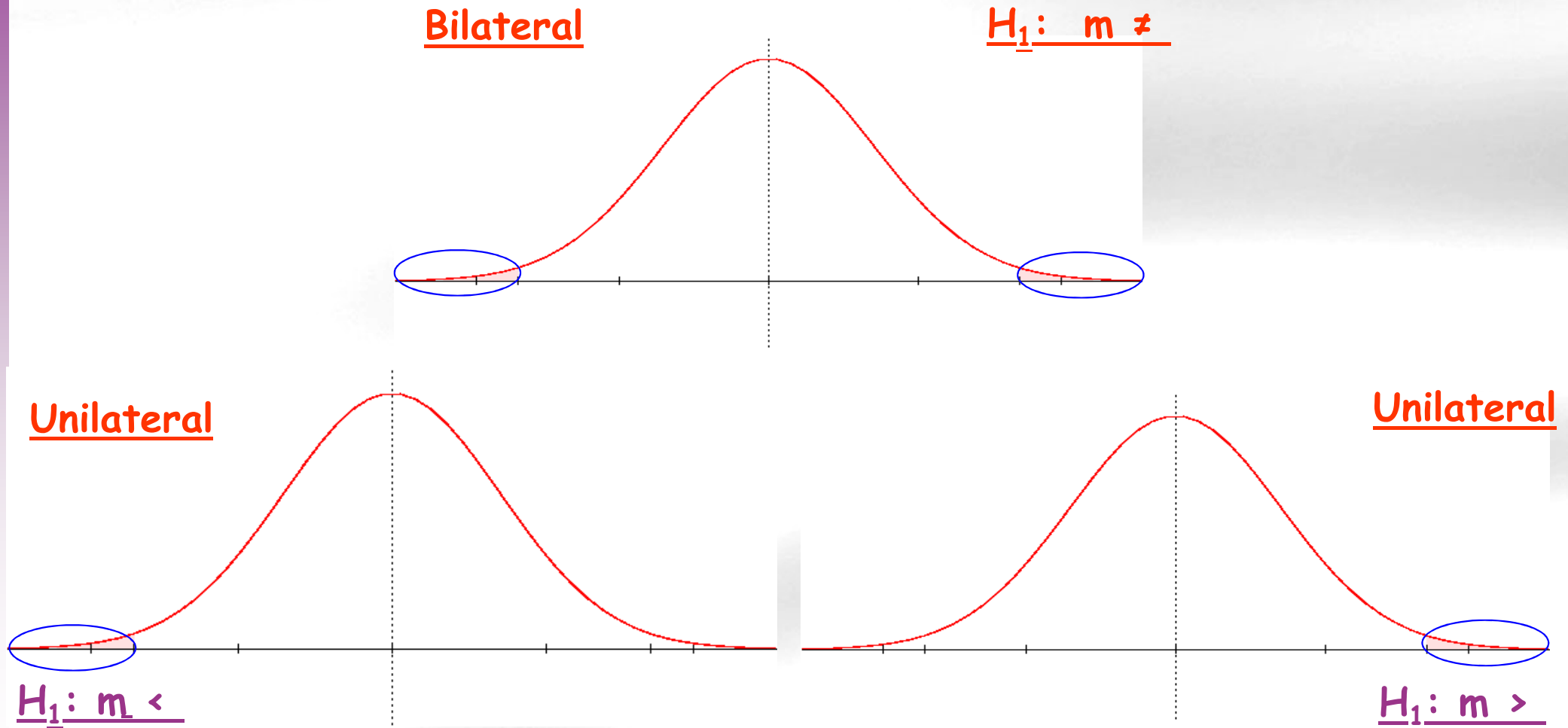
Summary: α vs p

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

Unilateral vs Bilateral

Critical value depends on the type of alternative Hypothesis

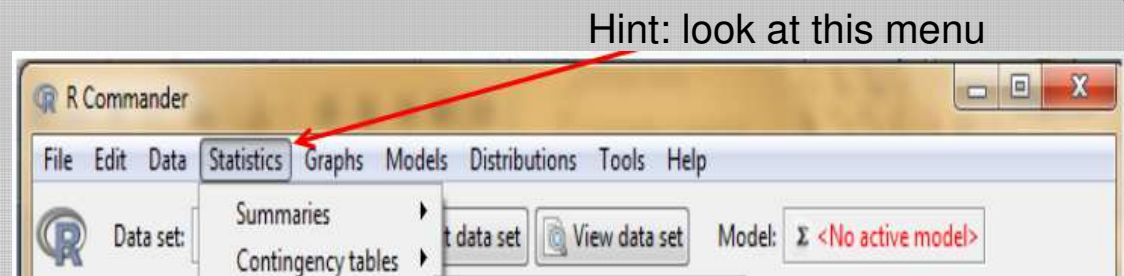


Example: hypothesis testing with R-commander

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"))
```

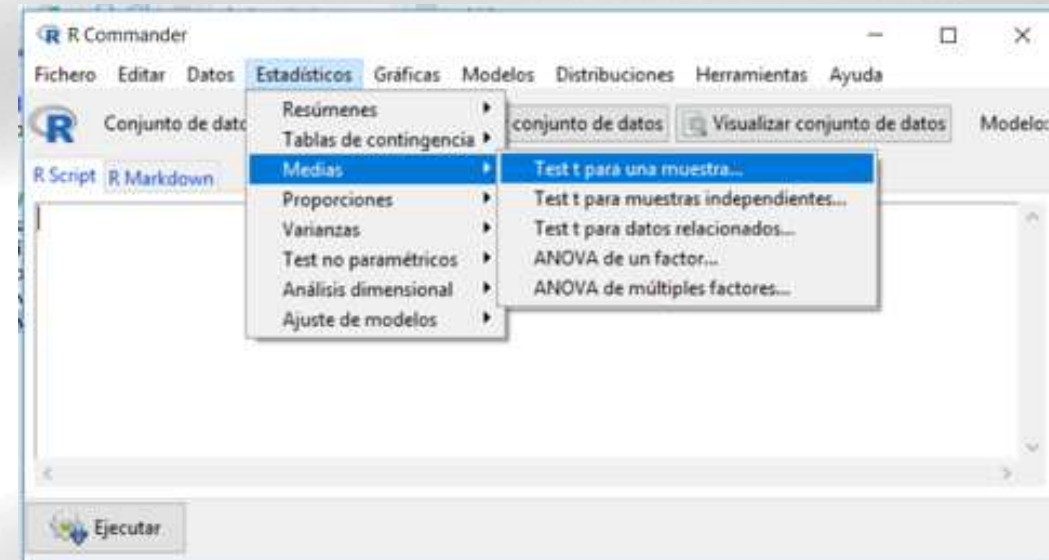
t.test(bua, mu=70)

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



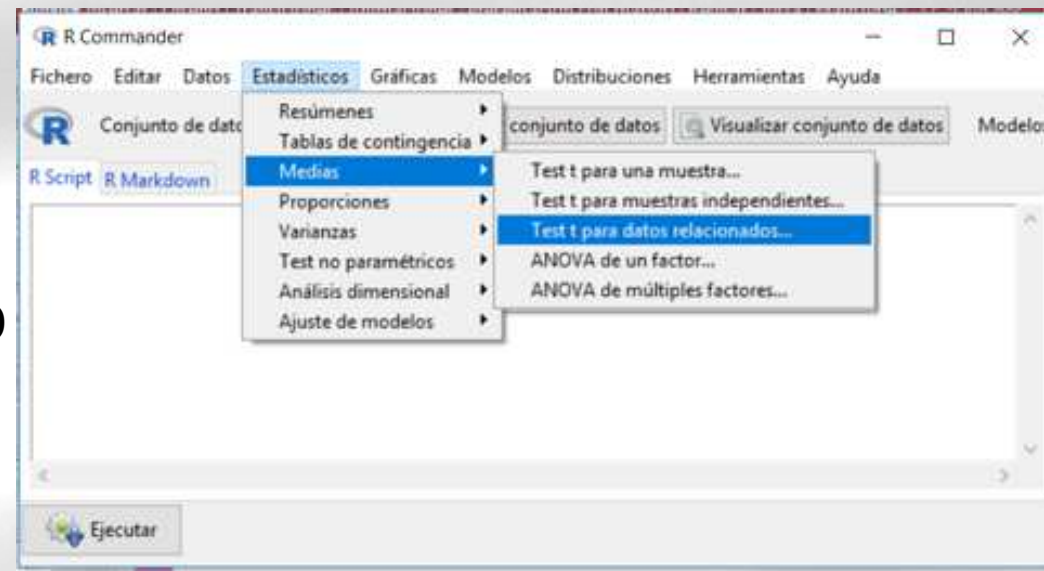
```
# osteoAll is the dataset obtained reading osteoporosis.csv
set.seed =123456 # change by your ID to obtain a distinct sample
index100<- sample( 1:nrow(osteoAll), 100)
osteo100 <- osteoAll[index100,]
with(osteo100, t.test (bua~menop, alternative="two.sided"))
```

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



Autoguardado EBB... Alexandre Sánchez Pla dia2... - Se h... Presentación con diaposi MathType ¿Qué des

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

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

Estadísticos Gráficas Modelos Distribuciones Herramientas Ayuda

Resúmenes

Tablas de contingencia

Medias

Proporciones

Varianzas

Test no paramétricos

Análisis dimensional

Ajuste de modelos

Conjunto de datos

Visualizar conjunto de datos

Test t para una muestra...

Test t para muestras independientes...

Test t para datos relacionados...

ANOVA de un factor...

ANOVA de múltiples factores...

Modelo:

R Commander

Fichero Editar Datos

Conjunto de datos

R Script R Markdown

Ejecutar

Master in Tr Errors and power (in hypothesis testing)

H_0
(innocent)
(not speculative)

H_1
(guilty)
(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

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

Diapositiva 27 de 75 Inglés (Reino Unido)

Diapositiva 28 de 38 Inglés (Reino Unido)

Diapositiva 27 de 75

Diapositiva 28 de 38

21:52 11/03/2018

ESP

Escribe aquí para buscar

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

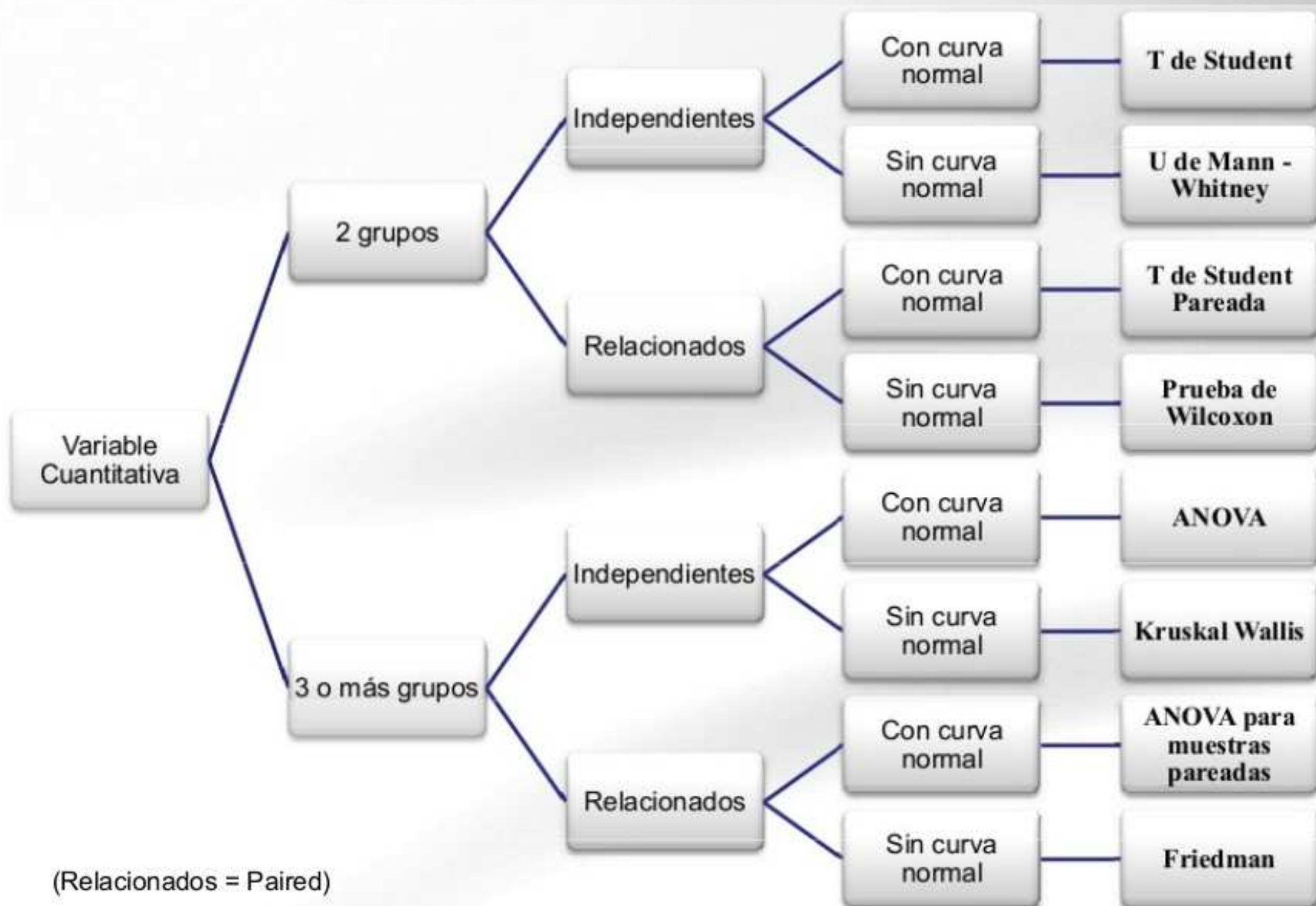
Types of error

	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)

Test selection diagrams: Qualitative vars



Test selection diagrams: Quantitative vars



Common misunderstandings about the p-value:

- The p-value is **not** the probability that the null hypothesis is true, nor is it 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 falsely 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.

To cross or not to cross?



Imagine you are an adventurer that has the option of to cross a bridge in order to escape from danger, find a treasure...

and that there is a post in front of the bridge stating:

"This bridge has broken only one out of 100 times"

To cross or not to cross?



Imagine you are an adventurer that has the option of to cross a bridge in order to escape from danger, find a treasure...

and that there is a post in front of the bridge stating:

"This bridge has broken only one out of 100 times"

So, the p-value of our metaphor is 0.01

You could accept that 1% is a risk small enough to pass the bridge and pursue your goal. OK.

To cross or not to cross?



Imagine you are an adventurer that has the option of to cross a bridge in order to escape from danger, find a treasure...

and that there is a post in front of the bridge stating:
"This bridge has broken only one out of 100 times"

So, the p-value of our metaphor is 0.01

You could accept that 1% is a risk small enough to pass the bridge and pursue your goal. OK.



But... what do you decide if, in order to reach your goal, you have to cross hundreds of bridges of that kind?

To cross or not to cross? :Multiple comparison



Imagine you are an adventurer that has the option of to cross a bridge in order to escape from danger, find a treasure...

and that there is a post in front of the bridge stating:
"This bridge has broken only one out of 100 times"

So, the p-value of our metaphor is 0.01

You could accept that 1% is a risk small enough to pass the bridge and pursue your goal. OK.



But... what do you decide if, in order to reach your goal, you have to cross hundreds of bridges of that kind?

In this case, the probability of falling while crossing one of the bridges is obviously too high ('cause we have just one life).

Therefore, in this case (multiple testing), the p-value by itself is not a good reference for accepting or not statistical significance. We must apply some type of correction to the p-values (allowing us to be safe in crossing all the bridges).