

Práctica Test de hipótesis

Introducción al análisis de datos con R A2B2C

Octubre 2020

Test de hipótesis

Esta práctica tiene como objetivo trabajar con conceptos de probabilidad, de distribuciones y aplicar test de hipótesis para poner a prueba diferentes hipótesis.

1. Distribuciones y probabilidad

- a) Se tira una moneda con dos posibles resultados, cara o ceca, n veces. La distribución binomial da la probabilidad de obtener exactamente m veces cara. La moneda puede estar cargada, por ejemplo, la probabilidad de que salga cara en cada tirada es de 0.75 y de que salga ceca es 0.25.
¿Cuál es la probabilidad de observar exactamente 5 caras en 10 tiradas con una moneda justa (no cargada)? Ayuda: `?dbinom`
Spoiler alert! `dbinom(5, size = 10, prob = 0.5)`
- b) Se tiene una secuencia de ADN de 50 nucleótidos. Si en cada posición la probabilidad de que salga alguna de las cuatro bases es igual (o sea, 0.25), ¿cuál es la probabilidad de observar exactamente 20 “A” en una secuencia? ¿Y de observar 20 o menos? Ayuda: `?dbinom` `?pbinom`
- c) Graficar la probabilidad de obtener exactamente 0, 1, 2...50 “A” en la secuencia.
- d) Generar 10000 secuencias de 50 nucleótidos de ADN (A, C, G o T) y contar cuantas veces sale “A” en cada una (Ayuda, `rowSums(secuencias == .^A)`) con secuencias el data.frame de 10000 secuencias de 50 nucleótidos). Luego graficar un histograma (de densidades, con `freq = FALSE`) de cuantas veces salió “A” en cada secuencia y superponerlo con el gráfico del punto 2. ¿Qué se observa? Ayuda: `?sample`, `?points`
- e) Generar 100000 números aleatorios con distribución uniforme. Realizar un histograma y mostrar que la probabilidad de que un número se encuentre entre 0 y p es p . Ayuda: `?runif`
- f) Obtener 1000 números aleatorios con distribución normal con media = 1 y desviación estandar 2. ¿Cuál es la media de la muestra? ¿Y la desviación estandar? ¿Por qué? Realizar un histograma. Ayuda: `?rnorm`
- g) Repetir el ejercicio anterior para 100000 números. ¿Cómo son ahora la media y desviación estandar de la muestra?

- h) Realizar un histograma de densidad (`freq = FALSE`) de la altura para adultos varones en el dataset de antropometria. ¿Qué tipo de distribución parece ser? Superponer en el histograma una gaussiana con media y desviación iguales a las del dataset, con `x` entre el mínimo y el máximo valor de altura. ¿Qué concluís? Ayuda: `?dnorm` `?lines`

2. Test de hipótesis

En todos los casos verificar que se cumplan las condiciones para poder realizar el test.

- a) El dataset `datasets::sleep` contiene los datos de 20 pacientes con problemas de sueño. Los del grupo 2 recibieron una droga nueva para tratar el problema mientras que los del grupo 1 un placebo. El resultado en la regulación del sueño se midió un mes después (la columna `'extra'`).
¿Fue efectiva la droga?
- b) El dataset `boot::cd4` contiene un estudio realizado en 20 pacientes con VIH de una droga experimental. Se midió el contenido de CD4 en sangre antes de administrar la droga (baseline) y un año después de administrarla (`oneyear`).
¿Fue efectiva la droga? Tener en cuenta que este test es pareado, es decir, no son individuos distintos antes y después si no que son los mismos individuos. Ayuda `?t.test paired = TRUE`
- c) El dataset `iris` contiene las mediciones de ancho y largo de pétalo y sépalo para 3 especies distintas de plantas iris. Mediciones anteriores del largo del pétalo de la variedad setosa indican que esas plantas tienen un largo de pétalo medio de 1.5 cm.
¿Estas nuevas mediciones coinciden?
¿Son significativamente distintas las medias del largo del sépalo entre setosa y virgínica?
- d) Instalar el paquete `titanic` usando `install.packages("titanic")`. El dataset `titanic::titanic_train` contiene datos sobre pasajeros embarcados en el Titanic. Decidir si existe relación entre las clases de pasajeros (1, 2, 3, baja, media y alta respectivamente) y la supervivencia de los pasajeros. Construir previamente la tabla de contingencia.
Realizar la misma prueba entre sexo y supervivencia y entre sexo y clase.
¿A qué conclusión se llega en cada caso?
- e) Se desea saber si la proporción de adolescentes que estudian para los exámenes en la secundaria es mayor entre mujeres que entre hombres. De una muestra de 24 estudiantes, se tiene que hay 10 que estudian (1 hombre y 9 mujeres) y 14 que no (11 hombres y 3 mujeres).
¿Cuál es la hipótesis nula? ¿A qué conclusión se llega?