UPES
MATH2058 : Applied Statistics and Probability
Unit-3: Short Notes
Date: September 21, 2025

Instructions

- These are very short notes only. For further and in depth study you can refer any textbook.
- The solutions to examples have already been discussed in the class, hence the solutions are not provided here.
- You can email me for any doubt or error.

1 Introduction

A massive quantitative data exhibit general characteristics as follows:

1. Tendency to concentrate at certain values somewhere in the centre of the distribution. These measures are called as *measures of central tendency* or *everages*.
2. variaton about measure of central tendency, called as *measures of variation or dispersion*.
3. The measures of the direcion or degree of asymmetry, also called as *measures of skewness*.
4. Measures of *peakedness* or *flatness* of the frequency curve, also called as *measures of kurtosis*.

2 Frequency Distribution

Discrete or Continuous observations on a single charactersitics of a large number of individuals requires condensation often without losing information of interest. For example: the marks of 250 students appearing in a certain exam: $10, 20, 15, 11, \ldots$.

A much better representation of data is:

Marks	Number of students
15	2
17	3
$\vdots$	$\vdots$

The above such representation is known as *frequency distribution.* In the above table marks are called as *variable(x)* and the number of students against the marks are called as *frequency(f).*

It is possible to further condense the data by dividing the observed range of variable into a suitable number of *class-intervals*, and to record the number of observations in each class-interval.

For the above example the frequency table can be as follows: The manner in which the

marks	No. of students (f)
$15-19$	9
$20-24$	11
$25-29$	10
$30-34$	44
$35-39$	45
$40-44$	54
$45-49$	37
$50-54$	26
$55-59$	8
$60-64$	5
$65-69$	1

Table 1: Frequency Distribution

class frequencies are distributed over the class interval is called as the *grouped frequency distribution* of the variable.

NOTE: The classes in which both the upper bound and lower bound are included are called as '*inclusive classes*'.

2.1 Continuous Frequency Distribution

The above such arrangement is not possible in case of continuous frequency distribution. For example: consider the distribution of ages in years. If the class-intervals are $15-19, 20-24$. Then, the persons whose age is between 19 and 20 years are not considered. For practical reasons we can rewrite the classes in the example as: $15-20, 20-25$. Frequency distribution that are formed from such classes are known as *continuous frequency distribution.* Generally, in such distributions the upper limit is excluded from each class. Such a class in which the upper bound is excluded from the class and included in the immeidate next class is known as '*exclusive classes*' and the classification is termed as '*exclusive type clasification*'.

If the grouped frequency distribtution is not continuous (see Table 1) then we first convert it into a continuous frequency distribution with exclusive type classes as follows:
– Let d be the gap between the uppper limit of any class and the lower limit of the succeeding class.
– Then, the class boundaries for any class ia given by:
Upper class boundary = Upper class limit + $d/2$;
Lower class boundary = Lower class limit - $d/2$.

3 Measures of Central Tendency

3.1 Arithmetic Mean

Arithmetic mean of a set of observations is given by the sum of all the observations divided by the number of observations. Let $\bar{x}$ be the arithmetic mean of n observations $x_1, x_2, \ldots, x_n$ then:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

Example 3.1. *Compute the arithmetic mean of the following set of observations:* $1, 2, 3, 4, 5$.

Ans: 3.

In case of frequency distribution, we compute the arithmetic mean as follows:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_n x_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{N} \tag{2}$$

where f_i is the frequency of the variable x_i, and $N = \sum_{i=1}^{n} f_i$.

Example 3.2. *Find the arithmetic mean of the following frequency distribution:*

x:	1	2	3	4	5	6	7
f:	5	9	12	17	14	10	6

Ans. 4.09

In case of grouped or continuous frequency distribution, x is taken as the *mid value* of the corresponding class.

Example 3.3. *Calculate the Arithmetic Mean of the marks from the following table:*

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Students	12	18	27	20	17	6

Ans. 28

Computation of arithmetic mean can be reduced by taking the deviations of values from any arbitrary point 'A'. Let $d_i = x_i - A$. Then, the mean $\bar{x}$ is computed as follows:

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N} \tag{3}$$

You can reduce the arithmetic further to a grater extend by taking $d_i = \frac{x_i - A}{h}$. Then, the mean $\bar{x}$ is computed as follows:

$$\bar{x} = A + \frac{h}{N} \sum_{i=1}^{n} f_i d_i \tag{4}$$

where A is an arbitrary point and h is the common magnitude of class interval.

Example 3.4. *Calculate the mean for the following frequency distribution:*
NOTE: *For practice you can solve this question first using Equation* (3) *and then using Equation* (4).

Ans. 25.404.
TOPICS FOR SELF-STUDY: *weighted mean.*

Class interval	$0-8$	$8-16$	$16-24$	$24-32$	$32-40$	$40-48$
Frequency	8	7	16	24	15	7

Merits	**Demerits**
It is rigidly defined	It cannot be determined by inspection nor it can be located graphically
It is easy to understand and easy to calculate	It cannot be used if we are dealing with qualitative characteristics which cannot be measures quantitatively, such as intelligence, honest etc.
It is based upon all the observations	It cannot be obtained if any of the observation is missing for some reason.
It is amenable to algebraic treatments	Affected by the presence of extreme values
Arithmetic mean is a stable average	Airthmetic mean may lead to wrong conclusions if the details of the data from which it is computed are not given.
	Arithmetic mean cannot be calculated if the extreme class is open
	Arithmetic mean is not suitable if the distribution is aymmetrical

Merits and Demerits of Mean

3.2 Median

– Median of a distribution is the value of a variable that divides the distribution into two equal parts.
– The number of observations before (or below) median is equal to the number of observations after (or above) median.
– Therefore, it is also called as *positional* average.

Process of obtaining median in case of ungrouped data.
– First, arranage the values in either non-increasing or non-decreasing order.
– Let n be the number of observations. If n is odd:
then median is $\lceil \frac{n}{2} \rceil$th value, else median is $\frac{\lfloor n/2 \rfloor + \lceil n/2 \rceil}{2}$th value.

Example 3.5. *Find the median of the below series:*

1. $25, 20, 15, 35, 18$
2. $8, 20, 50, 25, 15, 30$

Ans. (i) 20, (ii) 22.5

Steps to calculate the median in case of *discrete frequency distribution*:

1. Find $N/2$ where $N = \sum f_i$.
2. See the cumulative frequency just greater than $N/2$.
   NOTE: I already discussed in the class how to compute cumulative frequency.

x	1	2	3	4	5	6	7	8	9
f	8	10	11	16	20	25	15	9	6

3. The corresponding value of x is median.

Example 3.6. *Obtain the median for the following freuency distribution:*

Ans. 5.

Median for Continuous Frequency Distribution: In case of continuous frequency distribution the class corresponding to the cumulative frequency just greater than $N/2$ is called the median class.

We compute median as follows:

$$\text{Median} = l + \frac{h}{f}(\frac{N}{2} - c) \tag{5}$$

where
l is the lower limit of the median class,
f is the frequency of the median class,
h is the magnitude of the median class,
$'c'$ is the cumulative frequency of the class *preceding* the median class, and
$N = \sum f_i$.

Example 3.7. *Find the median wage of the following distribution:*

Wages(in Rs.)	2000 – 3000	3000 – 4000	4000 – 5000	5000 – 6000	6000 – 7000
No of workers	3	5	20	10	5

Ans. 4675

Merits and Demerits of Median See Table 3

Merits	**Demerits**
It is rigidly defined	Median is not exact in case of even number of observations
It is easy to understand and easy to calculate	It is not based on all the observations. This propoerty says that median is *insensitive*.
It is not affected by the presence of extreme values	It is not amenable to algebraic treatment
It can be calculated for distributions with open-end classes	in comparison with mean, it is much affected by the fluctuations of sampling.

Table 2: Merits and Demerits of Median

x	1	2	3	4	5	6	7	8
f	4	9	16	25	22	15	7	3

3.3 Mode

Mode is the value that occurs *most frequently* in a set of observations and around which the other items of the set cluster densely. In other words, mode value is predominant in the series.

Example 3.8. *Find the mode of the following frequency distribution.*

Ans. 4
NOTE: In any of the following cases:

1. the maximum frequency is repeated,
2. maximum frequency occurs in the begining or in the end,
3. irregularities in distribution

generally we use *method of grouping* to determine the mode. Interested reader can self-study this topic.

Mode for Continuous Frequency Distribution: Mode is given by the following formula:

$$\text{Mode} = l + \frac{h(f_1 - f_0)}{2f_1 - f_0 - f_2}$$

where,
l is the lower limit of the modal class,
h is the magnitude of the modal class,
f_1 is the frequency of the modal class,
f_0 frequency of the class preceding the modal class,
f_2 frequency of the class succeeding the modal class.

Example 3.9. *Find the mode for the following distribution:*

Class-interval	0 − 10	10 − 20	20 − 30	30 − 40	40 − 50	50 − 60	60 − 70	70 − 80
Frequency	5	8	7	12	28	20	10	10

Ans. 46.67

Merits and Demerits of Mode See **??**

3.4 Quartiles, Deciles and Percentiles

Median is a value such that it divides the set of observations in two parts such that half of the observations comes before median and remaining half comes after it. In the same line it is possible to divide the set of observations into 4 parts by using 3 points, into 10 parts using 9 points, into 100 parts using 99 points. Other divisions are also possible.

Merits	Demerits
it is readily comprehensible and easy to calculate	Mode is ill-defined. It is not always possible to find a clearly defined mode. Distributions can be *bi-modal*(two modes), and *multimodal*(more than two modes)
It is not affected by the presence of extreme values	It is not based upon all observations
Mode can be conveniently located even if the frequency distribution has class-intervals of unequal magnitude provided the modal class and the classes preceding and succeeding it are of the same magnitude	It is not capable of further mathematical treatment. Mode is greatly effected by fluctuations of sampling

Table 3: Merits and Demerits of Median

The 3 points that divide the set of observations (arranged in ascending order of value) into 4 parts in such a way that each position contains an equal number of items, are called as quartiles. Therefore, first quartile, denoted as Q_1, is the value such that 25% of the observations comes before it. Similarly, second quartile, denoted as Q_2, is the value such that 50% of the observations comes before it, and third quartile, denoted by Q_3, is the value such that 75% of the observations comes before it.

It is easy to interpret deciles and percentiles in the same way. We denote the ith decile by D_i, and $j$the percentile by P_j.

Finding Quartiles In case of ungrouped or discrete frequency series:
– $Q_1 =$ size of $\frac{N+1}{4}$th item,
– $Q_3 =$ size of $\frac{3}{4}(N+1)$th item,

Example 3.10. *Consider the following set of observations:*

$$1, 2, 3, 4, 5, \ldots, 15$$

Find Q_1 and Q_3.

In case of grouped data, we determine quartile, decile and percentile as follows:

$$Q_i = l + \frac{h}{f}(\frac{i \times N}{4} - c) \tag{6}$$

$$D_j = l + \frac{h}{f}(\frac{j \times N}{10} - c) \tag{7}$$

$$P_k = l + \frac{h}{f}(\frac{k \times N}{100} - c) \tag{8}$$

$$\tag{9}$$

where Q_i, D_j, P_k are the $i$the quartile, jth decile, and kth percentile respectively.

Example 3.11. *Find the first and third quartile and the* 90*th percentile of the following data (see Table 4):*

Ans. $Q_1 = 58.18, Q_2 = 79.5, P_{90} = 87.16$

Group No.	Monthly earnings	No of workers	cumulative frequency
1	27.5 – 32.5	120	120
2	32.5 – 37.5	152	272
3	37.5 – 42.5	170	442
4	42.5 – 47.5	214	656
5	47.5 – 52.5	410	1066
6	52.5 – 57.5	429	1495
7	57.5 – 62.5	568	2063
8	62.5 – 67.5	650	2713
9	67.5 – 72.5	795	3508
10	72.5 – 77.5	915	4423
11	77.5 – 82.5	745	5168
12	82.5 – 87.5	530	5698
13	87.5 – 92.5	259	5957
14	92.5 – 97.5	152	6109
15	97.5 – 102.5	107	6216
16	102.5 – 107.5	50	6266
17	107.5 – 112.5	25	6291

Table 4: Monthly earnings for worker table

4 Dispersion

– Meausres of central tendency give us an idea of the concentration of the observations about the cnetral part of the distribution.
– But they do not give us a complete picture about the distribution. Consider the following examples:

1. $7, 8, 9, 10, 11$
2. $3, 6, 9, 12, 15$
3. $1, 5, 9, 13, 17$

In all of the above examples the mean is 9. But, given this mean of 5 observations we can not tell which series has this mean. Therefore, it is necessary to supplement the mean with other measures. In this section we will discuss one such measure, that is *dispersion.* Literally dispersion means 'scatteredness'. It gives us an idea about the *homogeneity* (less dispersed or scattered) or *heterogeneity* (more dispesed or scattered) if the distribution.

Characteristics of an ideal dispersion: The following are the characrteristics of an ideal dispersion:

1. It should be rigidly defined.
2. It should be easy to calculate and easy to understand.
3. It should be based on all the observations.
4. It should be amenable to further mathematical treatment.
5. It should be affected as little as possible by fluctuations of sampling.

4.1 Measures of dispersion

There are two broad categories of dispersion:

1. The first category is based on distance between the values of selected observations. Such measures are termed as *distance measures.*

2. The second category is based on the average of the deviations of observations taken from some central value. Examples of such measures are mean deviation and standard deviation.

Range: Defined as the difference between two extreme values in the set of observations.

$$Range = X_{\max} - X_{\min} \tag{10}$$

It is subject to fluctuations as it is based on two extreme values only. Hence, it is not a reliable measure of dispersion.

Quartile Deviation: Also, called as semi-quartile range. Denoteed by Q. It is defined as:

$$Q = \frac{Q_3 - Q_1}{2} \tag{11}$$

Better than range as it makes use of the 50% of the observations. But since it ignores the remaining 50% of the data, it is not regarded as a reliable measure.

Mean Deviation: Mean deviation from the average A (usually mean, median and mode) is given by:

$$\text{Mean deviation about average } A = \frac{1}{N}\sum_{i=1}^{n} f_i|x_i - A| \tag{12}$$

where $N = \sum f_i$. Since it takes into account every measure it is considered a better measure of dispersion. The problem with the mean deviation is it ignores the sign of $(x_i - A)$. This introduces some artificiality, and makes it useless for further mathematical treatment.

Example 4.1. *Calculate (a) Quartile Deviation, (b) Mean Deviation from mean, for the following data:*

marks	$0-10$	$10-20$	$20-30$	$30-40$	$40-50$	$50-60$	$60-70$
No. of Students	6	5	8	15	7	6	3

Ans. (a) 11.23, (b) 13.184

Standard deviation: Standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. Arithmetic mean is denoted by σ. It is defined as:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2} \tag{13}$$

where $\bar{x}$ is the arithmetic mean of the distribution. It do not ignore the sign of $(x_i - \bar{x})$ and hence is amenable to further mathemetical treatment. Also, it is least affected by fluctuations of sampling.

The square of standard deviation is called as the *variance* and is given by,

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2 \tag{14}$$

Sometimes σ^2 is also written as σ_x^2 i.e. variance of x. To make calculations easier, an alternative formula for computing variance can be used. This formula is:

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2 - \Big(\frac{1}{N}\sum_{i=1}^{n} f_i x_i\Big)^2 \tag{15}$$

To further simplify the calculations for each value we generally take its deviation from some arbitrary value A. Therefore,

$$d_i = x_i - A$$

In that case it is true that

$$\sigma_x^2 = \sigma_d^2$$

and when

$$d_i = \frac{x_i - A}{h}$$

Then, it holds true that

$$\sigma_x^2 = h^2\sigma_d^2$$

Example 4.2. *Calculate the mean and standard deviation for the following table giving the age distribution of* 542 *members:*

Age(in years)	$20-30$	$30-40$	$40-50$	$50-60$	$60-70$	$70-80$	$80-90$
No of members	3	61	132	153	140	51	2

Ans. Mean is 54.72 years. Standard deviation is 11.88 years.

Variance of the combined series: Below are the sizes, means and standard deviation for two series, Then, the mean of the combines series is:

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} \tag{16}$$

Standard deviation of the combined series is given by:

$$\sigma^2 = \frac{1}{n_1 + n_2}[n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)] \tag{17}$$

where $d_1 = \bar{x}_1 - \bar{x}, d_2 = \bar{x}_2 - \bar{x}$.

Series	Size	Mean	Standard Deviation
Series 1	n_1	$\bar{x}_1$	σ_1
Series 2	n_2	$\bar{x}_2$	σ_2

Example 4.3. *The first of the two samples has* 100 *items with mean* 15 *and standard deviation* 3. *If the whole group has* 250 *items with mean* 15.6 *and standard deviation* $\sqrt{13.44}$. *Find the standard deviation of the second group.*

Ans. 4

4.2 Coefficient of Dispersion

To compare the variability of the two series that differ in average or are measured in different units we compute the coefficient of dispersion. The coefficient of dispersion (in short C.D.) based on different measures of dispersion are:

1. Based upon range C.D. $= \frac{X_{\text{max}} - X_{\text{min}}}{X_{\text{max}} + X_{\text{min}}}$.
2. Based unpon quartile deviation, C.D. $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$.
3. based upon mean deviation, C.D. $= \frac{\text{Mean deviation}}{\text{Average from which it is calculated}}$.
4. Based upon standard deviation, C.D. $= \frac{\sigma}{\bar{x}}$.

Coefficient of variation is defied as 100 times the coefficient of dispersion based upon standard deviation:

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

C.V. is the percentage variation in the mean, standard deviation being considered as the total variation in the mean.
The series having higher C.V. is said to be more variable than the other series having lesser C.V.

Example 4.4. *An analysis of monthly wages padi to the workers of two firms A and B belonging to the same industry gives the following results: Answer the following questions:*

	Firm A	Firm B
Number of workers	500	600
Average daily wage	Rs. 186	Rs. 175
Variance of distribution of wages	81	100

1. *Which firm, A or B has a larger wge bill?*
2. *In which firm, A or B, is there greater variability in individual wages?*
3. *Calculate a) average daily wage, b) the variance of the distribution of wages of all the workers in the firms A and B taken together.*

Ans. 1. Firm B has larger wage bill, 2. Firm B has greater variability, 3.a Rs. 180, 3.b 121.36

5 Moments

A moment is a specific quantitative measure of the shape of a function. All members of moment are measures based on deviations of values from some arbitrary point.

The $r$the moment of a variable x about any point A, usually denoted by ${\mu_r}'$ is given by:

$$ {\mu_r}' = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - A)^r \tag{18} $$

where $N = \sum f_i$. When $A = 0$ we call it the *raw moment.*

When $A = \bar{x}$, we call it moment about mean and denote it by μ_r. Therefore,

$$ \mu_r = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^r \tag{19} $$

Note that $\mu_0 = 1, \mu_1 = 0, \mu_2 = \sigma^2$. When $d_i = x_i - A$, then it is easy to verify that

$$ \bar{x} = A + \frac{1}{N}\sum_{i=1}^{n} f_i d_i = A + {\mu_1}' $$

Relation between Moments about mean in terms of Moments about Any Point and Vice-Versa

$$ \mu_2 = {\mu_2}' - {\mu_1'}^2 \tag{20} $$
$$ \mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2{\mu_1'}^3 \tag{21} $$
$$ \mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'{\mu_1'}^2 - 3{\mu_1'}^4 \tag{22} $$

$$ \mu_2' = \mu_2 + {\mu_1'}^2 \tag{23} $$
$$ \mu_3' = \mu_3 + 3\mu_2\mu_1' + {\mu_1'}^3 \tag{24} $$
$$ \mu_4' = \mu_4 + 4\mu_3\mu_1' + 6\mu_2{\mu_1'}^2 + {\mu_1'}^4 \tag{25} $$

Pearson's β and γ coefficients Karl Pearson defined the four coefficients based upon the first four moments about mean:

$$ \beta_1 = \frac{\mu_3^2}{\mu_2^3} \tag{26} $$
$$ \gamma_1 = \sqrt{\beta_1} \tag{27} $$
$$ \beta_2 = \frac{\mu_4}{\mu_2^2} \tag{28} $$
$$ \gamma_2 = \beta_2 - 3 \tag{29} $$

- These coefficients characterize distributions quantitatively beyond just mean and variance.
- β_1, γ_1 measures skewness (asymmetry of data distribution).

- β_2, γ_2 measures Kurtosis (peakedness or flatness realtive to normal distribution).

Example 5.1. *The first-four moments of a distribution about the value of* 4 *of the variable are* $-1.5, 17, -30$ *and* 108*. Find the moments about mean,* β_1, β_2*. Find also the moment about (i) the origin, (ii) the point* $x = 2$*.*

Ans. The moments about mean are: $\mu_2 = 14.75, \mu_3 = 39.75, \mu_4 = 142.3125$. (What will be μ_1?)
– β_1 is 0.4926 and β_2 is 0.6543.
– Moments about origin are: $\mu_1' = \bar{x} = 2.5, \mu_2' = 21, \mu_3' = 166, \mu_4' = 1132$
– Moments about point $A = 2$ is: $\mu_1' = 0.5, \mu_2' = 15, \mu_3' = 62, \mu_4' = 244$.

6 Skewness

For a given distribution of data, we can draw the shape of the curve. Skewness gives the idea about the shape of the curve. Skewness literally means 'lack of symmetry'. A distribution is said to be skewed, if:

1. Mean, Median and Mode fall at different places i.e. Mean $\neq$ Median $\neq$ Mode,
2. Quartiles are not equidistant from Median, and
3. The distribution curve is not symmetrical but is stretched on one side than to the other.

6.1 Measures of Skewness

1. $S_k = M - M_d$,
2. $S_k = M - M_0$
3. $S_k = (Q_3 - M_d) - (M_d - Q_1)$

where M, M_d, M_0 are Mean, Median and Mode of the distribution respectively. Above measures are the *absolute measures* of Skewness. For comparing two series we need *relative measures* called as the *coefficients of skewness* which are pure numbers independent of units of measurements.

6.2 Coefficients of Skewness

(I) Prof. Karl Pearson's Coefficients of Skewness: It is defined as:

$$S_{k_1} = \frac{(M - M_0)}{\sigma} \tag{30}$$

If mode is ill-defined, then using the empirical relation, $M_0 = 3M_d - 2M$, for moderately asymmetrical distribution, we get

$$S_{k_2} = \frac{3(M - M_d)}{\sigma} \tag{31}$$

Note that $S_{k_i} = 0$ if $M = M_0 = M_d$. Skewness is positive (i.e. tail to the right) if $M > M_0$ or $M > M_d$ and negaitve (i.e. tail to the left) if $M < M_0$ or $M < M_d$.
Both coefficients will have the same sign for the same data set if the mode is estimated using the empirical relation (Why so?).
Due to multomodality or data errors if the calcualted mode is far from the empirical value, then discrepencies might arise.
The limits for the Karl Pearson coefficient typically are ± 3 for most practical datasets.
– The coefficient is independent of the scale and is dependent on the divergence of mean from mode or median.
– Useful for quick identification of skewness direction.
– It is sensitive to extreme values and may not be reliable if the mode is poorly defined.

(II) Prof.Bowley's Coefficient of Skewness: It is based on quartiles, and is defined as:

$$S_k = \frac{Q_3 + Q_1 - 2M_d}{Q_3 - Q_1} \tag{32}$$

It is useful in situations
– When the mode is ill-defined and extreme observations are present in the data.
– When the distribution has open end classes or unequal class intervals.
Bowley's coefficient is:
– 0, when $Q_3 - M_d = M_d - Q_1$
– Positive, when $Q_3 - M_d > M_d - Q_1$
– Negative when $Q_3 - M_d < M_d - Q_1$
The limits for the Bowley's coefficients are ± 1. The only serious limitations of this coefficient is that it is based only on the central 50% of the data.

Coefficient of Skewness based upon Moment: It is defined as:

$$S_k = \frac{\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)} \tag{33}$$

It is zero when $\beta_1 = 0$.

Example 6.1. *Assume that a firm has selected a random sample of* 100 *from its production line and has obtain the data shown in the table below: Compute the following: (i) The*

Class Interval	Frequency
130 – 134	3
135 – 139	12
140 – 144	21
145 – 149	28
150 – 154	19
155 – 159	12
160 – 164	5

arithmetic mean, (ii) The standard deviation, and (iii) Karl Pearson's coefficient of skewness.

Ans. (i) 147.2, (ii) 7.2083, (iii) 0.0711

7 Kurtosis

Knowing measures of central tendency, dispersion, and skewness all together can not form a complete picture about the distribution. We need some measures that will give us an idea about the flatness or peakedness of the curve. Prof. Karl Pearson calls it '*Convexity of the frequency curve*' or '*Kurtosis*'. It is measures by the coefficients β_2 or its deriation γ_2 explained earlier. In Figure 1, there are three types of curve, all re symmetrical. However, they differ in terms of flatness or peakedness. Curve B is more flatter than other curves, and curve C is more peaked compared to other curves. Curve A is a normal curve.

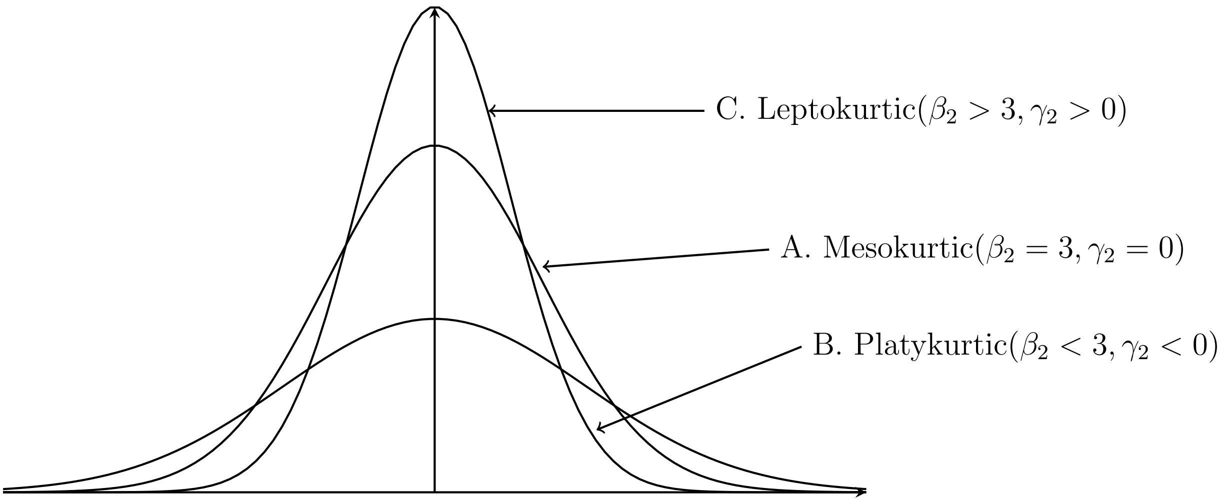

Figure 1: Different types of curve

Example 7.1. *The standard deviation of a symmetrical distribution is* 5*. What must be the value of the fourth moment about the mean in order that the distribution be (i) Leptokurtic, (ii) Mesokurtic, (iii) Platykurtic*

Ans. (i) $\mu_4 > 1875$, (ii) $\mu_4 = 1875$, (iii) $\mu_4 < 1875$

References

[GK16] Fundamentals of mathematical statistics: a modern approach, Gupta, S. C. and Kapoor, V. K., 10th rev. ed. (Greatly improved), ISBN 978-81-8054-969-4, Sultan Chand, New Delhi, 2016