

## GUIÓN PARA EL EXAMEN

IMPORTANTE - Cargar paquetes: epiDisplay, car y e1071

Justificar la elección del test y establecer hipótesis

Código de colores: fórmula, datos a introducir, ejemplo y explicación de la introducción de los comandos, explicación, atención

### Transformar variables

-Para utilizar categorías (ej hombre/mujer) en vez de los 0 y 1

`datos$variable=ifelse(datos$variable==0, "categoria0", "categoria1")`

Crear una nueva variable

`datos$nombre de la variable a crear=datos$combinación algebraica para la creación de la variable`

`datos$IMC = datos$peso/(datos$altura/100)^2`

### ESTADÍSTICOS DESCRIPTIVOS

Para trabajar más cómodamente cambiamos el nombre de la variable:

`x=datos$nombre de la variable`

Media: `mean(x)`

Mediana: `median(x)`

Desviación típica: `sd(x)`

Cuartiles: `quantile(x,porcentaje que queremos dejar por debajo)`

Rango intercuartílico: `IQR(x)`

Histograma: `hist(x)`

La variable media muestral es dato obtenido (media)

La mitad de observaciones de mi muestra tienen menos de dato obtenido (mediana)

El 25% de los individuos de la muestra tienen dato obtenido o menos (Q25)

El 75% de los individuos de la muestra tienen dato obtenido o menos (Q75)

El 50% central de la edad de los individuos de mi muestra, ofrecen una variación de variable de datos obtenido (IQR)

Obtenemos los estadísticos descriptivos de una variable con respecto a una variable con distintas categorías.

Para trabajar más cómodamente cambiamos el nombre de las dos variables que vamos a utilizar.

`x=datos$nombre de la variable 1`

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

Cartagena99

`supply(x,y,función)`  
`hist(x[y=="Categoría"])`

\*En categoría si hemos utilizado el ifelse introducimos hombre, mujer, pero si no hemos utilizado ifelse tendremos que introducir 0 y 1.

La variable media de los categoría1 de mi muestra es dato obtenido y de las categoría2 es dato obtenido

La mitad de categoría1 de mi muestra tienen menos de dato obtenido; y la mitad de categoría2 de mi muestra tienen menos de dato obtenido.

El 25% de categoría1 de mi muestra tienen menos de dato obtenido; y el 25% de categoría2 de mi muestra tienen menos de dato obtenido

El 50% central de la variable de categoría1 de mi muestra, ofrecen una variación de dato obtenido, mientras que el 50% central de la variable en categoría2 de mi muestra poseen una variación de dato obtenido

Lectura del histograma (ejemplo hombres o mujeres):

- La distribución parece simétrica o asimétrica.
- La media, mediana y moda coinciden o no coinciden.
- Posee una kurtosis normal o no.

## FRECUENCIAS DE UNA VARIABLE CUALITATIVA

Calcular frecuencias absolutas y relativas de una variable

Para esto debemos :

`a=table(datos$nombre de la variable) #frecuenciaabsoluta`

`prop.table(a) #frecuenciarelativa`

1° Ejecutar ambos #frecuencias relativas.

2° En la consola escribimos a y damos a intro. #frecuencias absolutas.

Comentario ejemplo para los grados

Con los resultados obtenidos podemos decir que el  $\frac{\%}{100}$  de mi muestra, es decir,  $\frac{\%}{100}$  individuos, pertenecen al categoría1; el  $\frac{\%}{100}$ ,  $\frac{\%}{100}$  individuos; pertenecen al categoría2; y el  $\frac{\%}{100}$  de mi muestra, es decir,  $\frac{\%}{100}$  individuos pertenecen al categoría3.

Calcular frecuencias absolutas y relativas de una variable con respecto a la categoría de otra

Adaptamos lo anterior: `a=table(datos$variable1,datos$variable2)`

Porcentaje por fila: `prop.table(a,1)`

Porcentaje por columna: `prop.table(a,2)`

Podemos decir que el  $\frac{\%}{100}$  de los individuos pertenecientes al categoría de la variable1 son categoría de la variable2; lo que explicado en función de la otra variable, significa que el  $\frac{\%}{100}$  de los categoría de la variable2 pertenecen al categoría de la variable1. Y así sucesivamente.

**DISTRIBUCIÓN NORMAL** R solo nos da  $P(x < n^\circ \text{ dado})$

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, blue, serif font. The '99' is significantly larger and more prominent than the 'Cartagena' part. The text is set against a light blue background with a subtle gradient and a soft shadow effect.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

pero debemos calcularlas. Para calcular  $P(x > 78)$ :

`x=datos$peso`

`media=mean(x)`

`des=sd(x)` `1-pnorm(78,media,des)`

La probabilidad de encontrar individuos de mi muestra con un peso mayor que el n° dado es del \_%.

### INTERVALO DE CONFIANZA PARA LA MEDIA - para la POBLACIÓN

Usando la variable

`ci.numeric(datos$nombre de la variable)`

Calcular un intervalo de confianza al 95% para la variable peso

`ci.numeric(datos$peso)`

Tengo la confianza del 95% de que la media de variable poblacional esté entre 📊 ➡

Usando valores

`media=mean(datos$ nombre de la variable)`

`des=sd(datos$nombre de la variable)`

`n=1500`

`ci.numeric(media,n,des)`

Tengo la confianza del 95% de que la media de variable poblacional esté entre 📊 ➡

Intervalo de confianza para una variable en función de otra

Ej.: calcular el intervalo de confianza para el peso en hombre y mujer

`x=datos$variable 1`

`y=datos$variable 2`

`tapply(x,y,ci.numeric)`

Tengo una seguridad del 95% de que la media poblacional del variable de categoría1 está entre 📊 ➡

Tengo una seguridad del 95% de que la media poblacional del variable de categoría2 está entre 📊 ➡

### INTERVALO DE CONFIANZA PARA LA PROPORCIÓN

`ci.binomial(datos$nombre de la variable)`

Utilizar los 0 y 1 de la variable (ej: 0=mujer y 1=hombre); por este motivo si al principio hemos utilizado el comando `ifelse`, debemos anularlo. Esto se puede hacer situando una # delante del comando y ejecutándolo, de este modo es como si no hubiésemos puesto el comando.

`#datos$Sex = ifelse(datos$Sex==0, "Mujer","Hombre")`

Con esto se calcula el 1 (en este caso hombre).

Por esto, para calcular el intervalo de confianza y la proporción de 0 (en este caso mujer) hay varias opciones:

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, outlined font. The '99' is significantly larger and more prominent than the rest of the text. The logo is set against a background of a blue and orange gradient with a subtle geometric pattern.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

`ci.binomial(1-datos$nombre de la variable)`

Hombres=1, por lo que es el evento  
ci.binomial(datos\$Sex)

Mi muestra está compuesta por  $n$  individuos, de los cuales  $n_1$  son categoría. La proporción de categoría es del  $\frac{n_1}{n}$ %, el ES (error estándar) es  $\sqrt{\frac{p(1-p)}{n}}$  y con una confianza del  $1-\alpha$ % puedo decir que la proporción poblacional de categoría está entre  $\frac{n_1}{n} \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$ .

Mujeres=0, por lo que es el complementario.

1. Los datos de las mujeres son complementarios a los de los hombres
2. ci.binomial(382,1500)
3. ci.binomial(1-datos\$Sex)

Por tanto podemos decir que de los  $n$  individuos de mi muestra  $n_2$  son categoría2. La proporción de ellas es de un  $\frac{n_2}{n}$ %, el ES es  $\sqrt{\frac{p(1-p)}{n}}$  y con una confianza del 95% puedo decir que la proporción poblacional de categoría2 estará entre  $\frac{n_2}{n} \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$ .

## ESTUDIO DE LA NORMALIDAD PARA UNA SOLA VARIABLE

$x = \text{datos}\$ \text{nombre de la variable}$

Estadísticos descriptivos (media, mediana, asimetría, kurtosis e histograma)

Media:	<code>mean(x)</code>
Mediana:	<code>median(x)</code>
Asimetría:	<code>skewness(x)</code>
Kurtosis	<code>kurtosis(x)</code>
Histograma:	<code>hist(x)</code>

Gráfico q-q

Dibuja los cuantiles:	<code>qqnorm(x)</code>
Dibuja la línea teórica ideal:	<code>qqline(x)</code>

Confirmación con contrastes de hipótesis: Kolmogorof Smirnov y Shapiro-Wilk

Los contrastes de hipótesis establecen como hipótesis nula que hay normalidad y como hipótesis alternativa que no la hay. Hay dos test que realizan este contraste de hipótesis:

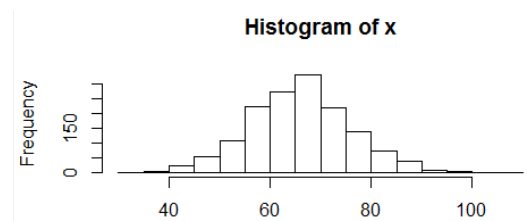
- Shapiro-Wilk (muestras pequeñas) `shapiro.test(x)`
- Kolmogorof Smirnov `b=rnorm(n,mean(x),sd(x))` `ks.test(x,b)`

Paso 1º evaluar la normalidad de la variable

$x = \text{datos}\$ \text{Age}$

→ Estadísticos descriptivos

`mean(x)`  
`median(x)`  
`skewness(x)`  
`kurtosis(x)`  
.....



CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

`qqline(x)`

Hay valores extremos

Viéndolo todo en conjunto puedo decir que hay valores extremos pero en su conjunto se observa normalidad, pese a la ligera asimetría y el ligero apuntamiento  
 Justificamos la decisión tomada con el contraste de normalidad

→ Paso 3º Contrastes de hipótesis : elegir test y aplicarlo

$H_0$ : hay normalidad (p.valor>0.05)

$H_1$ : no hay normalidad (p.valor<0.05)

Kolmogorof Smirnov (utilizamos este test porque n=1500)

b=rnorm (1500,mean(x),sd(x))

ks.test(x,b)

Como mi p-valor es 0.26, es decir, mayor que 0.05, podemos decir con una seguridad del 95% que no tenemos motivos para rechazar la hipótesis nula, por lo que aceptamos normalidad (ACEPTAMOS  $H_0$ )

**EVALUAR LA NORMALIDAD DE UNA VARIABLE CON RESPECTO A OTRA (DIFERENCIA UTILIZAR TAPPLY)**

**Estadísticos descriptivos (media, mediana, asimetría, kurtosis e histograma)**

- Media: `tapply(x,y,mean)`
- Mediana: `tapply(x,y,median)`
- Asimetría: `tapply(x,y,skewness)`
- Kurtosis: `tapply(x,y,kurtosis)`
- Histograma: `hist(x[y==categoría0])`  
`hist(x[y==categoría1])`

**Gráfico q-q**

- `qqnorm(x[y==categoría0])` `qqline (x[y==categoría0])`
- `qqnorm(x[y==categoría1])` `qqline (x[y==categoría1])`

**Confirmación con contrastes de hipótesis: Kolmogorof Smirnov y Shapiro-Wilk**

- Shapiro-Wilk `tapply(x,y,shapiro.test)`
- Kolmogorof Smirnov `b=rnorm(n,mean(x),sd(x))` `tapply(x,y,ks.test,b)`

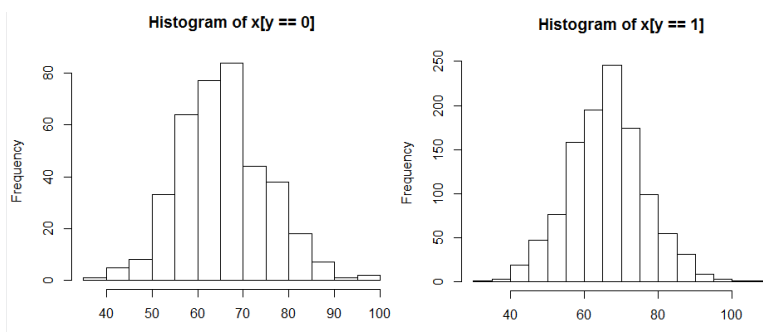
**Paso 1º evaluar la normalidad de la variable edad en hombres y mujeres**

x=datos\$Age  
 y=datos\$Sex

→ Estadísticos descriptivos

- `tapply (x,y,mean)`
- `tapply(x,y,median)`
- `tapply(x,y,skewness)`
- `tapply(x,y,kurtosis)`

`hist(x[y==0])` `#mujer`



**CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
 LLAMA O ENVÍA WHATSAPP: 689 45 44 70**

---

**ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
 CALL OR WHATSAPP:689 45 44 70**



x=datos\$variable1

y=datos\$variable2

a=table(x,y)

b=chisq.test(a) #hago el test chi cuadrado

b\$expected

(Presionas “b” en la pantalla inferior izquierda y presionamos ENTER → Nos dará el valor de “X-squared” > 3,84 entonces p.valor < 0,05 y se rechaza la Ho, sino (p.valor > 0,05) se acepta la H<sub>1</sub>)

Procedemos a realizar la  $\chi^2$  cuadrado de Pearson cuando trabajamos con variables nominales independientes entre si y queremos determinar si existen diferencias entre las proporciones de los valores observadas en la población en comparación a las esperadas.

(p.valor > 0,05) → Aceptamos la H<sub>0</sub>: Con una seguridad del 95% la proporción de variable en la categoría1 en la población es igual que la proporción de variable en categoría2 ya que el p-valor ha resultado ser  $\geq$  y este es mayor a 0,05.

(p.valor < 0,05) → Rechazamos H<sub>0</sub>, es decir, aceptamos H<sub>1</sub>: Con una seguridad del 95%, encontramos motivos para rechazar la hipótesis nula, es decir, la proporción de categoría1 en la población que variable es distinta a la proporción de categoría2 que variable; debido a que el p-valor hallado es  $\leq$ , es decir, menor que el 0.05.

-Test de Fisher = N < 5 Alternativa no paramétrica - No tiene condición, se puede aplicar siempre.

x=datos\$variable1

y=datos\$variable2

a=table(x,y)

fisher.test(a)

(Presionas “a” en la pantalla inferior izquierda y presionamos ENTER → Nos dará los valores observados. Si todos son mayores a 5 (n < 5) entonces procedemos a hacer Test de Fisher - entonces p.valor < 0,05 y se rechaza la Ho, sino (p.valor > 0,05) se acepta la H<sub>1</sub>)

Nuevamente la misma justificación que con el  $\chi^2$  cuadrado de Pearson

### COMPARACIÓN DE PROPORCIONES EN MUESTRAS DEPENDIENTES

- $\chi^2$  cuadrado de McNemar: Se suele utilizar para casos de variables cualitativas nominales con casos dicotómicos.

Ejemplo: Si dos fármacos causa efecto o no en una misma población / Tratamientos antes y después / Comparar dos colirios, uno distinto en cada ojo / Comparar mortalidad por cáncer entre los expuestos a radiaciones y los no expuestos.

Planteamos nuestra H<sub>0</sub> = proporción efectos adversos farmaco A = proporción efectos adversos B. H<sub>1</sub> = Las proporciones entre los farmacos es distinta, y no se debe al azar.

x=datos\$adverso1 #cambia variable1 por la que quieras

y=datos\$adverso2 #cambia variable2 por la que quieras

a=table(x,y) #me construyo la tabla

The logo for Cartagena99 features the word 'Cartagena' in a stylized, outlined font, followed by '99' in a similar style. The text is set against a background of a blue and orange gradient with a subtle pattern.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP: 689 45 44 70

## T-student para muestras independientes

Dos condiciones de aplicación: **normalidad** (Kolgomorov o Shapiro) y **homogeneidad** (test de Levene).

Utilizamos el **test de Levene**

$H_0$ = varianzas iguales

$H_a$ = varianzas distintas

`leveneTest(x~factor(y))`

p-valor mayor de 0.05, no hay evidencia suficiente para decir que las varianzas sea distintas, varianzas homogéneas

Por esto se puede utilizar el test de la t-student para muestras independientes

El comando para este test es:

`t.test(x~factor(y),paried=FALSE,var.equal=TRUE/FALSE)`

p-valor mayor del 0.05 no tenemos evidencias para rechazar  $H_0$ , con una confianza del 95% la diferencia media del tamaño del tumor entre los que viven y los que mueren está entre -0.03 y 0.06

Si no hay homogeneidad de las muestras en la fórmula del test ponemos `var.equal=FALSE`; sin embargo, si no hay normalidad, debemos utilizar el test de Wilconson o la U de Mann Whitney.

Tanto el test de la T-Student como el de Wilconson y la U de Mann Whitney, establecen como  $H_0$  que las medias son iguales y como  $H_a$  que son distintas.

## Alternativa no paramétrica: test de wiloxon para muestras independientes

`wilcox.test(x~factor(y),paried=FALSE)`

## COMPARACIÓN DE DOS MEDIAS DEPENDIENTES

Su única condición es que ambas sigan una normal.

El comando de este test es:

`t.test(x,y,paried=TRUE)`

Si falla la condición de la normal utilizamos el test de Wilconson:

`wilcox.test(x,y,paried=TRUE)`

\*\* En este caso utilizar la diferencia de ambas variables para estudiar su normalidad:

`x=datos$variable1`

`y=datos$variable2`

`dif=x-y`

`mean(dif)`

`median(dif)`

`skewness(dif)`

`kurtosis(dif)`

`hist(dif)`

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

- - -

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70

Establece como  $H_0$  que la media de los distintos grupos es la misma, y como  $H_a$  establece que alguna de ellas es distinta.

`a=aov(x-factor(y))`

`summary(a)`

si  $p\text{-valor} < 0.05 \rightarrow$  rechazo  $H_0 \rightarrow$  test de Bonferroni (estudia las medias dos a dos para averiguar cuáles son las distintas)

`pairwise.t.test(x,y,p.adj="bonf")`

si Levene  $\rightarrow p\text{-valor} < 0.05 \rightarrow$  no ANOVA  $\rightarrow$  alternativa no paramétrica  $\rightarrow$  test de Kruskal

`kruskal.test(x-factor(y))`

si  $p\text{-valor de Kruskal} < 0.05 \rightarrow$  diferencias de medias  $\rightarrow$  alternativa no paramétrica a Bonferroni

`pairwise.wilcox.test(x,y,p.adj="bonf")`

The logo for Cartagena99 features the text 'Cartagena99' in a stylized, dark blue font with a light blue shadow. The text is positioned above a blue and orange gradient arrow pointing to the right.

CLASES PARTICULARES, TUTORÍAS TÉCNICAS ONLINE  
LLAMA O ENVÍA WHATSAPP: 689 45 44 70

---

ONLINE PRIVATE LESSONS FOR SCIENCE STUDENTS  
CALL OR WHATSAPP:689 45 44 70