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?**

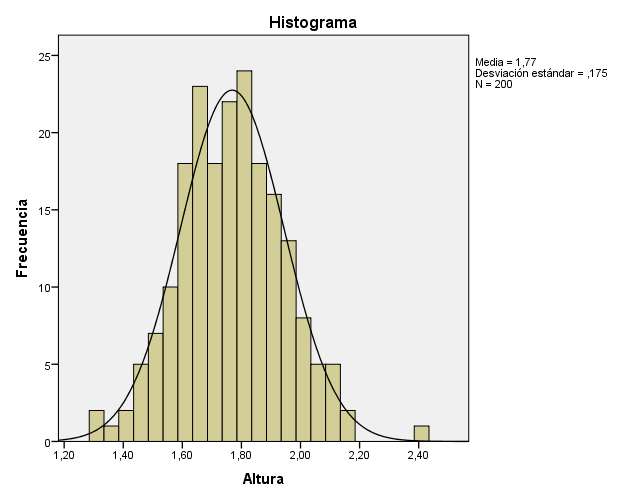
Primero calculando la media y la desviación estándar para calcular los parámetros del modelo.

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

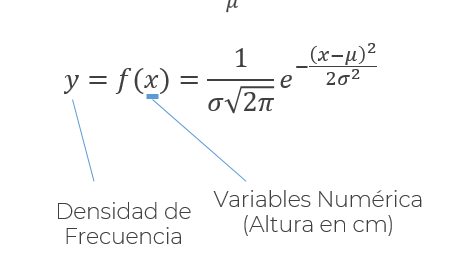
* Valor central: Media = 1.77m
* Dispersión: Desviación Estándar = 0.175 m

La variable se puede representar como la media ±desviación estándar

* 1.77± 0.175 m



**¿Cuál es nuestro modelo?**



¡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)

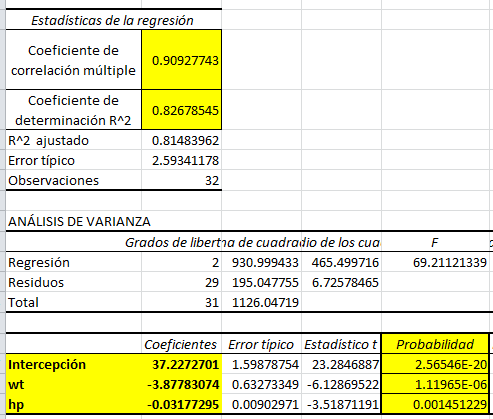
El modelo lineal será:

**mpg = a · hp + b · 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, WT |
| Variables de salida | MPG |
| Función matemática o modelo | **f(mpg) = -3.877(wt) -0.0317(hp) + 37.22** |
| Coeficientes del modelo | **-3.877; -0.0317; 37.22** |
| El objetivo del modelo | El objetivo es tratar de estimar aproximadamente cuanto seria el rendimiento del auto MPG dados un Peso(wt) y los caballos de fuerza del auto (HP).  Por los valores de coeficiente de Correlacion (R ) y R2 vemos que hay una fuerte asociación entre las variables entradas y salida.  Igualmente la probabilidad es menor a 5% para todos los coeficientes por lo que podemos aceptar dichos coeficientes |



# 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 | **log(PIB per cápita)** |
| Variables de salida | **Esperanza de Vida en Años** |
| Función matemática o modelo | **y = 7.2028x + 4.9496** |
| Coeficientes del modelo | **7.2028 ; 4.9496** |
| El objetivo del modelo | Por un lado se trata de ver si hay asociación entre la el PIB y la esperanza de vida y adicional estimar aproximadamente cual podría ser la esperanza de vida dado un PIB. Se tuvo que aplicar el Log Neperiano a PIB porque los valores del PIB son números muy grandes en comparación con la esperanza de vida y los residuales no mostraban un buen comportamiento a pesar de que el p-valor si era menos a 5%. Una vez usado el Log Neperiano las curvas de los residuales mostraron un mejor comportamiento y la gráfica de la correlación muestra una asociación fuerte y positiva |

¡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!