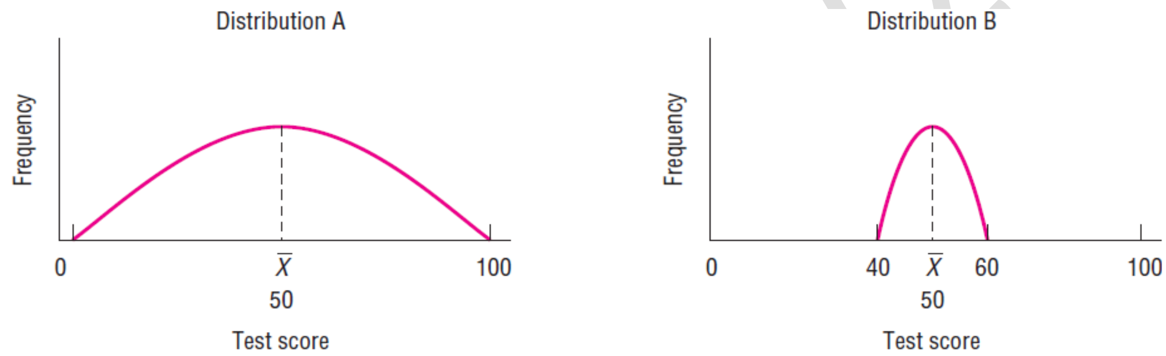
SES'S L.S.RAHEJA COLLEGE OF ARTS AND COMMERCE

Course: PSYCHOLOGICAL TESTING & STATISTICS Unit: 4 of Sem VI Prepared by: Ms. Shivani Chande

MEASURES OF VARIABILITY

Variability is an indication of how scores in a distribution are scattered or dispersed. Two or more distributions of test scores may have the same mean even though differences in the dispersion of scores around the mean can be wide.



As can be seen in distributions A and B, test scores could range from 0 to 100 and the mean in both distributions is 50. However, in distribution A the scores are widely distributed around the mean while in distribution B few people higher than 60 or lower than 40. Hence, individuals in distribution B are less variable than those in distribution A.

We may conclude that there is a tendency for data to be dispersed, scattered or to show variability around the average. The tendency of the attributes of a group to deviate from the average or central value is known as dispersion or variability.

The popular measures of variability are range, average deviation, quartile deviation and standard deviation.

1. RANGE

Range is the simplest measure of variability or dispersion. It is calculated as the difference between the highest and the lowest score in a distribution.

It is the simplest and the quickest measure of variability.

However, it's potential use is limited. One extreme value in the distribution of scores can radically alter the value of the range. When the range is based on extreme scores, the resulting description of variation may be understated or overstated.

The computation of Range is recommended when:

1. We need to know simply the highest and lowest scores of the total spread.
2. The group distribution is too small.
3. We want a quick measure of variability within the group.
4. We require speed and ease in the computation of a measure of variability.
5. The distribution is such that the computation of other measures of variability is not much useful.

2. AVERAGE DEVIATION (AD)

Average deviation is the mean of deviations of all separate scores in the series taken from their mean (occasionally, median or mode). It is the simplest measure of variability that takes into account the fluctuation or variation of all items in a series. It is calculated by the formula:

$$\mathrm{AD} = \frac{\sum |x|}{n}$$

Where, x is the score's deviation from the mean and is computed as (Raw score X – Mean = x), n is the total number of scores and AD is the average deviation. The bars on each side of x indicate that it is the absolute value of the deviation score and ignores the positive and negative value of the deviations. Hence, it does not regard whether the score is above or below the mean.

Merits:

- It is simple to understand.
- It is easy to calculate.
- It is based on all the observations of a series.
- It shown the dispersion, or scatter of the various items of a series from its central value.
- It is not very much affected by the values of extreme items of a series.
- It truly represents the average of deviations of the items of a series.

Limitations:

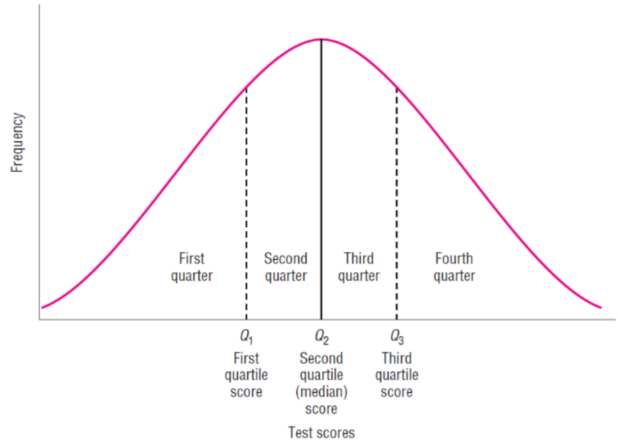
- It is not rigidly defined in the sense that it is computed from any central value viz. Mean, Median, Mode etc. and thereby it can produce different results.
- It ignores the algebraic signs of +ve and –ve while calculating the deviations of the different items from the central value of a series.
- Since it deletes the algebraic signs, the average deviation is not capable of further algebraic treatment.

The computation of AD is recommended when:

1. Distribution of the scores is normal or near to normal.
2. The standard deviation is unduly influenced by the presence of extreme deviations.
3. It is needed to weigh all deviations from the mean according to their size.
4. A less reliable measure of variability can be employed.

3. QUARTILE DEVIATION (QD)

A distribution of test scores can be divided into four equal parts such that 25% of the test scores occur in each quarter. The dividing points between the 4 quarters in the distribution are the quartiles. There are three quartile points, Q_1, Q_2, and Q_3.



Q_2 is the mid-point (or median) of the distribution while Q_1 and Q_3 are the quarter-points of the distribution.

The interquartile range is a measure of variability equal to the difference between Q_3 and Q_1 (Q_3 - Q_1). A related measure of variability is the quartile deviation or semi-interquartile range, which is equal to the interquartile range divided by 2 Hence, QD = (Q_3 - Q_1)/2.

The three quartile values are computed as follows:

1st quartile $$Q_1 = L + \frac{N/4 - F}{f} \times i$$

2nd quartile $$Q_2 = L + \frac{N/2 - F}{f} \times i$$

3rd quartile $$Q_3 = L + \frac{3N/4 - F}{f} \times i$$

Merits:

- It is rigidly defined.
- It is superior to range in as much as its calculation is based on middle 50% of the items of a series.
- It is useful in case of an open-end series when values at the extremes are not known.
- It is not very much affected by the extreme values of a series.
- It can be determined, even if, the first 25% and the last 25% of the items are replaced or deleted.

Limitations:

- It is not based on all the observations of a series.
- It is not capable of further algebraic treatment.
- It does not exhibit any scatter around an average for which is remarked as a measure of partition rather than a measure of dispersion.

The computation of quartile deviation is recommended when:

1. The distribution is skewed containing very few extreme values.
2. The measure of central tendency is available in the form of median.
3. The distribution is truncated (irregular) or has some indeterminate end values.
4. We have to determine the concentration around the middle 50 % of the cases.
5. The various percentiles and quartiles have been computed.

4. STANDARD DEVIATION (SD)

The computation of SD is recommended when:

1. We need the most reliable measure of variability.
2. There is a need of computation of correlation coefficients, significance of difference between the means and other statistic.
3. Measure of central tendency is available in the form of mean.
4. The distribution is normal or near to normal.

PERCENTILE AND PERCENTILE RANKS

Percentile is a measure use to indicate the relative position of a single item or an individual with reference to the group to which the item or individual belongs. It is used to tell the relative position of a given score among other scorers. In percentiles, the scores series is divided into 100 equal parts and each part is referred to as a percentile. Hence, the number of percentiles for a given score series or frequency distribution ranges from 1 to 100 (i.e, 1^{st} percentile to 100^{th} percentile).

A percentile maybe defined as a point below which a given percent of the cases lie. The 1^{st} percentile or P_1 will mean "a score point in the given series or distribution below which one percent of the cases lie and above which 99% of the cases lie". Similarly, P_{15} will indicate the point below which 15% of the cases lie and above which 85% of the cases lie. Likewise, P_{70} will indicate the point below which 70% of the cases lie and above which 30% of the cases lie.

Computation of a certain percentile is done by the formula:

$$\text{Percentile, } P = L + \frac{pN/100 - F}{f} \times i$$

where

L = Lower limit of the percentile class (the class in which th given percentile may be supposed to lie)
N = Total of all the frequencies
F = Total of the frequencies before the percentile class
f = Frequency of the percentile class
i = Size of the class interval
p = No. of the percentile which has to be computed

Percentile rank may be defined as the number representing the percentage of total number of cases lying below the given score. Hence, if we are provided with a specific score of a distribution and are required **to find out the percentage of cases** below that score, we are supposed to compute the percentile rank. On the other hand, if we want **to find out the score** under which a certain percent of the cases lie, we are supposed to compute percentiles.

Percentile ranks are computed as:

$$PR = \frac{100}{N}\left[F + \left(\frac{X - L}{i}\right) \times f\right]$$

where

PR = Percentile rank for the desired score X
F = Cumulative frequency below the interval containing score X
X = Score for which we want the percentile rank
L = Actual lower limit of the interval containing X
i = Size of the class interval
f = Frequency of the interval containing X
N = Total number of cases in the given frequency distribution