Statistics 101 :

"Academic discipline dealing with all aspects of learning from data."

quantification

Perspectives :

- art of summarizing data
  → make data comprehensible
- science of uncertainty
  → most information in the world is uncertain
- science of decisions
  → ultimate goal of statistics
- science of variation
  → central tendency and spread
- art of forecasting
- science of measurement and data collection. Ⓐ SH /Mgt

Foundations of data

Source of data

- Designed data — "artificially collected" (surveys, studies etc)
- Organic data (process generated)

For both, data needs to be i.i.d

"independent", "identically distributed".

more on this later!

Question: What is the source of NHANES data?

Types of data:

- Some data is not numeric!

  For instance race or gender

- Just as we have data types in programming languages, we have different types here.

- Weight — numeric, continuous
- # of kids — numeric, discrete
- Age group ($\overset{1}{\text{child}}$, $\overset{2}{\text{adult}}$, $\overset{3}{\text{elder}}$) — categorical, ordered
- Gender — categorical, unordered.

Practical Note:

$\begin{bmatrix}1\\0\\0\end{bmatrix} \overset{c}{\begin{bmatrix}0\\1\\0\end{bmatrix}}$

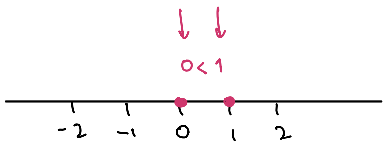
lf $\begin{bmatrix}0\\0\\1\end{bmatrix}$

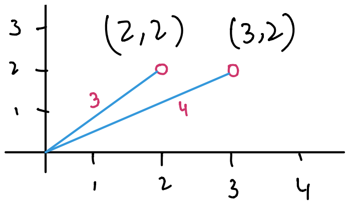
Gender represented as: M / F

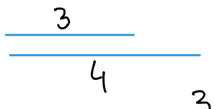
or: 0 / 1 But still unordered!

or: $\begin{bmatrix}1\\0\end{bmatrix}$ / $\begin{bmatrix}0\\1\end{bmatrix}$

"one-hot vector representation"



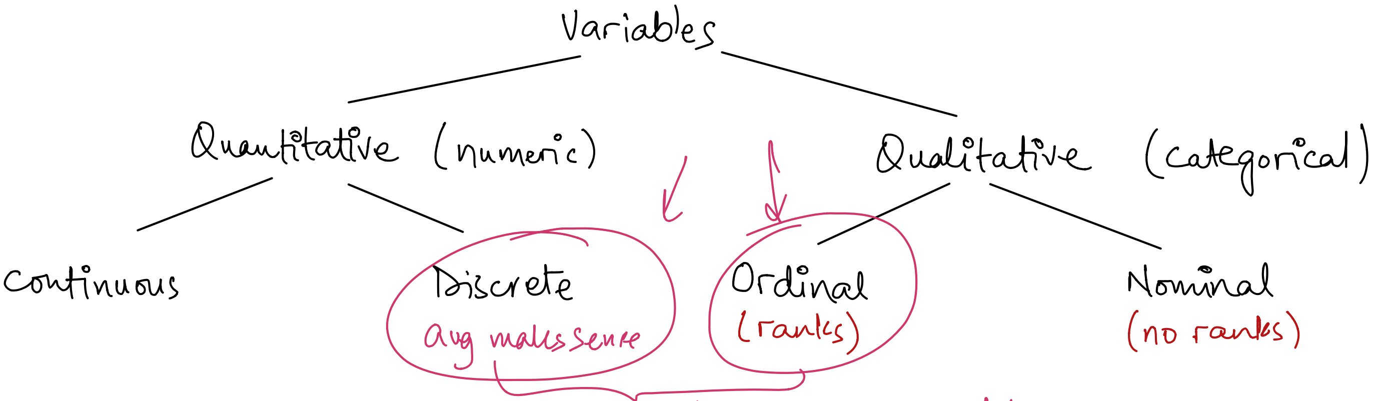




$3<4$

$\underbrace{(2,2)} < \underbrace{(3,2)}$

uncomparable
not ↗



Both discrete & ordered!
So, difference?