3. STATISTICS OF LOCATION

Statistics of location belong to the class of *descriptive statistics* (page 11). They serve to locate specific positions of the frequency distribution of a variable in a sample on the scale of scores of that variable. They include mean, median, mode, percentiles, deciles and quartiles.

3.1 CLASSIFICATION

Statistics of location are classified further into two main classes.

(*a*) *Measures for central values* :

They include mean, geometric mean, median and mode, and are also called *central tendencies*. They describe the locations of specific central positions of the frequency distribution of a variable in a sample on the scale of that variable.

(*b*) *Quantiles or fractiles* :

They are the scores below which lie specific fractions of the frequency distribution of a variable in a sample. They thus partition out specified fractions like specific numbers of quarters, one-tenths and one-hundredths of the distribution. They are also called *partition values* and include quartiles, deciles and percentiles. Median is both a measure of central value and a partition value.

3.2 MEAN

Mean is the arithmetic average of a set of scores. The mean of a sample (statistical mean) and that of a population (parametric mean) are represented by the symbols $\overline{X}$ and μ, respectively. Where X (or X_i) represents each individual score of a sample, ΣX (or ΣX_i) is the sum of all its scores, and n is the sample size or the total frequency of cases in the sample,

$$\overline{X} = \frac{\Sigma X}{n}; \quad \text{or, } \overline{X} = \frac{\Sigma X_i}{n}.$$

This is also how the mean is worked out for a small sample, with its scores few in number and not arranged in a frequency distribution or a frequency table. Properties of mean are as follows:

(*a*) The sum of all the scores of a sample is given by the product of their mean and the sample size.

$$\Sigma X = n\overline{X}\ ; \quad \text{or, } \Sigma X_i = n\overline{X}\ .$$

(*b*) Mean would be the score of each individual if the total score of the sample (ΣX) were equally distributed among all its individuals.

(*c*) The sum of positive deviations of some of the scores from the mean equals that of negative deviations of the remaining scores of the sample from it. So, the algebraic sum of the deviations of all the individual scores from the mean amounts to zero in any sample.

$$\Sigma(X - \overline{X}) = 0\ ; \quad \text{or, } \Sigma(X_i - \overline{X}) = 0.$$

(*d*) The *sum of squares* about the mean, i.e., $\Sigma(X - \overline{X})^2$ or $\Sigma(X_i - \overline{X})^2$, is the sum of squared deviations of all the scores of a sample from its mean and is the *lowest* of all such sums of squares about the respective measures for central values.

(*e*) If the individual scores of a sample are all multiplied or divided by a constant number (say, k), the mean also gets respectively multiplied or divided by the same number.

$$\frac{\Sigma(kX)}{n} = k\overline{X}\ ; \quad \frac{\Sigma X}{kn} = \frac{\overline{X}}{k}.$$

(*f*) If a constant number k is added to or subtracted from each score of a sample, its mean also gets respectively increased or decreased by the same number.

$$\frac{\Sigma(X+k)}{n} = \overline{X} + k\ ; \quad \frac{\Sigma(X-k)}{n} = \overline{X} - k.$$

(*g*) In a bilaterally symmetrical and unimodal distribution, with a single peak and neither of its tails longer or more tapering than the other, the mean is the exactly central score and identical with the median and the mode. For such a distribution of scores in a sample, mean is the most reliable, stable and widely applicable central value.

(*h*) The presence of a score, with an extreme positive or negative deviation from the mean and not counterbalanced by the presence of another score with an equal but opposite deviation, makes the distribution asymmetric, displaces the mean towards that extreme score, and causes the mean to differ from both the median and the mode of the distribution. In such cases, $\overline{X} > Mdn > Mo$, if the unbalanced extreme score/scores is/are in the high-value (positive) tail of the distribution to make that tail longer ; on the contrary, $\overline{X} < Mdn < Mo$, if such extreme scores occur in the low-value (negative) tail making the latter longer. This implies that the mean is unreliable as a central value in an asymmetric distribution which has one tail longer than the other due to a few scores with larger deviations in the longer tail.

(*i*) If the scores (*Y*) of a variable are the linear functions of the scores (*X*) of another variable, then the mean $\overline{Y}$ of the former is also a linear function of the mean $\overline{X}$ of the latter. Thus, if *a* is the vertical intercept and *b* is the slope of the straight line formed by plotting the *Y* scores against the *X* scores of the respective individuals in a sample,

$$Y = a + bX, \text{ and}$$
$$\overline{Y} = a + b\overline{X}.$$

Computation from frequency tables

In simple frequency tables with frequencies entered against single distinct scores, each forming a class by itself (Table 3.1), one or more individuals possess identical scores so that the scores are repeated in the data ; but the data are not classified into groups. The mean is computed here from the frequencies (repetitions) of the individual scores. Where $f_1, f_2, \ldots, f_k$ are the frequencies (f_i) of the respective individual scores (X_i) like $X_1, X_2, \ldots, X_k$, and *n* is the total frequency or sample size,

$$\overline{X} = \frac{\Sigma f_i X_i}{n} = \frac{f_1X_1 + f_2X_2 + \ldots + f_kX_k}{n}.$$

Example 3.2.1.

Compute the mean of the following interorbital width scores ...

12.8, 11.7, 12.3, 10.8 ...

3.6 MEDIAN

Median (*Mdn*) is that score in a frequency distribution, above and below which lie equal numbers, i.e., 50%, of the scores or cases of the sample. It is a partition value or fractile (§ 3.4) and is identical with D_5, P_{50} and Q_2. Some of its properties are given below.

(*a*) The ordinate on the X axis at the *Mdn* bisects the area of a frequency distribution into two equal halves.

(*b*) In symmetric unimodal distributions, (*i*) *Mdn* coincides with the mean and the mode, and (*ii*) the algebraic sum of deviations of the observed scores from the median amounts to 0.

$$Mdn = \overline{X} = M_0\,;\quad \Sigma(X - Mdn) = 0.$$

(*c*) In asymmetric distributions, (*i*) the median differs from the mean and the mode, the mean being located further than the median towards the longer tail of the distribution, and (*ii*) the algebraic sum of deviations of the scores from the median differs from 0 in value, indicating by its positive or negative sign respectively a longer positive or negative tail of the distribution. Thus for an asymmetric distribution with a longer positive tail, $\overline{X} > Mdn > M_0$, and $\Sigma(X - Mdn) > 0$; but for a distribution with a longer negative tail, $M_0 > Mdn > \overline{X}$, and $\Sigma(X - Mdn) < 0$.

(*d*) As the median is less deflected than the mean by extreme deviations of a few scores, it is a more reliable and representative measure of central value than the mean for an asymmetric distribution.

(*e*) Median can be computed even for frequency distributions with *open class intervals* or *unequal class sizes*, and also for ranked data like those of psychological and achievement tests.

Median is used in working out mean deviation, coefficient of mean deviation, coefficient of skewness and the median test.

Graphical determination

After drawing a *cP* ogive (*Example 2.7.1*, Fig. 2.12), a line is drawn parallel to its X axis from that point on the Y axis which corresponds to the *cP* of 50. From where this line meets the ogive, an ordinate is dropped to the X axis. The point of intersection of this ordinate with the X axis gives the *Mdn*.

Computation from ungrouped data

In an ungrouped set of data, median is the $(n + 1)/2$th score, counted from either the lowest or the highest score of the sample.

(*a*) If there is an odd number of scores in the sample, i.e., *n* is an odd number, *Mdn* coincides with that observed score which belongs to the $(n + 1)/2$th individual.

3.7 MODE

The mode (M_o) is that score of the variable which belongs to the largest number of individuals in a sample. It is, therefore, the most frequent score in the sample and coincides with that point on the X axis of a frequency distribution which corresponds to the peak of the latter. Some of its properties are as follows.

(*a*) A distribution may be *unimodal*, *bimodal* or *multimodal*, according to its one, two or more peaks and as many M_o values.

(*b*) There is no mode if all scores of the sample are either identical or have the same frequency.

(*c*) In a perfectly symmetric unimodal distribution, M_o, $\overline{X}$ and *Mdn* are identical.

(*d*) Mode, unlike median and mean, does not change even if some extreme scores occur in only one tail of the frequency distribution.

(*e*) In an asymmetric unimodal distribution, *Mdn* lies between M_o and $\overline{X}$ while M_o lies on that side of the *Mdn* which leads to the shorter tail of the distribution. Thus, for a distribution with a longer positive (high-value) tail, $\overline{X} > Mdn > M_o$; but if the negative (low-value) tail is longer than the positive one, $M_o > Mdn > \overline{X}$.

(*f*) The amount and algebraic sign of the deviation of the mean from the mode indicate respectively the degree and direction of asymmetry of the distribution.

Computations of mode

(*a*) In a simple series of scores or in the ungrouped data of a simple frequency distribution,