

# Práctica 11. Contrastes chi cuadrado. Enunciados.

Estadística (650008). Grado en biología sanitaria. UAH.

Actualizado: 2023-12-22

## Ejercicio 1

En el trabajo de G. Mendel con guisantes, hay un vector de frecuencias observadas (semilla lisa, semilla rugosa):

```
observados = c(5474, 1850)
```

y unas proporciones (distribución) teóricas

```
probEsperados = c(3/4, 1/4)
```

1. ¿Estaba Mendel en lo cierto?
2. Calcula las frecuencias absolutas esperadas

## Ejercicio 2

A partir de la siguiente tabla de contingencia, determina si hay asociación entre las variables “dar positivo en una prueba diagnóstico” y “estar enfermo”

	Enfermos	Sanos
Positivo	192	158
Negativo	10	9646

Calcula las frecuencias absolutas esperadas

## Ejercicio 3

En [este fichero](#) encontrarás un extracto de los datos de la [encuesta nacional de salud de 2012](#). Aquí tienes las primeras líneas de la tabla

```
df = read.table(file = "Practica11_ENSE_chi2.csv", sep = ";", header = TRUE)
head(df)
```

```
  Aperitivos Refrescos ActividadFisica Fumar
1           1         3                1     1
2           4         4                1     4
3           1         4                1     1
4           3         5                1     4
5           4         4                1     1
6           2         3                4     4
```

- Considera las variables las variables `Fumar` y `ActividadFisica`, que toman los valores
  1. Sí fuma, diariamente
  2. Sí fuma, pero no diariamente

3. No fuma actualmente, pero ha fumado antes
4. No fuma ni ha fumado nunca de manera habitual

y

1. No hago ejercicio. El tiempo libre lo ocupo de forma casi completamente sedentaria (leer, ver la televisión, ir al cine, etc.)
2. Hago alguna actividad física o deportiva ocasional
3. Hago actividad física varias veces al mes
4. Hace entrenamiento deportivo o físico varias veces a la semana

respectivamente. Analiza si hay relación entre las dos variables o, por contra, son independientes.

Ten en cuenta que los valores 8 y 9 en la tabla se refieren, respectivamente, a “no sabe” y “no contesta”, por lo que hay que eliminarlos.

#### Ejercicio 4

Vamos a seguir parte del tutorial 12 de [postdata](#). Este [fichero](#) contiene los datos del Lunar *Orbiter Laser Altimeter instrument* (LOLA) para determinar si los cráteres de la Luna están distribuidos de forma uniforme entre el hemisferio norte y el sur o si, por contra, hay más en uno de los hemisferios.

Las tres variables que aparecen en ese fichero:

```
crateres = read.table(file = "Cap09-LolaLargeLunarCraterCatalog.csv", sep = ",", header = TRUE)
colnames(crateres)
```

```
## [1] "Lon"      "Lat"      "Diam_km"
```

que se refieren a la latitud, longitud (ambas en grados) y diámetro (en km) de los cráteres lunares y son todas ellas cuantitativas continuas.

La pregunta a responder se puede formular como sigue: ¿hay diferencia entre los diámetros de los cráteres entre ambos hemisferios de la Luna?

- Usa la función `cut` para categorizar la variable `Lat` en “Norte” y “Sur”.
- Analiza si los diámetros de los cráteres contienen valores atípicos. Si los hay, eliminalos
- Agrupa los diámetros de los cráteres n(no atípicos) en clases, por ejemplo, usando los cortes: 20 - 40 - 60 - 80, mayor que 80
- Haz el correspondiente contraste Chi cuadrado.

#### Ejercicio 5

Sospechas que cierto caracter se hereda de forma independiente con una probabilidad de 0.2. Para comprobarlo, haces 100 experimentos, cada uno de ellos consiste en obtener 5 descendientes y contar el n'umero de ellos que presenta dicho caracter.

El resultado es

```
table(descendientes)
```

```
descendientes
 0  1  2  3  4
26 46 23  3  2
```

¿Se cumple tu hipótesis?

#### Ejercicio 6

**PARA HACER EN CASA:** En este [fichero de datos](#) está el genoma del bacteriófago  $\Phi$ X174, primer genoma basado en ADN secuenciado (1977).

Puedes leer el contenido del fichero con

```
phiX174 = read.table(file = "phix174.txt")
```

y convertir en tipo caracter (ahora es una palabra muy larga)

```
(phiX174 = as.character(phiX174$V1))
```

```
## [1] "GAGTTTTATCGCTTCCATGACGCAGAAGTTAACTTTTCGGATATTTCTGATGAGTCGAAAAATTATCTTGATAAAGCAGGAATTACTACTGCTT"
```

y separar en letras individuales

```
# Separamos la secuencia en caracteres (nucleótidos)
genoma = unlist(strsplit(phiX174, split = ""))
head(genoma)
```

```
[1] "G" "A" "G" "T" "T" "T"
```

¿Dirías que los cuatro nucleótidos están igualmente representados?

**Nota:** Si quieres hacer todo el proceso de adquisición de datos necesitarás instalar BioConductor. Hay que descargar varios paquetes, a continuación está el código necesario para ello: instalarás las herramientas necesarias y, con la función `getSEQ()` descargaras la secuencia de ADN directamente desde el repositorio oficial. Luego, Sólo necesitas su número de acceso en la base de datos GenBank del NCBI. Cambiando el número de acceso, podrás descargar otros genomas allí almacenados.

```
# Instalamos el software de BioConductor
# SOLO LA PRIMERA VEZ
# if (!requireNamespace("BiocManager", quietly = TRUE))
#   install.packages("BiocManager")
# BiocManager::install(version = "3.12")
#
# BiocManager::install("annotate")
if(!require(annotate))biocLite("annotate")

# Descargar el genoma
phiX174 = getSEQ("NC_001422.1")
```

## Ejercicio 7

**PARA HACER EN CASA:** En [este fichero](#) encontrarás un extracto de los datos de la [encuesta nacional de salud de 2012](#). Aquí tienes las primeras líneas de la tabla

```
df = read.table(file = "Practica11_ENSE_chi2.csv", sep = ";", header = TRUE)
head(df)
```

	Aperitivos	Refrescos	ActividadFisica	Fumar
1	1	3	1	1
2	4	4	1	4
3	1	4	1	1
4	3	5	1	4
5	4	4	1	1
6	2	3	4	4

- Considera las variables Consumo de aperitivos o comidas saladas de picar (patatas fritas, galletitas saladas) y consumo de refrescos azucarados, cada una de las cuales puede tomar valores del 1 al 5 de cuerdo con

1. A diario
2. Tres o más veces a la semana, pero no a diario

3. Una o dos veces a la semana
4. Menos de una vez a la semana
5. Nunca o casi nunca

8 y 9 corresponde con no sabe/no contesta

Analiza si hay relación entre las frecuencias con que se toman aperitivos y refrescos o, por contra, son independientes.