CHAPTER IV

MEASURES OF DISPERSION

MEASURES OF DISPERSION

- Two facts:
- The distribution may have similar averages, but the degree of scatterdness in the individual values is different.
- The distribution may have different averages, but the variation among the values may be quite less

Production of Teff (In Metric Tones Per hectare) in three fields.

Year	Field'A'(X1)	Field `B'(x2)	Field`C'(X3)
1984	8	6	2
1985	8	13	8
1986	8	8	30
1987	8	13	28
1988	8	15	10
1989	8	11	2
1990	8	7	0
1991	8	6	0
1992	8	0	0
1993	8	1	0
	$\Sigma X_1 = 80$	$\Sigma X_2 = 80$	$\Sigma X_3 = 80$

Arithmetic mean of:

Field `A' $= \frac{\sum x_1}{n} = \frac{80}{10} = 8$

Field `B' $= \frac{\sum x_2}{n} = \frac{80}{10} = 8$

Field `C' $= \frac{\sum x_3}{n} = \frac{80}{10} = 8$

Year / Firms	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
Yassem (X1)	6	10	14	18	22	26	30	34	38	42
Tianes (X2)	16	20	24	28	32	36	40	44	48	52
Winner Cap (X3)	150	154	158	162	166	170	174	178	182	186

$$\overline{x_1} = \frac{\sum x_1}{n} = \frac{240}{10} = 24$$

$$\overline{x_2} = \frac{\sum x_2}{n} = \frac{340}{10} = 34$$

$$\overline{x_3} = \frac{\sum x_3}{n} = \frac{1680}{10} = 168$$

Year	Firms					
	Yassem	**Variation**	**Tianes**	**Variation**	**WinnerCap**	**Variation**
	$X_1=24$		$X_2=34$		$X_3=168$	
1984	**6**	**-18**	**16**	**-18**	**150**	**-18**
1885	**10**	**-14**	**20**	**-14**	**154**	**-14**
1986	**14**	**-10**	**24**	**-10**	**158**	**-10**
1987	**18**	**-6**	**28**	**-6**	**162**	**-6**
1988	**22**	**-2**	**32**	**-2**	**166**	**-2**
1989	**26**	**+2**	**36**	**+2**	**170**	**+2**
1990	**30**	**+6**	**40**	**+6**	**174**	**+6**
1991	**34**	**+10**	**44**	**+10**	**178**	**+10**
1992	**38**	**+14**	**48**	**+41**	**182**	**+14**
1993	**42**	**+18**	**52**	**+18**	**186**	**+18**
	$\Sigma x_1 =$ **240**	**= 0**	$\Sigma x_2 =$ **340**	**=0**	$\Sigma x_3 =$ **1680**	**= 0**

the average production of the firm's Yassem, Tianes, and WinnerCap is 24,000 computers, 34,000 computers, and 168000 computers respectively.

Types of Measures of Dispersion

1. Distance Measures/Positional Measures
2. Measures of deviations

We can use

Numerical methods

(i) Range
(ii) Quartile deviation,
(iii) Mean deviation
(iv) Standard deviation.

Graphical Methods

(v) Normal Distribution Curve (Lorenz Curve)

The range

- Simply the difference between the largest and smallest values in a set of data
- Useful for: daily temperature fluctuations or share price movement
- Is considered primitive as it considers only the extreme values which may not be useful indicators of the bulk of the population.
- The formula is:

Range = largest observation - smallest observation

For ensuring comparability, relative measure of range, known as coef[illegible] d.

$$\text{Coefficient of Range } \frac{x_{max} - x_{min}}{x_{max} + x_{min}} = \frac{L - S}{L + S}$$

Interquartile range

- Measures the range of the middle 50% of the values only
- Is defined as the difference between the upper and lower quartiles

Inter-quartile range = upper quartile - lower quartile

$$= Q_3 - Q_1$$

QUARTILE DEVIATION

- Is the average absolute difference between the third (upper) quartile and the first (lower) quartile of the frequency distribution.

$$Q.D = \frac{Q_3 - Q_1}{2} \quad or \quad \frac{I.R}{2}$$

Where Q.D = Quartile deviation (Semi-Inter-quartile range)

Q_3 = third quartile (upper quartile)

Q_1 = first quartile (lower quartile)

Example 1 The following is data corrected on the result of 11 student in a final examination of statistics.

Result: 72 65 75 69 35 43 52 37 61 58 41

Solution The data when arranged in ascending order will read as follows:

35 37 41 43 52 58 61 65 69 72 75

Calculating the quartiles,

$$Q_3 = \text{value of the } 3\left(\frac{N+1}{4}\right)^{th}$$

$$Q_1 = \text{Value of the } \left(\frac{N+1}{4}\right)^{th}$$

Here N = 11

$$Q_3 = \text{value of the } 3\left(\frac{11+1}{4}\right)^{th}$$

$$= \text{Value of } 3 \times \left(\frac{12}{4}\right)^{th} \text{ item}$$

$$\therefore \quad \text{Value of the } 9^{th} \text{ item} = \underline{69}$$

$$Q1 = \text{the value of } \left(\frac{N+1}{4}\right)^{th} \text{ item}$$

$$= \text{Value of } \left(\frac{11+1}{4}\right)^{th} \text{ item}$$

$$= \text{Value of } \left(\frac{12}{4}\right)^{th} \text{ item}$$

$$\therefore \quad \text{Value of the } 3^{rd} \text{ item} = \underline{41}$$

Substituting 69 & 41 for the upper & lower quartiles respectively in the formula, we have:

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{69 - 41}{2} = \frac{28}{2} = 14$$

COEFFICIENT OF QUARTILE DEVIATION

- The coefficient of quartile deviation (sometimes called the quartile coefficient of dispersion) allows you to compare dispersion for two or more sets of data. The formula is:

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{69 - 41}{69 + 41} = \frac{28}{110} = 0.255$$

If one set of data has a larger coefficient of quartile deviation than another set, then that data set's interquartile dispersion is greater.

The mean deviation

- **`average deviation' is the absolute average of the deviations taken from a measure of central Tendency.**
- Measures the 'average' distance of each observation away from the mean of the data.
- Gives an equal weight to each observation
- Generally more sensitive than the range or interquartile range, since a change in any value will affect it.

Mean deviation

- The mean of the absolute deviations

1.	Calculate mean of data	Find $\bar{x}$
2.	Subtract mean from each observation Record the differences	For each x, find $x - \bar{x}$
3.	Record absolute value of each residual	Find $\lvert x - \bar{x} \rvert$
4.	Calculate the mean of the absolute values	Mean Deviation= $\frac{\sum \lvert x - \bar{x} \rvert}{n}$ Add up absolute values and divide by n

- **Example 2** Calculate the mean deviation from
  - Arithmetic mean, Mode and , Median
- Marks (out of 25) 7, 4, 10, 9, 15, 12, 7, 9, 7. If the marks are all doubled and taken

Solution

Calculation of mean deviation from mean, median, and mode.

$$\text{Mean} = \bar{x} = \frac{7+4+10+9+15+12+7+9+7}{9} = \frac{89}{9} = 8.9$$

$$\text{Median} = \tilde{x} = \text{size of } \frac{N+1}{2}th \text{ item} = \frac{9+1}{2} = 5^{\text{th}} \text{ item on the data arranged as}$$

4, 7, 7, 7, 9, 10, 10, 12, 15 so the 5^{th} value is 9. Hence $\tilde{x}$ = median = 9

Mode = $\hat{x}$ = 7 (since 7 is repeated the maximum number of times i.e. 3)

Computation of mean deviation

Marks X	Deviation from mean $/D\bar{x}/$	Deviation from median $/D\tilde{x}/$	Deviation from mode $/D\hat{x}/$
7	1.9	2	0
4	4.9	5	3
10	1.1	1	3
9	0.1	0	2
15	6.1	6	2
12	3.1	3	5
7	1.9	2	5
9	0.1	0	2
7	1.9	2	0
$\Sigma x = 80$	$\Sigma/D\bar{x}/ = 2101$	$\Sigma/D\tilde{x}/ = 21$	$\Sigma/D\hat{x}/ = 23$

$$\text{Mean Deviation (from mean)} = MD_{\bar{x}} = \frac{\sum|D\bar{x}|}{N} = \frac{21.1}{9} = 2.34$$

$$\text{Mean Deviation (from median)} = MD_{\tilde{x}} = \frac{\sum|D\tilde{x}|}{N} = \frac{21}{9} = 2.33$$

$$\text{Mean Deviation (from mode)} = MD_{\hat{x}} = \frac{\sum|D\hat{x}|}{N} = \frac{23}{9} = 2.56$$

The standard deviation

- Measures the variation of observations from the mean
- The most common measure of dispersion
- Takes into account every observation
- Measures the ‘average deviation’ of observations from mean
- Works with squares of residuals not absolute values—easier to use in further calculations

Standard deviation of a population δ

- Every observation in the population is used.
- The square of the population standard deviation is called the variance.

$$Standard\ deviation = \delta = \sqrt{\frac{\sum (x - \bar{x})^2}{N}}$$

$$\textbf{\textit{Variance}} = \delta^2$$

Standard deviation of a sample ---*s*

- In practice, most populations are very large and it is more common to calculate the sample standard deviation.

$$Sample\ standard\ deviation = s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

- Where: ***(n-1)*** is the number of observations in the sample

To calculate standard deviation

1. Calculate the mean	$\bar{x}$
2. Calculate the residual for each *x*	$x - \bar{x}$
3. Square the residuals	$(x - \bar{x})^2$
4. Calculate the sum of the squares	$\sum (x - \bar{x})^2$
5. Divide the sum in Step 4 by (*n*- 1)	$\frac{\sum (x - \bar{x})^2}{N}$
6. Take the square root of quantity in Step 5	$\sqrt{\frac{\sum (x - \bar{x})^2}{N}}$

- **<u>Example 1</u>** Find the standard deviations of the following.

14 15 17 22 23 25 24 16 18 26

Solution

- Computation of standard deviation and coefficient of standard deviation.

x	$(x-\bar{x})$	$(x-\bar{x})^2$
14	-6	36
15	-5	25
17	-3	9
22	2	4
23	3	9
25	5	25
24	4	16
16	-4	16
18	-2	4
26	6	36
$\Sigma x =$		
200		$\Sigma (x-\bar{x})^2 = 180$

$$Standard\ deviation = \delta = \sqrt{\frac{\sum (x-\bar{x})^2}{N}}$$

$$\sigma = \sqrt{\frac{180}{10}}$$

$$\sigma = \sqrt{18} = \underline{\underline{4.24}}$$

- **Group Discrete and Continuous Series:** When deviations are measured from Actual arithmetic mean:
- S[illegible]ulation

$$\delta = \sqrt{\frac{\sum f(x-\mu)^2}{N}}$$

Where

μ is the population arithmetic mean,

x is the class mark of the i^{th} class,

f is the frequency of the i^{th} class and

$N=\sum f_i$

Standard deviation from a sample

$$\delta = \sqrt{\frac{\sum f(x-\bar{x})^2}{n-1}}$$

- The standard deviation is the absolute measure of dispersion. Its relative measure is called the standard coefficient of dispersion or coefficient of standard deviation.

$$\text{Coefficient of standard deviation} = \frac{\sigma}{\bar{x}}$$

- A Distribution is consistent and less variable, if its coefficient of standard deviation is low.

COEFFICIENT OF VARIATION

- The relative measure of standard deviation
- Is the percentage variation in the mean the standard deviation (Coefficient of Standard Deviation) being treated as the total variation in the mean.
- It is used to measure the consistency and variability or uniformity of two or more distributions.
- Lower value of the coefficient of variation indicates higher consistency, more uniformity, and lower variability.

$$C.V = \frac{\sigma}{\bar{x}} \times 100\%$$

Where C.V = coefficient of variation

σ = standard deviation

$\bar{x}$ = arithmetic mean

VARIANCE

- It is the square of standard deviation (σ^2).
- The variance is “The mean of the squared deviations about the mean of a series” Thus symbolically variance may be expressed as:

$$V = \frac{\sum (x - \bar{x})^2}{N}$$

Where, V = variance

COMBINED STANDARD DEVIATION

- Is a single standard deviation representing two or more distributions.

  - *and number of observation for each population*

Example: From the analysis of monthly wages paid to employees in two service organizations *X* and *Y*, the following results were obtained

	Firm X	Firm Y
Number of Employees	550	650
Average monthly wage	Br 5000	Br 4500
Variance of the distribution of wages	Br 900	Br 1600

i. Which organization pays a larger amount as monthly wages?

ii. In which organization is there greater variability in individual wages of all the wage earners taken together?

Solution:

i. For finding out which organization *X* or *Y* pays larger amount of monthly wages, we have to compare the total wages:

- Total wage bill paid monthly by *X* and *Y* is

- Organization Y pays a larger amount as monthly wages as compared to organization X.

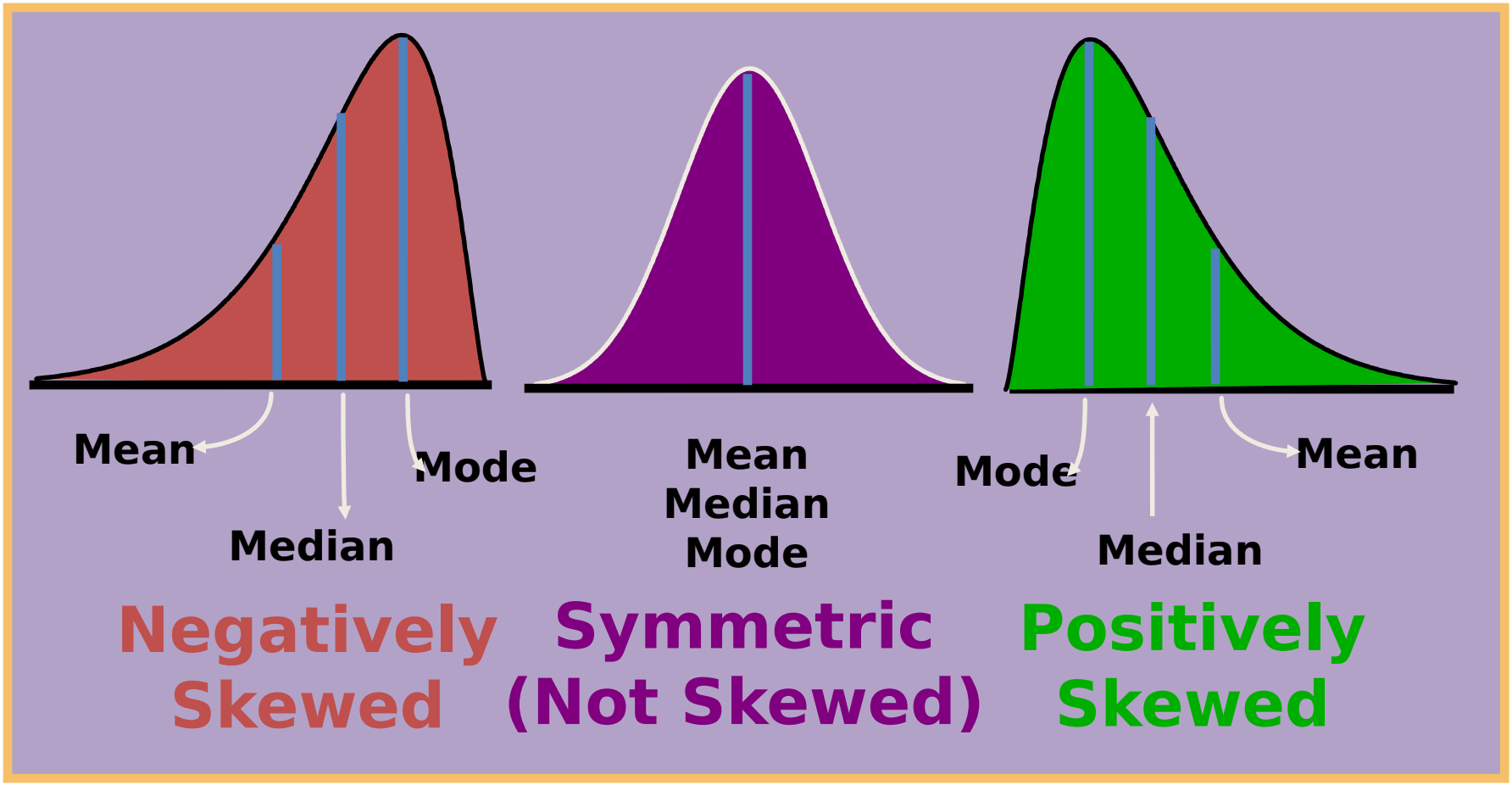
- For calculating the combined variance, we will first calculate the combined mean
- **Combined Mean**

- **Combined Standard Deviation**

Measures of Shape

- **Skewness**
  - Absence of symmetry
  - Extreme values in one side of a distribution
- **Kurtosis**
  - Peakedness of a distribution
- **Box Plots**
  - Graphic display of a distribution
  - Reveals skewness

Skewness


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

- μ_1 = First moment is mean
- μ_2 = Second moment is the variance
- μ_3 = Third moment about mean
- μ_4 = Forth moment about mean

- Measure of Skewness:

- Coefficient of skewness

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

Coefficient of Skewness

$\mu_1 = 23$

$M_{d1} = 26$

$\sigma_1 = 12.3$

$$S_1 = \frac{3(\mu_1 - M_{d1})}{\sigma_1}$$

$$= \frac{3(23 - 26)}{12.3}$$

$$= -0.73$$

$\mu_2 = 26$

$M_{d2} = 26$

$\sigma_2 = 12.3$

$$S_2 = \frac{3(\mu_2 - M_{d2})}{\sigma_2}$$

$$= \frac{3(26 - 26)}{12.3}$$

$$= 0$$

$\mu_3 = 29$

$M_{d3} = 26$

$\sigma_3 = 12.3$

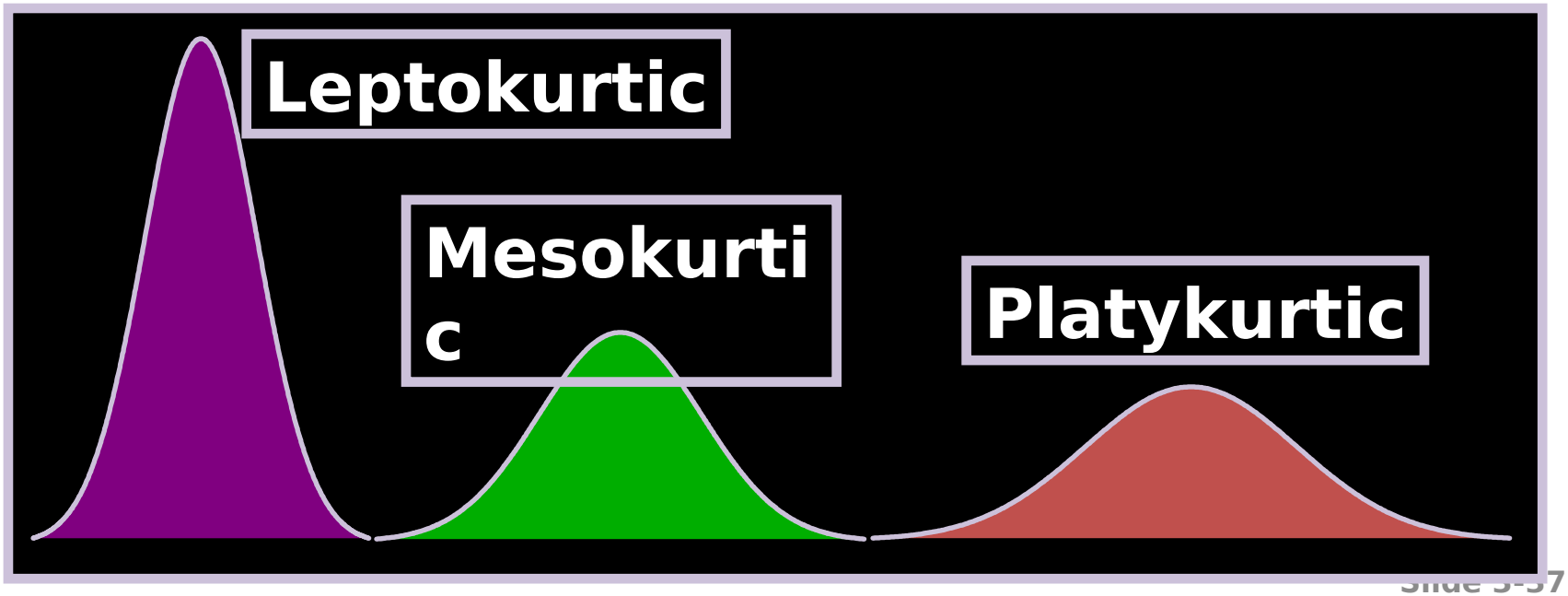
$$S_3 = \frac{3(\mu_3 - M_{d3})}{\sigma_3}$$

$$= \frac{3(29 - 26)}{12.3}$$

$$= +0.73$$

KURTOSIS

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



Measures of Kurtosis

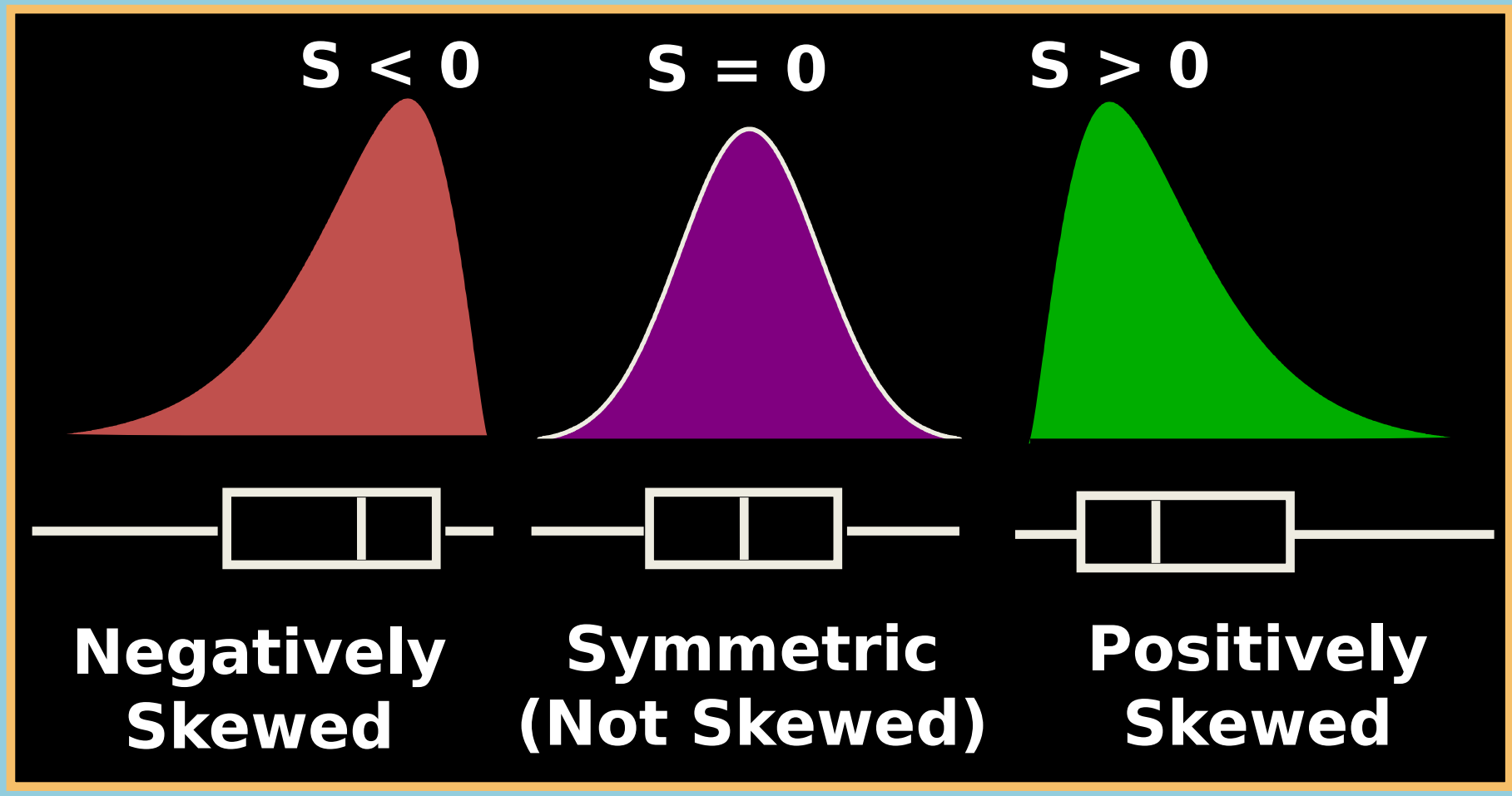
$\beta_2 = \mu_4 / \mu_2^2$

- μ_4 = Forth moment about mean
- μ_2 = Second moment about mean
- Interpretation
  - If $\beta_2 < 3$ Curve is Leptokurtic
  - If $\beta_2 = 3$ Curve is Mesokurtic
  - If $\beta_2 > 3$ Curve is PlatyKurtic

Further Moments of the Distribution

- There are **further statistics** that describe the **shape** of the distribution, using formulae that are similar to those of the mean and variance
- 1st moment - **Mean** (describes **central value**)
- 2nd moment - **Variance** (describes **dispersion**)
- 3rd moment - **Skewness** (describes **asymmetry**)
- 4th moment - **Kurtosis** (describes **peakedness**)

Skewness: Box Plots, and Coefficient of Skewness




Thanks!