IIT ROORKEE

NPTEL ONLINE CERTIFICATION COURSE

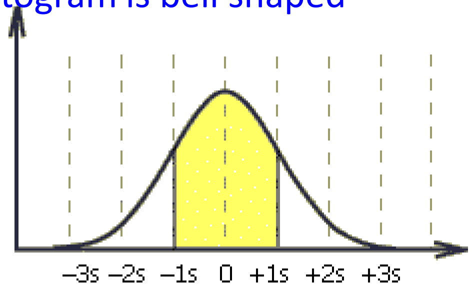
Lecture 5: Central Tendency and Dispersion- II

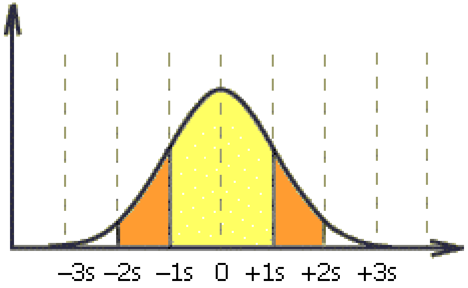
Dr. A. Ramesh

Department of Management Studies

The Empirical Rule… If the histogram is bell shaped

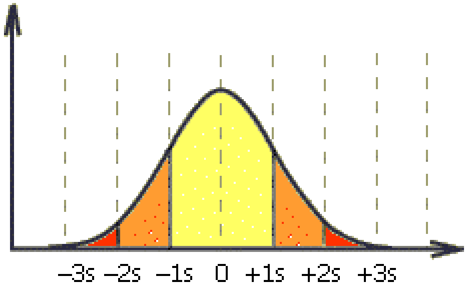
- Approximately 68% of all observations fall within **one** standard deviation of the mean.





- Approximately 95% of all observations fall within **two** standard deviations of the mean.

- Approximately 99.7% of all observations fall within **three** standard deviations of the mean.



Empirical Rule

- Data are normally distributed (or approximately normal)

Distance from the Mean	Percentage of Values Falling Within Distance
$\mu \pm 1\sigma$	68
$\mu \pm 2\sigma$	95
$\mu \pm 3\sigma$	99.7

Chebysheff's Theorem…Not often used because interval is very wide.

- A more general interpretation of the standard deviation is derived from ***Chebysheff's Theorem***, which applies to all shapes of histograms (not just bell shaped).
- The proportion of observations in any sample that lie within **k** standard deviations of the mean is *at least:*

$$1 - \frac{1}{k^2} \quad for \quad k > 1$$

For k=2 (say), the theorem states that at least 3/4 of all observations lie within 2 standard deviations of the mean. This is a "lower bound" compared to Empirical Rule's approximation (95%).

Coefficient of Variation

- Ratio of the standard deviation to the mean, expressed as a percentage
- Measurement of relative dispersion

$$C.V. = \frac{\sigma}{\mu}(100)$$

Coefficient of Variation

$$\mu_1 = 29$$

$$\sigma_1 = 4.6$$

$$C.V._1 = \frac{\sigma_1}{\mu_1}(100)$$

$$= \frac{4.6}{29}(100)$$

$$= 15.86$$

$$\mu_2 = 84$$

$$\sigma_2 = 10$$

$$C.V._2 = \frac{\sigma_2}{\mu_2}(100)$$

$$= \frac{10}{84}(100)$$

$$= 11.90$$

Variance and Standard Deviation of Grouped Data

Population	Sample
$\sigma^2 = \dfrac{\sum f\left(M-\mu\right)^2}{N}$	$S^2 = \dfrac{\sum f\left(M-\bar{X}\right)^2}{n-1}$
$\sigma = \sqrt{\sigma^2}$	$S = \sqrt{S^2}$

Population Variance and Standard Deviation of Grouped Data(mu=43)

Class Interval	f	M	fM	$M-\mu$	$(M-\mu)^2$	$f(M-\mu)^2$
20-under 30	6	25	150	-18	324	1944
30-under 40	18	35	630	-8	64	1152
40-under 50	11	45	495	2	4	44
50-under 60	11	55	605	12	144	1584
60-under 70	3	65	195	22	484	1452
70-under 80	1	75	75	32	1024	1024
	50		2150			7200

$$\sigma^2 = \frac{\sum f(M-\mu)^2}{N} = \frac{7200}{50} = 144$$

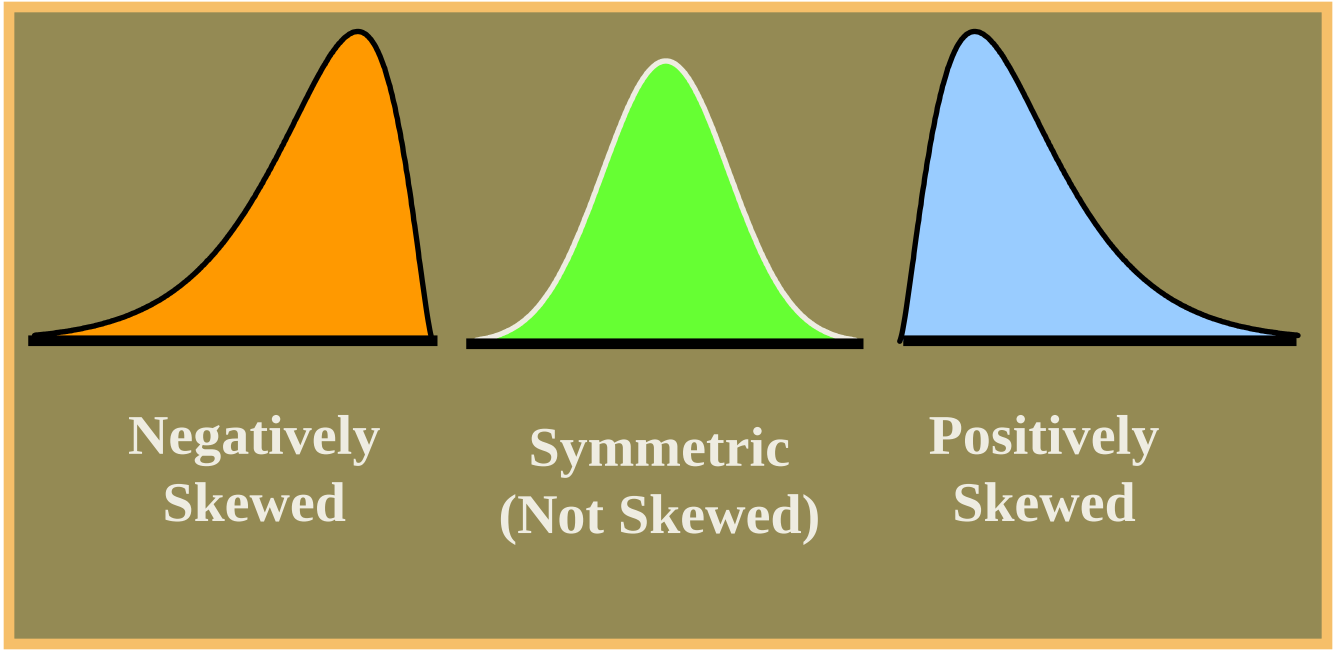
$$\sigma = \sqrt{\sigma^2} = \sqrt{144} = 12$$

Measures of Shape

- **Skewness**
  - Absence of symmetry
  - Extreme values in one side of a distribution
- **Kurtosis**

  Peakedness of a distribution
  - Leptokurtic: high and thin
  - Mesokurtic: normal shape
  - Platykurtic: flat and spread out
- **Box and Whisker Plots**
  - Graphic display of a distribution
  - Reveals skewness

Skewness

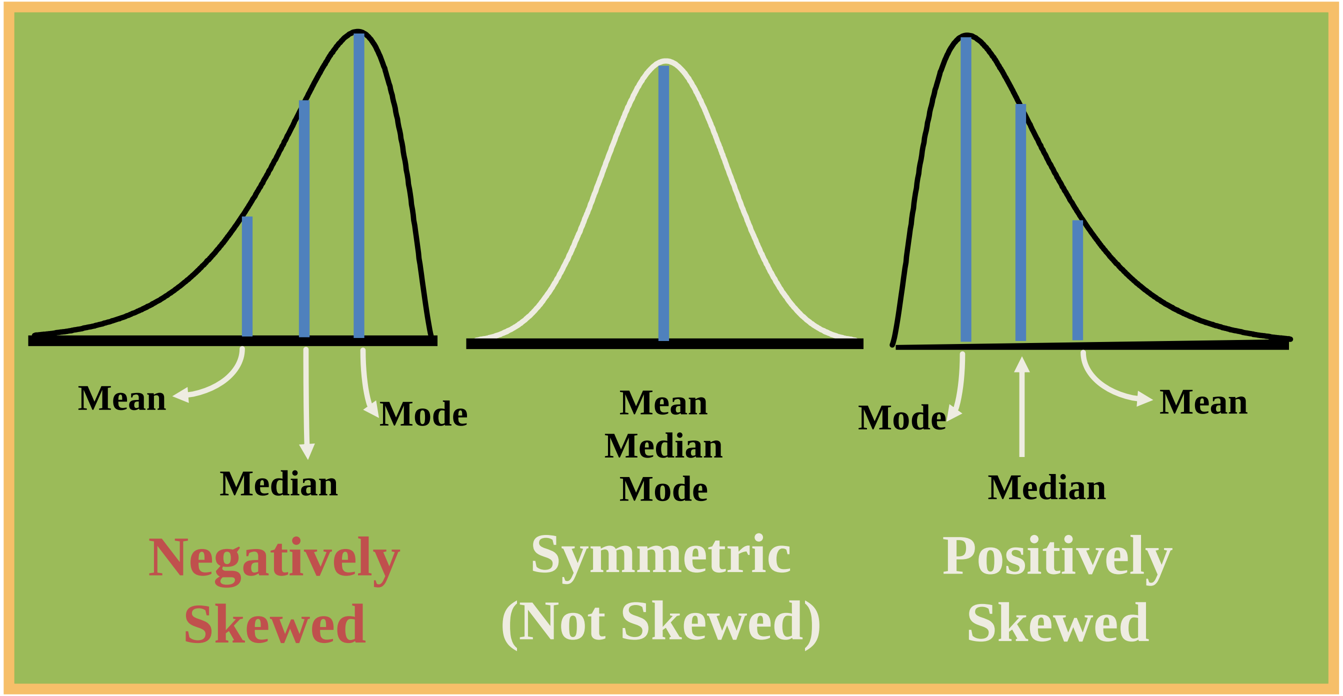

Negatively Skewed
Symmetric (Not Skewed)
Positively Skewed

Skewness..

The *skewness* of a distribution is measured by comparing the relative positions of the mean, median and mode.

- Distribution is *symmetrical*
  - ***Mean = Median = Mode***
- Distribution *skewed right*
  - ***Median lies between mode and mean, and mode is less than mean***
- Distribution *skewed left*
  - ***Median lies between mode and mean, and mode is greater than mean***

Skewness


Mean
Mode
Median
Mean
Median
Mode
Mode
Mean
Median
Negatively Skewed
Symmetric (Not Skewed)
Positively Skewed

Coefficient of Skewness

- Summary measure for skewness

$$\boxed{S = \frac{3(\mu - M_d)}{\sigma}}$$

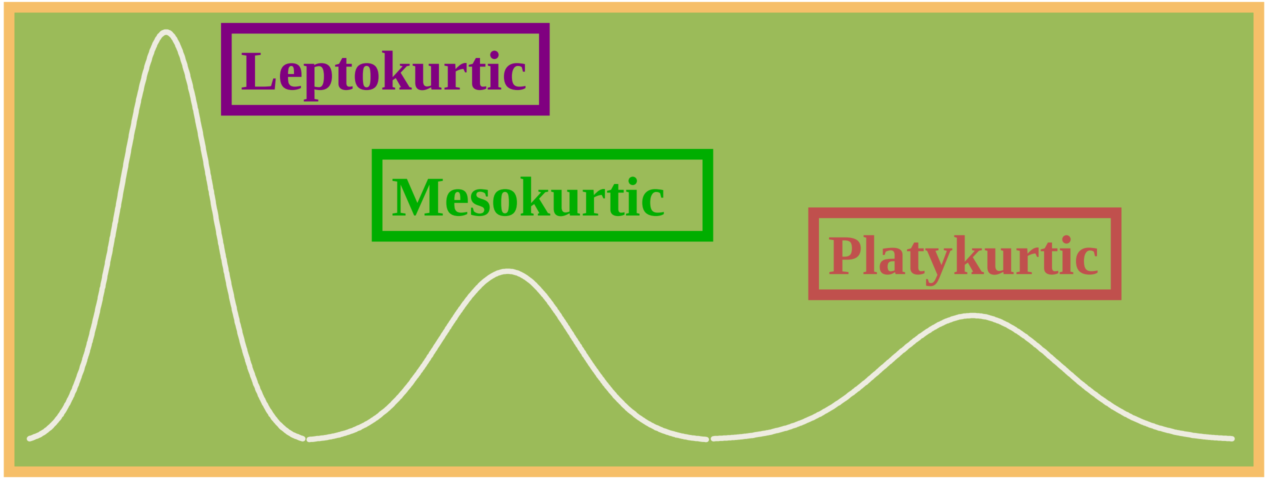
- If S < 0, the distribution is <u>negatively skewed</u> (skewed to the left)
- If S = 0, the distribution is <u>symmetric</u> (not skewed)
- If S > 0, the distribution is <u>positively skewed</u> (skewed to the right)

Coefficient of Skewness

$\mu_1 = 23$	$\mu_2 = 26$	$\mu_3 = 29$
$M_{d1} = 26$	$M_{d2} = 26$	$M_{d3} = 26$
$\sigma_1 = 12.3$	$\sigma_2 = 12.3$	$\sigma_3 = 12.3$
$S_1 = \dfrac{3(\mu_1 - M_{d1})}{\sigma_1}$	$S_2 = \dfrac{3(\mu_2 - M_{d2})}{\sigma_2}$	$S_3 = \dfrac{3(\mu_3 - M_{d3})}{\sigma_3}$
$= \dfrac{3(23-26)}{12.3}$	$= \dfrac{3(26-26)}{12.3}$	$= \dfrac{3(29-26)}{12.3}$
$= -0.73$	$= 0$	$= +0.73$

Kurtosis

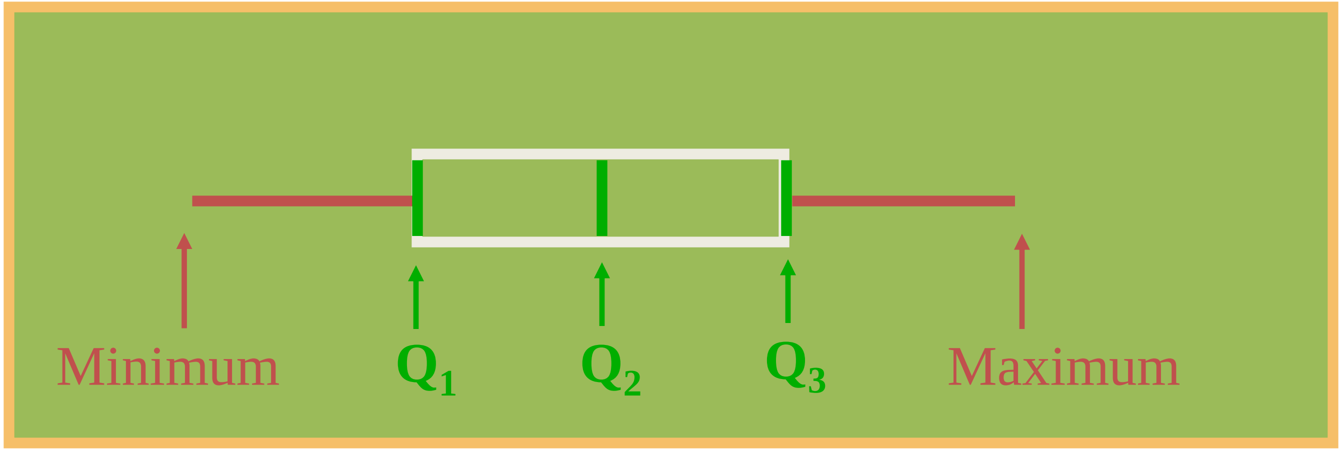
- Peakedness of a distribution
  - Leptokurtic: high and thin
  - Mesokurtic: normal in shape
  - Platykurtic: flat and spread out



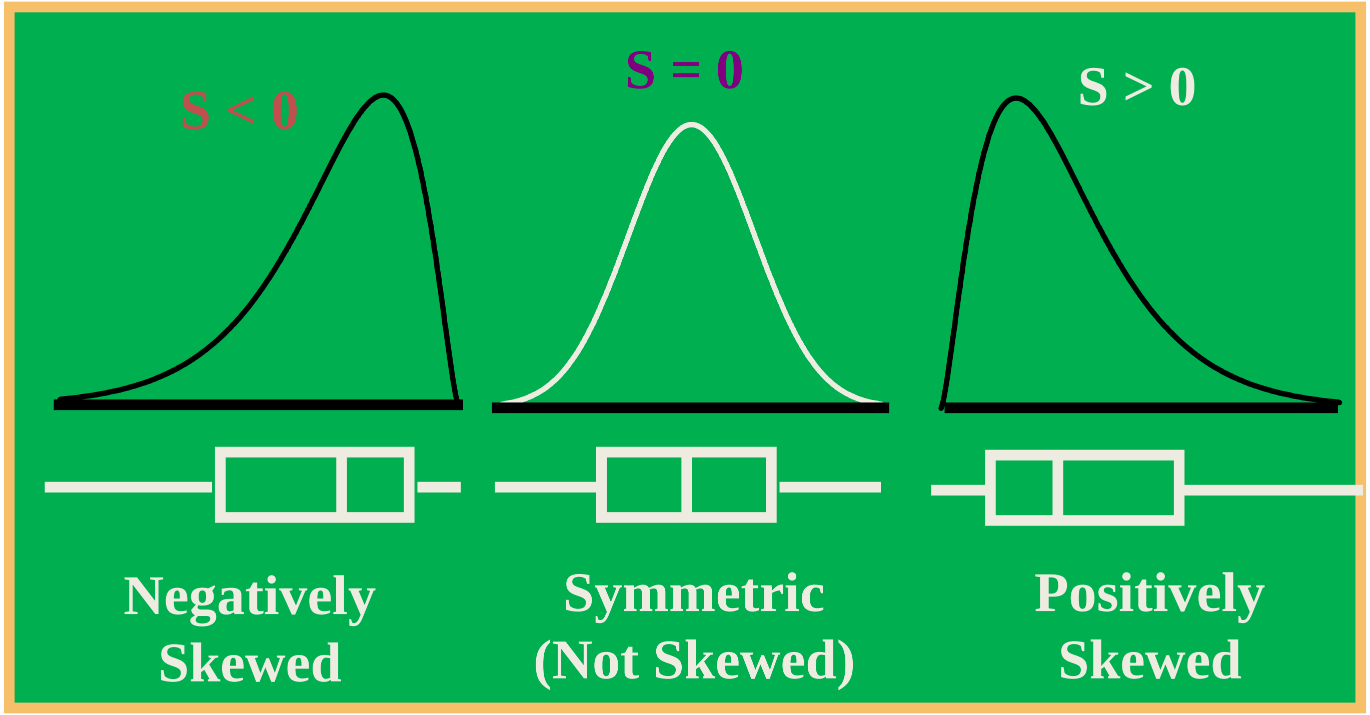
Box and Whisker Plot

- Five specific values are used:
  - Median, Q_2
  - First quartile, Q_1
  - Third quartile, Q_3
  - Minimum value in the data set
  - Maximum value in the data set

Box and Whisker Plot



Skewness: Box and Whisker Plots, and Coefficient of Skewness



THANK YOU