La correlación es muy utilizada como herramienta descriptiva como ya has visto en el bloque anterior.

Y se utiliza para encontrar las variables relacionados entre sí para poder después entender mucho mejor qué variables contienen información similar.

Pero también se puede utilizar des del punto de vista analítico y poder aportar conclusiones importantes en estudios científicos.

En esta hoja de trabajo practicarás las tres situaciones más comunes paso a paso para que entiendas cómo hacerlo en la práctica.

¡Let’s go!

# LA CORRELACIÓN DE DOS MEDIDAS

Con los datos Espalda como siempre vamos a tratar de ver si están relacionadas dos variables. El peso y la altura y el ODImes0.

Las variables a trabajar son:

* Peso
* Altura
* ODImes0

Los pasos serán los siguientes.

## Paso 1 – Define la preguntas u objetivo

El objetivo lo acabamos de definir pero escríbelo aquí para seguir el orden de la hoja de trabajo de los test estadísticos.

|  |
| --- |
| Escribe el objetivo:  Se intenta ver si las variables Peso, Altura y OdiMes0 tienen algún grado o no de relación. |

## Paso 2 – Describe la pregunta con estadística descriptiva

Antes de nada, es muy interesante visualizar la información que queremos analizar.

En el test de normalidad nos interesa dibujar:

* El matrixplot mixto
* El diagrama de correlación

|  |
| --- |
| Copia los gráficos aquí |

## Paso 3 – Rellena la plantilla de contraste

Ahora calcula para el coeficiente de Pearson y Spearman:

* El Coeficiente de correlación
* La tabla de p-valores

|  |
| --- |
| **Define la H1 o hipótesis de investigación:**  La H1 nos sugiere que hay relación entre las variables Peso, Altura y Odimes0 |
| **Define la H0 o hipótesis nula:**  La H1 nos sugiere que **NO** hay relación entre las variables Peso, Altura y Odimes0 |
| **Umbral de contraste (alpha):**  5% = 0.05 |
| **Coeficiente utilizado: spearman y Pearson**    **Comprueba las restricciones:**  Normalidad de las variables para Pearson  Revisando la tabla o matriz de coef correlacion y p-valor tanto con Spearman y Pearson se ve que las variables Peso y Altura tiene un tipo de relación y viendo el p-valor el cual es menor a 5% indica que es significativa.  Como se observa la variable **peso** al test de normalidad de **shapiro-wilk** da un valor **< 0.05** por lo que **rechazamos la H0**, es decir, los datos observados no provienen de una distribución Normal**.** Por lo tanto debemos usar el coeficiente de correlación de **Spearman** |
| **Significación o p-valor:**  Matriz de p-valores y Matriz de correlación: |
| **Respuesta:**  Qué variables están relacionada y qué grado de asociación presentan  Las variables relacionadas son **Peso y Altura**, su asociación parece lineal y positiva. Su coeficiente o grado de asociación es de 0.55 casi Alta. |

## PASO 4 - CONCLUSIÓN

Juntamos la información del test y el de la descripción de datos.

|  |
| --- |
| *Utiliza la información de test y los gráficos y estadísticos que has calculado*  Podemos concluir que las variables Peso, Altura y OdiMes0, solo presentan algún grado de asociación lineal y positiva las variables Peso y Altura. Luego de descartada la H0 al test de shapiro-wilk de la variable peso nos quedamos con el coeficionte de correlacion de Sperman que fue de 0.55 |

# UNA VARIABLE ORDINAL y UNA MEDIDA

Comprueba la asociación de la variable diff\_ODI =ODImes0-ODImes1 con el número de hernias discales (NHD)

Es decir,

* Medida = diff\_ODI
* Factor = NHD

Utiliza Spearman como coeficiente de correlación

## Paso 1 – Define la preguntas u objetivo

|  |
| --- |
| *Define el objetivo en clave a las variables*  Se trata de validar si la variable diff\_odi tiene alguna asociación con la variable de numero de hernias. Es decir si la variable de mejoría de alguna forma puede ser impactada por el Numero de Hernias del paciente. |

## Paso 2 – Describe la pregunta con estadística descriptiva

|  |
| --- |
| *Utiliza las herramientas gráficas para conseguir tu objetivo*   * *Diagrama de errores*     Al parece a nivel del descriptivo box plot no pareciera haber alguna asociación, parece no afectar el resultado de la mejoría.  Le aplique un log() a diff odi pues pensé que podría haber alguna mejoría en la linealidad. |

## Paso 3 – Rellena la plantilla de contraste

|  |
| --- |
| **Define la H1 o hipótesis de investigación: (hipótesis de diferencias)**  H1: Existe una asociación o relacion entre diff\_oddi y NHD |
| **Define la H0 o hipótesis nula: (la contraria a la H1)**  H0: NO Existe una asociación o relacion entre diff\_oddi y NHD |
| **Umbral de contraste (alpha):**  5% = 0.05 |
| **Test Estadístico (selección del test):** Al comparar una medida y una variable ordinal, automáticamente y si existe cierta linealidad debo usar SPEARMAN  **Comprueba las restricciones:** Por aquello valido la normalidad de la variable diff\_odi  Como se ve la variable no es normal |
| **Significación o p-valor: (el resultado del test)**    p-valor:0.528  Coeficiente de correlación:0.044 |
| **Respuesta: (si el p-valor <5% 🡪 te quedas con la H1)**  El p-valor es > 5%. Por lo que no puedo rechazar la H0. No podemos aceptar la H1 |

## Paso 4 – Conclusión

|  |
| --- |
| *Utiliza la información de test y los gráficos y estadísticos que has calculado*  Luego de realizado el test correlacion de spearman, y ver que el coeficiente de correlacion es muy bajo casi 0 y ver el p-valor que es > a 5% podemo concluir que no existe una asociación o relación entre el NHD y la mejoría reflejada en diff\_oddi |

Con esto creo que ya puedes con muchos de los problemas que te encontrarás en el sentido relacional.

¡Vamos a seguir avanzando!

¡A por ello Valiente!