Empezamos con el primer paso de la descripción o exploración de variables cuantitativas o medidas.

Durante todo el bloque vamos a trabajar con la exploración.

Yo te propongo una tabla de datos en cuestión, pero tú puedes aplicar esta misma hoja de trabajo con tu propia tabla de datos.

Te doy opciones de tablas de datos para que puedas practicar:

[>> Repositorio de datasets muy muy interesantes](https://archive.ics.uci.edu/ml/datasets.html)

# Abre la tabla de datos “espalda”

Si no quieres pensar en base de datos vamos a seguir trabajando con la tabla de datos espalda.

Más que nada para seguir un orden en la aplicación de la metodología ☺

Abre la tabla de datos “espalda.xlsx” y crea una nueva variable llamada diff\_ODI. Calcula esta variable de esta forma:

* diff\_ODI = ODI\_mes0 – ODI\_mes1

Esta variable será reflejo de la mejora de los pacientes.

Recuerda que ODI es sinónimo de “lo mal” que está el paciente por el dolor de espalda. Y con diff\_ODI calculamos la mejora de este paciente.

## Explora todas las variables cuantitativas (medidas) de la tabla de datos

Para ayudarte te he creado una guía con los resultados más interesantes para la descripción de estas variables cuantitativas.

Repite este flujo de trabajo para cada una de ellas.

Empieza por la edad por ejemplo.

## Estadísticos – resumen numérico

|  |
| --- |
| Copia la tabla de estadísticos de centralidad, dispersión y forma. Inspírate de los que calculo en la zona Tech Zen.  HIPOTECARIOS Y CONSUMO PERSONALES |
|  |

¿Qué conclusiones obtienes?

Piensa en:

* Centralidad
* Dispersión
* Simetría

Imagínate la distribución dibujada.

|  |
| --- |
| Lista las conclusiones que has podido observar con los índices  Viendo la desviación standard en los tres casos, tiene bastante dispersión. Lo saldos están bastante desplazados de su media  Viendo la asimetría que es positiva pero > 1 lo que indica que esta sesgada hacia la izquierda de la media.  Viendo los Q1 y Q3 para el caso de los créditos hipotecarios el monto máximo de saldos de los créditos es menor a 38K y esto implica una muy baja exposición por lo atomizada que esta. |

## Visualiza la distribución

El siguiente paso es darle forma a la distribución con la ayuda de:

* Histograma o histograma de densidad
* Boxplot
* Qqplot

|  |
| --- |
| Copia los gráficos de la distribución (histograma, boxplot y qqplot)        Hipotecas Personales |

¿Qué conclusiones obtienes?

Piensa en:

* Forma de la distribución
  + ¿Es centrada la distribución?
  + ¿Es dispersa?
  + ¿Dónde se concentran los valores?
  + ¿Hay grupos de puntos separados o es uniforme?
* ¿Es parecida a una distribución normal o es asimétrica? ¿Hacia dónde?
* ¿Observas valores atípicos con el boxplot?

|  |
| --- |
| Lista las conclusiones que has podido observar con los gráficos de distribución  En ninguno de los casos parece una distribución Normal es asimétrica.  No está centrada, esta sesgada a la izquierda.  Si hay valores atípicos, sobre todo, en CFC, SAL, Pan en la parte del personal.  Si hay valores atípicos, muchos diríamos. |

## Repite este procedimiento para todas las variables cuantitativas:

* Edad
* Altura
* Peso
* ODI mes0
* ODI mes1
* Diff\_ODI

Resume las conclusiones utilizando esta guía.

## [Para los que quieren más]: Compara por grupos de tratamiento

Si quieres darle más sentido al ejercicio te invito a que compares las medidas por grupos de tratamiento (variable, Grupo)

De esta manera gráficamente podemos comparar los dos grupos de tratamiento.

¡A por ello!

Nota: este tipo de ejercicios lo aprenderás a hacer paso a paso en siguientes lecciones. Pero es mejor que lo intentes antes ;)

¡Enhorabuena por el trabajo!

¡Sigue avanzando como lo estás haciendo!