

DESCUBRIENDO LOS INTERVALOS DE CONFIANZA

El concepto de intervalo de confianza es muy utilizado en la etapa del análisis.

Por ejemplo, en la comparación de medias el intervalo de confianza es una gran herramienta fácil y efectiva.

O los coeficientes de la regresión lineal que calculaste en el pre training también pueden acompañarse con sus intervalos de confianza. (lo verás más adelante)

En esta hoja de trabajo vas a calcular los intervalos de confianza de las variables de la tabla de datos “espalda”.

Y de esta manera te vas a introducir en la comparación de medias.

¡Empezamos!

Abre la tabla de datos “espalda”

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

- $\text{diff_ODI} = \text{ODI_mes0} - \text{ODI_mes1}$

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.

Filtrando la variable por la variable “grupo”

Ahora se trata de utilizar los filtros de Excel para que consigas la variable diff_ODI por grupos de tratamiento.

Recuerda que la variable “Grupo” tiene dos posibles números:

- 0 = Tratamiento convencional
- 1 = Tratamiento innovador / avanzado

Puedes copiar la variable diff_ODI en dos columnas reflejando la información de cada tratamiento.

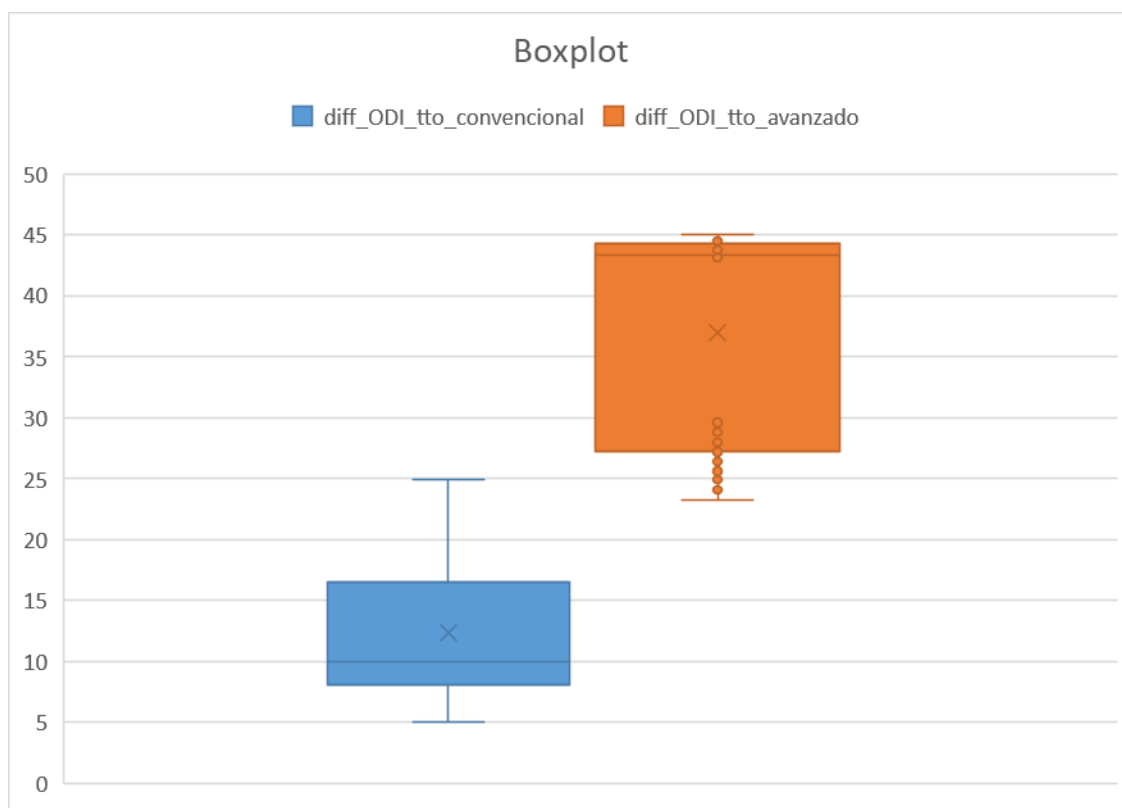
El encabezado te quedaría así:

diff_ODI	diff_ODI
Tratamiento convencional	Tratamiento avanzado

Describiendo la variable diff_ODI

Crea un boxplot o un histograma para visualizar las dos distribuciones y comparar los dos tratamientos gráficamente.

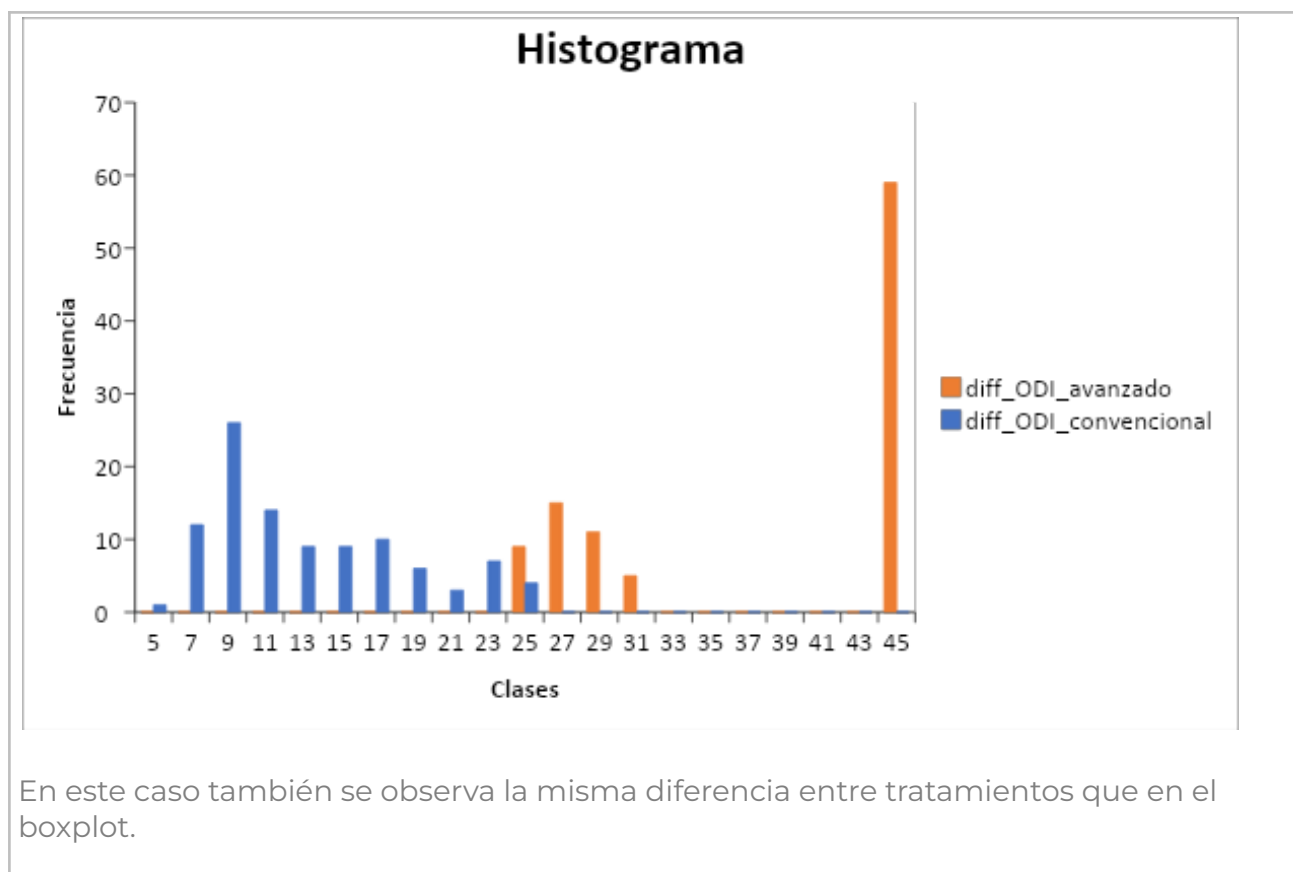
Copia el boxplot de los dos grupos o los dos histogramas. El que prefieras ☺



En el boxplot se observan diferencias entre los niveles de mejoría de los dos tratamientos. Donde el tratamiento avanzado parece ejercer un efecto más fuerte disminuyendo el dolor de espalda de los pacientes que el convencional.

A continuación se representan también los histogramas de ambas variables:

DESCUBRIENDO LOS INTERVALOS DE CONFIANZA



Calcula los intervalos de confianza

Después de la preparación que hemos hecho es momento de calcular los intervalos de confianza de la media al 95% y al 99%.

Rellena esta tabla para lograrlo.

"Grupo" Tratamientos	Media o Promedio	Desviación estándar	n: número de pacientes	Error Estándar EE	IC 95%		IC 99%	
					Inferior	Superior	Inferior	Superior
Tratamiento convencional	12.34	5.43	101.00	0.54	11.28	13.40	11.08	13.60
Tratamiento avanzado	36.97	8.72	99.00	0.88	35.25	38.69	34.93	39.01

Recuerda que para calcular los intervalos de confianza tienes las siguientes pistas:

$$EE = \frac{\text{Desv. Estándar}}{\sqrt{n}}$$

$$IC\ 95\% = Media \pm 1.96 \cdot EE$$

$$IC\ 99\% = Media \pm 2.33 \cdot EE$$

Ahora que ves los resultados de los intervalos de confianza. ¿Qué puedes decir de los tratamientos?

Pista: si los intervalos no se solapan entre tratamientos quiere decir que hay diferencias significativas al 95% o al 99% dependiendo de cuál mires.

DESCUBRIENDO LOS INTERVALOS DE CONFIANZA

Si te ves con fuerzas dibuja los IC con un diagrama (googlea: gráfico intervalos de confianza). Te ayudará a verlo gráficamente.

[Duda del foro]:

Tengo una duda sobre el último punto de la tarea. Cuando pregunta por la interpretación de los IC de las dos variables, por ejemplo si me fijo en la variable diff_ODI_avanzado su IC95% superior hace que excluyamos la mayoría de los datos (porque los datos están muy dispersos). He decidido poner el histograma para que vieras mejor mi duda, porque se ve (o al menos yo lo interpreto así), que los datos de la variable diff_ODI_avanzado, no siguen una distribución normal, entonces pregunto, ¿es correcto que si nuestra muestra no sigue una normal, podamos inferir que la población de donde se muestrea sí lo sea? Entiendo que estoy podría darse si no sacamos una muestra lo suficientemente grande pero, entonces, ¿sería correcto aplicar los cálculos de IC (1.96 en el ejemplo que describo)? Por que de nuevo, si los usamos, estaríamos excluyendo la mayoría de los datos de diff_ODI_AVANZADO.

[R_JORDI]:

Hay dos claros clusters. En un proyecto real tendríamos que agrupar los pacientes que les ha funcionado bien el tratamiento y los pacientes que no les ha funcionado tan bien el tratamiento.

Y hacer la comparativa con cada grupo diferente para sacar conclusiones en ambos grupos diferentes.

Tendríamos que ver qué cumplen los pacientes que no ha funcionado tan bien los resultados y qué características tienes lo que sí para dar más información médica al proyecto.

Y es correcto. El IC tiene mucho más sentido cuando tu variable es normal. Cuando no lo es puede que nos perdamos información como este caso.

Sin darte cuenta acabas de aprender dos aspectos:

- El cálculo de los intervalos de confianza
- La comparación de medias con la ayuda de los IC

No ha sido tan difícil, ¿verdad?

Con esta estrategia estás preparado para presentar resultados estadísticamente relevantes cuando quieras comparar variables cuantitativas por grupos.

¡Sigue avanzando!