SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

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

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

Coefficient of Variation

Department of Computer Applications

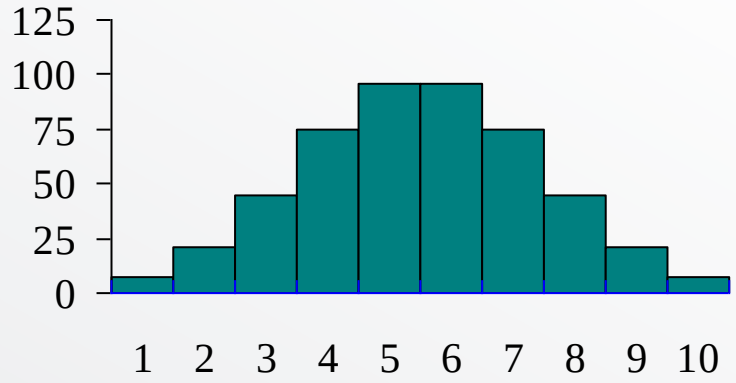
Course : 23CAT705- RESEARCH METHODOLOGY

UNIT II : STATISTICS

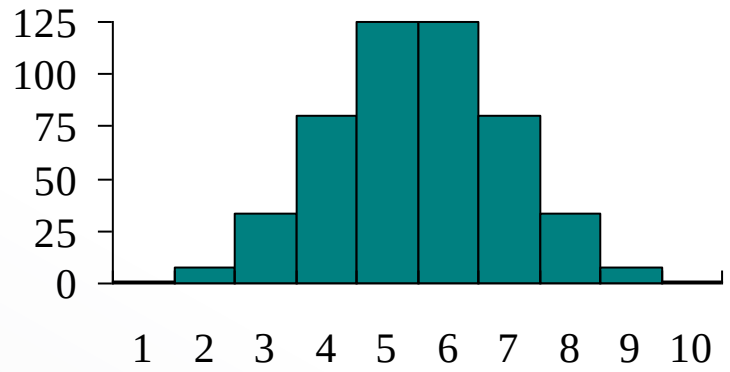
Class : II MCA / III SEMESTER

Measure of Dispersion

- *It* describe how similar a set of data are to each other
  - more similar the data are to each other, the lower the measure of dispersion
  - less similar the data are to each other, the higher the measure of dispersion



It has more dispersion because the data are more spread out. ie they are less similar to each other



Measure of Dispersion

Range	Difference between the largest score in the set of data
Semi-interquartile range	Difference of the first and third quartiles divided by two
Variance/ Standard Deviation	Average of the square deviations

Range

Example

- Range of the data:

  4 8 1 6 6 2 9 3 6 9

- The largest is 9; the smallest is 1;

  the range is 9 - 1 = 8

Purpose

- It is used when
  - If you have ordinal data or
  - Presenting finds to people with little/no knowledge of statistics
- The range is rarely used in scientific work as it is fairly insensitive because
  - It depends on only two scores in the set of data, X_L and X_S
  - Two very different sets of data can have the same range:

Semi-interquartile range (SIR)

Example

25 % of the scores are below 5
5 is the first quartile
25 % of the scores are above 25
25 is the third quartile
SIR = $(Q_3 - Q_1) / 2 = (25 - 5) / 2 = 10$

2
4 ← 5 = 25th %tile
6
8
10
12
14
20 ← 25 = 75th %tile
30
60

SIR = $(Q_3 - Q_1) / 2$

Purpose

- is often used with skewed data as it is insensitive to the extreme scores

Variance

- It is defined as the average of the square deviations
- **Deviate** tells us how far a given score is from the typical, or average, which can be calculated by subtracting the mean from each of the data
- Deviate is a measure of dispersion for a given score

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

Standard Deviation

- Standard deviation is the square root of variance
- Standard deviation = √variance
- Variance = standard deviation²

$$\sigma^2 = \frac{\sum(X-\mu)^2}{N}$$

$$= \frac{12}{6}$$

$$= 2$$

Example

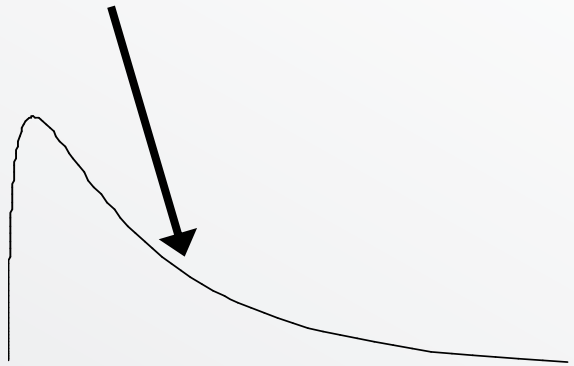
X	X^2	X-μ	$(X-\mu)^2$
9	81	2	4
8	64	1	1
6	36	-1	1
5	25	-2	4
8	64	1	1
6	36	-1	1
Σ = 42	Σ = 306	Σ = 0	Σ = 12

Measures of Skew

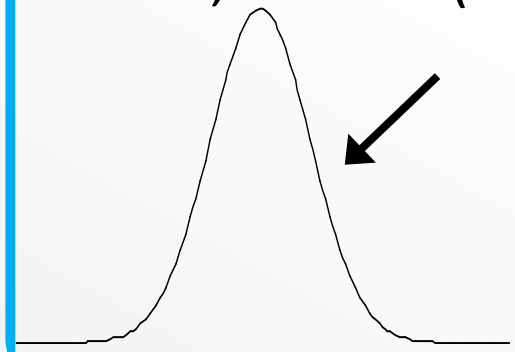
- A measure of symmetry in the distribution of data

$$S^3 = \frac{\frac{\sum(X-\bar{X})^3}{N}}{\sqrt{\frac{\sum(X-\bar{X})^2}{N}}}$$

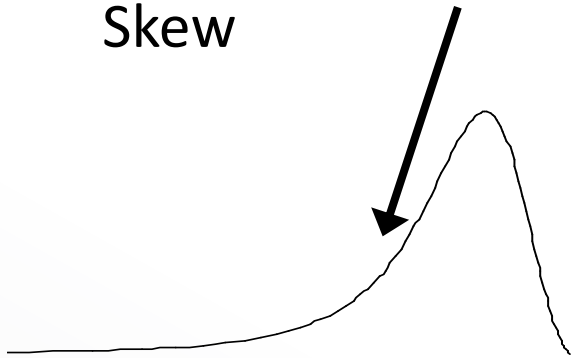
If $s^3 > 0$, Positive Skew



If s=0, Normal (skew = 0)



If $s^3 < 0$, Negative Skew

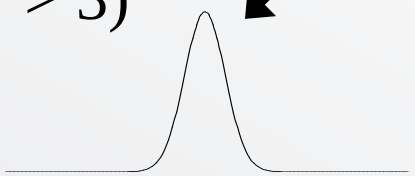


Kurtosis

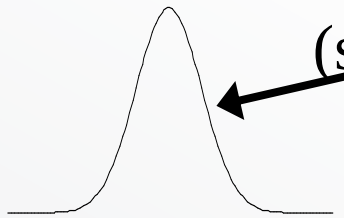
- It measures whether the scores are spread out more or less than they would be in a normal (Gaussian) distribution

$$s^4 = \frac{\sum \left(\frac{X - \bar{X}}{\sqrt{\frac{\sum (X - \bar{X})^2}{N}}} \right)^4}{N}$$

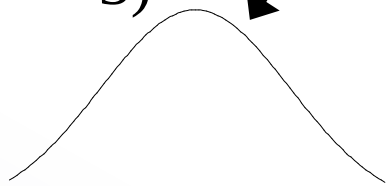
Leptokurtic ($s^4 > 3$)



Mesokurtic ($s^4 = 3$)

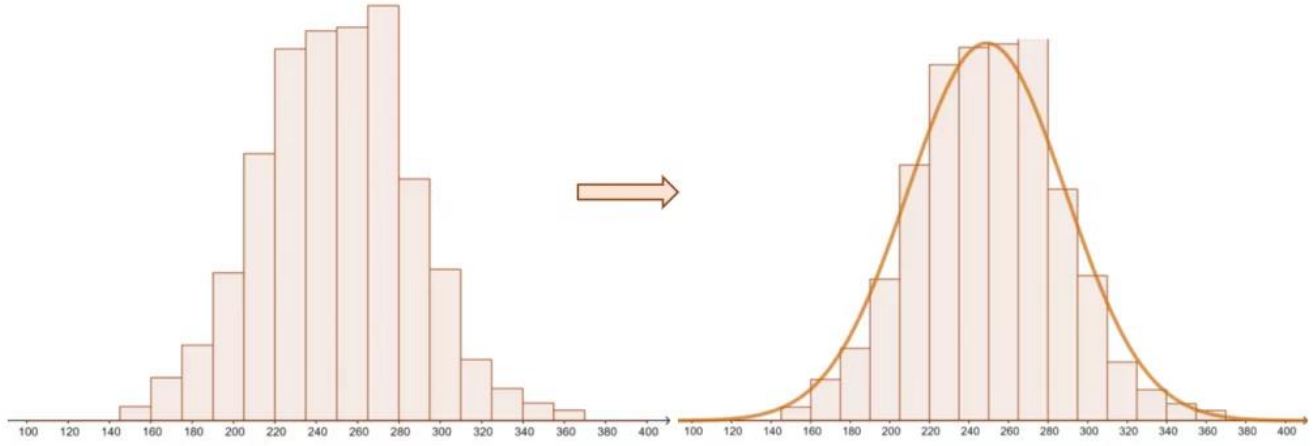


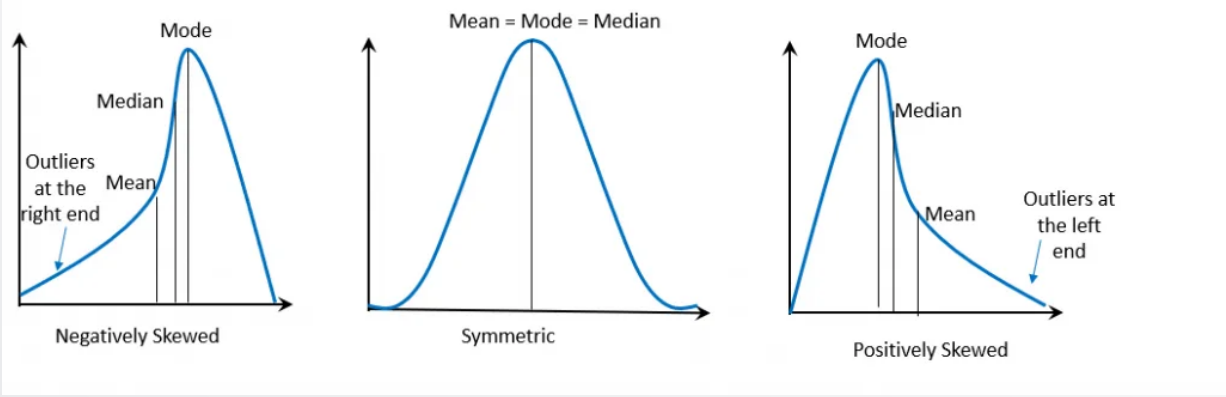
Platykurtic ($s^4 < 3$)

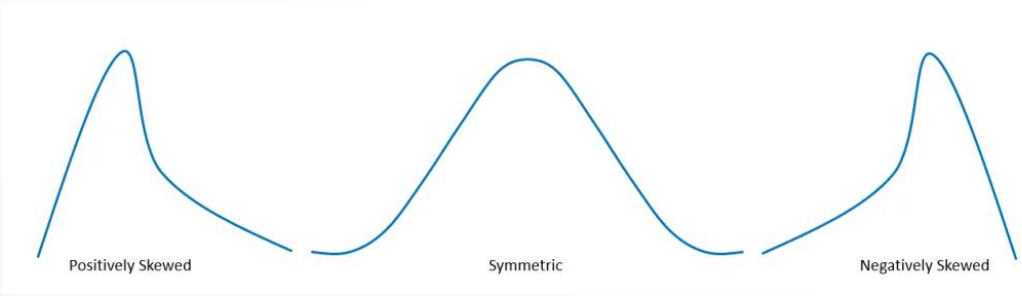


Shape of Distribution

Collectively, the variance (s^2), skew (s^3), and kurtosis (s^4) describe the shape of the distribution







References

1. Kothari, C.R. &Garg, G. (2019). *Research Methodology: Methods and Techniques*. New Age International Publishers, New Delhi

2. Goode, W.J. &Hatt, P.K. (2022). Methods in Social Research. McGraw Hill, London

3. Bhandarkar, P.L. & Wilkinson, T.S. (2016). Methodology and Techniques of Social Research. Himalaya Publishing House, Mumbai.

Thank You