En esta hoja de trabajo conceptual vamos a trabajar distintos tipos de modelos para captar la idea y la esencia que hay detrás.

No te preocupes por si no entiendes la fórmula matemática en sí. Quiero que te sirva para que veas ejemplos de modelos y captes la idea de fondo. Estamos nutriendo las bases.

La idea es que identifiques tres partes:

1. Las variables de entrada (var. independientes)
2. La variable de salida (var. dependientes)
3. Los coeficientes o parámetros del modelo
4. El modelo o función matemática
5. El objetivo del modelo

EJEMPLO: UN MODELO DE UNA VARIABLE – DISTRIBUCIÓN GAUSSIANA

Como ya has visto una distribución de probabilidad se puede expresar como una función matemática. ¿Te acuerdas de la distribución normal?

La distribución normal es básicamente una función matemática que representa un histograma de densidad.

Puedes inferir o calcular los parámetros de un modelo de distribución de probabilidad conocida con los datos de una muestra.

Aquí tienes un ejemplo utilizando la variable altura [m] de la base de datos “espalda.xlsx”

El objetivo es calcular un modelo de distribución normal con los datos que disponemos.

¿Cómo calculamos un modelo normal a partir de los datos?

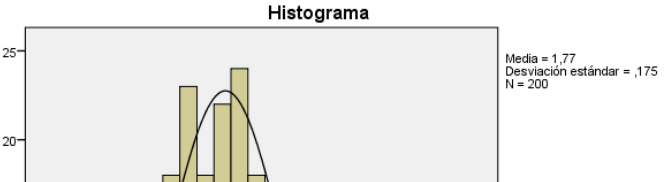
Primero calculando la media y la desviación estándar para calcular los parámetros del modelo.

Describiendo la variable como has aprendido: media y desviación estándar.

- Valor central: Media = 1.77m
- Dispersión: Desviación Estándar = 0.175 m

La variable se puede representar como la media ±desviación estándar

- 1.77± 0.175 m



¿Cuál es nuestro modelo?

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Densidad de Frecuencia

Variables Numérica (Altura en cm)

¡Exacto! La función de densidad de probabilidad normal. ¡Toma!

Cuando calculemos test estadísticos paramétricos son test que utilizan modelos matemáticos para las distribuciones.

Por ello mucha de las restricciones de aplicación de test paramétricos es que las variables sean normales.

¿Cómo sabes si una variable es normal?

Con un test estadística llamado Kolmorov-Smirnov. ¡Lo veremos, tranquilo!

Encuentra las partes del modelo Normal:

Característica de los modelos	Respuesta
Variables de Entrada	En modelos de una variable no tenemos variables de entrada. Solamente estamos trabajando con una variable de salida y calcularemos el modelo de esa variable
Variables de salida	La variable peso [kg]
Función matemática o modelo	La función Gaussiana
Coeficientes del modelo	Mu = media Sigma = desviación estándar
El objetivo del modelo	Tratar la población de la muestra como una función Normal. Y conocer los parámetros de valor central y dispersión

UN MODELO PREDICTIVO – REGRESIÓN LINEAL

Ahora es el turno de la regresión lineal.

Este modelo te lo he explicado en el vídeo de la lección y lo has trabajado en el pre training.

El objetivo de este modelo es relacionar la variable de salida en función de la de entrada con un modelo lineal. Con una recta.

Es decir que tenemos relación de proporcionalidad.

Ahora vamos a calcular una regresión lineal con dos variables que influyen en la respuesta.

Abre la tabla de datos "mtcarts.xls" y calcula un modelo lineal con dos variables independientes (de estudio o de entrada) y una variable dependientes (respuesta o de salida):

- La variable y "mpg" millas por galón
- Las variables x "hp" (caballos) y "wt" (peso)

El modelo lineal será:

$$mpg = a \cdot hp + b \cdot wt + c$$

Los coeficientes los puedes calcular siguiendo el tutorial Tech Zen.

Con los resultados que has obtenido rellena la tabla:

Característica de los modelos	Respuesta
Variables de Entrada	"hp" (caballos) y "wt" (peso)
Variables de salida	"mpg" (millas por galón)
Función matemática o modelo	mpg = a · hp + b · wt + c mpg = -0.0318*hp + -3.8778*wt + 37.2273
Coeficientes del modelo	a = -0.0318 b = -3.8778 c = 37.2273
El objetivo del modelo	Relacionar el peso y los caballos de coches con la variable millas por galón

UN MODELO PREDICTIVO – REGRESIÓN LOGÍSTICA

Con el ejemplo de modelo logístico que te he explicado en el vídeo de la lección trata de identificar las partes de este modelo.

Vuelve a revisar el vídeo para rellenar la tabla. ¡Utiliza el modelo de dos variables de entrada!

Característica de los modelos	Respuesta
Variables de Entrada	- Presión sistólica - Sexo - Edad
Variables de salida	IAH30 (Indica la gravedad de la apnea del sueño de un paciente. 0 = no es grave, se producen menos de 30 apneas por hora; 1 = grave, se producen más de 30 apneas por hora)
Función matemática o modelo	logit(IAH30) = d + a · TAS_m + b · Sexo + c · Edad □ logit(IAH30) = -7.32477 + 0.04029*TAS_m + 0.66036*Sexo + 0.04302*Edad
Coeficientes del modelo	- a = 0.04029 - b = 0.66036 - c = 0.04302 - d = -7.32477
El objetivo del modelo	Relacionar la presión sistólica, el sexo y la edad con el grado de gravedad de la apnea del sueño.

¡Buen trabajo!

Con esta hoja de trabajo entiendes la esencia de los modelos matemáticos.

Existen muchos modelos y el principal reto es conocer qué tipo de ellos se adapta mejor a tus datos.

Y el segundo reto, es saber escoger las variables que necesitas y las más importantes.

Durante el bloque 5 y 6 vas a ver tipos de modelos y cómo utilizarlos en la práctica.

Poco a poco iremos construyendo este conocimiento práctico que necesitas ☺

¡Nos vemos pronto!