Vamos a seguir con el caso práctico.

Ahora que nos hemos situado, hemos entendido la tabla de datos y las variables.

Es momento de pasar a una fase extremadamente importante, la descripción de los datos o la exploración.

Y es precisamente lo que has hecho sin darte cuenta en el primer ejercicio. Has creado gráficos y los has interpretado. Es precisamente la exploración de datos.

Vamos a describir dos situaciones:

1. Una comparación de medias con SPSS
2. Una exploración de la correlación con Excel

*Aviso: el ejercicio lo vamos a hacer en Excel y SPSS. Pero puedes utilizar RStudio, RCommander o cualquier otro software. No hay ningún problema.*

*El ejercicio trata de entender conceptos estadísticos de forma práctica y entender un método de trabajo ☺*

*Puedes conseguir una copia de prueba SPSS durante un mes en esta página:*

[*https://www.ibm.com/analytics/es/es/technology/spss/spss-trials.html*](https://www.ibm.com/analytics/es/es/technology/spss/spss-trials.html)

*O utilizar PSPP que es el software libre y gratuito casi igual que SPSS:*

[*https://www.gnu.org/software/pspp/*](https://www.gnu.org/software/pspp/)

*Excel supongo que lo tendrás instalado ☺*

¡A por ello!

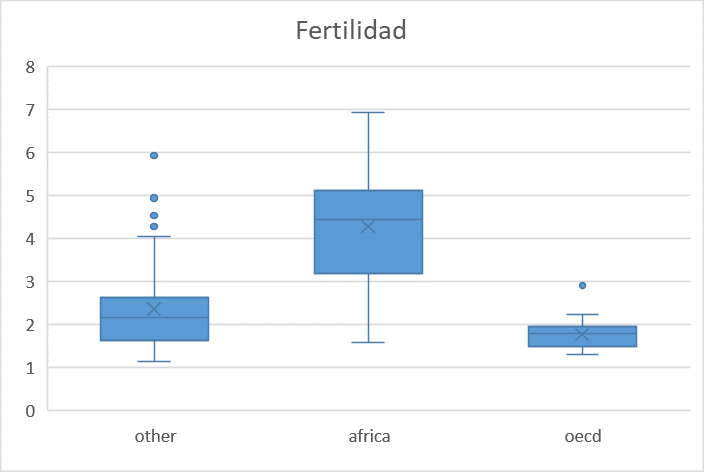
# Explorando la comparación de medias con SPSS

En el primer ejercicio creaste unos boxplots que nos permiten comparar distribuciones de forma sencilla y fácil.

Te recuerdo el gráfico que creaste.

Fíjate que es de gran ayuda este gráfico. Nos situar muy bien las distribuciones y las podemos comparar.

Aun así, podemos ir un paso más allá. Calculando características de la centralidad y la dispersión.



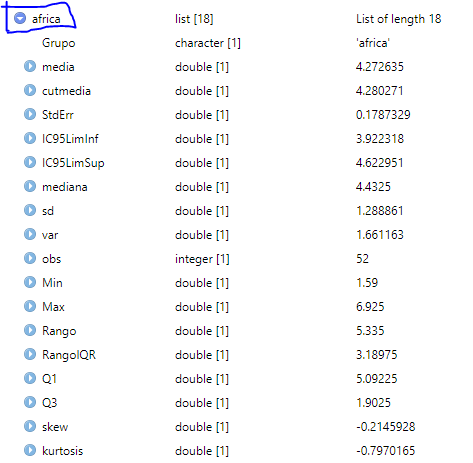
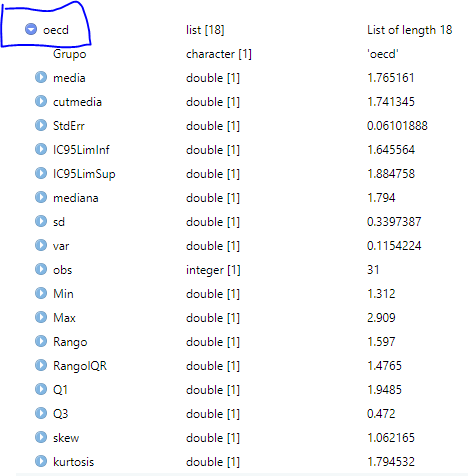
El objetivo del ejercicio era poder comparar los grupos según la fertilidad, la mortalidad infantil la esperanza de vida.

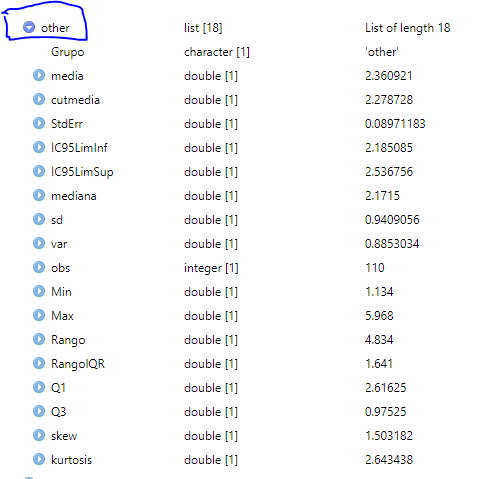
Ahora quiero que completes estos boxplots con resultados descriptivos (de centralidad y de dispersión). Más info aquí:

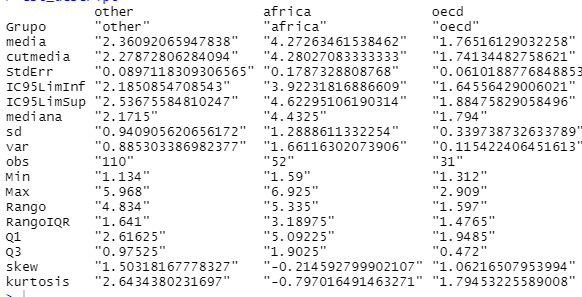
<https://conceptosclaros.com/pro-statistics-leccion-5/>

<https://conceptosclaros.com/pro-statistics-leccion-6/>

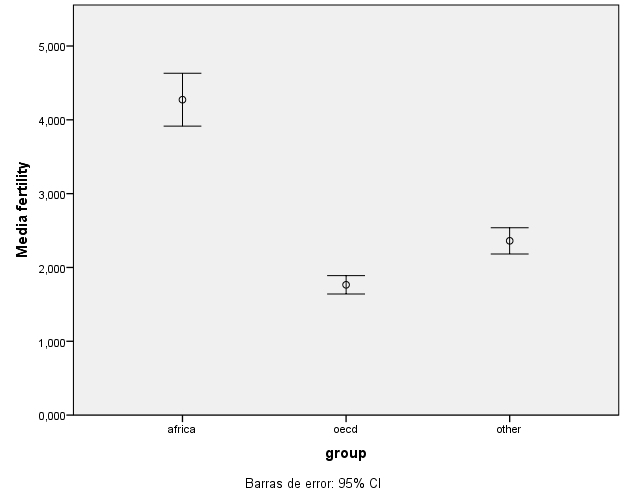
Rellena la tabla siguiente con los resultados que obtengas. Sigue los tutoriales tech zen para poderlo hacer:







Crea un diagrama de error o diagrama de medias para poder comparar el valor central y la dispersión por grupos. Más abajo un ejemplo:



Sigue los tutoriales Tech Zen de la página del ejercicio

|  |
| --- |
| Copia las tablas de comparación de descriptivos y los diagramas de error que has obtenido de la variable PIB per cápita ppgdp  Fertility por Group (Diagramas de Error)  Me parece que hay grandes diferencias entre el grupo África y los demás… Entre oecd y other hay pero no son tan evidentes a-priori    PIB por grupo |

¿Qué conclusiones puedes obtener?

|  |
| --- |
| PIB VS GRUPOS  A nivel del PIB por grupo se observa diferencias de PIB marcadas entre ocde y el resto (áfrica, other), Entre other y áfrica hay diferencias también ya que la media de other es mayor a la de áfrica. |

Repite el ejercicio para las otras variables: lifeExpF, pctUrban, infantMortality en función del grupo de OECD o no.

|  |
| --- |
| Copia las tablas de comparación de descriptivos y los diagramas de error para el resto de variables.  lifeExpF por Grupo.    Esperanza de vida por grupo veo que Africa esta muy seperada negativamente de Otros y Ocde..Se nota que hay diferencias ocde y Africa. Donde no parece haber muchas diferncias es entre ocde y otros          Se observan claras diferencias de población urbana de ocde y lo de otros y áfrica. |

¿Qué conclusiones puedes obtener?

|  |
| --- |
| Infant Mortality muestra grandes diferencias en los tres grupos |

# Explorando la correlación con Excel

La correlación entre medidas es la relación o asociación de distintas variables. Aquí tienes más información:

<https://conceptosclaros.com/importancia-correlacion-estadistica/>

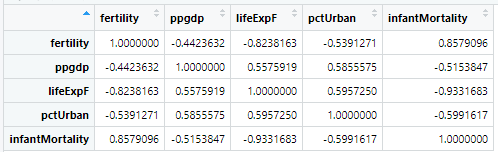
Dibuja los diagramas de dispersión con una línea de tendencia. Aquí un ejemplo:

Fíjate aquí que la línea de tendencia nos da información de la tendencia lineal de la relación.

Con Excel es sencillo calcularlo. Mira el vídeo Tech Zen.

|  |
| --- |
| Copia los diagramas de dispersión con la línea de tendencia de la mortalidad infantil en función también del resto de variables: ppgdp, pctUrban, fertilidad.  Son tres diagramas de dispersión con la línea de tendencia:   * Mortalidad infantil vs ppgdp * Mortalidad infantil vs pctUrban * Mortalidad infantil vs fertilidad |

Y calcula la correlación entre las medidas. Rellena esta tabla de coeficientes de correlación. Recuerda que es una matriz simétrica y la correlación es por parejas.



**Se observa una fuerte asociación**

**InfantMortality vs fertility (0.857) y tiene sentido positivo.**

**InftantMortality vs lifeExpf (-0.933) tiene sentido negativo.**

Sigue los tutoriales tech Zen para poder rellenar esta hoja de trabajo.

¡Enhorabuena!

Has terminado la exploración de dos objetivos que nos hemos planteado al inicio.

Ahora solo, queda un pasito más

Aplicar la estadística inferencial. Lo verás en los siguientes ejercicios.

¡Ciao ciao!