

Examen parcial

Estadística para Ciencia de datos

Alumno 1: _____

Alumno 2: _____

En este ejercicio trabajaremos con los siguientes conjuntos de datos

Florida, del paquete carData.

UScereal, del paquete MASS

Vitamin3.csv

starwarsSmall.csv

mtcars

1.-Del conjunto de datos starwarsSmall, describe que tipo de dato son las siguientes columnas:

Height,

mass,

hair_color,

gender

birth_year

Especifica entre numéricos continuos y discretos, y entre categóricos nominales, ordinales y binarios

2.-calcula la tabla de frecuencias para la variable eye_color de StarwarsSmall, y crea su gráfica de barras, y despliega el nombre de todas las categorías que tienen la menor cantidad de repeticiones en ambos casos.

- 3.-Grafica el histograma de las variables protein, sugars y potassium de UScereal, usando 15 bins. Encuentra sus medidas de tendencia central, media, moda y mediana.
- 4.-Crea la gráfica de tallo y hoja de las variables protein de UScereal, y effort de Vitammin3. Calcula su varianza y desviación standard y explica si sus valores coinciden con la grafica de tallo y hoja y por qué. Explica que nos dicen las medidas de dispersión acerca de los datos.
- 5.-Calcula el coeficiente de variación para las variables protein de UScereal, y effort de Vitammin3. Y explica si significa una gran variación en los datos, o muy poca variación, explica por qué.
- 6.-De la variable effort de vitamin3, Crea una tabla que contenga su frecuencia, frecuencia acumulada, frecuencia relativa y frecuencia acumulada relativa. Despliega la tabla
- 7.-Usando la fórmula de cuantiles, encuentra el cuantil que le corresponde al 33% y al 67% de los datos del ejercicio anterior, comprueba el resultado usando la función de R de cuantiles.
- 8.-Del conjunto de datos de Florida, calcula el rango IQR para las columnas de Bush, Gore y Moorehead, despliega las gráficas de caja para cada uno de ellos y encuentra los outliers de cada columna. Explica que significa cada gráfica usando la relación entre el IQR, los cuartiles, la mediana y la cantidad de outliers en los datos.
- 9.-Encuentra los valores de skewness y kurtosis, y su histograma para las variables protein de UScereal, y effort de Vitammin3, explica el significado de los valores, y utiliza el histograma de los datos para apoyar tu explicación.
- 10.-Del conjunto de datos mtcars, crea la gráfica de dispersión entre hp y carb, y la grafica entre wt y mpg. Encuentra la covariancia entre cada par de variables y su coeficiente de correlación. Explica si el resultado coincide o no con la gráfica de dispersión. Describe que pasa con la variable carb cuando hp incrementa su valor, y que pasa con la variable wt cuando mpg incrementa su valor. Como describirías la relación entre cada par de variables.