BRANCH OF MATHS

STATISTICS

WHY STATS ?

Discipline consists of:
- → COLLECTION
- → ORGANISATION
- → ANALYSIS
- → INTERPRETATION
- → PRESENTATION

of DATA.

TOPICS

1. SAMPLE
2. POPULATION
3. MEAN, MEDIAN, MODE
4. STANDARD DIEVIATION
5. VARIATION / VARIANCE
6. IQR → Inter Quartile range
7. Central Limit Theorem
8. Z-Score
9. p-value, Significance value
10. Probability

---

1. Descriptive Stats

2 TYPES

→ Analyze, Summarize data

Central tendency → Mean, Median, Mode

→ SKEWNESS, KURTOSIS ← SHAPE

→ PLOTS → BAR, HISTOGRAM, PIE etc.

→ PDF → Prob. Density function

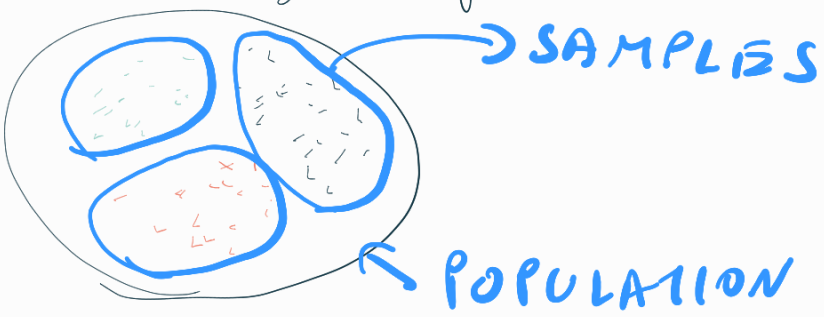
→ CDF → Cumulative Density function

DISPERSION → S.D, Variance, Range

(2) INFERENTIAL STATS

→ Inferencing some info from the entire population.

→ We take some sample and do EDA to get Inference & Conclusion

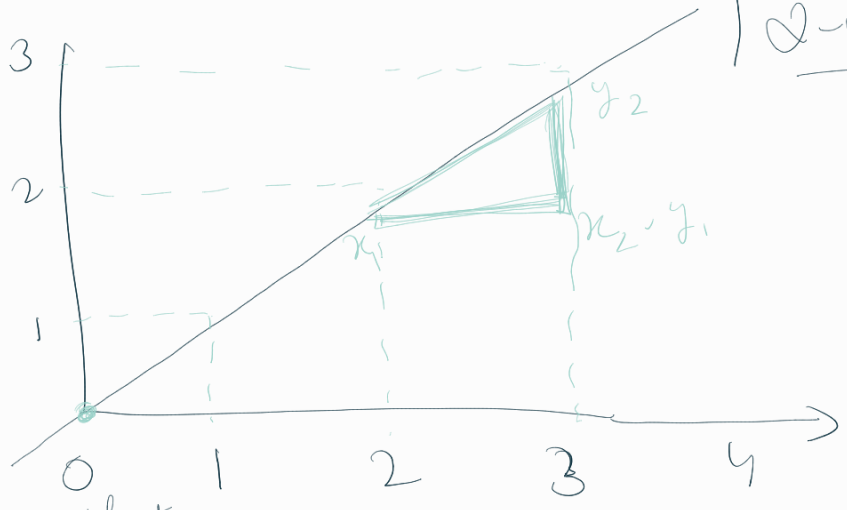


TECHNIQUES
① HYPOTHESIS TESTING
② CONFIDENCE INTERVALS
③ REGRESSION

02/10/2024

LINEAR REGRESSION

CLT
Z score
Standard score
Normal Dist
(P & C)
Q-Q plot, Outliers



$C = 0$

$y = mx + c$

$y = mx$

$m = \frac{dy}{dx}$

$= \frac{y_2 - y_1}{x_2 - x_1}$

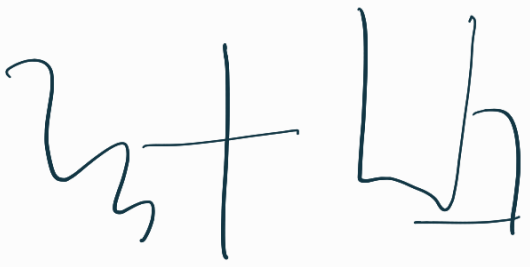
$= \frac{3-2}{3-2} = \frac{1}{1} = 1$

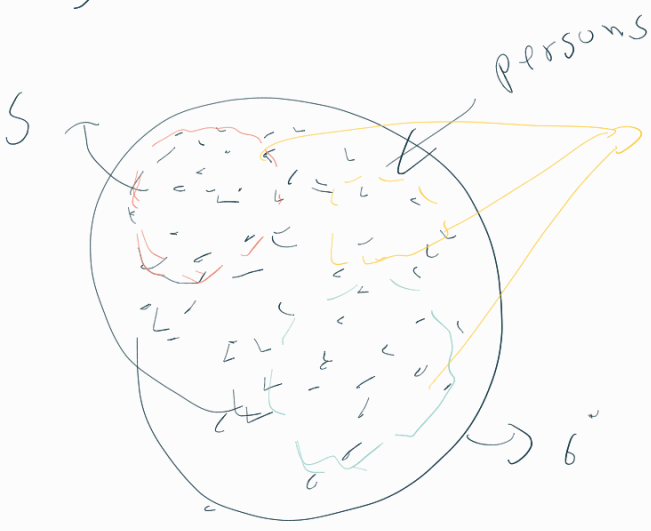
Slope = 1

→ Probability
→ Sample Dist
→ Central Limit Theorem
→ Margin of Error
→ P-value
→ Significance Level
→ Confidence Level
→ Null Hypothesis,
→ Type-I, Type-II Error
→ Linear Reg Correlation, Residuals

SAMPLING

- RANDOM
- STRATIFIED
- CLUSTERING
- SYSTEMATIC





SAMPLES
↳ Avg (μ) of samples

A

SAMPLES

- RANDOM
- STRATIFIED
-

→ Computer science
- DBMS
- O.S
- D.S.A

- MATHS
  - STATS
- ML & D.S

- VISUALIZATIONS
- DATA ANALYSIS