Este es el último ejercicio y es para mí, el más interesante.

¡Así que vamos a ello!

Vamos a crear un modelo que nos permita explicar la mortalidad infantil en función de otras variables.

Es importante entender qué es un modelo y cómo trabajo.

Para ello tienes un artículo que te puede interesar:

<https://conceptosclaros.com/que-es-modelo-matematico/>

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>

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

<https://www.gnu.org/software/pspp/>

Excel supongo que lo tendrás instalado ☺

¡A por ello!

# Un modelo lineal simple

La mortalidad infantil en función del PIB per cápita.

* La variable respuesta o de salida es la mortalidad infantil
* La variable de entrada o independientes es el PIB per cápita

Los pasos que vamos a seguir son los siguientes:

1. Crear un diagrama de dispersión con línea de tendencia de la mortalidad infantil vs PIB per cápita
2. Transformar el PIB per cápita a logarítmico ya que presenta no linealidades
3. Crear el modelo siguiente:

Mortalidad Infantil = a\*LOG\_PIBCápita + b

Crea el modelo con la ayuda de Excel o SPSS (como prefieras)

Tienes el vídeo Tech Zen para lograrlo:

## Un modelo lineal simple para cada grupo

Vamos a crear un modelo lineal para cada grupo:

* Other
* OECD
* Africa

Para hacerlo simplemente ordena la tabla según la variable group y separa los grupos.

Crea el modelo lineal igual que antes.

|  |
| --- |
| Copia los diagramas de dispersión con las líneas de tendencias y los coeficientes de los modelos (la ecuación de la recta)  Aquí se observa que el p-valor es < al 5%. El coeficiente de relación es normal e inverso que indica que a mayor PIB menor Mortalidad Infantil. La fórmula del modelo y = 41 – 0.00087x sugiere lo mismo.  Se calcula el LN de PIB para ver el cambio en la forma del diagrama de dispersión.    Aquí se observa que el p-valor es < al 5%. El coeficiente de relación es normal e inverso que indica que a mayor PIB menor Mortalidad Infantil. La fórmula del modelo y = 160 – 34x sugiere lo mismo.  Al correlacionar con ln(pib) se ve más claramente la tendencia y asociación inversa.    En este caso obtenemos digramas de dispersión de infantMortality vs Pib pero con el ln(pib) y agrupado por el grupo al que pertenece al país. En todos los caso hay asociación. Todas inversas y los modelos también los confirman. A menor PIB mayor Mortalidad Infantil. |

## Crea un modelo lineal múltiple

Ahora lo que puedes es calcular un modelo múltiple.

Añade una variable de entrada:

Mortalidad Infantil = a1\*LOG\_PIBCápita + a2\*pctUrban +b

|  |
| --- |
| Copia la tabla del modelo con los coeficientes    Y = 153.95 + (-32.9268)lneppib + (-0.05083)pctUrban |

¿Qué conclusiones puedes obtener?

|  |
| --- |
| A priori veo que el p-valor del intercepto y el del pib son menores a 0.05 por los pasan la prueba de significancia.. Sin embargo el de pctUrban no cumple el p-valor, por lo que no creo que deba incluirse en el modelo. A pesar de el R2 es de 0.61 por lo que indica que el modelo es aceptable….  Lo que si me dice el modelo es que a mayor pib y pctUrban hay menos mortalidad infantil. Ahora puedo creer que en este caso el pctUrban no aporta mucho ya que posiblemente esta muy correlacionado con pib por lo que con solo la variable pib el modelo bastaría. Abajo se muestra diagrama de dispersión entre pib y pcturban, en la que se nota que el nivel de asociación es alto . |

¡Has hecho un gran trabajo de análisis!

Estas captando la esencia de los modelos lineales

Es el principio para poder trabajar con modelos lineales generalizados. ¡Que molan mucho!

1 fuerte abrazo!

Jordi