UNIT 6 MEASURES OF SKEWNESS AND KURTOSIS

Structure

6.0 OBJECTIVES

After going through this Unit, you will be able ta :

distinguish between a symmetrical and a skewed distribution;

compute various coefficients to measure the extent of skewness in a distribution;

distinguish between platykurhc, mesokurtic and leptokurtic distributions; and compute the coefficient of kurtosis.

6.1 INTRODUCTION

In this Unit you will learn various techniques to distingush between various shapes of a frequency distribution. This is the final Unit with regard to the summarisation of univariate data. This Unit will make you familiar with the concept of skewness and kurtosis. The need to study these concepts arises fiom the fact that the measures of central tendency and dispersion fail to describe a distribution completely. It is possible to have fkquency distributions which differ widely in their nature and composition and yet may have same central tendency and dispersion. Thus, there is need to supplement the measures of central tendency and dispersion. Consequently, in th~s Unit, we shall discuss two such measures, viz, measures of skewness and kurtosis.

6.2 CONCEPT OF SKEWNESS

The skewness of a distribution is defined as the lack of ***symmetry.*** In a symmetrical distribution, the Mean, Median and Mode are equal to each other and the ordinate at mean divides the distribution into two eual arts such that one

part is mirror image of the other (Fig. 6.1). If some observations, of very

high (low) magnitude, are added to such a distribution, its right (left) tail gets elongated.

I Symmetrical Distribution

Fig. 6.1

, Positively Skewed Distribution | Negatively Skewed Distribution

Fig. 6.2

These observations are also known as extreme observations. The presence of extreme observations on the right hand side of a distribution makes it positively skewed and the three averages, **viz.,** mean, median and mode, will no longer be equal. We shall in fact have Mean > Median > Mode when a distribution is positively skewed. On the other hand, the presence of extreme observations to the left hand side of a distribution make it negatively skewed and the relationship between mean, median and mode is: Mean < Median < Mode. In Fig. 6.2 we depict the shapes of positively skewed and negatively skewed distributions.

The direction and extent of skewness can be measured in various ways. We shall discuss four measures.@skewness in this Unit.

6.2.1 Karl ~earhn's Measure of Skewness

In Fig. 6.2 you noticed that the mean, median and mode are not equal in a skewed distribution. The Karl Pearson's measure of skewness is based upon the ***divergence of mean from mob*** in a skewed distribution.

Since Mean = Mode in a symmetrical distribution, (Mean - Mode) can be taken as an ***absolute measure of skewness.*** The absolute measure of skewness for a distribution depends upon the unit of measurement. For example, if the mean = 2.45 qetre and mode = 2.14 metre, then absolute measure of skewness will be 2.45 aetre - 2.14 metre = 0.31 metre. For the same distribution, if we change the dt of measurement to centimetres, the absolute measure of skewness is 245 centimetre - 2 14 centimetre = 3 1 centimetre. In order to avoid such a problem

Karl Pearson takes a relative measure of skewness.

A relative measure, independent of the units of measurement, is defined as the Karl Pearson ь Coeficient of Skewness Sk, given by

$$S_k = \frac{\text{Mean} - \text{Mode}}{\text{s. d.}}$$

The sign of Sk gives the direction and its magnitude gives the extent of skewness.

If Sk > 0, the distribution is positively skewed, and if S, < 0 it is negatively skewed.

So far we have seen that Sk is strategically dependent upon mode. If mode is not defined for a distribution we cannot find Sk . But empirical relation between mean, median and mode states that, for a moderately symmetrical distribution, we have
Mean - Mode = 3 (Mean - Median)

Hence Karl Pearson's coefficient of skewness is defined in tms of median as

Example 6.1: Compute the Karl Pearson's coefficient of skewness fiom the following data:
Table 6.1

Height (in inches)

58
59
60
61
62
63
64
65
Number of Persons

10
18
30
42
3 5
28
16
8

Table for the computation of mean and s.d.

Height (X)

5 8

5 9

60

61

62

63
64
65

Total

u = X - 61

- 3
- 2
- 1

0
1
2
3
4

No. of persons V)

10
18
30
42
35
28
16
8

187

fi

- 30 -3 6
- 30 0

3 5
5 6
48
32

75

fu2   90

72
30
0
35
112
1 44
128

611

Summarisation of

 Mean ^{75} = 61 + -= 61.4

To find mode, we note that height is a continuous variable. It is assumed that the height has been measured under the approximation that a measurement on height that is, e.g., greater than 58 but less than 58.5 is taken as 58 inches while a measurement greater than or equal to 58.5 but less than 59 is taken as 59 inches. Thus the given data can be written as

Height (in inches) No. of persons

57.5 - 58.5

58.5 - 59.5

59.5 - 60.5

60.5 - 61.5

61.5 - 62.5

62.5 - 63.5

63.5 - 64.5

64.5 - 65.5

By inspection, the modal class is 60.5 - 61.5. Thus, we have

12 . Mode = 60.5 + *12+7* x 1 = 61.13

61.4 - 61.13 - 0.153.
Hence, the Karl Pearson's coeficient of skewness Sk = -

Thus the distribution is positively skewed.

6.2.2 Bowley's Measure of Skewness

This measure is based on quartiles. For a symmetrical distribution, it is seen that ***Q,*** and ***Q3 are*** equidistant ftom median. Thus ***(Q3 - Md)*** - (M, - ***Q,)*** can be taken as an absolute measure of skewness.

A relative measure of skewness, known as Bowley's coefficient (SQ), is given by

The Bowley's coefficient for the data on heights given in Table 6.1 is computed below.

Height (in inches) 57.5 - 58.5

58.5 - 59.5

59.5 - 60.5

60.5 - 61.5

61.5 - 62.5

62.5 - 63.5

63.5 - 64.5

64.5 - 65.5

Computation of Q, :

No. of persons V)

10

18

3 0

42	Cumulative Frequency	135
35	10-	163
2 8	2 8	179
16	5 8	187
8	100	

^{N} Since p $= 46.75$, the first quartile class is 59.5 - 60.5.

Thus $l_{a,} = 59.5$, $C = 28$, $f_{a,} = 30$ and $h = 1$.

***Computation of** M,* (Q,) :

^{N} Since 2 $= 93.5$, the median class is 60.5 - 61.5. Thus

$l_m = 60.5$, $C = 58$, $f_m = 42$ and h = 1.

Computation of Q, :

^{3N} Since $= 140.25$, the third quartile class is 62.5 - 63.5. Thus

$l_{a,} = 62.5$, $C = 135$, $f_Q = 28$ and $h = 1$.

Hence, Bowley's

62.688 - 2 x 61.345 + 60.125 = 0.048 .

coefficient $S_Q =$

62688 - 60.125

6.2.3 Kelly's Measure of Skewness

Bowley's measure of skewness is based on the middle 50% of the observations because it leaves 25% of the observaticins on each extreme of the distribution. As an improvement over Bowley's measure, Kelly has suggested a measure based on P,, and, P,, so that only 10% of the observations on each extreme are ignored.

Kelly's coefficient of skewness, denoted by S, is given by

Note that $P_{50} = M$, (median).

The value of S,, for the data given in Table 6.1, can be computed as given below.

Computation of P_{10} :

Since $\frac{10N}{100} = \frac{10 \times 187}{100} = 18.7$. 10th percentile lies in the class 58.5 - 59.5. Thus

Computation of P_{90} :

Since $\frac{90N}{100} = 168.3$, 90th percentile lies in the class 63.5 - 64.5. Thus

$l_{90} = 63.5$, $C = 163$, $f_{90} = 16$ and $h = 1$.

$$P_{90} = 63.5 + \frac{168.3 - 163}{16} \times 1 = 63.831.$$

Hence, Kelly's coefficient

It may be noted here that although the coefficient S,, So and S, are not comparable, however, in the absence of skewness, each of them will be equal to zero.

Check Your Progress 1

1) Compute the Karl Pearson's coefficient of skewness from the following data :

Daily Expenditure (Rs.) :	0-20	20-40	40-60	60-80	80-100
No. of families :	13	25	27	19	16

2) The following figures relate to the size of capital of 285 companies :

Capital (in Ks. lacs.)	1-5	610	11-15	1620	21-25	2630	31-35	
No. of companies	20	27	29	38	48	53		ibtal 285

Compute the Bowley's and Kelly's coefficients of skewness and interpret the results.

3) The following measures were computed for a frequency distribution

: Mean = 50, coefficient of Variation = 35% and

Karl Pearson's Coefficient of Skewness = - 0.25.

Compute Standard Deviation, Mode and Median of the distribution.

6.3 MOMENTS

The rth moment about mean of a distribution, denoted by p,, is given

by P he - where r = 0, 1, 2, 3, 4, ... N i=1

Thus, **rth** moment about mean is the mean of the rth power of deviations of observations from their arithmetic mean. In particular,

1'' if r = 0, we have PO =-~h(xi -x)O =I, N i=1

1'' if r = 1, we have PIXi -x)=o, N ,=1

1'' if *r* = 2, we have~l =--~f;(xi -zy =a2, N i=1 .

if *r* = 3, we have P 3 = L?h (xi - x) ' and so on. N i=1

In addition to the above, we can define ***raw moments*** as moments about any arbitrary mean.

Let **A** denote an arbitrary mean, then uth moment about *A* is defined as

When $A = 0$, we get various moments about origin.

Moment Measure of Skewness

The moment measure of skewness is based on the property that, for a symmetrical distribution, all odd ordered central moments are equal to zero.

We note that $\mu_1 = 0$, for every distribution, therefore, the lowest order moment that can provide an absolute measure of skewness is μ_3.

Further, a coefficient of skewness, independent of the units of measurement, is given by

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \sqrt{\beta_1} = \gamma_1,$$

where β_1 and γ_1 are defined as the ***first beta*** and ***first gamma*** coefficients respectively. β_2 is measure of kurtosis as you will come to know in the next Section.

Very often, the skewness is measured in terms of $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$, where the sign of skewness is determined by the sign of μ_3.

Example 6.2: Compute the Moment coefficient of skewness (β_1) from the following **data.**

Marks Obtained : 0-10 10-20 20-30 30-40 40-50 50-60 60-70

Frequency : 6 12 22 24 16 12 8

Table for the computations of mean, s.d. and β_1.

Class Intervals	Frequency (f)	Mid values (X)	$u = \frac{X-35}{10}$	fu	fu2	fu3
0 - 10	6	5	- 3	- 18	54	- 162
10 - 20	12	15	- 2	- 24	48	- 96
20 - 30	22	25	- 1	- 22	22	-22
30 - 40	24	35	0	0	0	0
40 - 50	16	45	1	16	16	16
50 - 60	12	55	2	24	48	96
60 - 70	8	65	3	24	72	216
Total	**100**			**0**	**260**	**48**

Since $\Sigma fu = 0$, the mean of the distribution is 35.

The second moment μ_2 is equal to the variance (σ^2) and its positive square root is equal to standard deviation (σ).

$\frac{260}{100}$ **p2** =-~100=260, and

$$1'' = -Eh[(xi\ _{-A)-piIr}\ 1\ _N$$ (Since *pi*

=GZh(Xi-~)=z-~) Expanding the term within brackets by binomial theorem, we get

Since the sign of **p3** is positive and p, is small, the distribution is slightly positively skewed.

If the mean of a distribution is not a convenient figure like 35, as in the above example, the computation of various central moments may become a cumbersome task. Alternatively, we can first compute raw moments and then convert them into central moments by using the equations obtained below.

Conversion of Raw Moments. into Central Moments

We can write

From the above, we can write

pr = - *fCIp~-lpi* + *r~2p:-2pi2- rC3p;-3p1* +"".
,3

In particular, taking ***r*** = 2, **3,** 4, etc., we get

p2 = p;-2~lp;2+2~2p;p~2 = pi- *pi2* (since *p;* = 1)

Example 6.3: Compute the first four moments about mean from the following data.

ClassIntervals :	**0-10**	**10-20**	**20-30**	**30-40**
Frequency V) :	1	**3**	**4**	2

Table for computations of raw moments (Take ***A*** = 25).

Class Intervals	f	Mid-Value	$u = \frac{X-25}{10}$	fu	fu2	fu3	fu4
0 - 10	1	5	- 2	**- 2**	**4**	**-8**	**16**
10 - 20	**3**	15	- 1	**- 3**	**3**	**-3**	**3**
20 - 30	**4**	25	**0**	**0**	**0**	**0**	**0**
30 - 40	**2**	**35**	**1**	2	2	2	**2**
Total	**10**			**-3-**	**9**	**-9**	**21**

From the above table, we can write

-9x lo3
pi= I(, = 900 and

Moments about Mean

By definition,

Check Your Progress 2

1) Calculate the first four moments about mean for the following distribution. Also calculate 9, and comment upon the nature of skewness.

Marks :	0 - 20	20 - 40	40 - 60	60 - 80	80 - 100 '
Frequency :	8	28	35	17	12

2) The first three moment of a distribution about the value 3 of a variable are **2,10** and 30respectively. Obtain **F, p2, p3** and hence **P,.** Comment upon the nature of skewness.

6.4 CONCEPT AND MEASURE OF KURTOSIS

Kurtosis is another measure of the shape of a distribution. Whereas skewness measures the lack of symmetry of the frequency curve of a distribution, kurtosis is a measure of the relative peakedness of its fi-equency curve. Various frequency curves can be divided into three categories depending upon the shape of their peak. The three shapes are termed as Leptokurtic, Mesokurtic and Platykurhc as shown in Fig. 6.3.

Leptokurtic

Mesokurtic

Platykurtic

Fig. 6.3

P_4 A measure of kurtosis is given by $P_2 = 2$, a coefficient given by Karl Pearson. P_2

The value of $p_2 = 3$ for a mesokurtic curve. When $P_2 > 3$, the curvt: is more peaked than the mesokurtic curve and is tenned as leptokurtic. Similarly, when $p_2 < 3$, the curve is less peaked than the mesokurtic curve and is called as platykurtic curve.

Example 6.4: The first four central moments of a distribution are 0,2.5,0.7 and 18.75. Examine the skewness and kurtosis of the distribution.

To examine skewness, we compute **p,.**

Since $p_3 > 0$ and **p,** is small, the distribution is moderately positively skewed.

P4 18.75 - 3.0 , Kurtosis is given by the coefficient **P2** = ---i- = - -

P2 (q2

Hence the curve is mesokurtic.

Check Your Progress 3

1) Compute the first four central moments hm the following data. **Also** find the two beta coefficients.

Value 5 10 15 20 25 30 35

Frequency : 8 15 20 32 23 17 5

2) The first four moments of a distribution are 1,4, 10 and 46 respectively. Compute the moment coefficients of skewness and kurtosis and comment upon the nature of the distribution.

6.5 LET US SUM UP

In this Unit you have learned about the measures of skewness and kurtosis. These two concepts are used to get an idea about the shape of the fiequency curve of a distribution. Skewness is a measure of the lack of symmetry whereas kurtosis is a measure of the relative peakedness of the top of a fiequency curve.

6.6 KEY WORDS

Skewness: Departure from symmetry is skewness.

Moment of Order r: It is defined as the arithmetic mean of the **rth** power of deviations of observations.

Coefficient of Kurtosis: It is a measure of the relative peakedness of the top of a frequency curve.

6.7 SOME USEPUL BOOKS

Elhance, D. N. and V. lhance, 1988, *Fundamentals of Statistics,* Kitab Mahal, Allahabad.

Nagar, **A.** L. and R. K. Dass, 1983, *Basic Statistics,* Oxford University Press, DeIhi .

Mansfield, E., 199 1, *Statistics for Business and Economics: Methods and Applications, W.W.* Norton and Co.

Yule, G U. and M. G Kendall, 1991, *An Introduction to the Theov of Statistics,* Universal Books, Delhi.

6.8 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) 0.237

2) - 0.12, - 0.243

3) 17.5, 54.38, 51.46

Check Your Progress 2

1) 0,499.64, 2579.57, 5891 11.61, 0.053, skewness is positive.

2) 5, 6, -14,0.907, since **p3** is negative the distribution is negatively skewed.

Check Your Progress 3

1) 0,59.99, - 50.18, 8356.64,0.012 (negatively skewed), 2.32 (platykurtic).

2) 0,3. Thus thedistribution is symmetrical and mesokurtic. Such a distribution is also known as a Normal Distribution.