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

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.

## Describe las relaciones entre medidas – primer paso

Utiliza todas las variables cuantitativas a la vez. Describe su relación por parejas.

Utiliza las herramientas que te propongo

## Dibuja el matrixplot y la matriz de correlación

|  |
| --- |
| Copia el matrixplot y la matriz de correlación entre todas las  variables      El Matriz Plot muestra que hay posibles asociaciones entre: Grupo-diff\_odi, diff\_odi – ODIMes1, Grupo-OdiMes1, diff\_odi-LC, ODIMes0-LC, Altura Peso, Edad-LC, Edad-OdiMes0, Edad-LC, LC-OdiMes0  ,  Matriz de Coorrelación  Matrix p-values    Otra forma de Ver la Coorrelacion por Parejas y el P-Valor |

## ¿Qué conclusiones obtienes?

Identifica las relaciones por parejas.

* ¿Dónde ves relaciones importantes? Observa el matrixplot y los valores de correlación.
* ¿Cuáles son lineales y cuáles no? Observa la forma de los diagramas de puntos
* ¿Qué parejas de medidas están relacionadas? Observa el p-valor también

|  |
| --- |
| Lista las conclusiones que has podido observar con los índices  Peso-Altura (Directa)  Edad-LC. (directa)  Edad-NHD. (No Clara pero directa)  LC-NHD. (directa)  NHD – OdiMes0(directa)  Edad-OdiMes0(directa)  LC-OdiMes0(directa)  LC-DiffOdi(directa) |

## Diagrama de dispersión de las variables con mayor relación (segundo paso)

¡Bien!

Ahora que has identificado las parejas con mayor relación ya las puedes visualizar más de cerca.

Utiliza el diagrama de dispersión y la línea de regresión o tendencia.

![Table

Description automatically generated]()Selecciona las parejas con mayor relación y dibuja el matrixplot mixto. Te ayudará a entender mejor los resultados.

Como se observa Ahora están mas claramente las correlaciones. El mas evidente dice que hay asociación entre mejoría (diff\_oddi) y el grupo, Peso y Altura, Edad y LC, LC y Mejoria (diff oddi)

## Dibuja el diagrama de dispersión de cada pareja o el matrixplot mixto

|  |
| --- |
| Es bueno que pintes la línea de tendencia lineal para entender si la relación es o no lineal. Con el diagrama mixto lo puedes hacer en R  pairs(cconti, data=cconti,  lower.panel=panel.smooth, upper.panel=panel.cor,  pch=20, main="Matrix Plot Mixto")    #Diagrama de Corrleacion que coloca SCATTERPLOT, HISTOGRAMAS Y COEF CORRELACION..  chart.Correlation(cconti,histogram = TRUE,pch=19,method = "spearman") |

## ¿Qué conclusiones obtienes?

Visualiza las relaciones con un poco de cariño identificando:

* Valores atípicos al resto (que están bastante alejados de la línea de tendencia. Intente explicarlo
* Observa si tienes grupos de puntos diferenciados y porque puede ser. (a veces es por un factor
* ¿De las relaciones no lineales que has identifica son no lineales realmente en el detalle?

|  |
| --- |
| Lista las conclusiones que has podido observar con los índices  A priori se puede observar que NHD, Grupo y LC muestran que hay como subgrupos dentro de ellos, ya que los scatter plot se notan agrupaciones de puntos en un mismo plot. Pareciera que existiesen algún tipo de agrupación dentro de estos.  Parece que hay Factores |

## Correlación por factores de estudio

Ahora que has obtenido mucha información de los patrones de asociación, te toca ir un paso más allá.

Identifica las variables de estudio que pueden influir en las relaciones. Es decir, las variables cualitativas que pueden influir en la correlación.

En nuestro caso nos interesa ver en relación al número de hernias discales.

Analiza la correlación con lo que has aprendido en los pasos anteriores poniendo foco en el factor de estudio NHD.

¡A por ello!

* Dibuja el matrixplot con el factor de estudio
* Visualiza la línea de regresión por factor de estudio para poder comparar relaciones y observar si influye el factor en esta relación
* Puedes calcular también la correlación diferenciando por factor de estudio

|  |
| --- |
| Copia el matrixplot por factores, el diagrama de dispersión por factores y las líneas de tendencia linear para observar si este factor influye en las relaciones    En este matrix plot agrupado por Lumbociatica y se nota que claramente el factor lumbociatica esta claramente definido en la matriz. Parece haber una asociación entre Edad y LC y es directa.  Entre LC y la mejoría.  Aparente asociación entre Grupo y Diff\_Odi y se nota que el factor de si hay o no LC si se nota en el matrix plot    También se notan los grupos (factores) de No de Hernias . |

## Repite este procedimiento para los factores que creas importantes:

Recuerda:

* Dibuja el matrixplot con el factor de estudio
* Visualiza la línea de regresión por factor de estudio para poder comparar relaciones y observar si influye el factor en esta relación
* Puedes calcular también la correlación diferenciando por factor de estudio
* Compara las líneas de tendencia y el valor de la correlación por factor ☺

¡Bien!

Ahora ya sabes muy bien cómo ordenar la descripción de la correlación.

¡Lo puedes implementar en cualquier contexto!