

Variables

¿Qué es una variable?

Una variable es una característica o propiedad que se puede medir o categorizar, y cuyos valores pueden cambiar entre individuos, objetos o momentos.

Ejemplo:

- La **Edad** puede tomar valores como 20, 30, 40.
- El **Género** puede tomar valores como Masculino, Femenino, Otro.
- El **Ingreso Mensual** puede tomar valores como 500, 1000, 10000.
- El **Estado Civil** puede tomar valores como Soltero, Casado, Divorciado.

Comúnmente asociamos el término “**columna**” de una base de datos a “**variable**”, es decir, cada “**columna**” es una “**variable**”.

Tipos de Variables

Existen diversos tipos de variables

Variable cualitativas (categóricas)

- a) **Nominal:** No tienen orden lógico. Colores (Verde, Rojo, Azul)
- b) **Ordinal:** Tienen un orden lógico. Nivel educativo (Básico/Medio/Superior)

Variable cuantitativas (numéricas)

- a) **Discretas:** Se obtienen al contar. Números enteros (Número de hijos/mascotas)
- b) **Continuas:** Se obtienen al medir. Número decimales posibles (Peso/Altura)

Conocer estas diferencias es fundamental, sobre todo para implementar modelos de machine learning y análisis estadístico.

Ejemplo de Base de datos y tipos de Variables

Veamos un ejemplo de base de datos con diversos tipos de variables.

Cualitativa
Nominal

Cuantitativa
Discreta

Cualitativa
Nominal

Cuantitativa
Continua

Cualitativa
Ordinal

Cuantitativa
Continua

ID	Género	Edad	Ciudad	Ingreso Mensual (CLP)	Nivel Educativo	Horas de estudio semanales
A1	Femenino	44	Valparaíso	700000	Secundaria	8,5
A2	Masculino	50	Concepción	1600000	Universitario	7
A3	Masculino	58	Temuco	400000	Postgrado	2
A4	Otro	40	Concepción	1900000	Secundaria	17,6
A5	Otro	22	Santiago	1800000	Postgrado	17
A6	Masculino	56	Concepción	900000	Postgrado	1,5
A7	Masculino	36	Temuco	1600000	Secundaria	3

En diversos libros de ciencia de datos podemos encontrar nombres como “**features**”, “**atributos**”, “**campos**”, entre otros.

Tipos de Variables

¿Qué ocurre en Python y R?

Mediante funciones podemos ajustar para que una columna sea de cierto tipo, ya que muchas otras funciones, como por ejemplo de machine learning, lo exigen.

Tipo de Variable	Ejemplos	Tipo de dato en R	Tipo de dato en Python (pandas)
Cualitativa nominal	Sexo, color, país	factor o character	object o category
Cualitativa ordinal	Nivel educativo, satisfacción	ordered factor	category con ordered=True
Cuantitativa discreta	Nº de hijos, cantidad de compras	integer	int64
Cuantitativa continua	Peso, ingreso, temperatura	numeric (double/float)	float64
Binaria / Dummy	0/1, Sí/No	logical (TRUE/FALSE) o integer	bool o int64 (dependiendo del enfoque)

Tipos de Variables y Machine Learning

Los modelos de Machine Learning más cercanos a nuestra realidad diferencian fuertemente lo cualitativo de lo cuantitativo.

- Cuantitativo se asocia a modelos de Regresión (Números, enteros o decimales)
- Cualitativo se asocia a modelos de Clasificación (Categorías, ordenadas y no ordenadas)
- Existen otros tipos de modelos, como por ejemplo lectura de imágenes, pero eso es otro mundo.
- Para nuestros casos trabajaremos con modelos de regresión y clasificación.

Cuando lo veamos aplicado entenderás mejor varios conceptos, ya que los ejemplos ayudan bastante.

Conclusiones

Conocer las diferencias entre los tipos de variables es fundamental, sobre todo para implementar modelos de machine learning y análisis estadístico.

Python y R nos ofrecen todo para ajustar las columnas a su formato correcto, para su uso posterior en diversos análisis.