

# STAT 231

- STAT 230 Review Video posted —.

Lectures — Banerjee

Week 1.

---

## Roadmap

- 5 min recap
- Some terminology of Statistical Inference
- Numerical data summaries

We have a sample. from a <sup>unknown</sup> population  
that we have an interest in.

Based on that sample, we "infer"  
properties of the population

EMPIRICAL STUDIES study is

conducted using data either through  
observation or through experimentation

Different studies will have different  
results.

Every ~~a~~ empirical study involves uncertainty  $\rightarrow$  we analyze this using STAT 230 techniques.

Example 1 We are trying to find the proportion of women with at least a high school degree & ~~and~~ an US voter who supports Trump.

Example 2 Medical study: Suppose we want to check whether Drug A reduces B.P. among patients with a heart disease.

Typically, we use "BLIND" studies  
"DOUBLE BLIND" "

- Remove bias and the PLACEBO effect.

Example 3      2016 Subaru Outback.

average # of kilometres the car  
will go before the 1<sup>st</sup> tuneup.

---

SOCIAL STUDIES, EMPIRICAL

MEDICAL  
SCIENCES  
PHYSICAL

↑  
OBSERVATIONAL

↑  
EXPERIMENTAL

we have no  
control over  
the variables

• data collector controls  
some of the variables

Sample Survey <sup>Obs. study</sup>  $\rightarrow$  B. Underlying  
population of interest is finite.  
~~and~~

Trump example  $\rightarrow$  Sample Survey

---

Each member  $\rightarrow$  UNIT  
of the population

VARIATE: The property of the  
unit you are interested in.

TRUMP problem  $\rightarrow$  Unit ?  
Variate of interest  $\leftarrow$   $\begin{matrix} Y_i \\ N \\ \text{Undecided.} \end{matrix}$

Attribute : A function of the  
variables.

$y_1, \dots, y_n$

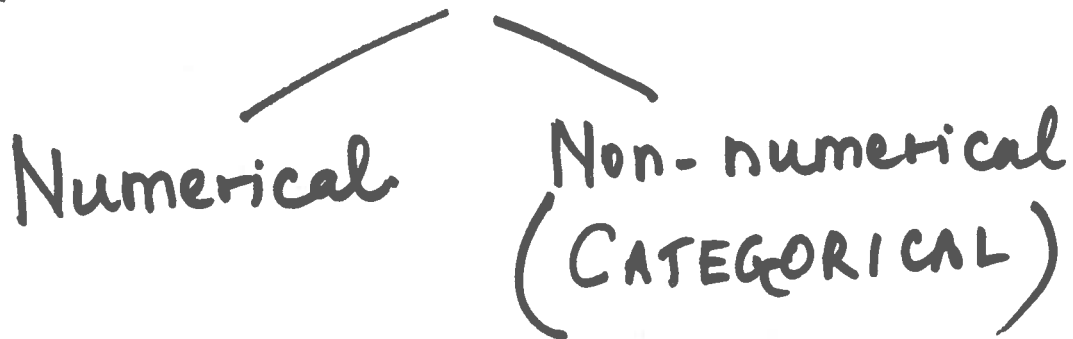
$f(y_1, \dots, y_n) = \text{attribute}$ .

---

Proportion, averages, maximum, etc.  
are attributes.

---

TYPES OF DATA.



# NUMERICAL DATA

DISCRETE

"integers"

things we  
count.

CONTINUOUS

could potentially  
take any real value.

Things we measure,  
height, weight.

# CATEGORICAL DATA

ORDINAL

that has an  
underlying order.

NON-ORDINAL  
DATA.

# Numerical Data Summaries

## Important properties:

- (i) Measures of central tendency  
(centre of the data)
- (ii) Measures of variability
- (iii) Measures of symmetry (SKEWNESS)
- (iv) How fat the tails are.  
(How frequent are extreme observations)

KURTOSIS



Two reasons why we summarise data

- (i) Extract the important properties
- (ii) Identify the likely distribution from which the data is drawn.

## MEASURES OF CENTRAL TENDENCY

$\{y_1, \dots, y_n\} \rightarrow$  DATA SET

Sample mean:  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

Arithmetic mean: A.M.

Geometric mean:  $G.M. = (y_1 \dots y_n)^{1/n}$

16, 9

Sample mean = 12.5

$$G.M. = (.16.9)^{1/2} = 12$$

Example 1 \$100.

Yr 1  $\rightarrow$  4% ✓

Yr 2  $\rightarrow$  8%

Yr 3  $\rightarrow$  12%

What is the average interest rate?

$x$  = av. int. rate

$$\frac{100(1+x)^3}{(1+x)^3} = 100(1+0.04)(1+0.08)(1+0.12)$$
$$= \sqrt[3]{1.04 \times 1.08 \times 1.12}$$