

Trabajo Práctico 4 en R

Test de hipótesis

Importante: se deberá entregar un informe que incluya el código con la resolución, las salidas solicitadas y las respuestas a las preguntas.

Sólo se aceptará en formato pdf. Puede obtenerlo compilando el R Markdown a pdf o compilándolo a word y luego transformando el word en un pdf.

Incluya el nro de TP y de grupo y los apellidos de los integrantes en el nombre del archivo (por ejemplo, “TP4_G3_Fernandez_Gimenez.pdf”) y también en el título del informe.

Justifique todas sus respuestas.

Fecha máxima de entrega: lunes 17/10.

1. El responsable de bromatología de un restaurante desea averiguar si los pollos que vende el proveedor A están contaminados (es decir si su concentración media de arsénico es mayor a las 80 ppb) en cuyo caso cancelará la relación con el proveedor. Para ello selecciona una muestra de pollos de dicho proveedor a los que les mide la concentración de arsénico. Supongamos que la concentración de arsénico (en ppb) de los pollos del proveedor sigue una distribución normal con varianza 16. Si la cadena está dispuesta a correr un riesgo del 10 % de cancelar al proveedor cuando en realidad los pollos no están contaminados, plantee el test de hipótesis que debe realizar el responsable de bromatología del restaurante. Para ello...
 - a) Defina (con palabras) el parámetro de interés.
 - b) Plantee las hipótesis nula y alternativa del test. Asuma que H_0 es simple.
 - c) Describa las variables de la muestra aleatoria, es decir, defínalas con palabras y especifique todo lo que sabe de su distribución
 - d) Indique cuál es el nivel del test
 - e) Indique cuál es el estadístico del test y qué distribución tiene bajo H_0 . Justifique.
 - f) Plantee la región de rechazo.
2. Guarde en un vector llamado **concentraciones**, 5 datos generados con distribución $\mathcal{N}(\mu, \sigma^2)$ con $\mu = 80$ y $\sigma^2 = 16$ usando como semilla a su número de DNI. En base a estos datos...
 - a) Calcule el valor del estadístico observado, definido en 1. e).
 - b) ¿Qué decisión toma según el test planteado en 1)? ¿Rechaza o no H_0 ?
 - c) Interprete la conclusión del test en el contexto del problema.

- d) En base a la decisión del test, ¿cancela la cadena la relación con el proveedor?
3. Implemente una función llamada **estadistico** que tenga por argumento un conjunto de datos y devuelva el estadístico del test definido en 1. e). Utilice la función para calcular el valor del estadístico en base a los datos que guardó en **concentraciones** para chequear que esté bien programada.
 4. Genere $Nrep = 10000$ muestras normales de tamaño 5 con media 80 y varianza 16 y guarde en el vector **muchos_estadisticos** los valores de los estadísticos observados en cada muestra. ¿Qué hipótesis satisfacen los datos de las muestras generadas? ¿La nula o la alternativa?
 5. Realice un histograma de densidad con los valores de **muchos_estadisticos** y superponga la curva de la densidad de una $\mathcal{N}(0, 1)$.
 - a) ¿Qué observa?
 - b) ¿A qué se debe lo que observa?
 6. Calcule la proporción de muestras generadas en 4 para las cuales se rechazaría H_0 en base al test propuesto en 1.
 - a) ¿A qué valor debería acercarse esa proporción? Justifique
 - b) ¿Efectivamente se acerca a ese valor?
 7. Ahora genere $Nrep = 10000$ muestras normales de tamaño 5 con media 84 y varianza 16 y guarde en el vector **muchos_estadisticos2** los valores de los estadísticos observados en cada muestra. ¿Qué hipótesis satisfacen los datos de las muestras generadas? ¿La nula o la alternativa?
(Ojo: el estadístico que debe calcular en cada muestra es el correspondiente al test planteado en 1, es decir el que calcula la función **estadistico** que implementó en el ítem 3)
 8. Realice un histograma de densidad con los valores de **muchos_estadisticos2** y superponga la curva de la densidad de una $\mathcal{N}(0, 1)$.
 - a) ¿Qué observa?
 - b) ¿A qué se debe lo que observa?
 - c) Vuelva a realizar el histograma de densidad con los valores de **muchos_estadisticos2**, pero ahora superponga la curva de la verdadera densidad del estadístico.
(Hint: para darse cuenta de cuál es la verdadera densidad del estadístico en este caso, piense cuál es la distribución con la cual fueron generadas las muestras en 7. En base a esto deduzca la distribución de \bar{X}_5 y luego use las propiedades que conoce sobre la distribución de combinaciones lineales de variables normales para hallar la distribución (o densidad) del estadístico.)
 9. Calcule la proporción de muestras generadas en 7 para las cuales NO se rechazaría H_0 en base al test propuesto en 1. ¿Qué probabilidad está estimando con esa proporción? Justifique.

10. Considere un test idéntico al propuesto en el ítem 1 pero de nivel 5 %.
- a) Indique la región de rechazo de este test.
 - b) Calcule la proporción de muestras generadas en 7 para las cuales NO se rechazaría H_0 en base a este test.
 - c) Compare este valor con el obtenido en el ítem 9, ¿a qué cree que se debe lo que observa?
11. Se extrae una muestra de tamaño 7 en la cual las concentraciones de arsénico observadas resultaron:

81,12 82,87 82,08 81,19 78,31 82,26 87,85

- a) Calcule el p-valor de esta muestra.
- b) Utilizando el p-valor, diga qué decisión tomaría en base estos datos a nivel 0,05 y a nivel 0,10. Justifique.
- c) Proponga una simulación que permita estimar dicho p-valor. Implementela con un $N_{rep} = 10000$ y corrobore que efectivamente el valor obtenido está cerca del p-valor calculado.