

PRÁCTICAS Nº 2 Y Nº3

Estadística descriptiva. Ejercicios

EJERCICIOS DE CLASE

Ejercicio 1. Importación de datos

- Importar el fichero "*Práctica de ordenador nº2 (R).xlsx*" publicado en eGela y asignar a la variable datos la información contenida
- Analizar la estructura de la variable datos
- Guardar el fichero *Excel* anterior como *Texto (delimitado por tabulaciones)* con el nombre datos2.txt e importar dicho fichero
- Analizar la estructura de la variable datos2

Ejercicio 2. Datos categóricos

Se considera la variable Grado del *data frame* datos.

- Obtener la tabla de frecuencias absolutas y relativas para las tres diferentes categorías de la variable Grado: "Mec", "Elec" y "EAut"
- Ampliar la tabla de frecuencias anterior con las frecuencias acumuladas (absoluta y relativa)
- Trazar un diagrama de barras de frecuencias absolutas y otro de frecuencias relativas
- Trazar un diagrama de barras de frecuencias absolutas acumuladas y otro de frecuencias relativas acumuladas
- Representar las frecuencias absolutas de la variable en un diagrama de sectores
- Trazar un polígono de frecuencias absolutas y otro de frecuencias relativas

Ejercicio 3. Datos numéricos

Se considera la variable Altura del *data frame* datos.

- Determinar el número de intervalos de clase y su amplitud
- Obtener la tabla de frecuencias con los valores de la variable agrupados en intervalos de clase (se aconseja ordenar previamente los valores)
- Representar un diagrama de tallo y hojas
- Trazar el histograma de las frecuencias absolutas de los valores de la variable agrupados en intervalos de clase
- Trazar el histograma de las frecuencias relativas de los valores de la variable agrupados en intervalos de clase
- Trazar un polígono de frecuencias absolutas y otro de frecuencias relativas

Ejercicio 4. Estadísticos de posición central

Se consideran las variables `Altura`, `Peso` y `Edad` del *data frame* `datos`.

- a) Obtener las medias de las variables
 - b) Repetir el apartado anterior haciendo uso de la función `apply()`
 - c) Calcular de nuevo las tres medias anteriores eliminando el 5% de los valores más extremos de cada una de las variables
 - d) Obtener las medianas de las variables
 - e) Repetir el apartado anterior haciendo uso de la función `apply()`
 - f) Obtener la moda de las variables `Altura` (cuantitativa) y `Población` (cualitativa)
-

Ejercicio 5. Estadísticos de dispersión

Se consideran las variables `Altura`, `Peso` y `Edad` del *data frame* `datos`.

- a) Obtener el recorrido de las variables
 - b) Calcular su varianza
 - c) Determinar su desviación típica
 - d) Calcular el coeficiente de variación de Pearson, ¿cuál de las tres variables presenta mayor dispersión?
-

Ejercicio 6. Estadísticos de posición no central

Se considera la variable `Altura` del *data frame* `datos`.

- a) Calcular los siguientes cuantiles: $C_{0.10}$, $C_{0.25}$, $C_{0.50}$, $C_{0.75}$
 - b) Comparar los resultados obtenidos con la tabla de frecuencias de la variable
 - c) Obtener todos los deciles
 - d) Hallar el rango intercuartílico
-

Ejercicio 7. Números resumen

Se consideran las variables `Altura`, `Peso` y `Edad` del *data frame* `datos`.

- a) Determinar los cinco números resumen de las variables junto con su media
 - b) Obtener los números resumen de la variable `Altura` clasificados según los niveles de la variable `Grado`
-

Ejercicio 8. Diagramas de caja

Se considera la variable `Altura` del *data frame* `datos`.

- a) Trazar un diagrama de caja básico
 - b) Repetir el apartado anterior dando color a la caja y añadiendo algún título al gráfico
 - c) Etiquetar en el gráfico los valores representados en el diagrama (hacer uso de la información adicional que facilita `boxplot()`)
 - d) Representar el *boxplot* en forma horizontal
 - e) Trazar los diagramas de caja para la variable numérica `Altura` clasificados según los niveles del factor `Grado`
-

Ejercicio 9. Estadísticos de forma

Se consideran las variables `Altura`, `Peso` y `Edad` del *data frame* `datos`.

- a) Determinar la simetría de sus distribuciones
 - b) Obtener la curtosis de sus distribuciones
 - c) Comparar los resultados anteriores con las gráficas de densidad de las variables
-

Ejercicio 10. Eliminación de datos atípicos

Se considera la variable `Edad` del *data frame* `datos`. Se ha detectado la introducción errónea de un dato ($x_i = 10$) que resulta atípico.

- a) Hallar la posición del dato
 - b) Asignar a la variable `Edad2` los datos de `Edad` eliminando el valor erróneo
 - c) Realizar un estudio descriptivo completo de `Edad2`
-

EJERCICIOS COMPLEMENTARIOS

Ejercicio 11

Se utiliza una variable, llamada `grietas`, para almacenar las longitudes, en micras, de noventa grietas medidas en una estructura de hormigón. Se pide:

- a) usando la función `sample()`, generar de forma aleatoria los noventa valores de la variable `grietas` de forma que pertenezcan al intervalo $[30, 120]$
 - b) calcular la media, la mediana y la desviación típica
 - c) agrupar los datos en intervalos de clase con amplitud constante, $A = 10$
 - d) presentar una tabla de frecuencias con los intervalos y las marcas de clase, las frecuencias absolutas y relativas y las correspondientes frecuencias acumuladas
 - e) trazar el histograma de frecuencias absolutas
 - f) añadir a la muestra un valor mayor que 120 que resulte atípico, añadir el intervalo de clase $[120, \infty]$ y trazar el *boxplot*
-

Ejercicio 12

Se considera el marco de datos `mtcars` del paquete `datasets`. Obtener un vector que contenga el tiempo que cada modelo de vehículo tarda en recorrer un kilómetro partiendo de cero. Hacer un estudio descriptivo de esta variable.

Ejercicio13

El siguiente conjunto de datos refleja las victorias, empates y derrotas (denotadas, respectivamente, como 1, X y 2) que un equipo ha obtenido en los 19 partidos que ha jugado en su estadio durante una temporada: $\{2, X, 1, 1, 1, 1, X, 1, 1, X, X, 2, 1, X, 2, 1, 1, 1\}$.

Se pide:

- a) obtener una tabla de frecuencias
 - b) representar gráficamente los datos de dos maneras distintas, usando un diagrama de barras y un diagrama de sectores
-

Ejercicio 14

Se facilita la siguiente tabla de frecuencias en la que los datos de la variable observada se encuentran agrupados en intervalos de clase:

<i>Intervalos</i>	x_i	n_i
[10.5,21.5)	16	6
[21.5,32.5)	27	8
[32.5,43.5)	38	7
[43.5,54.5)	49	17
[54.5,65.5)	60	18
[65.5,76.5)	71	10
[76.5,87.5)	82	3
[87.5,98.5]	93	3

Se pide:

- calcular los estadísticos de centralización y dispersión presentados en el guión de la práctica
 - obtener los cinco números resumen y el rango intercuartílico
 - trazar el polígono de frecuencias
-

Ejercicio15

Los siguientes valores son los precios, en miles de euros, de las viviendas vendidas en una ciudad en el último mes: 311, 241, 180, 163, 154, 139, 173, 168, 129, 680, 224, 425, 333, 245, 410, 489, 438, 269, 151, 321, 330, 400, 312, 266, 388, 299, 425, 323, 291, 611, 332, 313, 299, 300, 313, 222, 181, 155, 200, 155.

Se pide:

- representar el histograma
 - trazar el diagrama de caja
 - detectar datos atípicos
 - calcular la media, mediana, desviación típica y recorrido intercuartílico antes y después de eliminar los datos atípicos, en caso de existir
-

Ejercicio16

Un encuestador, entre otras cuestiones, ha preguntado por la edad de las personas entrevistadas. Se han obtenido los siguientes datos: 18, 24, 18, 63, 54, 39, 73, 68, 29, 43, 72, 50, 33, 45, 56, 48, 53, 56, 51, 21, 70, 51, 66, 38, 39, 44, 52, 61, 32, 22.

Se pide:

- generar una tabla de frecuencias sin agrupar los datos y otra con los datos agrupados en intervalos de clase
 - representar gráficamente las tablas anteriores
 - obtener el diagrama de tallos y hojas
 - representar el histograma por defecto y compararlo con el obtenido anteriormente
-