Page No.
Date

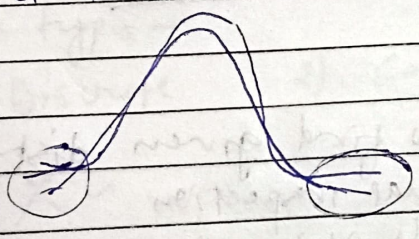
Session -42

- Statistical moment -
  - numerical measures that describe shape & characteristics of data

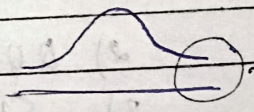
They are of 4 types-

1) First moment → mean
2) Second " → variance
3) Third " → skewness
4) fourth " → kurtosis

- Kurtosis -
  - It tells about intensity of taildness



- Its use -
  - Profit dist. =>



means lots of outliers. Lots of people have made profit & lots of people has made loss

- Excess kurtosis
  - It compares Normal dist. tale with the given graph tale
  - Sample - 3

i) Lepto kurtic

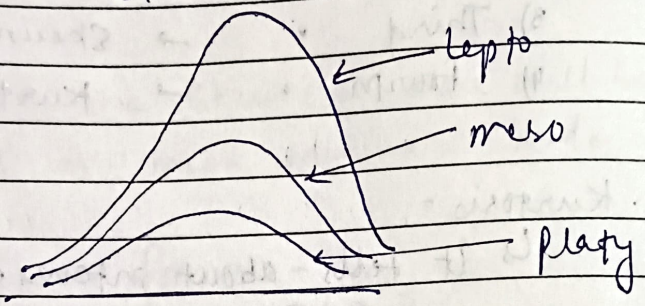
Sample - 3 > 0

→ more outlier (as fatter tale)

2) Platykurtic Platykurtic -
Sample - 3 < 0

→ less outlier

3) Mesokurtic
is excess kurtosis = 0



. QQ Plot :-
. How to find given dist is normal or not?
1) Visual inspection
is plot histogram
2) QQ plot.
3) Statistical tests

→ Quantile - Quantile plot is used to see the similarities of dist. of two sets of data

[x , y]

x ↑ given data

y ↑ theoritical data (Assume normal) distribution

Page No.
Date

Steps -

- Take theoritical data (Assume Normal)
  - Sort it and find quantile
- Take actual data (X)
  Sort it & find quantile

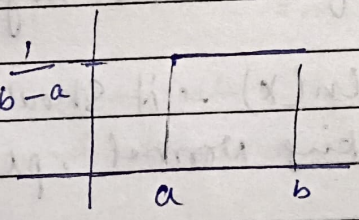
Plot graph using (Xquantile, Yquantile)

→ After plotting if they both form a straight line then dist. of x is same as theoritical (Normal)

- Uniform distribution -
  - probability of all outcomes is equal
  - 2 types -
    1) Discrete 2) Continuous

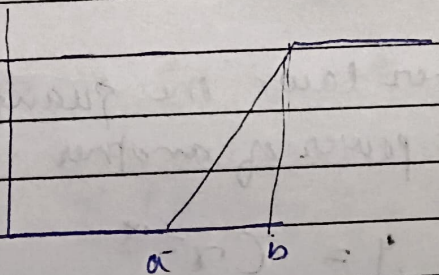
$X \sim u(a, b)$ a → lower value, b → higher

- PDF



$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

- CDF

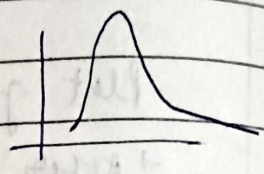


- Skewness = 0 (as it is symmetric)

• Log Normal Distribution -

↳ a Right skewed dist. whose logarithmic is Normal dist. is called log Normal dist

→ parameter - μ, σ



eg/ ↳ The comments in social media platform.

$X \sim \text{lognormal}(\mu, \sigma)$

$\ln(X) \sim N(\mu', \sigma')$

$$\text{PDF} \rightarrow \frac{1}{x\sigma\sqrt{2\pi}} e^{\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)}$$

- It is skewed.

• How to check if a R.V is log normally dist?

↳ take $\ln(x)$ - it should be $N(\mu, \sigma)$

- for checking Normal, plot QQ graph

• Pareto Distribution -

→ Power law - one quantity changes as a power of another

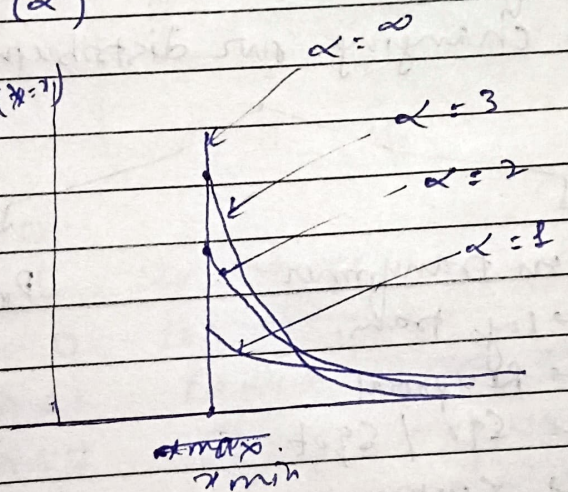
$$y = Cx^{-\alpha}$$

- Pareto dist. is based on power law.
  - This dist. is used when small no. of cases account for large effect

→ 80:20 (for $\alpha = 1.16$)
- 80% of wealth is to 20% of people
& 20% of wealth is to 80% of people

$X \sim Pr(\alpha)$

- PDF $Pr(x=?)$



more the value of α → more the height and thinner the tail.

- at $\alpha = \infty$ we get Pareto graph.

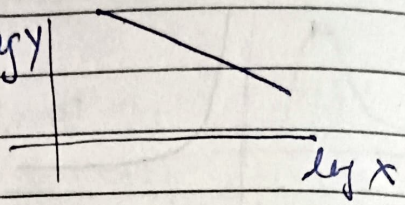
eg:- human settlement in areas

$$P.D.F = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}} \qquad x \geq x_m$$

- Its skewed.

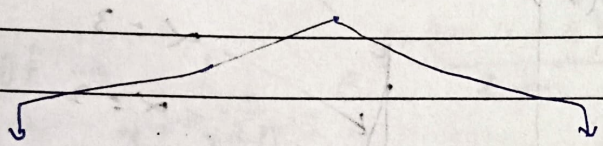
- How to detect Pareto dist ?
  ↳ use log-log plot

$$y = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}$$



↳ using Q-Q plot :-

- Transformations -
  ↳ changing our distribution to normal distribution



function transformer
- log trans.
- Reciprocal
- sqr / sqrt
- custom

Power trans.
- Box - cox
- Yeo - Johnson

- Log transform -
  ↳ take log of all the values and plot it
  ↳ it will be like normal dist

  ↳ It will not applied on -ve value
  ↳ It's good (applied) in right skewed graph

- Reciprocal transform -
  - $1/x$ we do reciprocal of each value
- Square (x^2) transform.
  - good for left skewed
- Square-root -
  - do square root of each value
- Box-Cox -

$$x_i^{\lambda} = \begin{cases} \dfrac{x_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(x_i) & \text{if } \lambda = 0 \end{cases}$$

  - $x > 0$
  - $\lambda = 1$ No change
  - $\lambda = 0$ log transform
  - $0 < \lambda < 1$ Reduces right skew
  - $\lambda < 0$ Strong compression
  - $\lambda > 1$ Expand large values

$-5 \leq \lambda \leq 5$

- Yeo-Johnson
  - can work for -ve values of x also