Entorno de programación RStudio: inferencia (2/2)

Román Salmerón Gómez

Abril de 2017

El objetivo del presente documento es el de crear un breve informe con **RStudio** mediante el cual se muestre al alumno como realizar inferencia en un conjunto de datos referentes a proporciones.

## Funciones a usar

Para comparar prorciones se dispone del comando *prop.test*:

**prop.test(x, n, p = NULL, alternative = c("two.sided", "less", "greater"), conf.level = 0.95, correct = TRUE)**

Sin embargo, nostros usaremos la siguiente:

proporcion.test <- function(muestra1, muestra2 = 0, p = 0, alternative = "two.sided", conf.level = 0.95)  
 {  
 alfa = 1 - conf.level  
 if (length(muestra2)==1)  
 {  
 P = mean(muestra1)  
 n = length(muestra1)  
 z.exp = (P-p)/sqrt((P\*(1-P))/n)  
 resultado1="No rechazo hipótesis nula"  
 if (alternative == "two.sided")  
 {  
 if(abs(z.exp)>qnorm(1-alfa/2)){resultado1="Rechazo hipótesis nula"}  
 resultado2 = abs(z.exp)  
 resultado3 = qnorm(1-alfa/2)  
 }  
 if (alternative == "greater")  
 {  
 if(z.exp>qnorm(1-alfa)){resultado1="Rechazo hipótesis nula"}  
 resultado2 = z.exp  
 resultado3 = qnorm(1-alfa)  
 }  
 if (alternative == "less")  
 {  
 if(z.exp<qnorm(alfa)){resultado1="Rechazo hipótesis nula"}  
 resultado2 = z.exp  
 resultado3 = qnorm(alfa)  
 }  
 extremo.inf = P - qnorm(1-alfa/2)\*sqrt((P\*(1-P))/n)  
 extremo.sup = P + qnorm(1-alfa/2)\*sqrt((P\*(1-P))/n)  
 resultado4 = c(extremo.inf, extremo.sup)  
 }  
 else  
 {  
 P1 = mean(muestra1)  
 n1 = length(muestra1)  
 P2 = mean(muestra2)  
 n2 = length(muestra2)  
 z.exp = (P1-P2)/sqrt(((P1\*(1-P1))/n1)+((P2\*(1-P2))/n2))  
 resultado1="No rechazo hipótesis nula"  
 if (alternative == "two.sided")  
 {  
 if(abs(z.exp)>qnorm(1-alfa/2)){resultado1="Rechazo hipótesis nula"}  
 resultado2 = abs(z.exp)  
 resultado3 = qnorm(1-alfa/2)  
 }  
 if (alternative == "greater")  
 {  
 if(z.exp>qnorm(1-alfa)){resultado1="Rechazo hipótesis nula"}  
 resultado2 = z.exp  
 resultado3 = qnorm(1-alfa)  
 }  
 if (alternative == "less")  
 {  
 if(z.exp<qnorm(alfa)){resultado1="Rechazo hipótesis nula"}  
 resultado2 = z.exp  
 resultado3 = qnorm(alfa)  
 }  
 extremo.inf = (P1-P2) - qnorm(1-alfa/2)\*sqrt(((P1\*(1-P1))/n1)+((P2\*(1-P2))/n2))  
 extremo.sup = (P1-P2) + qnorm(1-alfa/2)\*sqrt(((P1\*(1-P1))/n1)+((P2\*(1-P2))/n2))  
 resultado4 = c(extremo.inf, extremo.sup)  
 }  
 resultados = list(resultado1, resultado2, resultado3, resultado4)  
 names(resultados) = c("Decisión", "Valor experimental", "Valor teórico", "Intervalo de confianza bilateral")  
 resultados  
 }

Los argumentos son similares a los de la función *t.test*.

## Conjunto de datos

En los siguientes conjuntos de datos se tienen la proporción de individuos que recomiendan comprar cierta marca de bebida tras probarla:

muestra1 = c(0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0)  
 mean(muestra1)

## [1] 0.36

muestra2 = c(0, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1)  
 mean(muestra2)

## [1] 0.4333333

En el primer caso la muestra de individuos que recomiendan la compra es del 36% y en la segunda del 43.3%.

### Inferencia para una población

Para contrastar si la proporción de la primera muestra es igual, mayor o menor que el 45% usaríamos:

proporcion.test(muestra1, p=0.45) # por defecto es a dos colas

## $Decisión  
## [1] "No rechazo hipótesis nula"  
##   
## $`Valor experimental`  
## [1] 0.9375  
##   
## $`Valor teórico`  
## [1] 1.959964  
##   
## $`Intervalo de confianza bilateral`  
## [1] 0.1718435 0.5481565

proporcion.test(muestra1, p=0.45, alternative="greater")

## $Decisión  
## [1] "No rechazo hipótesis nula"  
##   
## $`Valor experimental`  
## [1] -0.9375  
##   
## $`Valor teórico`  
## [1] 1.644854  
##   
## $`Intervalo de confianza bilateral`  
## [1] 0.1718435 0.5481565

proporcion.test(muestra1, p=0.45, alternative="less")

## $Decisión  
## [1] "No rechazo hipótesis nula"  
##   
## $`Valor experimental`  
## [1] -0.9375  
##   
## $`Valor teórico`  
## [1] -1.644854  
##   
## $`Intervalo de confianza bilateral`  
## [1] 0.1718435 0.5481565

Atendiendo al p-valor, no rechazo que la media sea igual al 45% frente a la alternativa de que sea distinta ( = 0.9375 < 1.96), no rechazo que la media sea igual al 45% frente a la alternativa de que sea mayor ( = -0.9375 < 1.64) y no rechazo la hipótesis de que sea igual al 45% frente a la alternativa de que sea menor ( = -0.9375 > -1.64).

### Inferencia para dos poblaciones

Empecemos viendo si *speed* y *dist* tienen la misma varianza. Observando las cuasivarianzas muestrales se debería verificar que la primera es menor que la segunda, por tal motivo nos interesamos por esta hipótesis:

proporcion.test(muestra1, muestra2) # por defecto es a dos colas

## $Decisión  
## [1] "No rechazo hipótesis nula"  
##   
## $`Valor experimental`  
## [1] 0.5559196  
##   
## $`Valor teórico`  
## [1] 1.959964  
##   
## $`Intervalo de confianza bilateral`  
## [1] -0.3318791 0.1852125

proporcion.test(muestra1, muestra2, alternative="greater")

## $Decisión  
## [1] "No rechazo hipótesis nula"  
##   
## $`Valor experimental`  
## [1] -0.5559196  
##   
## $`Valor teórico`  
## [1] 1.644854  
##   
## $`Intervalo de confianza bilateral`  
## [1] -0.3318791 0.1852125

proporcion.test(muestra1, muestra2, alternative="less")

## $Decisión  
## [1] "No rechazo hipótesis nula"  
##   
## $`Valor experimental`  
## [1] -0.5559196  
##   
## $`Valor teórico`  
## [1] -1.644854  
##   
## $`Intervalo de confianza bilateral`  
## [1] -0.3318791 0.1852125

Atendiendo a los resultados, no se rechaza la hipótesis nula en ninguno de los casos.

## Ejercicios propuestos

* Estudiar si la proporción de éxitos de la segunda muestra puede ser igual, mayor o menor que el 25%.
* Genera una muestra de tamaño 45 que tenga sólo los valores 0 y 1 y analiza si:
* La proporción de éxitos es igual mayor, menor o igual al 75%.
* La proporción de éxitos es la misma que en las muestras 1 y 2.