

Statistics



Summary of any dataset

① Descriptive Statistics

→ Description of the entire dataset

(Measures of central tendency)

↳ Mean,

Median,

Mode)

② Inferential Statistics



Inference

↳ Conclusion

(Measures of spread)

→ SD, Variance,

IQR, Range)

Inter-Quartile Range

↳ Max-

Min

Population correlation coefficient

Sample

↳ Experiment

Conclusions

(Data Distribution)

(Plots)

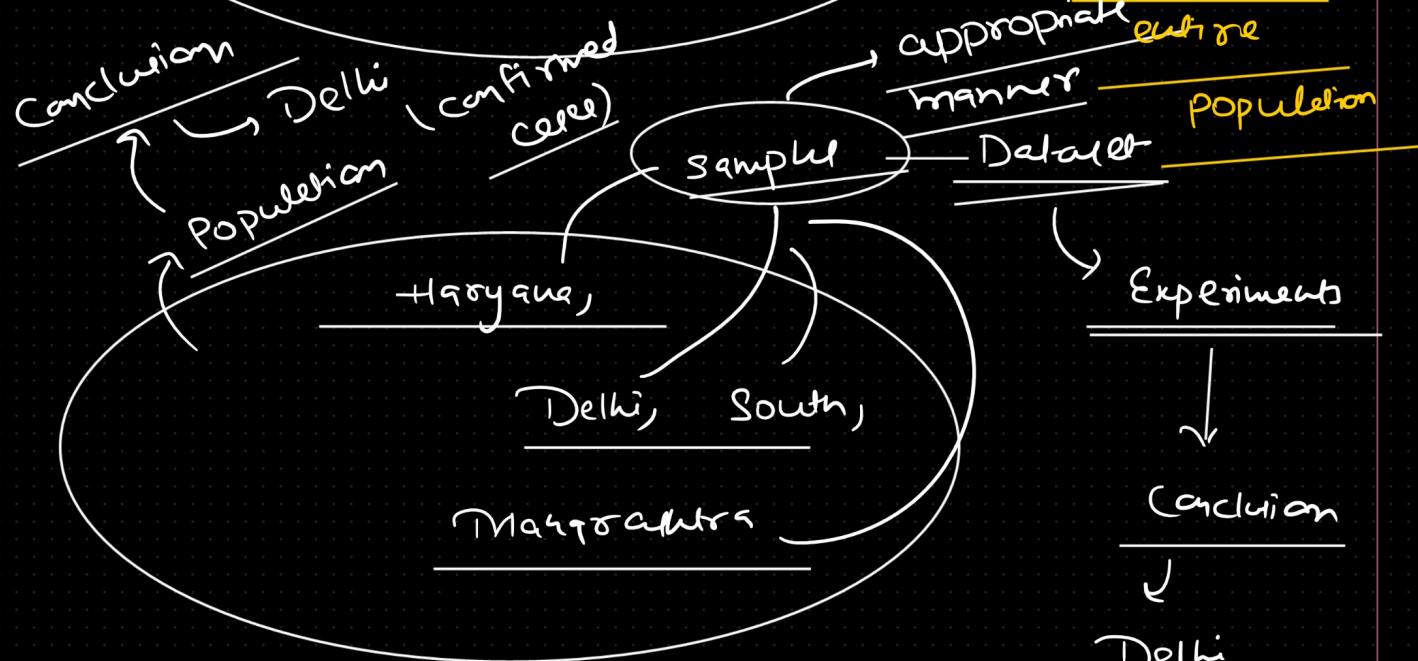
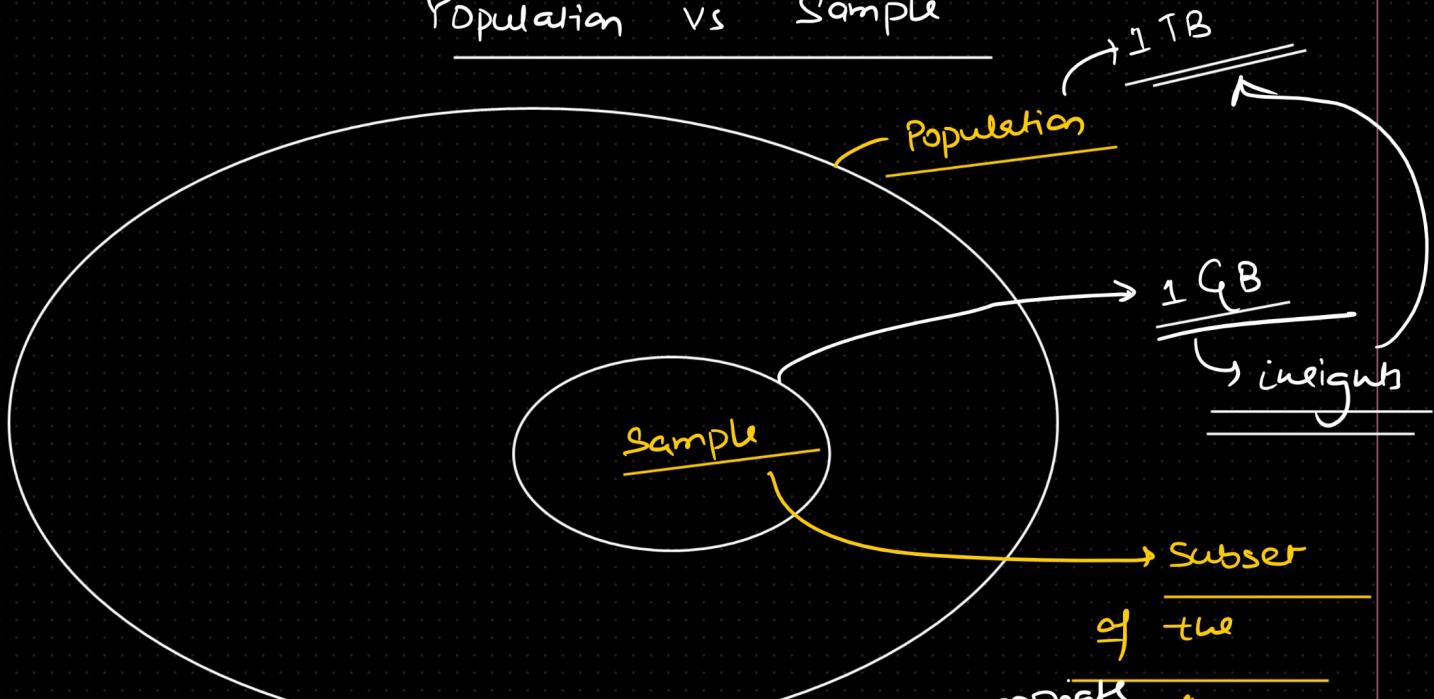
Hypothesis Testing

Statistical Testing

→ t-test, Z-test,

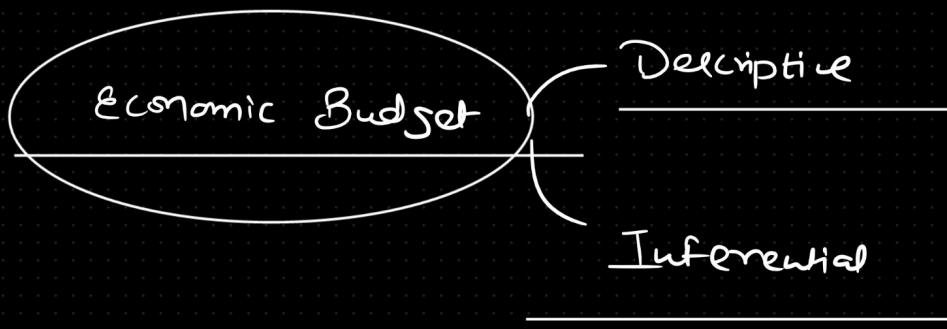
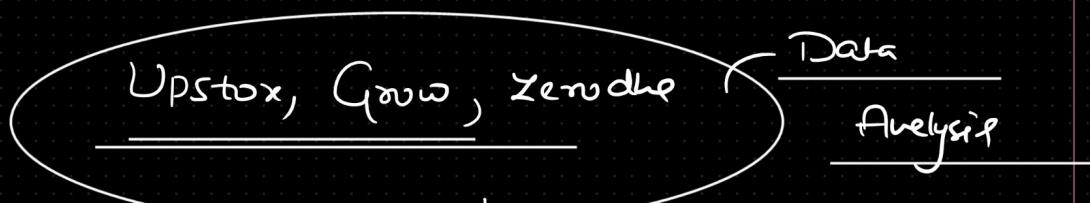
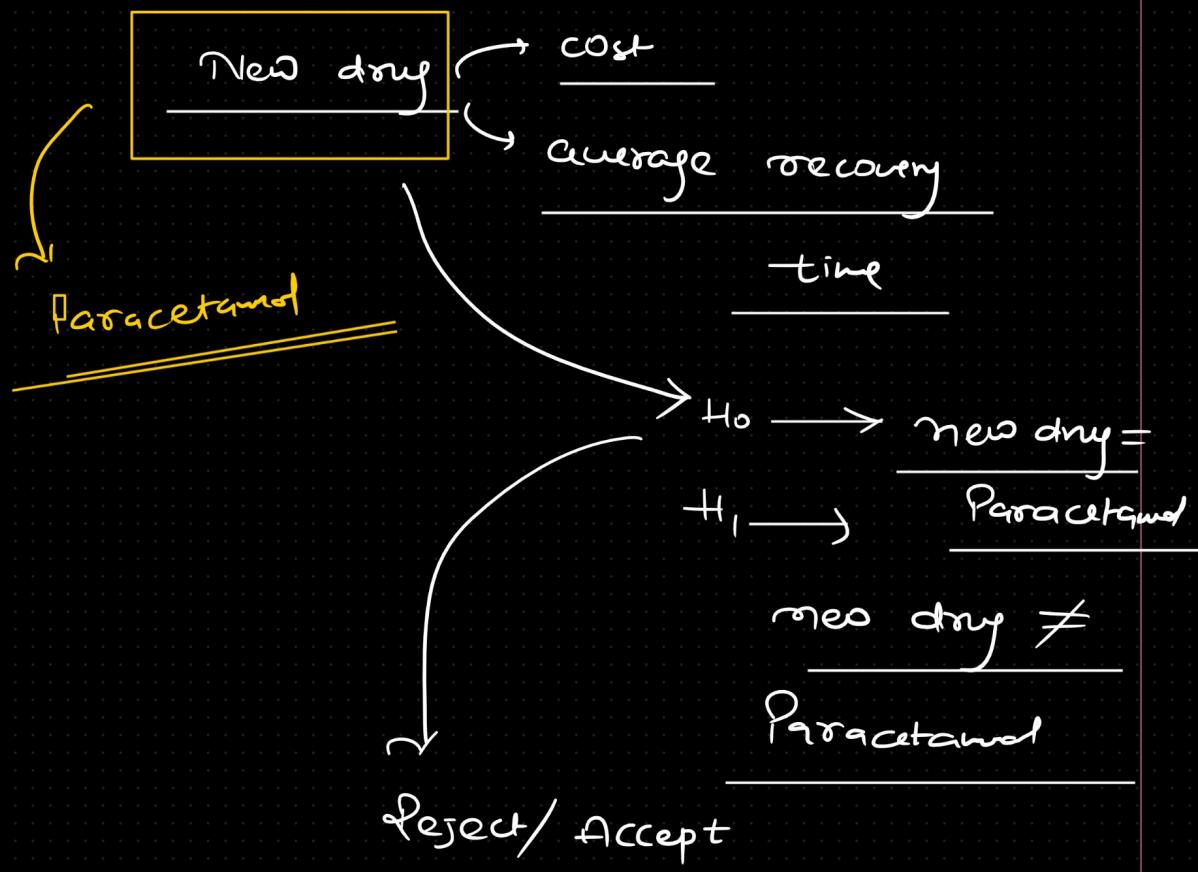
chi-Square test, ANOVA

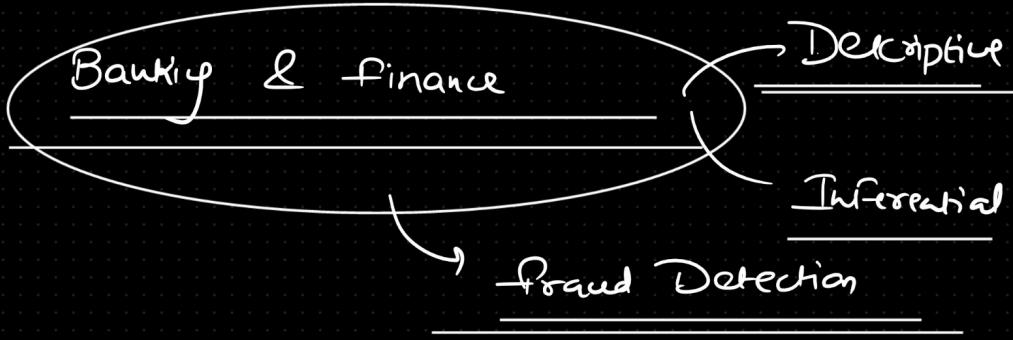
Population vs Sample



↓
Conclusion
↓
Delhi
(high
Confirmed
cases)

Real time → Healthcare Domain

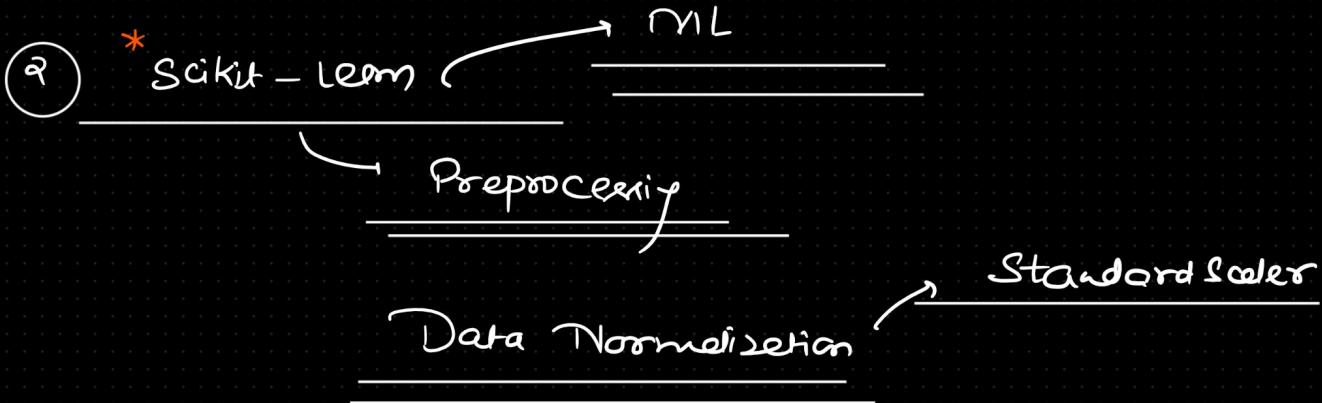
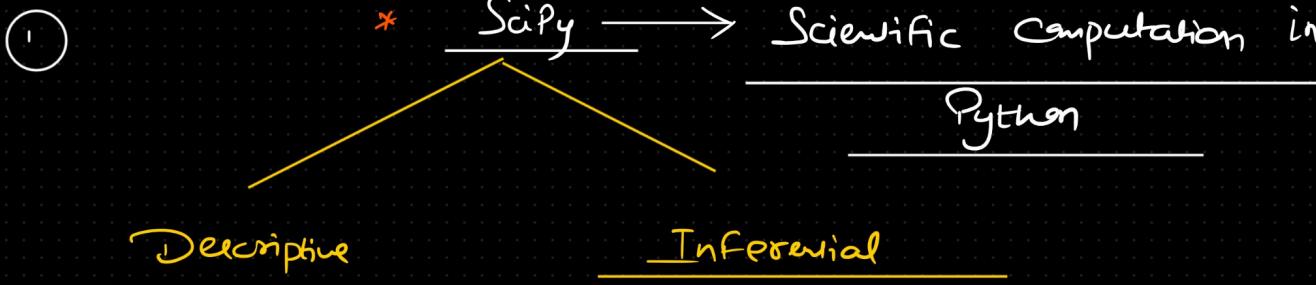




Pole of Statistics → Decision Making

- ① Informed choices
- ② Risk assessment
- ③ Performance Evaluation
- ④ Predictive Analysis

Backend of the data-driven world.



3 StatsModels

Data Science Pipeline

EDA → Exploratory Data Analysis



Dirty data

Types of Data in Real World

Qualitative Data :- Nominal & Ordinal Data

Categorical

data

Color

meaningful

order

④

B.Tech

Higher

⑤

12th

level

③

M.Tech

education

②

Ph.D

Postdoc

①

Quantitative Data → Interval Data

F
=

vs Ratio

Data

equal intervals

But no true

zero

(temperature)

equal

intervals

& a

true zero

point

[In case]