

# Test estadísticos para variables cualitativas: Test binomial exacto, test multinomial y test chi-cuadrado goodnes of fit

Joaquín Amat Rodrigo [j.amatrodrigo@gmail.com](mailto:j.amatrodrigo@gmail.com)

Enero 2016

## Índice

Introducción.....	2
Datos independientes.....	2
Datos pareados.....	2
Test exactos y test aproximados.....	3
Test binomial exacto.....	4
Condiciones.....	4
Ejemplo.....	4
Test multinomial exacto.....	6
Ejemplo.....	6
$\chi^2$ Goodness of fit.....	9
Hipótesis.....	9
Condiciones del test.....	9
Estatístico $\chi^2$ .....	10
Grados de libertad.....	10
Cálculo de $p$ -value.....	10
Ejemplo 1.....	11
Ejemplo 2.....	12
Solución manual.....	13
Solución con R.....	14
Ejemplo 3.....	15

## Introducción

Los contrastes de hipótesis para variables cualitativas se realizan mediante test de frecuencia o proporciones. Dentro de esta categoría existen distintos tipos de test, la utilización de uno u otro depende de qué tipo de información se quiera obtener:

- Test de distribución esperada o "*goodness of fit*": Se emplean para comparar la distribución observada frente a una distribución esperada o teórica.
- Test de diferencia de frecuencias o test de independencia: Se emplean para estudiar si la frecuencia de observaciones es significativamente distinta entre dos o más grupos.

En los test de "goodness of fit" solo hay una variable asociada a cada observación, mientras que en los test de independencia hay dos variables asociadas a cada observación. También se emplean distintos test dependiendo del tipo de datos (independientes o dependientes) con los que se vaya a trabajar. Las siguientes tablas muestran algunos de los más empleados.

## Datos independientes

Tipo de test	Distribución esperada	Comparación de grupos
Exacto	Test binomial exacto / Test multinomial exacto	Fisher's exact
Aproximado	Chi-square goodness of fit / G-test goodness of fit	Chi-square test of Independence

## Datos pareados

Tipo de tabla	test
Tablas 2x2	McNemar
Tablas 2Xk	Q de Cochran

## Test exactos y test aproximados

Los test exactos calculan la probabilidad de obtener los resultados observados de forma directa generando todos los posibles escenarios y calculando la proporción en los que se cumple la condición estudiada (son test de permutaciones). Los test aproximados calculan primero un estadístico y luego emplean la distribución teórica de dicho estadístico para obtener la probabilidad de que adquiera valores iguales o más extremos.

Existe bastante controversia en cuanto a si se deben de utilizar test exactos o aproximados. En la era pre-computacional, los test exactos se complicaban mucho cuando el tamaño total de muestras aumentaba, sin embargo, por medio de la computación esta barrera se ha eliminado. Los test exactos son más precisos cuando el tamaño total de observaciones es bajo o alguno de los grupos tiene pocas observaciones, una vez alcanzado un número alto de observaciones las diferencias son mínimas. En el libro *Handbook of Biological Statistics John H. McDonald* se recomienda utilizar test exactos cuando el número total de observaciones es menor a 1000 o cuando, aunque el número total sea mayor a 1000, haya algún grupo cuyo número de eventos esperados sea pequeño (normalmente menor que 5). En el caso de aplicar test aproximados sobre tamaños pequeños se suelen emplear correcciones, las más frecuentes son la *corrección de continuidad de Yate* o la *corrección de William*.

Se puede considerar a los test basados el estadístico  $\chi^2$  como una generalización del contraste de proporciones basado en la aproximación a la normal (*Z-test de una proporción* y *Z-test de dos proporciones*) cuando hay 2 o más variables cualitativas o alguna de ellas tiene 2 o más niveles. En aquellos casos en los que ambos test se pueden aplicar, el resultado de un *Z-test* y un *test*  $\chi^2$  es equivalente. Esto es debido a que en la distribución chi-cuadrado con 1 grado de libertad el estadístico  $\chi^2$  es igual al estadístico Z de una distribución normal, elevado al cuadrado.

## Test binomial exacto

El test binomial exacto se emplea para estudiar si la proporción de eventos verdaderos de una variable de tipo binomial se diferencia significativamente de la frecuencia teórica con la que se esperaría que apareciesen. En R puede realizarse un test binomial exacto mediante la función `binom.test()`, obteniendo tanto el *p-value* como el intervalo de confianza del valor de *p* en la población.

## Condiciones

- La variable tiene que ser de tipo binomial.
- Los eventos deben de ser independientes unos de otros.

## Ejemplo

*Un estudio sugiere que en una ciudad están naciendo más hombres que mujeres. Para determinar si es cierto se selecciona una muestra aleatoria de los niños nacidos durante los últimos 2 años y se identifica para cada uno el género. ¿Existen diferencias significativas para un nivel de significancia del 5%?*

```
datos <- c("masculino", "masculino", "masculino", "masculino", "femenino",  
"masculino", "masculino", "masculino", "femenino", "femenino", "femenino",  
"femenino", "femenino", "femenino", "masculino", "masculino", "femenino",  
"masculino", "femenino", "masculino", "femenino", "masculino", "masculino",  
"masculino", "femenino", "masculino", "masculino", "femenino", "masculino",  
"femenino")
```

## Hipótesis

$H_0$ : la proporción de nacimientos es igual (50%) para ambos sexos, por lo tanto  $p = 0.5$ .

$H_a$ : la proporción de nacimientos no es igual entre ambos sexos  $p \neq 0.5$ .

## Contar el número de eventos

```
tabla <- table(datos)
tabla
```

```
## datos
##  femenino masculino
##           13         17
```

## Test binomial exacto

Se va a considerar como verdadero si el sujeto es femenino y falso si es masculino.

```
binom.test(x = c(13, 17), alternative = "two.sided", conf.level = 0.95)
```

```
##
##  Exact binomial test
##
## data:  c(13, 17)
## number of successes = 13, number of trials = 30, p-value = 0.5847
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2546075 0.6257265
## sample estimates:
## probability of success
##           0.4333333
```

```
# También se puede pasar directamente la tabla como argumento. binom.test(x
# = tabla, alternative = 'two.sided', conf.level = 0.95)
```

## Conclusión

El test binomial no encuentra diferencias significativas entre las observaciones y lo que cabría esperar si la verdadera probabilidad de nacimiento es del 50% para ambos sexos.

## Test multinomial exacto

El test multinomial exacto es una extensión del test binomial que permite comparar la distribución de eventos observados para una variable con más de dos niveles respecto a una distribución teórica.

$$P(m_1, m_2, \dots, m_k) = \frac{n!}{m_1! m_2! \dots m_k!} p_1^{m_1} p_2^{m_2} \dots p_k^{m_k}$$

Este método, a pesar de ser un test exacto, es sustituido con frecuencia por su equivalente aproximado  $\chi^2$  *goodnes of fit* debido a que requiere mucha capacidad de cálculo cuando aumenta el número de observaciones o de grupos. Si es posible su computación, es más recomendable ya que no se basa en una aproximación mediante un modelo asintótico.

En R se puede realizar un test multinomial con la función `xmult()` del paquete *XNomial*. Si el número de observaciones o número de grupos genera un exceso de requisitos de cálculo, la función `xmulti()` devuelve un mensaje de error, por lo general, alcanzado dicho tamaño los test aproximados son prácticamente igual de buenos. Otra alternativa es recurrir a la simulación de Monte Carlo empleando la función `xmonte()` también del paquete *XNomial*.

Si a la función `xmulti()` se le indica como argumento *detail=3*, devuelve el *p-value* obtenido por multibinomial junto con el obtenido por las aproximaciones  $\chi^2$  y *the likelihood ratio (LLR)*. Para información más detallada consultar

<https://cran.r-project.org/web/packages/XNomial/vignettes/XNomial.html>,

donde se explican bien las ventajas de emplear un test multinomial frente a la aproximación por  $\chi^2$ .

## Ejemplo

En los estudios genéticos de Mendel se establece que la descendencia de un determinado cruce de plantas da lugar a cuatro tipos de semillas (amarillas-lisas, verdes-lisas, amarillas-rugosas y verdes-rugosas) con una proporción teórica de 9:3:3:1. Para demostrar si es cierto se recoge una muestra de 556 semillas y se observa que las cantidades

para cada grupo son de 315, 108, 101, 32 respectivamente. ¿Se corresponden las observaciones con el modelo esperado?

```
observaciones <- c(315, 108, 101, 32)
freq_teoricas <- c(9, 3, 3, 1)
```

### Solución mediante test multinomial exacto

```
require(XNomial)
xmulti(obs = observaciones, expr = freq_teoricas, statName = "Prob")

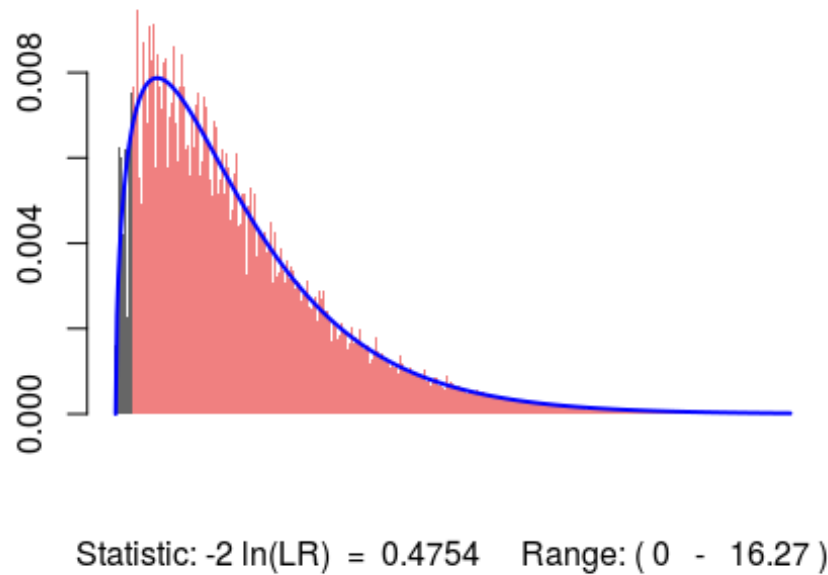
# Las frecuencias se pueden pasar también como porcentajes. En este caso no
# se supera el límite de computación.
```

```
##
## P value (Prob) = 0.9382
```

La función `xmulti()` también calcula el *p-value* a partir de modelos aproximados (explicados más adelante) y crea una representación gráfica que permite ver como de bien se ajusta la distribución chi-cuadrado a las probabilidades exactas.

```
xmulti(obs = observaciones, expr = freq_teoricas, detail = 3, histobins = TRUE)
```

```
##
## P value (LLR) = 0.9261
## P value (Prob) = 0.9382
## P value (Chisq) = 0.9272
##
## Observed: 315 108 101 32
## Expected ratio: 9 3 3 1
## Total number of tables: 28956759
```



## Conclusión

El *p-value* obtenido indica que las observaciones se asemejan mucho a las esperadas según la distribución de frecuencias del modelo teórico ( $H_0$ ). No hay evidencias para rechazar la hipótesis nula.



## $\chi^2$ Goodness of fit

El test  $\chi^2$  *goodness of fit* es la alternativa aproximada al test binomial (dos niveles) o multinomial (>2 niveles) normalmente utilizada cuando el número de observaciones o de grupos es demasiado alto para poder emplear los test exactos. Permite comparar la distribución de las observaciones en los diferentes niveles de una variable cualitativa con lo esperado acorde a una distribución hipotética  $H_0$ . Se trata por lo tanto de una expansión del *Z-test para una proporción* cuando la variable estudiada tiene dos o más niveles. Cuando la variable cualitativa tiene dos niveles, el test  $\chi^2$  *goodness of fit* y el *Z-test para una proporción* son equivalentes.

El funcionamiento del test consiste en cuantificar y sumar las diferencias entre el número de eventos observadas en cada nivel con respecto al número esperado acorde a  $H_0$ . El valor de este sumatorio se asigna a un estadístico llamado  $\chi^2$  y se emplea su relación matemática con la distribución  $\chi^2$  para estimar la probabilidad de obtener un valor igual o más extremo.

### Hipótesis

**$H_0$ :** Las observaciones siguen una distribución acorde a la distribución del modelo de referencia. Las desviaciones son debidas al azar.

**$H_a$ :** Las observaciones no siguen una distribución acorde a la distribución del modelo de referencia. Las desviaciones son mayores a las esperadas por azar.

### Condiciones del test

**Independencia:** las observaciones de la muestra deben ser independientes unas de otras.

- Muestreo aleatorio
- Tamaño de la muestra < 10% población
- Cada observación contribuye únicamente a uno de los niveles de la variable estudiada.

**Tamaño:** cada nivel debe tener al menos 5 eventos **esperados** (acorde a  $H_0$ ) y el número de observaciones totales ( $n$ )>30.

- En el libro *Bioestadística de Francisca Ruis Díaz* consideran que esta regla es muy estricta y rara vez se cumple en la práctica. Proponen unas condiciones más relajadas con las que no se pierde demasiada precisión: para ningún nivel el número de eventos esperados acorde a  $H_0$  es menor de 1 y como máximo un 20% de los niveles tiene menos de 5 eventos esperados.
- En caso de no cumplirse esta condición o estar en el límite es mejor recurrir a los test exactos o, si no es posible por requerimientos de computación, a la simulación de Monte Carlo.
- Aun cuando se cumplen las condiciones, son más precisos los test exactos y por lo tanto más recomendables.

## Estadístico $\chi^2$

$$\chi^2 = \sum_{i,j} \frac{(\text{observado} - \text{esperado})^2}{\text{esperado}}$$

Elevar al cuadrado las diferencias permite hacerlas todas positivas y además magnificar aquellas más grandes.

## Grados de libertad

La distribución chi-cuadrado tiene un único parámetro, los grados de libertad, que determinan su forma, centro y dispersión.  $df = k - 1$ , siendo  $k$  el número de niveles de la variable.

## Cálculo de $p$ -value

La distribución chi-cuadrado es siempre positiva, por lo que para calcular el  $p$ -value solo se tiene en cuenta la cola superior. Los  $p$ -values se pueden obtener de tablas tabuladas o mediante software.

## Ejemplo 1

Este ejemplo es el mismo que el propuesto para el test binomial exacto pero solucionado mediante  $\chi^2$  *goodness of fit*. Dado que la variable cualitativa tiene dos niveles, también podría resolverse mediante un contraste de hipótesis para una proporción (*Z-test de una proporción*).

*Un estudio sugiere que en una ciudad están naciendo más hombres que mujeres. Para determinar si es cierto se selecciona una muestra aleatoria de los niños nacidos durante los últimos 2 años y se identifica para cada uno el género. ¿Existen diferencias significativas?*

```
datos <- c("masculino", "masculino", "masculino", "masculino", "femenino",  
"masculino", "masculino", "masculino", "femenino", "femenino", "femenino",  
"femenino", "femenino", "femenino", "masculino", "masculino", "femenino",  
"masculino", "femenino", "masculino", "femenino", "masculino", "masculino",  
"masculino", "femenino", "masculino", "masculino", "femenino", "masculino",  
"femenino")
```

### Hipótesis

***H<sub>0</sub>***: la proporción de nacimientos es igual (50%) para ambos sexos.  $p_{\text{masculino}} = p_{\text{femenino}} = 0.5$ .

***H<sub>a</sub>***: la proporción de nacimientos no es igual entre ambos sexos.  $p_{\text{masculino}} \neq p_{\text{femenino}}$

### Contar el número de eventos

```
tabla <- table(datos)  
tabla
```

```
## datos  
##   femenino masculino  
##        13         17
```

## Test $\chi^2$ goodness of fit

Se va a considerar como verdadero si el sujeto es femenino y falso si es masculino.

```
chisq.test(x = c(13, 17), p = c(0.5, 0.5), correct = TRUE)
# También se puede pasar directamente la tabla como argumento. chisq.test
# (tabla,p = c(0.5, 0.5))
```

```
##
## Chi-squared test for given probabilities
##
## data:  c(13, 17)
## X-squared = 0.53333, df = 1, p-value = 0.4652
```

## Conclusión

El test  $\chi^2$  *goodness of fit* no encuentra diferencias significativas entre las observaciones y lo que cabría esperar si la verdadera probabilidad de nacimiento es del 50% para ambos sexos.

## Ejemplo 2

*Un estudio intenta comparar si existe discriminación racial en la formación de jurados populares. Para ello se compara la distribución racial en el país con la distribución de los miembros de los últimos 500 jurados.*

*Distribución racial en la población*

Raza	blanca	negra	latina	asiática	otra
% en la población	80.29	12.06	0.79	2.92	3.94

*Distribución racial en los jurados*

Raza	blanca	negra	latina	asiática	otra
n individuos	1920	347	19	84	130

## Solución manual

Como en todo contraste, el primer paso es establecer las hipótesis.

$H_0$ : Los jurados son muestras aleatorias de la población. La distribución de razas en los jurados es igual a la distribución en la población (las diferencias son las esperadas por azar).

$H_a$ : La distribución de raza en los jurados no es igual a la distribución de raza en la población.

A continuación se calcula el número de eventos que cabría esperar en cada nivel de la variable cualitativa, en este caso raza, si fuese cierta la hipótesis nula. Para ello se multiplica el total de observaciones por el % de cada nivel considerado en la distribución  $H_0$ .

Raza	blanca	negra	latina	asiática	otra
nº esperado	2007.25	301.50	19.75	73.00	98.50

Para que los resultados de un test  $\chi^2$  sean válidos se han de cumplir sus condiciones:

- Independencia: Sí
- Tamaño: todos los niveles tienen más de 5 eventos **esperados**.

Finalmente se calcula el valor del estadístico  $\chi^2$ , los grados de libertad y se obtiene el *p-value*.

$$\chi^2 = (1920-2007.25)^2/2007.25 + \dots + (130-98.5)^2/98.5 = 22.42$$

$$df = 5-1 = 4$$

$$p\text{-value} = \boxed{\text{pchisq}(q= 22.42, df= 4, lower.tail = FALSE)} = 1.653006310^{-4}$$

Existen evidencias significativas para rechazar  $H_0$  en favor de  $H_a$ . La distribución de razas en los jurados no es la que cabría esperar si la selección se hiciese de forma aleatoria.

## Solución con R

R tiene una función propia `chisq.test()` que permite realizar un test de *goodness of fit*, tanto teórico como por simulación de Monte Carlo.

```
chisq.test(x = c(1920, 347, 19, 84, 130), p = c(80.29, 12.06, 0.79, 2.92, 3.94)/100)
```

```
##  
## Chi-squared test for given probabilities  
##  
## data:  c(1920, 347, 19, 84, 130)  
## X-squared = 22.419, df = 4, p-value = 0.0001654
```

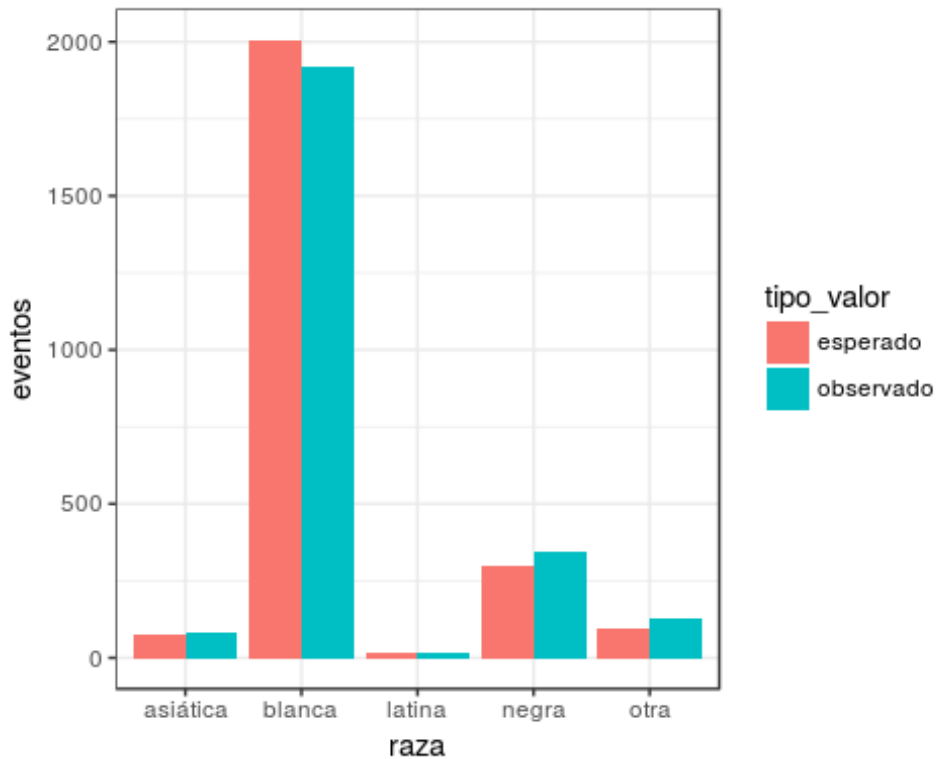
En caso de que no se cumpla el número mínimo de eventos por nivel, se recurre a la simulación:

```
chisq.test(x = c(1920, 347, 19, 84, 130), p = c(80.29, 12.06, 0.79, 2.92, 3.94)/100,  
  simulate.p.value = TRUE, B = 5000)
```

```
##  
## Chi-squared test for given probabilities with simulated p-value  
## (based on 5000 replicates)  
##  
## data:  c(1920, 347, 19, 84, 130)  
## X-squared = 22.419, df = NA, p-value = 0.0005999
```

Además del valor del *p-value* se recomienda incluir una representación gráfica de los valores observados y esperados.

```
require(ggplot2)  
datos <- data.frame(raza = rep(c("blanca", "negra", "latina", "asiática", "otra"),  
  2), tipo_valor = rep(c("observado", "esperado"), each = 5), eventos = c(1920,  
  347, 19, 84, 130, 2007.25, 301.5, 19.75, 73, 98.5))  
ggplot(data = datos, aes(x = raza, y = eventos)) + geom_bar(stat = "identity",  
  aes(fill = tipo_valor), position = "dodge") + theme_bw()
```



### Ejemplo 3

*Un estudio pretende demostrar si la tendencia diaria del mercado de valores es independiente del comportamiento observado en el mercado el día previo. Para ello se dispone de un conjunto de datos que muestran la tendencia diaria (al alza o a la baja) del mercado durante los últimos años y los días transcurridos desde la última vez que el mercado estuvo al alza.*

```
library(gridExtra)
grid.table(data.frame(variacion = c("2.52", "-1.46", "0.51", "-4.07", "3.36",
  "1.10", "-5.46", "-1.03", "-2.99", "1.71", "..."), tendencia = c("Up", "Down",
  "Up", "Down", "Up", "Up", "Down", "Down", "Down", "Up", "..."),
  dias_hasta_subida = c("0",
  "-", "1", "-", "1", "0", "-", "-", "-", "3", "..."))))
```

	variacion	tendencia	dias_hasta_subida
1	2.52	Up	0
2	-1.46	Down	-
3	0.51	Up	1
4	-4.07	Down	-
5	3.36	Up	1
6	1.10	Up	0
7	-5.46	Down	-
8	-1.03	Down	-
9	-2.99	Down	-
10	1.71	Up	3
11	...	...	...

Si la tendencia del mercado es independiente del día anterior, el número de días hasta la siguiente subida debería seguir una distribución geométrica con probabilidad igual al porcentaje de días que han sido al alza en todo ese tiempo, en este caso supongamos que es 0.532. *La distribución geométrica describe la probabilidad de esperar  $k$  intentos hasta observar el primer evento positivo*

La siguiente tabla recoge el número total de observaciones para cada número de días de espera hasta el siguiente día de subida:

Dias espera	0	1	2	3	4	5	6+	Total
Eventos observados	1532	760	338	194	74	33	17	2948

Las proporciones de días esperados acorde a una distribución geométrica son:

```
c(dgeom(x = 0:5, prob = 0.532), 1 - sum(dgeom(x = 0:5, prob = 0.532)))
```

```
## [1] 0.53200000 0.24897600 0.11652077 0.05453172 0.02552084 0.01194376
## [7] 0.01050691
```



Mediante `chisq.test()` se comparan la distribución observada con la esperada.

```
chisq.test(x = c(1532, 760, 338, 194, 74, 33, 17), p = c(0.532, 0.248976,  
0.1165208, 0.0545317, 0.0255208, 0.0119438, 0.0105069))
```

```
##  
## Chi-squared test for given probabilities  
##  
## data: c(1532, 760, 338, 194, 74, 33, 17)  
## X-squared = 15.189, df = 6, p-value = 0.01883
```

Para un nivel de significancia del 5% hay evidencias de que los datos no siguen una distribución geométrica, por lo tanto no son independientes de un día a otro.