

Práctica 1 MEE - Análisis exploratorio de datos

Nota: todos los resultados de esta práctica deben redondearse correctamente utilizando **2 cifras decimales**.

1. Utiliza R a modo de calculadora para obtener:

- $\sqrt{(1+2^2+3^3+4^4+5^5)} =$ _____ $\sqrt{5^9} + \sqrt{12^{11}} =$ _____
- $\log_2(2^1+2^2+2^3+\dots+2^{100}) =$ _____ $\ln(e^{1/3})+\log_5(15625) =$ _____
- $e^1 + e^{1/3} + e^{1/5} + e^{1/7} + \dots + e^{1/299} =$ _____

2. Según la revista *The Digital Bits*, las cifras de ventas mensuales (en miles de dólares) de reproductores de DVD en 2003 fueron:

ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
790	740	1610	1270	1330	2100	990	1430	2490	2490	3680	3020

Crea un vector llamado `ventas.dvd` con estos datos y calcula:

- Obtén un diagrama de caja de estas cifras. ¿La distribución es simétrica? _____
¿Existen valores aislados? _____
- La media _____ y la mediana _____ de la cifra de ventas.
- La desviación típica de la cifra de ventas: _____ (miles de dólares)

3. Maratón de NY 2002.

El archivo `nym2002.csv` contiene datos de una muestra de corredores que completó la maratón de Nueva York (42 km) en el año 2002. Consideraremos que esta muestra representa a la población de corredores que finalizaron esta maratón.

Utiliza las opciones de menú del programa RSTUDIO para importar estos datos.

Realizar los análisis que corresponda para responder las siguientes preguntas.

- ¿Qué nombre reciben las variables de este conjunto de datos?

- ¿Cuántos hombres hay en la muestra? _____ ¿Cuántas mujeres? _____

- En la muestra, ¿cuántos corredores son del estado de NY? _____

- En la muestra, ¿cuántos corredores son españoles (ESP)? _____

¿Cuántos corredores españoles son hombres? _____

¿Cuántos corredores españoles son mujeres? _____

- En la muestra, ¿cuántos hombres finalizaron la carrera en menos de 3 horas? _____

- En la muestra, ¿cuántas mujeres finalizaron la carrera en menos 3 horas? _____

- ¿Cuánto tiempo (en minutos) necesitó el español más rápido de la muestra para completar la carrera? _____

- ¿En qué posición finalizó el español más rápido de la muestra? _____

- En la muestra, ¿qué equipo de corredores masculino fue en promedio más rápido: el equipo español (ESP) o el francés (FRA)? _____

- En la muestra, ¿qué equipo masculino presenta menor dispersión (variabilidad) en cuanto a los tiempos de finalización de la carrera: el español (ESP) o el francés (FRA)? _____

- ¿Qué tipo de distribución presenta la variable *tiempo que tardan los hombres en finalizar la carrera*? ¿Podemos asumir que estos datos siguen una distribución normal?

- ¿Qué relación existe entre el tiempo que se tarda en finalizar la carrera y el sexo de los corredores?

- ¿Qué relación existe entre el tiempo que se tarda en finalizar la carrera y la edad de los corredores?

¿Existe algún punto en el diagrama de dispersión que se aparte del resto de los puntos?
¿Qué característica sorprendente presenta este punto?

Coef. de correlación lineal entre estas variables: _____

Conclusión: _____