

Ejercicios Tema 4 - Contraste hipótesis. Taller 2

Ricardo Alberich, Juan Gabriel Gomila y Arnau Mir

Curso completo de estadística inferencial con R y Python

Contenidos

1 Contraste hipótesis taller 2.	1
1.1 Librerías y datos necesarios	1
1.2 Ejercicio 1	1
1.3 Ejercicio 2	2
1.4 Ejercicio 3	3
1.5 Ejercicio 4	3
1.6 Ejercicio 5	3

1 Contraste hipótesis taller 2.

1.1 Librerías y datos necesarios

Para este taller necesitaremos los siguientes paquetes: `faraway`, `nortest`, `car` si no los tenéis instalados podéis ejecutar lo siguiente:

```
install.packages("faraway")
install.packages("nortest")
install.packages("car")
```

Para utilizarlos, deberéis cargarlos ejecutando las siguientes instrucciones:

```
library("faraway")
library("nortest")
library("car")
```

También necesitáis el fichero “zinc.txt”.

1.2 Ejercicio 1

El [iris data set](#) es una tabla clásica de datos que recopiló [Ronald Fisher](#) publicó en 1936. En este data set hay 150 flores de tres especies las que se mide la longitud y anchura de sus pétalos y sépalos.

La medias globales de toda la población son

```
library(tidyverse)
resumen1=iris %>% summarise(Media_muestral_Sepal.Length=mean(Sepal.Length),
                           Desviacion_muestral_Sepal.Length=sd(Sepal.Length))
resumen1
```

```
##   Media_muestral_Sepal.Length Desviacion_muestral_Sepal.Length
## 1                5.843333          0.8280661
```

Consideremos una muestra de tamaño $n = 50$ de la longitud del sépalo del data set iris que generamos con el siguiente código

```
set.seed(333)# por reproducibilidad, para fijar la muestra
muestra_50=sample(iris$Sepal.Length,size=50,replace = TRUE)
```

1. Contrastar si podemos aceptar que la media de la muestra es igual a la media poblacional es igual a 5.5 contra que es distinta, resolver utilizando el p -valor.
2. Calcular un intervalo de confianza del tipo $(-\infty, x_0)$ para la media poblacional de la muestra al nivel de confianza del 95%

1.3 Ejercicio 2

Si consideramos el data set iris la población la proporción poblacional p de flores que tienen la longitud del sépalo mayor que 5 es

```
Sepalo_mayor_5=prop.table(table(iris$Sepal.Length>5))["TRUE"]
Sepalo_mayor_5
```

```
##      TRUE
## 0.7866667
```

Tomamos una muestra de tamaño $n = 30$ de la población de iris y calculamos en ella la proporción de flores con sépalo mayor que 5.

```
set.seed(44)
muestra_30=sample(iris$Sepal.Length,size=30,replace = TRUE)
x=table(muestra_30>5)["TRUE"]
x
```

```
## TRUE
##   25
```

```
phat=as.numeric(prop.table(table(muestra_30>5))["TRUE"])
phat
```

```
## [1] 0.8333333
```

1. Queremos contrastar si esta muestra confirma la proporción de flores con sépalo mayor que 5 es 0.75 contra que es mayor de 0.75 con el test exacto y el test aproximado.
2. Extraer de los dos test los intervalos de confianza asociados al contraste y decir qué fórmula se utiliza para cada intervalo.
3. Extraer de los dos test el valor del estadístico de conytraste y el p valor.

1.4 Ejercicio 3

Concentración de zinc

El rastro de metales en el agua potable afecta el sabor y una concentración inusualmente alta puede representar un riesgo para la salud. El fichero *zinc.txt* contiene la concentración de zinc en el fondo y en la superficie de botellas de agua.

Se cree que la concentración media de zinc del agua en el fondo de la botella es mayor que la de la superficie. Suponiendo que los datos siguen una ley normal, ¿hay evidencia suficiente para asegurarlo con un nivel de confianza del 95%?

Plantea un contraste de hipótesis para estudiar si existe dicha evidencia. En particular, especifica la hipótesis nula H_0 , la hipótesis alternativa H_1 , la expresión del estadístico que se calculará a partir de los datos observados y la distribución de dicho estadístico.

1.5 Ejercicio 4

Continuación ejercicio concentración de zinc

Carga el fichero en la variable `conc.zinc`. Utiliza la función de R que calcule el test de hipótesis definido. Interpreta los resultados.

1.6 Ejercicio 5

Continuación ejercicio concentración de zinc.

Encuentra la región crítica (es decir el intervalo en el cual se rechaza la hipótesis nula) y la región de aceptación (es decir el intervalo de no rechazo de la hipótesis nula)