Bioestadística

Práctica de Ordenadores 8

José Aurelio Pina Romero
Ja.pina@ua.es
Bioestadística – Grado Enfermería
UA- Departamento de Enfermería

Universitat d'Alacant
Universidad de Alicante

¿Qué vamos hacer?

- Repaso clase teórica.
- Ejercicio 4.1 (apartado iii).
- Ejercicio 4.3 (apartado ii y iii).
- Contrastes de hipótesis con SPSS
- Chi-cuadrado (Repaso – Ej 1.21).
- Ejercicio 4.4 – 4.6 (finalizar la clase).

CONTRASTES DE HIPÓTESIS

- **Contraste de Hipótesis:** Procedimiento estadístico mediante el cual se investiga la aceptación o rechazo de una afirmación acerca de una o varias carácterísticas de una o más variables en una o más poblaciones.

A mayor discrepancia entre la hipótesis realizada y la información proporcionada por la muestra observada, mayor se la evidencia en contra de la hipótesis

Creo que el porcentaje de enfermos será el 5%

- **¿Qué es una hipótesis?**

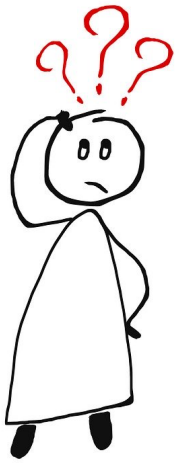
  Una creencia o afirmación sobre la población, principalmente sus parámetros:

  - Media
  - Varianza
  - Proporción/Tasa

Identificación de hipótesis

- Hipótesis nula H_o
  - La que contrastamos
  - Los datos pueden refutarla
  - Será mantenida a menos que los datos observados en la muestra indiquen una fuerte evidencia que no es cierta. – está es la razón de que, a pesar que el contraste conduzca a aceptarla NUNCA SE CONSIDERA PROBADA.

- Hip. Alternativa H_a
  - Niega a H_0
  - Los datos pueden mostrar evidencia a favor
  - No debería ser aceptada sin una gran evidencia a favor.



$$\begin{cases} H_0: & p = 50\% \\ H_a: & p \neq 50\% \end{cases}$$

$=, \leq, \geq$

$\neq, <, >$

Hipótesis

Hipótesis simples

Contrastes bilaterales

Se especifica un valor concreto para el parámetro o parámetros.

Ejemplo: Un investigador pretende, a partir de la información de 40 pacientes intervenidos quirúrgicamente de cierta afección, si el tiempo medio de supervivencia es exactamente de 15 años

Hipótesis compuestas

Contrastes unilaterales

Se especifican varios valores, habitualmente un intervalo.

Ejemplo: Un investigador pretende, a partir de la información de 40 pacientes intervenidos quirúrgicamente de cierta afección, si el tiempo medio de supervivencia sea superior a 15 años.

¿Quién es H_0?

- **Problema:** ¿En un estudio realizado para determinar el estado de salud de una comunidad se entrevistó a 82 personas, preguntándoles acerca de su actividad física habitual. De las 82 personas encuestadas, 36 de ellas declararon practicar algún deporte de forma regular ?

  Contraste la hipótesis de si la proporción poblacional de práctica de deporte de forma habitual <u>puede ser diferente de 0,50</u>

- Solución:
  - Traducir a lenguaje estadístico: $p = 50\%$
  - Establecer su opuesto: $p \neq 50\%$
  - Seleccionar la hipótesis nula

$$H_0 : p = 50\%$$

¿Quién es H_0?

- Problema: ¿El colesterol medio para la dieta mediterránea es 6 mmol/l?

- Solución:

  - Traducir a lenguaje estadístico: $\mu = 6$
  - Establecer su opuesto: $\mu \neq 6$
  - Seleccionar la hipótesis nula

$$H_0 : \mu = 6$$

Pasos

Problema: Un investigador pretende, a partir de la información de 40 pacientes intervenidos quirúrgicamente de cierta afección en los que a media de supervivencia es de 17 años con una desviación típica de 1,5 años, de si el tiempo medio de supervivencia es o no distinto de 15 años con un nivel de significación del 5%.

1) Definir las hipótesis: H_o - H_a

$$H_0 : \mu = 15$$
$$H_a : \mu \neq 15$$

2) Definir el estadístico de contraste (EC)

$$t = \frac{\overline{x} - \mu_0}{s / \sqrt{n}}$$

Si EC está cerca de 0 favorecerá H_0

EC: Medida de discrepancia entre lo que afirma Ho y la información que proporciona la muestra

3) Distribución de probabilidad asociada al EC t_{n-1}

4) Nivel de significación (probabilidad de rechazar H_0 cuando es cierta)

$$\alpha = 0{,}05$$

Pasos

5) Construcción de un regla de decisión

Región crítica
- Valores improbables
- Conocida antes de realizar el experimento
- Probabilidad rechazar Ho cuando es cierta.

$$t_{n-1} \rightarrow t_{39}$$

Si EC >2,021 o EC <-2,021 → Rechazamos H_0
Si -2,021<EC <2,021 → No rechazamos H_0

6) Cálculo del EC - Aplicación de la regla de la decisión - Conclusión

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

$$t = \frac{17 - 15}{1{,}5 / \sqrt{40}} = 8{,}43$$

Como t =8,43 > 2,021 → Rechazo H_0

Conclusión: Existe evidencia de que el tiempo medio de supervivencia es *significativamente* distinto de 15 años.

Pasos

Problema: Un investigador pretende, a partir de la información de 40 pacientes intervenidos quirúrgicamente de cierta afección en los que a media de supervivencia es de 17 años con una desviación típica de 1,5 años, de si el tiempo medio de supervivencia es superior de 15 años con un nivel de significación del 5%.

1) Definir las hipótesis: H_o - H_a

$$H_0 : \mu \leq 15$$
$$H_a : \mu > 15$$

Si EC está cerca de 0 favorecerá H_0

2) Definir el estadístico de contraste (EC)

$$t = \frac{\overline{x} - \mu_0}{s / \sqrt{n}}$$

3) Distribución de probabilidad asociada al EC t_{n-1}

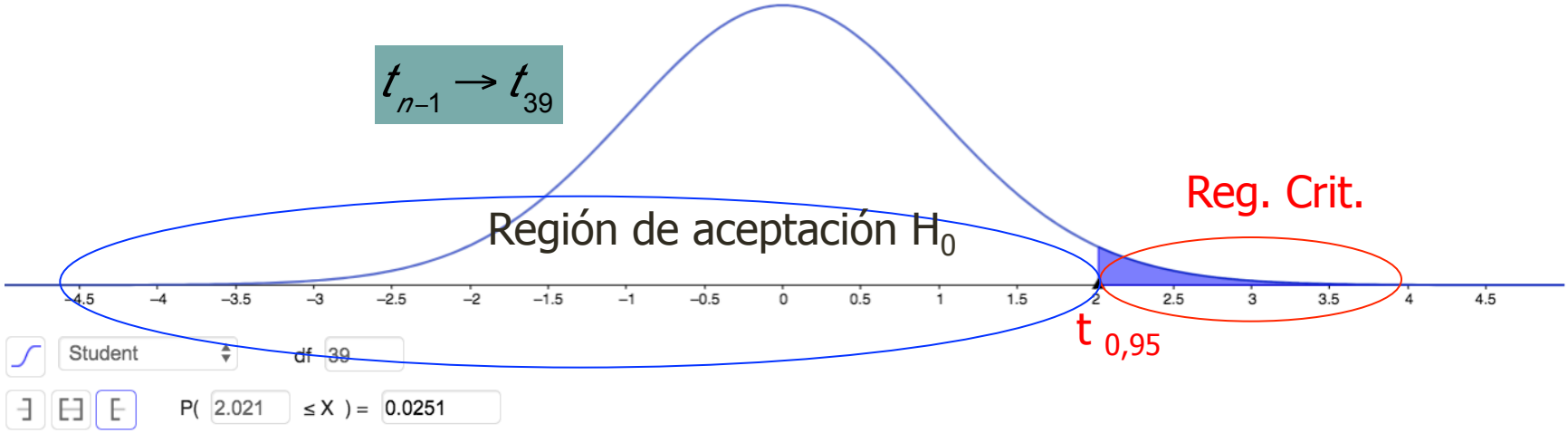
4) Nivel de significación (probabilidad de rechazar H_0 cuando es cierta)

$$\alpha = 0{,}05$$

5) Requerimientos n>=30

Pasos

6) Construcción de un regla de decisión



Si EC >1,684 → Rechazamos H_0
Si EC <1,684 → No rechazamos H_0

7) Cálculo del EC - Aplicación de la regla de la decisión - Conclusión

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

$$t = \frac{17 - 15}{1{,}5 / \sqrt{40}} = 8{,}43$$

Como t =8,43 > 2,021 → Rechazo H_0

Conclusión: Existe evidencia de que el tiempo medio de supervivencia es *significativamente* superior de 15 años.

Riesgos al tomar decisiones

Ejemplo 1: Se juzga a un individuo por la *presunta* comisión de un delito

- H_0: Hipótesis nula
  - Es inocente

Los datos pueden refutarla

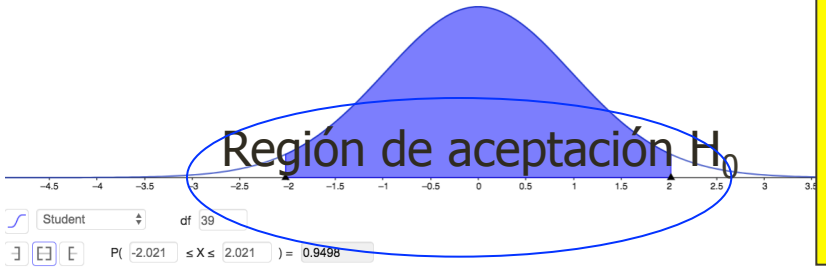
La que se acepta si las pruebas no indican lo contrario

Rechazarla por error tiene graves consecuencias

- H_1: Hipótesis alternativa
  - Es culpable

No debería ser aceptada sin una gran evidencia a favor.

Rechazarla por error tiene consecuencias consideradas menos graves que la anterior.



Riesgos al contrastar hipótesis

Ejemplo 2: Se cree que un nuevo tratamiento ofrece buenos resultados

Ejemplo 3: Parece que hay una incidencia de enfermedad más alta de lo normal

- H_0: Hipótesis nula
  - (Ej.1) Es inocente
  - (Ej.2) El nuevo tratamiento no tiene efecto
  - (Ej.3) No hay nada que destacar

No especulativa

- H_1: Hipótesis alternativa
  - (Ej.1) Es culpable
  - (Ej.2) El nuevo tratamiento es útil
  - (Ej. 3) Hay una situación anormal

Especulativa

Tipos de error al tomar una decisión

		Realidad	
		Inocente	Culpable
veredicto	Inocente	OK	Error tipo II Menos grave
	Culpable	Error tipo I Muy grave	OK

$$\alpha = P\left(\text{cometer error tipo I}\right) = P\left(\textit{culpable}/\text{inocente}\right)$$

$$\beta = P\left(\text{cometer error tipo II}\right) = P\left(\textit{inocente}/\text{culpable}\right)$$

$$\text{POTENCIA} = P\left(\textit{culpable}/\text{culpable}\right) = P(\textit{inocente}\,/\,\text{inocente})$$

$$1-\beta = 1-P\left(\textit{inocente}/\text{culpable}\right)$$

Tipos de error al tomar una decisión

	Realidad	
Decisión	H_0 cierta	H_0 Falsa
No Rechazo H_0	**Correcto** El tratamiento no tiene efecto y así se decide.	**Error de tipo II** El tratamiento si tiene efecto pero no lo percibimos. **Probabilidad β**
Rechazo H0 Acepto H_a	**Error de tipo I** El tratamiento no tiene efecto pero se decide que sí. **Probabilidad α**	**Correcto** El tratamiento tiene efecto y el experimento lo confirma.

$$\alpha = P\left(\text{cometer error tipo I}\right) = P\left(\text{Rechazar Ho/Ho cierta}\right)$$

$$\beta = P\left(\text{cometer error tipo II}\right) = P\left(\text{Aceptar Ho/Ho falsa}\right)$$

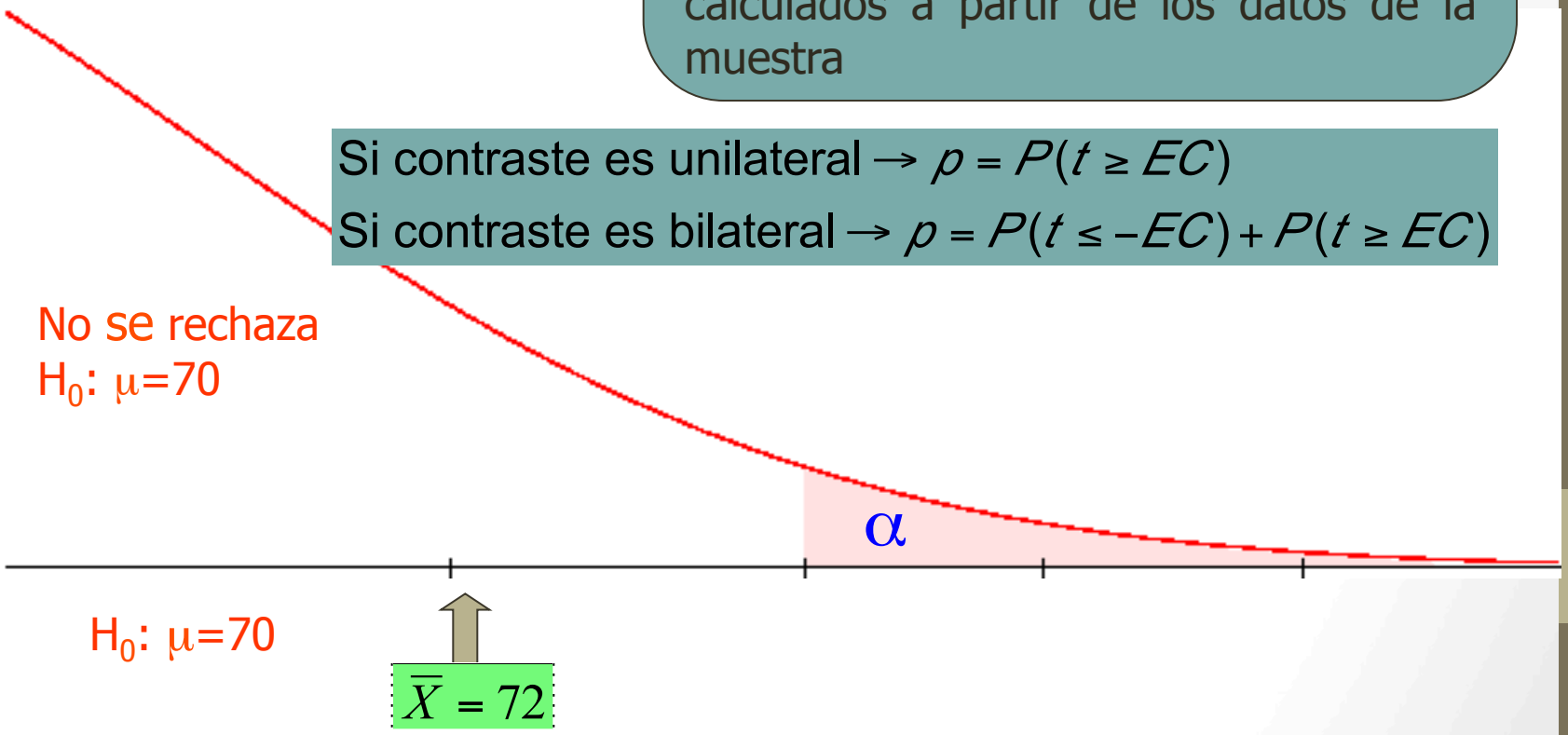
Significación: p (Nivel de significación a posteriori)

Se define como la **probabilidad de observar**, suponiendo que Ho es cierta, un valor del EC igual o mayor que el observado en la muestra.

P no se fija a priori, sino que es calculados a partir de los datos de la muestra

Si contraste es unilateral $\rightarrow p = P(t \geq EC)$

Si contraste es bilateral $\rightarrow p = P(t \leq -EC) + P(t \geq EC)$



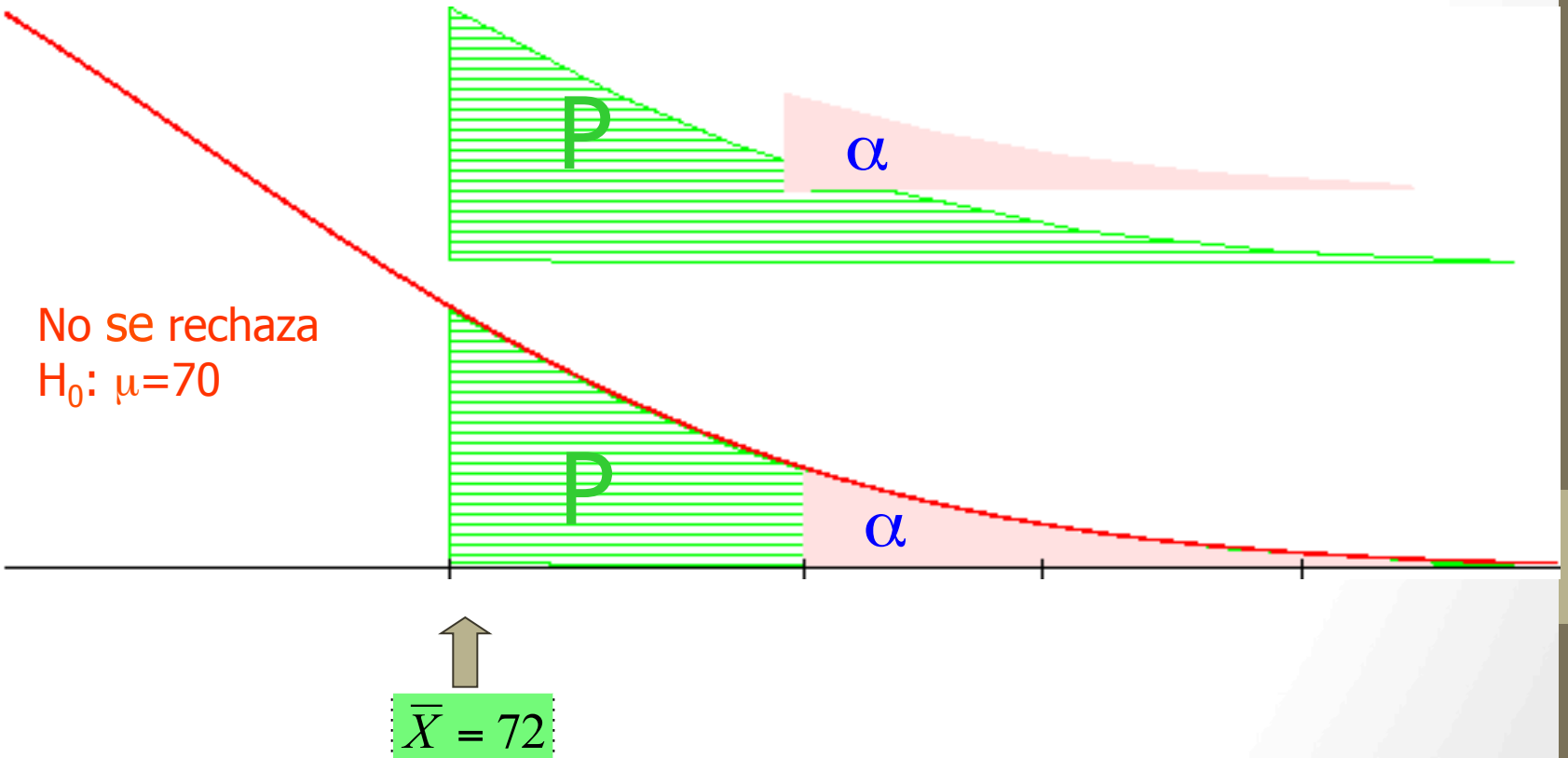
Significación: p (Nivel de significación a posteriori)

Es la probabilidad que tendría una región crítica que comenzase exactamente en el valor del estadístico obtenido de la muestra.

Es la probabilidad de tener una muestra que discrepe aún más que la nuestra de H_0.

p es conocido después de realizar el experimento aleatorio

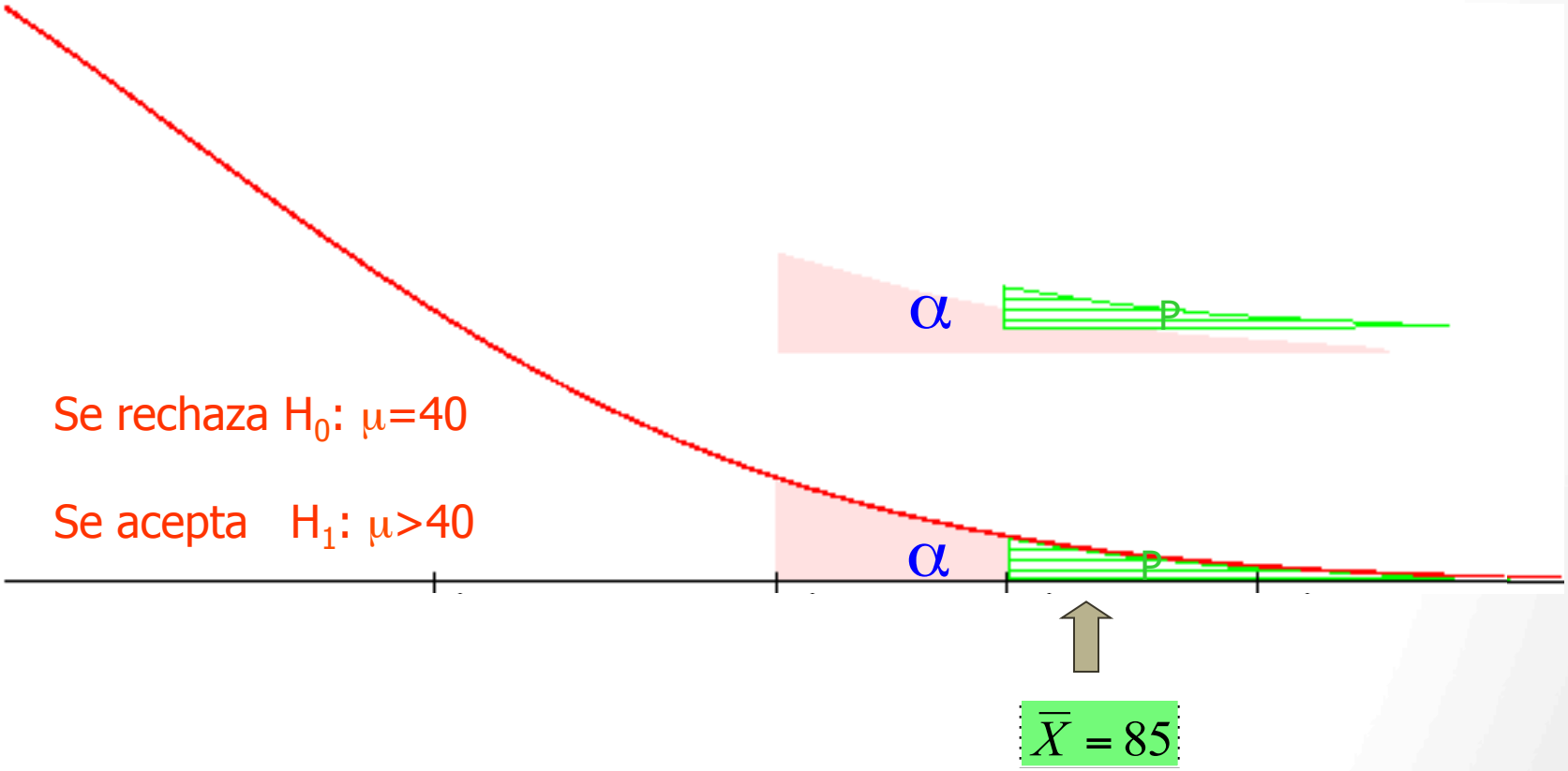
El contraste es no significativo cuando $p>\alpha$ → NO SE RECHAZA H_0



Significación : p

El contraste es estadísticamente significativo cuando $p<\alpha$

Es decir, si el resultado experimental discrepa más de "lo tolerado" *a priori*.



Resumen: α, p y criterio de rechazo

- Sobre α
  - Es número pequeño, preelegido al diseñar el experimento
  - Conocido α sabemos todo sobre la región crítica
- Sobre p
  - Es conocido tras realizar el experimento
  - Conocido p sabemos todo sobre el resultado del experimento

- Sobre el criterio de rechazo
  - Contraste significativo = p menor que α

Conclusiones

- Las hipótesis no se plantean después de observar los datos.
- En ciencia, las hipótesis nula y alternativa no tienen el mismo papel:
  - H_0 : Hipótesis científicamente más simple.
  - H_1 : El peso de la prueba recae en ella.
- α debe ser pequeño
- Rechazar una hipótesis no prueba que sea falsa. Podemos cometer error de tipo I
- No rechazar una hipótesis no prueba que sea cierta. Podemos cometer error de tipo II
- Si decidimos rechazar una hipótesis debemos mostrar la probabilidad de equivocarnos.

Ejercicio 4.1 -iii

En un estudio realizado para determinar el estado de salud de una comunidad se entrevistó a 82 personas preguntándoles acerca de su actividad física habitual. De las 82 personas encuestadas, 36 de ellas declararon practicar algún deporte de forma regular.

Contraste la hipótesis de si la proporción poblacional de práctica de deporte de forma habitual puede ser diferente de 0,50 con $\alpha = 0,05$

1) Definir las hipótesis: $H_o - H_a$
2) Definir el estadístico de contraste (EC)
3) Distribución de probabilidad asociada al EC
4) Nivel de significación (probabilidad de rechazar H_0 cuando es cierta)
5) Requerimientos ($np>=5$, $n(1-p)>=5$)
6) Cálculo del EC - Aplicación de la regla de la decisión – Conclusión

Ejercicio 4.1 -iii

Distribución | Estadísticas

Test Z de una proporción

Hipótesis nula p = 0.5

Hipótesis alternativa ○ < ○ > ● ≠

Muestra

Èxitos 36

N 82

Resultado

Test Z de una proporción

Successes	36
N	82
Z	-1.1043
P	0.2695

Z Estimada de una Proporción

Nivel de confianza 0.95

Muestra

Èxitos 36

N 82

Resultado

Z Estimada de una Proporción

Successes	36
N	82
SE	0.0548
Lower Limit	0.3316
Upper Limit	0.5464
Interval	0.439 ± 0.1074

Ejercicio 4.4

A partir de los datos obtenidos de una muestra de 120 alumnos de Enfermería se obtuvo que el tiempo medio que un alumno necesita para realizar un examen de la asignatura de Bioestadística es de 100 minutos con una desviación típica de 15.

Contraste la hipótesis de que el tiempo medio para la realización del examen pueda ser superior a 110 minutos. Utilice α =0.01.

Usar Geogebra

Ejercicio 4.4 (Con Geogebra)

Distribución | Estadísticas

Test T de una media

Hipótesis nula μ = 110

Hipótesis alternativa ○ < ● > ○ ≠

Muestra

Media 100

s 15

N 120

Resultado

Test T de una media

Mean	100
s	15
SE	1.3693
N	120
df	119
t	-7.303
P	1

Distribución | Estadísticas

Interavlo T de una media

Nivel de confianza 0.99

Muestra

Media 100

s 15

N 120

Resultado

Interavlo T de una media

Mean	100
s	15
SE	1.3693
N	120
df	119
Lower Limit	96.4155
Upper Limit	103.5845
Interval	100 ± 3.5845

Ejercicio 4.6

A partir de los datos del ejercicio 4.4, se sabe que 80 de los 120 alumnos eran mujeres y el resto hombres. El tiempo medio para la realización del examen en mujeres fue de 95 minutos con una desviación típica de 12, mientras que el tiempo medio en el caso de los hombres fue de 105 minutos con una desviación típica de 17.

Construya un intervalo de confianza al nivel 0.99 para cada uno de los tiempos medios de realización del examen en hombres y mujeres

A nivel de significación α=0.05, ¿proporcionan estos datos evidencia de que el tiempo medio para la realización del examen sea superior en hombres que en mujeres?

NO SE PUEDE HACER TODAVÍA

Ejercicio 4.3 (No se puede hacer todavía)

En un estudio realizado para determinar el estado de salud de una comunidad se por las horas de sueño de los encuestados. Los resultados expresados en media y desviación típica se recogen en la tabla adjunta y de forma separada para aquellos que declararon realizar ejercicio físico y para los que no:

	Realizan ejercicio	No realizan ejercicio
Nº individuos	36	46
Media horas de sueño	8.5 horas/día	7.2 horas/día
D. Típica horas de sueño	0.9 horas	0.8 horas

ii A nivel de significación α =0.05, ¿existen diferencias significativas en los tiempos medios de sueño entre los individuos que realizan ejercicio físico y los que no?

iii ¿Cuánto vale la diferencia de medias de horas de sueño entre los que hacen ejercicio físico y los que no? Utilice el nivel de significación 0.95.

BBDD LITIABIS

1) Averigua el nivel promedio y la desviación típica de COLESTEROL en hombres y mujeres.

2) Calcule el intervalo de confianza al 95% del nivel promedio de COLESTEROL en hombres. ¿Sería admisible, a la vista del resultado obtenido, un valor promedio de COLESTEROL superior a 235?

3) Calcule el intervalo de confianza al 95% del nivel promedio de COLESTEROL en mujeres. ¿Sería admisible, a la vista del resultado obtenido, un valor promedio de COLESTEROL superior a 235?

BBDD LITIABIS

1) Calcule el intervalo de confianza al 99% del nivel promedio de COLESTEROL en hombres y mujeres. ¿Qué ocurre con el tamaño del intervalo?

2) Contraste la hipótesis de que el nivel medio de COLESTEROL en hombres sea igual o diferente a 210. Utilice $\alpha=0.05$.

3) Contraste la hipótesis de que el nivel medio de COLESTEROL en hombres sea superior a 210. Utilice $\alpha=0.05$.

BBDD LITIABIS

1) Contraste la hipótesis de que el nivel medio de COLESTEROL en mujeres sea igual o diferente a 210. Utilice $\alpha=0.05$.

2) Contraste la hipótesis de que el nivel medio de COLESTEROL en mujeres sea igual o diferente a 210. Utilice $\alpha=0.05$.

3) Deduzca si existen, al nivel de significación $\alpha=0.05$, diferencias significativas entre los niveles promedio de COLESTEROL entre hombres y mujeres.

4) ¿Cuánto vale la diferencia de los niveles promedio de COLESTEROL entre hombres y mujeres? Utilice el nivel de confianza 0.95

BBDD LITIABIS

1) Averigua el nivel promedio y la desviación típica de TRIGLECERIDOS en hombres y mujeres.

2) Calcule el intervalo de confianza al 95% del nivel promedio de TRIGLECERIDOS en hombres. ¿Sería admisible, a la vista del resultado obtenido, un valor promedio de TRIGLECERIDOS superior a 135?

3) Calcule el intervalo de confianza al 95% del nivel promedio de TRIGLECERIDOS en mujeres. ¿Sería admisible, a la vista del resultado obtenido, un valor promedio de colesterol superior a 135?

BBDD LITIABIS

1) Calcule el intervalo de confianza al 99% del nivel promedio de TRIGLICERIDOS en hombres y mujeres. ¿Qué ocurre con el tamaño del intervalo?

2) Contraste la hipótesis de que el nivel medio de TRIGLICERIDOS en hombres sea igual o diferente a 135. Utilice $\alpha=0.05$.

3) Contraste la hipótesis de que el nivel medio de TRIGLICERIDOS en hombres sea superior a 135. Utilice $\alpha=0.05$.

BBDD LITIABIS

1) Calcule el intervalo de confianza al 99% del nivel promedio de TRIGLICERIDOS en hombres y mujeres. ¿Qué ocurre con el tamaño del intervalo?

2) Contraste la hipótesis de que el nivel medio de TRIGLICERIDOS en hombres sea igual o diferente a 210. Utilice $\alpha=0.05$.

3) Contraste la hipótesis de que el nivel medio de TRIGLICERIDOS en hombres sea superior a 210. Utilice $\alpha=0.05$.

BBDD LITIABIS

1) Averigua la proporción individuos con cálculos.

2) Si trabajamos al nivel de confianza 0.95, ¿cuál será la proporción de individuos con cálculos en la población de la que partió la muestra?

3) A nivel de significación $\alpha=0.01$ contraste la hipótesis de que la proporción de individuos en la población sea igual o diferente al 10%

4) A nivel de significación $\alpha=0.01$ contraste la hipótesis de que la proporción de individuos en la población con presencia de cálculos sea superior al 10%

BBDD LITIABIS

1) Averigua el valor promedio, y la desviación típica del índice de Quetelet en hombres y mujeres.

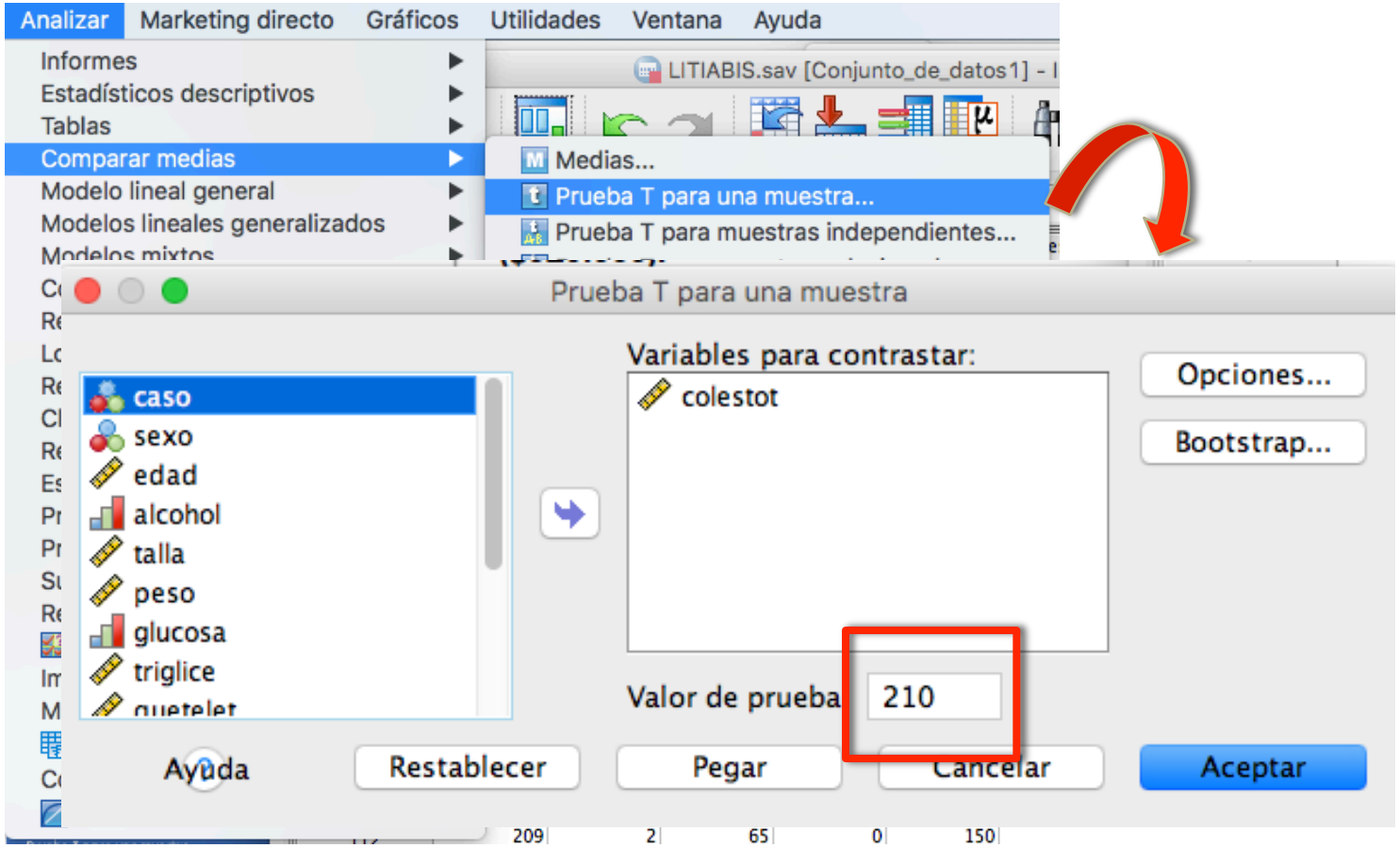
2) Al 95% de confianza ¿cuánto vale el valor medio poblacional del Indice de Quetelet en la población de mujeres de la que partió la muestra? ¿Sería admisible un valor promedio del Indice de Quetelet de 30?

3) Al 95% de confianza ¿cuánto vale el valor medio poblacional del Indice de Quetelet en la población de hombres de la que partió la muestra? ¿Sería admisible un valor promedio del Indice de Quetelet de 30?

4) Realize los apartados 3 y 4 con un nivel de confianza al 90%.

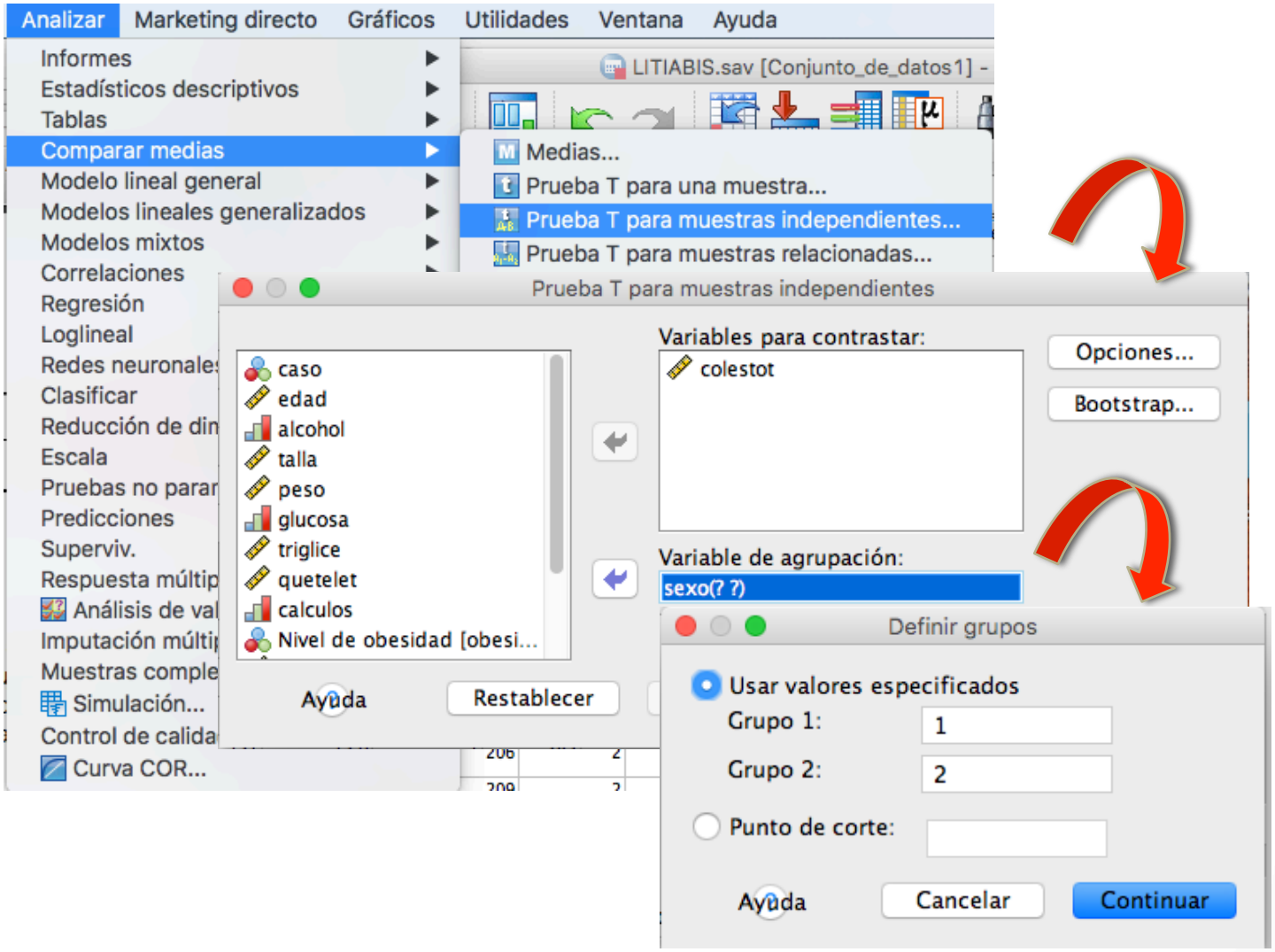
SPSS – Prueba de medias para una muestra

Analizar → Comparar medias → Prueba de medias para una muestra



SPSS – Prueba de medias para muestras independientes

Analizar → Comparar medias → Prueba T para muestras independientes



SPSS – Prueba de medias para muestras independientes

Group Statistics

	sexo	N	Mean	Std. Deviation	Std. Error Mean
colestot	Hombre	92	226,58	49,395	5,150
	Mujer	108	228,07	51,866	4,991

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
colestot	Equal variances assumed	,299	,585	-,208	198	,835	-1,498	7,200	-15,696	12,700
	Equal variances not assumed			[illegible]209	195,525	,835	-1,498	7,171	-15,641	12,645

H0: varianzas iguales
Ha: variables diferentes

p-valor >0,05 no rechazamos H0
(asumimos varianzas iguales)

Utilizando el criterio del valor p, se puede afirmar que, estadísticamente, no hay diferencias entre el nivel de colesterol de las mujeres en relación al de los hombres (valor p = 0,835), por lo que esa acusación resulta falsa (Se debe aceptar H_0)

Ejercicio 1.21 – Chi - Cuadrado

Construya las tablas de distribución de frecuencias conjuntas para las parejas de variables SEXO y CALCULOS, ALCOHOL y CALCULOS, y SEXO y GLUCOSA. Calcule e interprete las medidas de asociación entre estas variables.

Ho: No existe asociación

Ha: Existe asociación