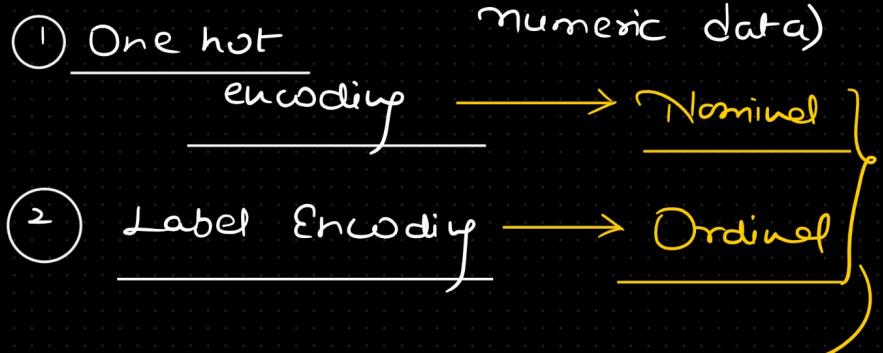


Summary

↳ Encoding Techniques (categorical data to numeric data)



more than
2 categorical/
discrete

↳ Correlation coefficient
(feature selection technique)

heatmap ↪ -1 to +1

$r = 0$ → No correlation among
two features

$f_1 \ \ \ \ \ f_2 \rightarrow 0.95$

↳ Outliers Detection

↳ Box Plot $Q_1 - (1.5 * IQR)$
 (5 Points) $\uparrow (1.5 * IQR)$
 $Q_3 +$

Q_1, Q_2, Q_3 , lower, upper
whisker whisker

Visualizations

→ Matplotlib & seaborn

Key Points :-

→ D outlier Detection → Box Plot, Violin Plot

→ Relationship among two variables → Scatterplot, relplot

→ Relationship among all the variables in one go → Pairplot

→ Correlation Coefficient → Heatmap
(Feature Selection)

→ Compare two numeric data Points → Bar Plot

→ Distribution of → distplot
any feature

(Symmetric or
skewed)

→ Compare the
Performance of → catplot
Categorical variables
in the x-axis

Random Variables

↳ variables that can take
different values depending
upon the outcome of
any experiment.

Rolling a

↳ Discrete Random Variables ↳ dice

↳ Continuous Random Variables

1)

temp of a

room

Probability Distribution Functions

PDF

PMF

* Probability

Density function

Continuous RV

Probability

Mass

function

Discrete RV

{ * Normal Distribution

→ Bernoulli Distribution

* Log Normal Distribution

↳ Binomial

Distribution

PDF

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

→ Normal Distribution

$\mu \rightarrow$ mean

$\sigma \rightarrow$ Standard deviation

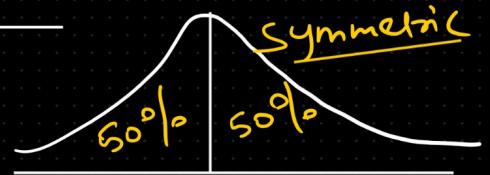
age

24
27
23
34
47

Normal →

Distribution

Normal
Distribution



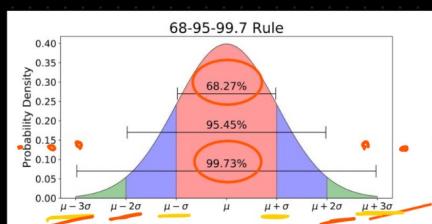
mean =

median =

mode

Empirical Rule

detect the outliers



z-score

Population

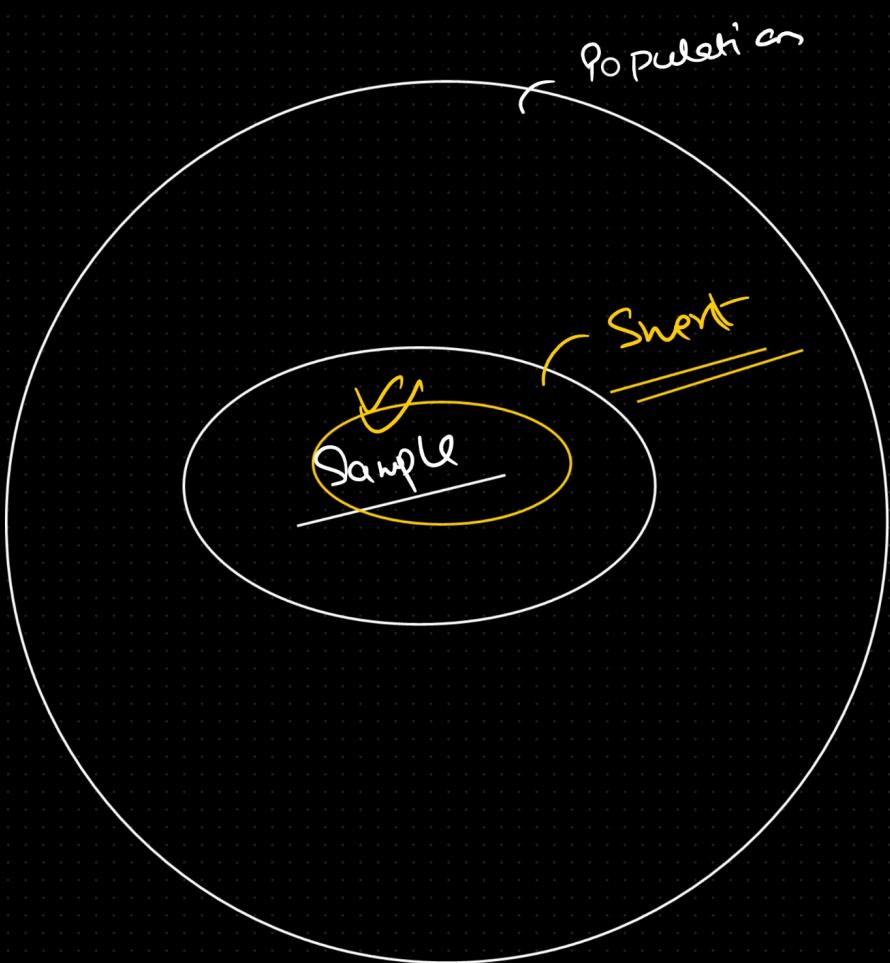
μ

Sample

σ

s

\bar{x}



- {
- Population & Sample
 - CLT
 - Hypothesis Testing