Entorno de programación RStudio: normalidad

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 comprobar si un conjunto de datos procede de una población normal.

Para ello, consideraremos una vez el conjunto de datos de la base *cars*:

attach(cars) # speed y dist

En prácticas anteriores se ha usado esta base de datos para realizar inferencia, siendo una de las hipótesis que se han de cumplir que las poblaciones se distribuyen como una normal. ¿Se distribuye la variable *speed* según una normal?:

# en todos los contrastes, H0: hay normalidad  
 ks.test(speed, pnorm, mean(speed), sd(speed)) # en este caso se especifica la media y varianza que ha de tener la normal

## Warning in ks.test(speed, pnorm, mean(speed), sd(speed)): ties should not  
## be present for the Kolmogorov-Smirnov test

##   
## One-sample Kolmogorov-Smirnov test  
##   
## data: speed  
## D = 0.068539, p-value = 0.9729  
## alternative hypothesis: two-sided

shapiro.test(speed)

##   
## Shapiro-Wilk normality test  
##   
## data: speed  
## W = 0.97765, p-value = 0.4576

# install.packages("nortest") # ejecutarlo sólo una vez  
 library(nortest) # cargamos paquete  
 lillie.test(speed)

##   
## Lilliefors (Kolmogorov-Smirnov) normality test  
##   
## data: speed  
## D = 0.068539, p-value = 0.8068

Si el p-valor asociado a cada uno de los contrastes es mayor que 0.05 no se rechaza la hipótesis nula de normalidad. Es decir, no se rechaza que el conjunto de datos almacenado en *speed* procede de una normal. En este caso, todos los p-valores son mayores que 0.05 (0.9729, 0.4576 y 0.8068, respectivamente), luego no se rechaza la hipótesi nula de normalidad.

Más información sobre la normalidad de un conjunto de datos en <https://rstudio-pubs-static.s3.amazonaws.com/122473_d6856ba024cc45abbb2cc4102471e362.html>.

## Ejercicios propuestos

* Estudiar si la variable aleatoria *dist* procede de una distribución normal.
* Ejecutar el código *runs.test(as.factor(speed>median(speed)))* (para ello es necesario instalar el paquete *tseries*, *install.packages("tseries")*, y cargarlo posteriormente, *library(tseries)*). ¿Qué hipótesis se estaría comprobando con este test? (pista: usar *help(runs.test)*)
* Ejecutar el código *wilcox.test(speed,dist)*. ¿Qué hipótesis se estaría comprobando con este test? (pista: usar *help(wilcox.test)*). Adviértase que dicha hipótesis también se puede comprobar ejecutando *ks.test(speed, dist)*.

Para las dos últimas preguntas, si las pistas no son suficiente, consultar <https://ocw.uca.es/pluginfile.php/283/mod_resource/content/1/ebrcmdr.c6.pdf>.