



# Descriptores de datos

Semana 02 - Clase 04

## Descriptores

---

Cuando trabajamos con datos, es recomendable que al pasar a tabla se comiencen con estas cosas como mínimo:

- Numero de variables.
- Numero de observaciones.
- Número de datos que faltan.
- Número de observaciones completas.
- Clasificación del tipo de cada variable.

Nota. Existe un tercer tipo de variables categóricas, denominadas dicotómicas, estas variables tienen que ver con preguntas (si/no), por lo que estas variables son binarias, sólo pueden tener dos posibles respuestas.

## Variables

---

**Numérica:** Es aquella variable, definida por un número, existen dos casos: las continuas y discretas, una continua es el tiempo (está definido por horas, minutos, etc infinitamente), mientras que la edad es una variable discreta (un número entero define la cantidad).

**Categórica:** Estas variables definen grupos específicos de cosas. Se encuentran categóricas nominales (no tienen orden, por ejemplo los colores) y ordinales (tienen orden). Por ejemplo un código universitario es una variable categórica ordinal, pues se sigue un orden de acuerdo al ingreso del alumno.

Ejemplo. Si tenemos una encuesta para puntuar del 1 al 5, donde 1 es la peor, y 5 la mejor, esta es una variable categórica, puesto que si fuese numérica haríamos un promedio y nos daría un puntaje de 3, cosa que no ha sucedido (esto para un 50%: 1, 50%: 5, seleccionado de la población).

Nota. Numéricas: es un número, discretas: "pocos", continuas: "muchos" valores contables. Categórica: no necesariamente un número, ordinales: valores tienen un orden natural, nominales: valores no tienen un orden natural (prestar atención a cómo se ordenan).

## Descriptores numéricos

---

Los descriptores numéricos de posición me permiten realizar estadística descriptiva, me indican dónde están ubicados la mayoría de los datos. Para ello se toma en cuenta:

**Promedio:** Es la sumatoria de todos los valores observados dividido entre el tamaño de la muestra.

**Mediana:** Una mediana es un valor tal que al menos la mitad de los valores es menor o igual a ese valor y al menos la mitad de los valores es mayor o igual a ese valor (el valor "del medio").

**Moda:** Valor o valores que más se repiten.