

1. Introducción a la ciencia de datos.

Tipos de variables:

1. Variables Independientes (entradas)

a) Cualitativas

1) Texto

a' Nominales (categorías, Por ejemplo: países, sexo)

b' Ordinales (poco, mucho, muchísimo, Por ejemplo: nivel de tabaquismos)

2) Numéricas

a' Nominales

b' Ordinales

b) Cuantitativas: cuando hablamos de cantidad

1) Discretas: Por ejemplo año, mes, edad, etc.

2) Continuas: Por ejemplo altura, peso, etc.

2. Variables Dependientes (salidas, categorías)

Las variables **cualitativas** son aquellas que describen características o cualidades y no pueden ser medidas en términos numéricos. Las variables **cuantitativas**, por otro lado, son aquellas que pueden ser medidas en términos numéricos y tienen valores numéricos.

Las **variables nominales** son aquellas que no tienen un orden natural, como el género o el color de ojos. Las **variables ordinales** son aquellas que tienen un orden natural, como el nivel de educación (primaria/secundaria/universidad), nivel de tabaquismos: Clasificamos como leve (1), moderado (2), nivel medio (3), importante (4) y muy importante (5).

Variables y tipos de problemas (video 02a min 9:00)

1. Si la variable dependiente es **cualitativa**, el tipo de problema es de **clasificación**.
2. Si la variable dependiente es **cuantitativa**, el problema es de **regresión**. Por ejemplo si quiero predecir el precio de una propiedad.
3. Si **NO hay variable** dependientes, el problema es agrupamiento.

La variable pueden presentar:

1. Outliers (valores atípicos). Pueden ser erroreso un dato que se sale de la norma.
2. Correlación:
 - a) Positiva
 - b) Negativa
 - c) Sin correlación.
3. Varianza
4. Covarianza