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 cualitativas por separado

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

Repite este flujo de trabajo para cada una de ellas.

Empieza por la variable sexo, por ejemplo.

## Tabla de frecuencias absolutas y porcentajes y el diagrama de barras y sectores

|  |
| --- |
| Copia la tabla de frecuencias y el gráfico de barras o sectores  TABLA DE FRECUENCIAS DE RIESGO POR EMPLEADOR 1  tbfreqEmpl1 <- tab1(df$RiesgoEmp1,cum.percent = TRUE,sort.group = "decreasing",main="Distribución Riesgo Empleador 1")        TABLA DE FRENCUENCIAS DE RIESGO EMPLEADOR 2  tbfreqEmpl2 <- tab1(df$RiesgoEmp2,cum.percent = TRUE,sort.group = "decreasing",main="Distribución Riesgo Empleador 2") |

## ¿Qué conclusiones obtienes?

Piensa en:

* Repartición de las proporciones por grupos
* Fíjate dónde tienes más observaciones y dónde no

Imagínate la distribución dibujada.

|  |
| --- |
| Lista las conclusiones que has podido observar con los índices |

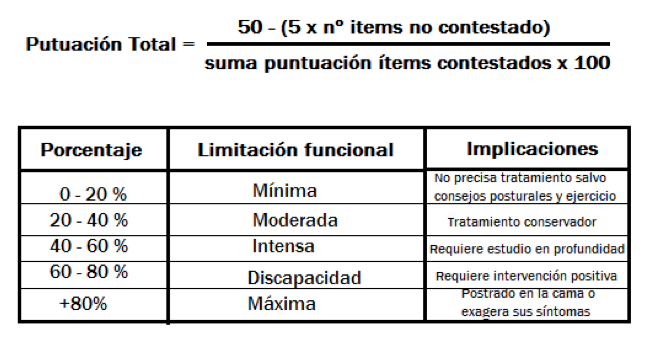
## Repite este procedimiento para todas las variables cualitativas:

* LC
* PC
* Grupo (tratamiento)

Resume las conclusiones utilizando esta guía.

## Describe por parejas de factores – 2 variables cualitativas

Por ejemplo, en este caso es interesante quizá observar dos variables cualitativas. La limitación funcional (variable ordinal) calculado a partir de ODI de esta forma:



Crear una variable ordinal llamada LF\_mes0:

* Donde de 0-20% de ODI mes 0 sea el grupo mínima.
* Donde de 20-40% de ODI mes 0 sea el grupo moderada
* Donde de 40-60% de ODI mes 0 sea el grupo intensa.
* Donde de 60-80% de ODI mes 0 sea el grupo discapacidad
* Donde de 80-100% de ODI mes 0 sea el grupo máxima.

Al final tendrás dos variables más llamadas:

* LF\_mes0 (ordinal a partir de ODI\_mes0)
* LF\_mes1 (ordinal a partir de ODI\_mes1)

Lo que vamos a hacer es describir los dos factores LF\_mes0 y LF\_mes1.

## Crea las tablas cruzadas o de contingencia de los dos factores.

El factor de estudio será el tiempo así que en columnas pondrás LF\_mes0 y LF\_mes1.

En cambio los grupos del nivel de limitación funcional son las columnas

|  |
| --- |
| Copia las tablas de contingencia de frecuencias absolutas, relativas y condicionadas por columnas (al grupo de estudio)   * Riesgo Empleado1 vs Riesgo Empleado 2 * Tabla de Contingencias DOS FACTORES |

## Crea los gráficos de barras de dos factores para visualizar las tablas

|  |
| --- |
| Copia las tablas de contingencia de frecuencias absolutas, relativas y condicionadas por columnas (al grupo de estudio)  MOSAIC PLOT PARA VARIABLES CATEGORICAS   * Riesgo Emp1 vs Riesgo Emp2     Riesgo Emp2  Riesgo Emp1 |

## ¿Qué conclusiones obtienes?

Piensa en:

* Compara los dos factores de estudio
* La parejas de grupos con mayor importancia
* ¿Están ordenadas de menor a mayor las frecuencias condicionadas, o no?

|  |
| --- |
| Lista las conclusiones que has podido observar con los gráficos de distribución |

## Repite este procedimiento para parejas de factores:

* LF\_mes0 vs NHD
* NHD vs LC

El procedimiento es:

* Identifica el factor de estudio
* Calcula la tabla de contingencias (el factor de estudio en columnas)
* Calcula las frecuencias condicionadas al grupo de estudio (por columnas)
* Dibuja el diagrama de barras de dos factores
  + Frecuencias absolutas
  + Frecuencias condicionadas
* Resume las conclusiones

El procedimiento que estamos siguiendo en esta hoja de trabajo es genérico.

Lo puedes aplicar siempre que quieras a tus variables cualitativas.

¡Buen trabajo!

A seguir avanzando como lo estás haciendo ☺