

Métodos Estadísticos en Economía. Grado en Administración y Dirección de Empresas

PRÁCTICA 4. Inferencia estadística

Alumnes:

Consejos:

Antes de iniciar la práctica recuerda lo siguiente:

1. Crea una carpeta vacía que se llame practica4_MEE
2. Abre Rstudio
3. Crea un nuevo proyecto con el menú de Rstudio

File->Create Project-> Existing Directory

Busca la carpeta creada en el punto 1.

Ya podemos empezar una sesión nueva de R que guardaremos al final en esta carpeta cuando salgamos. La ventaja es que podemos continuar esta sesión cuando queramos accediendo a la carpeta y ejecutando el proyecto.

4. Crea un fichero R script con el menú de Rstudio

File->New File->Rscript

Ponle nombre practica4 su extensión será .R

En este fichero guardaremos todas las órdenes de R de manera que podamos repetir las o modificarlas cuando lo deseemos.

Nota: NO UTILICES ACENTOS EN LOS NOMBRES DE LOS FICHEROS

Ejercicio 1

Importa a R los archivos de datos 'colhombres.xlsx' y 'colmujeres.xlsx' con la que trabajamos en la práctica 2 (disponibles en Poliformat en la carpeta de Práctica 2). Recuerda que estas muestras están formadas por personas que voluntariamente acudieron a una carpa con el reclamo de un panel que decía: "Sabéis cómo afecta el colesterol en su corazón?". Por tanto, las muestras no son representativas de la

población, pero en este caso actuaremos como si hubieran sido obtenidas de forma correcta, es decir, utilizando un método de muestreo probabilístico.

Recordatorio Leer Ficheros

Podemos leer los archivos a través del *paquete read_xlsx* utilizado en prácticas anteriores.

```
colhombres = read.csv(file = "colhombres.csv", header = TRUE, sep = ";", dec = ",")
```

```
colmujeres = read.csv(file = "colmujeres.csv", header = TRUE, sep = ";", dec = ",")
```

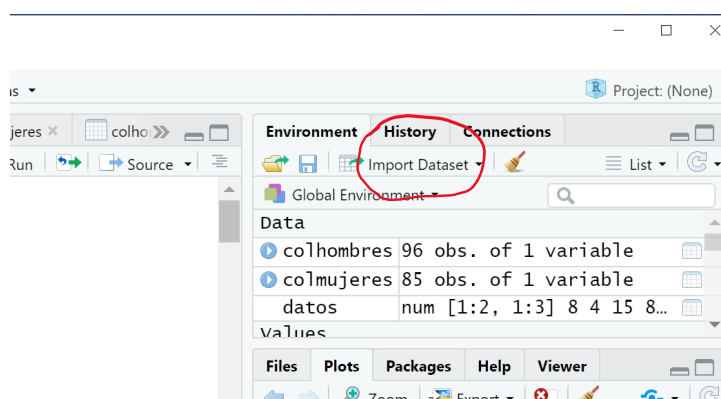
```
View(colhombres)
```

```
View(colmujeres)
```

```
attach(colhombres)
```

```
attach(colmujeres)
```

En esta práctica podemos aprovechar para leer los datos de otra manera, utilizando el **Import Data Set** que tenéis en la parte derecha de Rstudio (ver Figura).



Una vez lo tenemos importado, hacemos attach para acceder a las columnas:

```
attach(colhombres)
```

```
attach(colmujeres)
```

Cálculo e interpretación Intervalos de Confianza

Utilizando la función `t.test()`, tal y como se muestra a continuación

```
t.test(x, conf.level = 1-alpha)$conf.int
```

calcular con R los siguientes intervalos de confianza:

(a) Intervalo de confianza al 95% para μ_H , el colesterol total medio en sangre para la población de hombres:

- (b) Intervalo de confianza al 95% para μ_M , el colesterol total medio en sangre para la población de mujeres:

- (c) Justifica si la siguiente afirmación es correcta: "El 95% de los hombres y el 95% de las mujeres de la población estudiada tendrá valores de colesterol total en sangre dentro de los límites que definen los respectivos intervalos de confianza que acabamos de calcular".

- (d) ¿Se Incumple alguna de las condiciones necesarias para que estos intervalos sean válidos? En caso afirmativo, cuál/cuáles.

- (e) Un grupo de alumnos propuso aumentar el nivel de confianza de los intervalos al 99%, todo dando como argumento que, de este modo, se obtendrían estimaciones más precisas. Indica si este razonamiento es correcto, y por qué.

PISTA: Puedes calcular los intervalos en R, antes de contestar la pregunta.

Cálculo e interpretación Contraste de Hipótesis dos poblaciones independientes

- (f) Plantea un contraste de hipótesis para comprobar si el colesterol medio en sangre es el mismo en la población de hombres y de mujeres. Plantea las hipótesis del contraste.

Previamente puedes ejecutar el test de comparación de varianzas para verificar si puedes admitir varianzas iguales (`var.test()`)

Comandos de R para realizar ambos contrastes:

Igualdad de Varianzas: `var.test(x, y, conf.level = 1-alpha, alt = "two.sided")`

Igualdad de Medias: `t.test(x, y, conf.level = 1-alpha, alt = "two.sided")`

(g) Explica de forma clara las conclusiones de este estudio.

Ejercicio 2

Para realizar este ejercicio recordar los siguientes comandos de R:

`summary(x); mean(x); sd(x); hist(x); boxplot(x, horizontal = TRUE); qqnorm(x)`

El archivo 'stocks.xlsx' es una muestra aleatoria de 30 valores del NASDAQ (la bolsa de comercio electrónico más grande de EE.UU.), y recoge información sobre la variación (en porcentaje) que experimentaron estos títulos durante la 2ª semana del mes de octubre.

El NASDAQ cuenta con más de 3.300 compañías y en él cotizan más de 7.000 valores de mediana y pequeña capitalización.

(a) Identifica la población, la muestra y la variable aleatoria implicadas en este estudio.

(b) Realiza un análisis descriptivo de los datos e indica las conclusiones más importantes.

(c) Por término medio, ¿tipo de variación (positiva o negativa) se obtuvo en este mercado durante las citadas semanas? Obtén un intervalo de confianza al 95% para la variación media que experimentaron los títulos operados por este mercado e interpreta correctamente el significado.

Intervalo:

Interpretación:

(d) ¿Se cumplen las condiciones necesarias para obtener de manera correcta el intervalo de confianza anterior? ¿Por qué?