

# La estadística para Ciencia de Datos

Héctor de la Torre Gutiérrez  
[hdelatorreg@up.edu.mx](mailto:hdelatorreg@up.edu.mx)

# Inicio

- Vamos a conocer, comprender y aplicar diversos conceptos y herramientas de estadística aplicada y su manejo con R/Python.
- ¿Qué es estadística aplicada?  
.... es hacer estadística.... con datos reales obtenidos de encuestas, censos, etc. ... es un proceso total de organización, resumen, presentación e inferencias.

Es importante comentar que ante **una mala planeación o diseño**, **una deficiente toma de datos (sesgada)** o no validación de supuestos, **NO EXISTE** técnica estadística que pueda corregir los resultados **erróneos**.

# ¿Qué es estadística?

Es una ciencia universal, que se ocupa de:  
recolectar (con diseño, planeación y/o estructura),

organizar,

presentar,

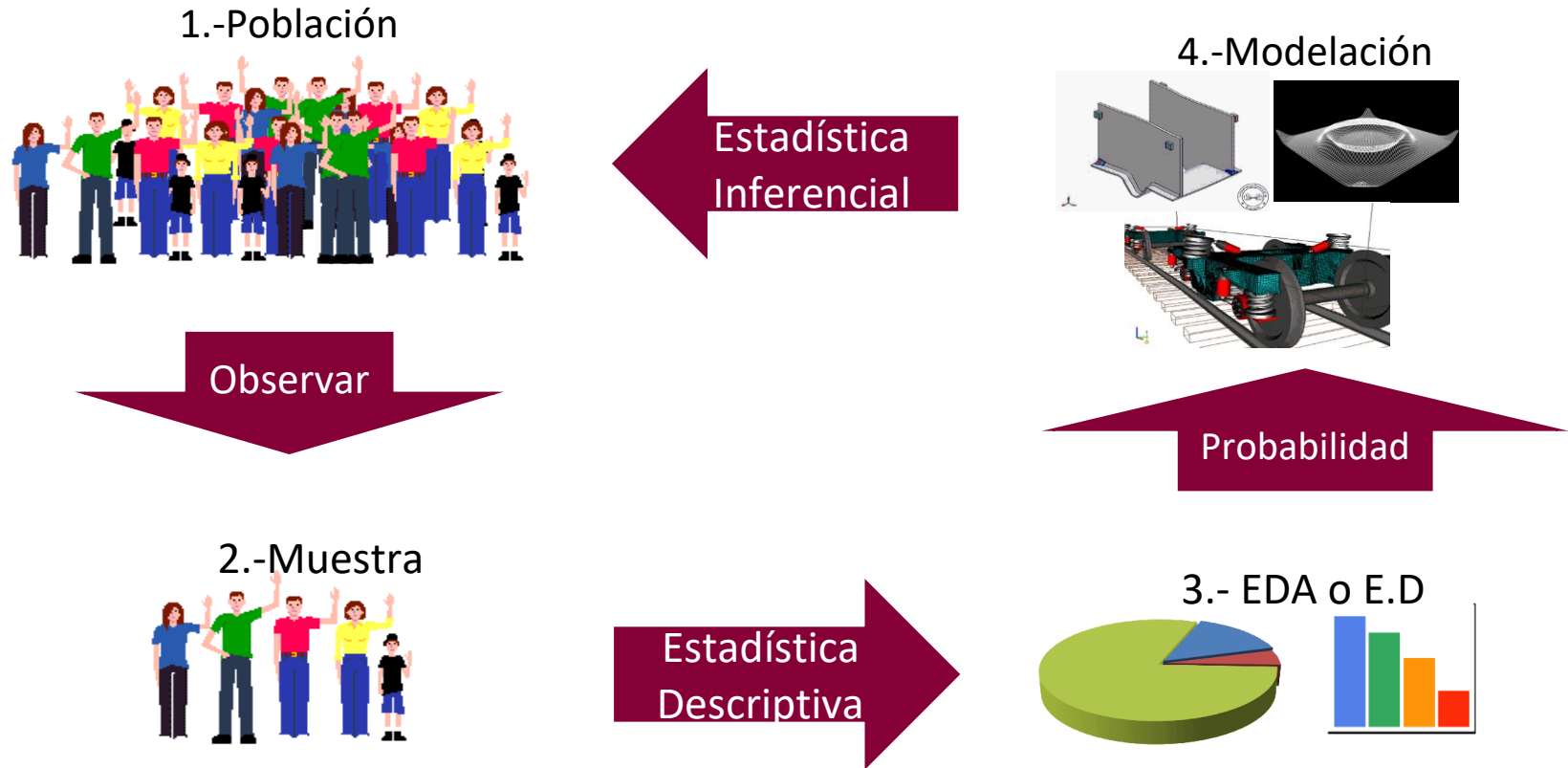
resumir ,

modelar,

La estadística es una herramienta útil para entender el mundo que nos rodea.

extraer e interpretar la información contenida en un conjunto de datos, con el fin de obtener conocimiento para la toma de decisiones de forma oportuna y eficiente; todo ello bajo incertidumbre y variabilidad.

# Proceso general de tratamiento de un problema con estadística aplicada



# Probabilidad & estadística

**Probabilidad** es el **vehículo de la estadística** (base teórica de la estadística), es decir, de NO DISPONER de las leyes de la probabilidad ... la ciencia de la estadística no sería posible.

Una forma de entender **probabilidad** y **estadística** es con el siguiente ejemplo: pensemos que tenemos una caja que contiene 15 fichas (**5 azules** , 5 negras y **5 rojas**).

Ahora usando probabilidad podemos responder la pregunta: ¿Cuál es la probabilidad de seleccionar una **ficha roja**?, etc.; con estadística lo que hacemos es extraer una **muestra aleatoria** y basados en la muestra realizamos **conjeturas** sobre cómo está conformado la **población**, por ejemplo, usando estadística descriptiva (en la muestra) y con inferencia trasladamos los resultados a la población.

En estadística un problema típicamente tiene que ver con fenómenos aleatorios.

# Definiciones

- **Población.** Es el conjunto “**completo**” de individuos u objetos sobre el que estamos interesados en obtener conclusiones (hacer inferencia).
  - Normalmente es demasiado grande para poder estudiarla por completo.
- **Muestra.** Es “**un**” subconjunto de la población al que tenemos acceso y sobre el que realmente hacemos las observaciones (mediciones).
  - Debería ser “representativo”
  - Está formada por miembros “seleccionados” de la población (individuos, unidades experimentales).



- **Parámetro:** Es una cantidad numérica calculada sobre una población

$\mu$

- La altura media de los individuos de un país
- La idea es resumir toda la información que hay en la población en unos pocos números (parámetros).

$\sigma$

- **Estadístico:** Es una cantidad numérica calculada sobre una muestra.

$s$

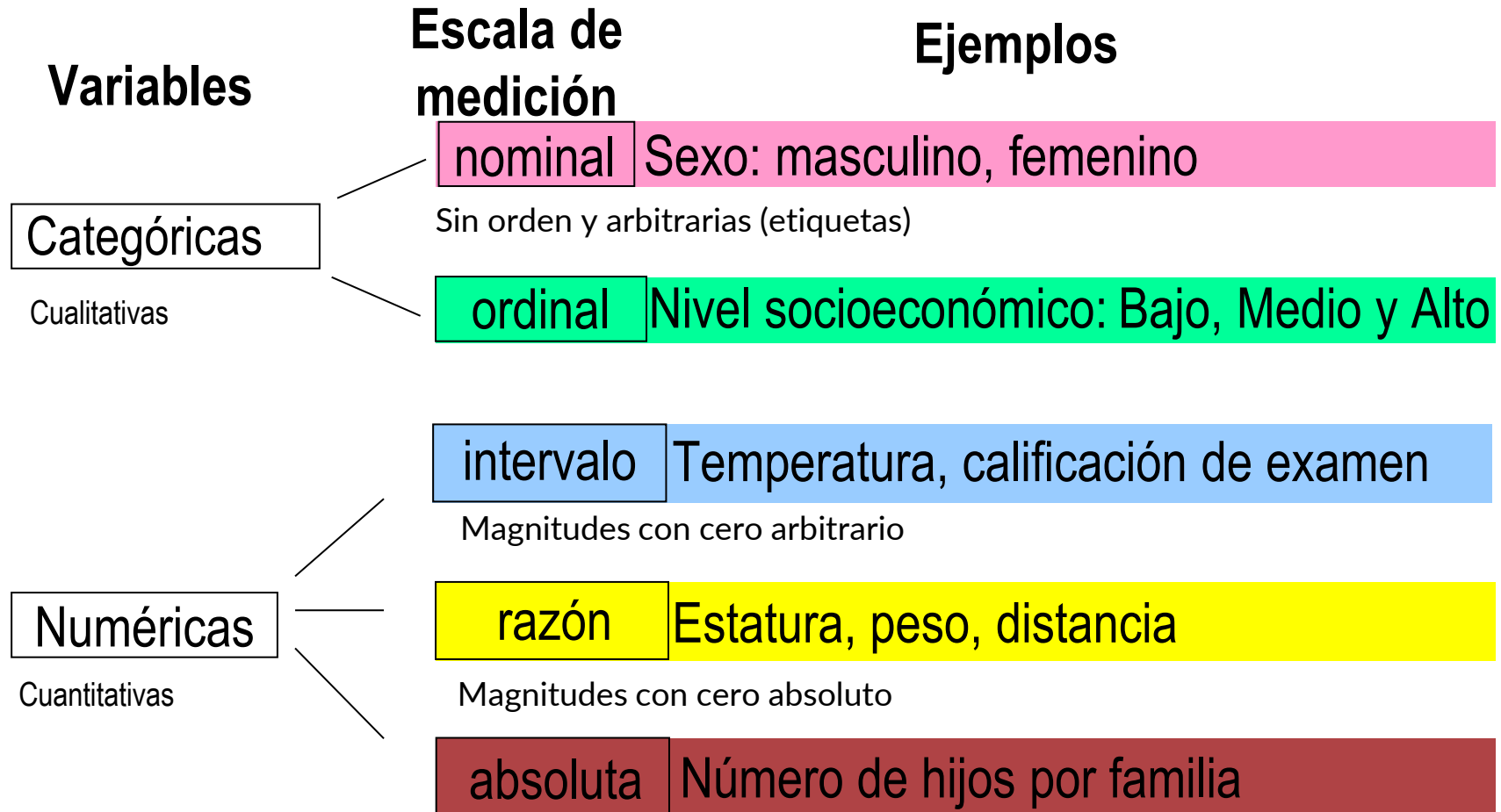
- La altura media de los que estamos en esta sesión.
- Si un estadístico se usa para aproximar un parámetro también se le suele llamar estimador.

$\bar{x}$

- **Variable:** Es la característica que se desea estudiar en la población
- **Dato u observación:** Es la realización (medición) de la variable de interés
- **Experimento (proceso generador de datos):** Actividad o proceso realizado “con un plan definido” cuyos resultados producen un conjunto de datos



# Escalas de medición y tipos de variables



# Integrando...

**Probabilidad.** Es una medida de la **incertidumbre** asociada al resultado de un **experimento aleatorio** (A). Es un valor entre cero y uno que resulta del cociente entre el **número de casos favorables al evento** (m) y el **número de casos totales posibles del experimento** (n).  
 $P(A)=m/n$ .

**Experimento aleatorio.** Tipo de experimento cuyo resultado es incierto (no se conoce) a pesar de realizar dicho experimento bajo las mismas condiciones, pero se conocen todos los resultados posibles.

**Espacio muestral.** Conjunto de “**todos**” los resultados posibles de un experimento aleatorio. Se denota con “ $\Omega$ ”, “S”, “U”.

**Evento.** Objeto de estudio de la teoría de probabilidades y son cualquier resultado posible de un experimento, es decir **un** valor o subconjunto del espacio muestral.

# Adicional

Con el recurso WEB viendo la teoría (<https://seeing-theory.brown.edu/es.html>),  
revisar ejemplos de:

- A. Formular el experimento
- B. Identificar el espacio muestral
- C. Definir algún evento de interés
- D. Calcular la probabilidad del evento