# SNS COLLEGE OF TECHNOLOGY

**(An Autonomous Institution)**

Re-accredited by NAAC with A+ grade, Accredited by NBA(CSE, IT, ECE, EEE & Mechanical)

Approvedy by AICTE, New Delhi, Recognized by UGC, Affiliated to Anna University, Chennai

## Department of MCA

## Topic: Statistics Inferences



**Course**

**16CA817**
**Big Data Analytics**

**Unit I**

**Introduction to Big data**

**Elective**

**V Semester / III MCA**

# Problem

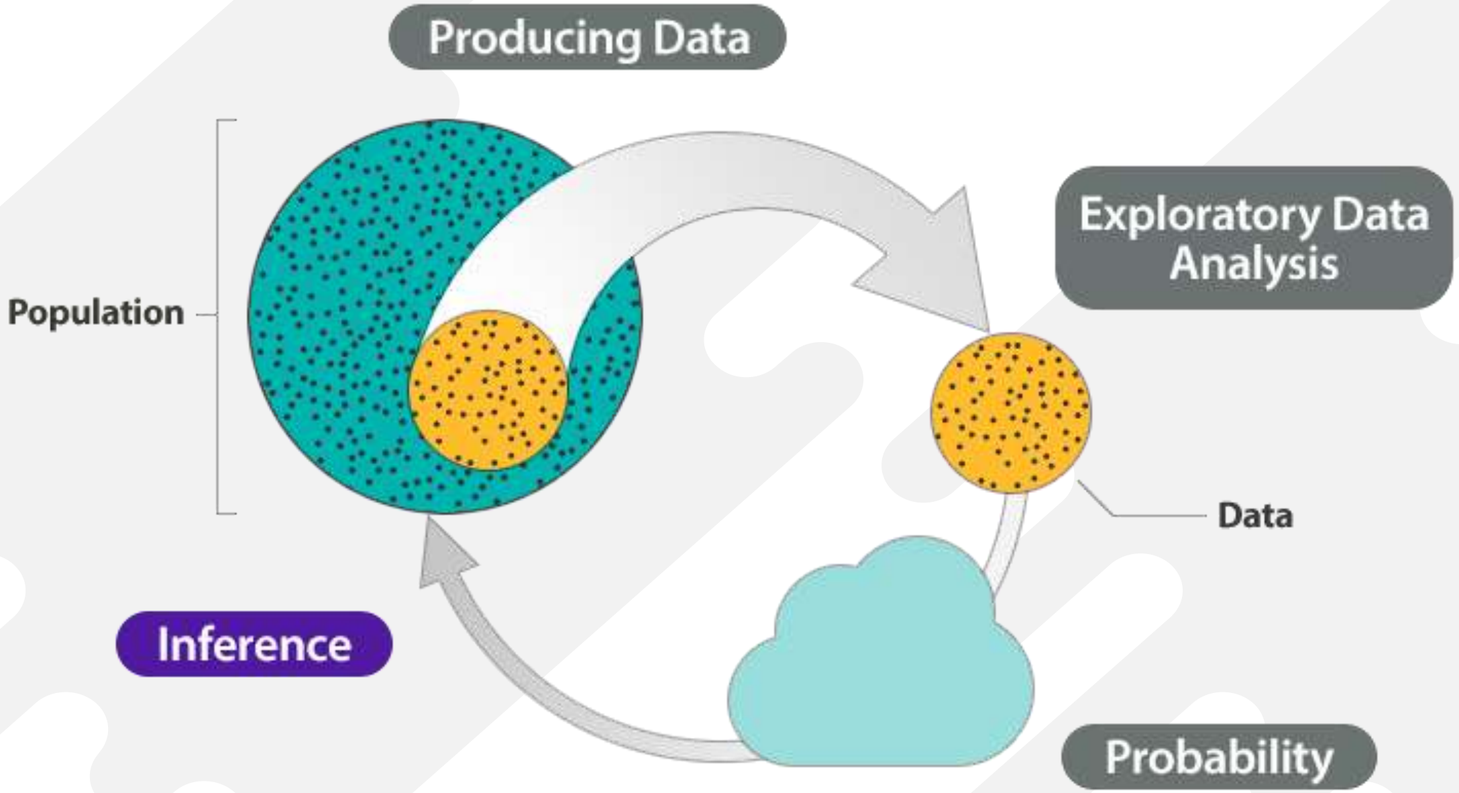
As per instruction from her professor, Deena, a statistical course student collected survey from the people about the influence of social media in the rural area of Tamilnadu.

Now she would like to apply analytical methods /statistical approach to find the outcome of the survey

# Statistics Inference

- A process of drawing conclusions about populations / scientific truths from data
- Estimate of the characteristics or properties of a population, derived from the analysis of a sample drawn from it.
- Two ways to make inference
  - Estimation of parameters
    - Point estimation
    - Intervals estimation
  - Hypothesis testing

# Statistics Inference



# Statistics Inference Methods

- ***Interval estimation***
  - Estimation of the amount of variability in a sample statistic when many samples are repeatedly taken from a population
- ***Hypothesis testing***
  - The comparison of sample results with a known or hypothesized population parameter

# Central Limit Theorem



| Statistics | Parameter |
| --- | --- |
| Mean X’ | Mean μ |
| Standard Deviation **S** | Standard Deviation σ |
| Proportion **P** | Proportion π |

# Point Estimation

- Let Population parameter θ and G denotes estimator, then estimation error would be G- θ, probably close to 0.

- Bias of an estimator
  - $\mathbf{B_\theta = E_\theta(G-\theta) = E_\theta(G) - \theta}$
  - If E θ (G) = θ, i.e. the expected value of the estimator is equal to the value of the population parameter, then the estimator G is called unbiased

Courtesy: https://data-flair.training.com

# Point Estimation

Point estimator of the population mean $\mu$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(where $X_i = 0$ or 1) is a point estimator of the population proportion $p$

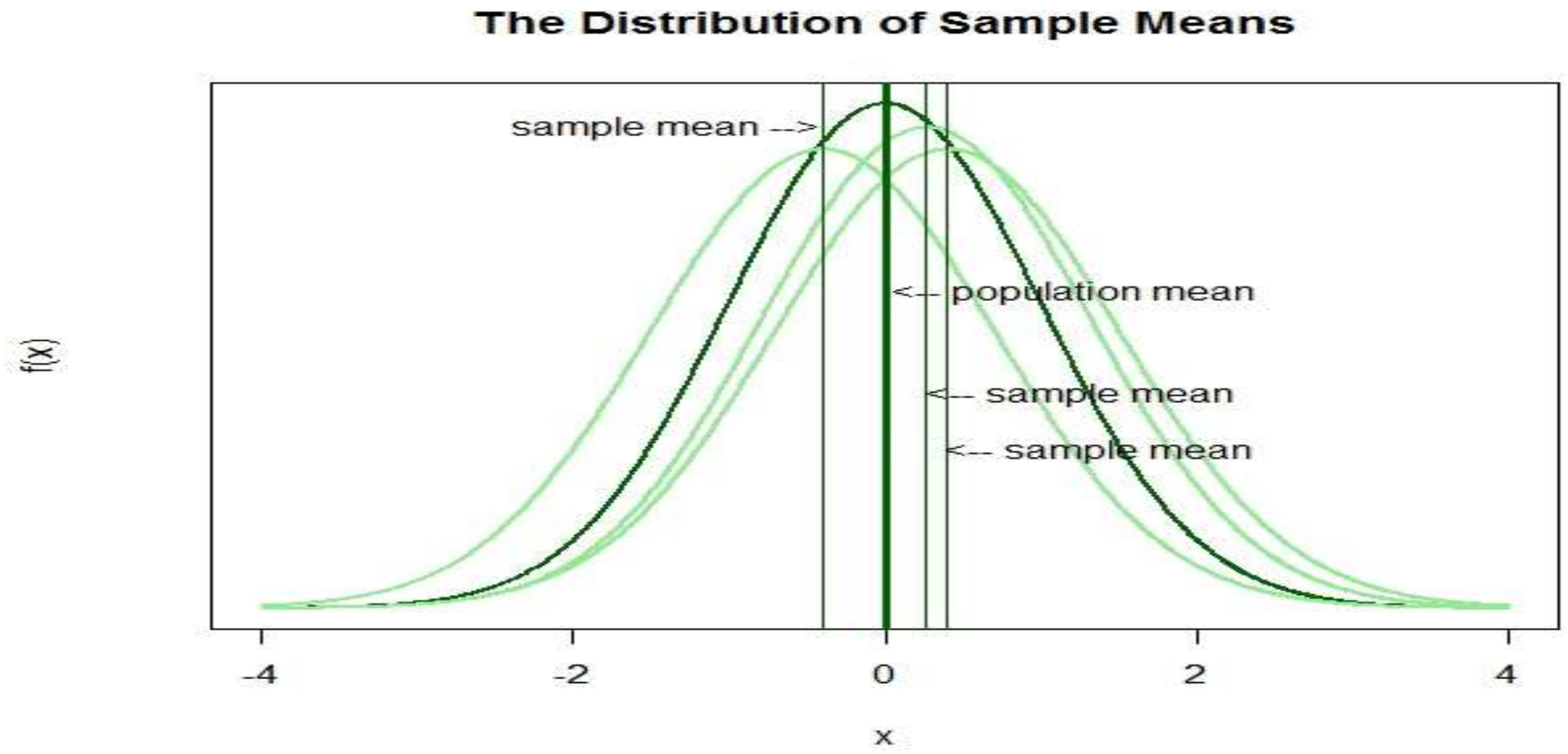
Point estimator of the population variance $\sigma^2$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Courtesy: https://data-flair.training.com

# Sampling Distribution

The distribution of possible outcomes of a sample statistic that would result from repeated sampling from the population



# Assessment (Identify inferences)

# Assessment (Identify inferences)

# References


https://online.stat.psu.edu/stat504/node/16/

http://www.bristol.ac.uk/medical-school/media/rms/red/4_ideas_of_statistical_inference.html

https://bolt.mph.ufl.edu/6050-6052/unit-4/


thank
you!