Norms

Definition and Nature

A norm is the average or typical score (mean or median) on a particular test made by a specified population, for example; the mean intelligence test score for a group of ten-year olds.

The raw score (that is, the actual number of units or points) obtained by an individual on a test does not in itself have much, if any, significance. One test may yield a score of 43 and cannot be directly compared with a score of 43 on another test. Furthermore, the average scores for each of the two tests will in all probability be different, as the degree of variation of scores (called the deviation) both above and below the average.

However, such score-for-score comparisons would be extremely cumbersome and would, in each instance, have to be interpreted in terms some common, meaningful index. It is thus, clear that if scores obtained on each of the several tests are to be compared, indexes must be used which will express the relative significance of any given score, or what is known as *relative rank*. Hence, to facilitate interpretation, sound psychological tests provide tables of age norms, or grade norms, or percentile ranks, or decile ranks, or standard scores, depending upon the instrument's purpose.

Reference to a tests table of norms enables us to rank an individual's performance relative to his own and other age or grade groups. For example, a child of 10 might have an intelligence test score that is average for his own age, or for a population of 9 year olds, or for those who are 11 years of age. Since it is desirable to locate an individual's score and relative rank not only with reference to an average, but also with reference to other levels in the scale, table of norms should include the frequency distribution of the scores, from which percentile ranks and standard scores may be readily calculated.

The characteristics of any table of norms depend on a number of factors affecting the individuals who make up the group:

1. In standardizing a psychological test, the norm and the distribution of scores are influenced by the representativeness of the population sample, that is, the proportion from each sex, their geographic distribution, their socioeconomic status, and their age distribution.

2. In devising a test of educational achievement, factors influencing the normative data, in addition to the above, are the quality of the schools and the kinds of curricula from which the standardization population is drawn.

3. Norms of tests of aptitude, like, clerical or mechanical, are influenced by the standardization population's degree of experience, the kind of work they have been doing and by the representativeness of the group.

Therefore, norms derived for several tests classified under the same name and intended for the same purpose are not necessarily comparable. It is necessary to know the characteristics of the standardization population before deciding on the selection and use of a test. This information is essential to decide whether the instrument is appropriate in the given situation or not.

Types and Methods

1. *Percentile Rank*: An individual's percentile rank on a test designates the percentage of cases or scores that fall below it. In other words, this statistical device determines at which one hundredth part of the distribution of scores or cases an individual is located. For example; a person having a percentile rank of 20 (P_{20}) is situated above twenty percent of the group of which he is a member; or, otherwise stated, twenty percent of the group fall below this persons rank. By this means a person's relative status or position in the group, can be established with respect to the traits or functions being tested. It is a known fact that psychological measurement, unlike physical measurement, derives its significance principally from relative ranks ascribed to individuals rather than from quantitative units of measurement.

A table of norms and frequency distribution often provides percentile ranks. In case the percentile ranks are not given in a table, it is possible to calculate them easily from the frequency distribution.

Percentile method is a technique whereby scores on two or more tests, given in units that are different, may be transformed into uniform and comparable values. This method has the advantage of being easily calculated, easily understood, and of making no assumptions with regard to the characteristics of the total distribution. It answers the question, "Where does an individual's score rank him in another group whose members have taken the same test?"

Fig. 10.4 Percentage Cumulative Frequency

Percentile Ranks

If k percent of members of a sample have scores less than a particular value of x then x is the kth percentile and k is the percentile rank of X.

Suppose a psychologist takes an intelligence test and finds that 60 percent of the sample score less than 72, then score 72 is the 60th percentile. In other words, if $P_{20} = 38$ then percentile rank of score 38 is 20. If $P_{75} = 105$ and $P_{25} = 78$, then percentile ranks of 78 is 25 and that of 105 is 75.

The computation of the percentile ranks is the reverse process of computation of percentile points. For this we have to compute ranks corresponding to particular scores. In case of individual scores first we have to find the rank of individual scores from the bottom, that is, rank in ascending order. If R is this rank and N is the total number of cases then Percentile Rank,

$$PR = 100R/N$$

In case of grouped data, the following steps are involved in computing the percentile rank are:

Step 1: Find the lower limit l of the class containing the score x whose percentile rank is required.

Deciles

Deciles are the points which divide the scale of measurement into ten equal parts. Thus there will be in all ten deciles, namely first decile to ninth decile. These deciles are denoted by D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8 and D_9. The first decile may be defined as a point on the scale of measurement below which one-tenth (1/10) or

Fig. 10.5: Cumulative frequency curve-Decile points.

The term *decile* is used to mean a dividing point. *Decile Rank* signifies a range of scores between two dividing points. For example, a testee who has a decile rank of 10 (D_{10}) is located in the highest 10 percent of the group; one whose decile rank is 9 (D_9) is in the second highest 10 percent; one whose decile rank is 1 (D_1), is in the lowest 10 percent of the group.

The decile rank is the same in principle as the percentile; but instead of designating the one-hundredth part of a distribution, it designates the one-tenth part of the group ($N/10$) in which any tested person is placed by his score. When the number of scores in a distribution is small, percentiles are not used, because there is little or no significance in making fine distinctions in rank. The decile-ranking method may be used instead.

3 Standard Score

This index too designates the individuals position with respect to the total range and distribution of scores, but its index is less obvious than that of percentile and

decile ranks. The standard score indicates, in terms of standard deviation, how far a particular score is removed from the mean of the distribution. The mean is taken as the zero point, and standard scores are given as plus or minus. If the distributions of scores of two or more tests are approximately normal, standard scores derived from one distribution may be compared with those derived from the others. The formula is,

$$z = \frac{X - M}{SD}$$

in which X is an individual score, M is the mean of the distribution and SD its standard deviation.

Fig. 10.6: The normal curve and derived scores.

T-Score

This variant of standard score was suggested by McCall (1922). In *T*-score method the mean is set at 50, unlike in standard score where the value of mean is zero. To obtain a *T*-score, the standard score is multiplied by 10 and then added to or subtracted from the mean *T*-score of 50. Thus, a standard score of +1.00 becomes a *T*-score of 60, while that of –1.00 becomes 40. The assumption in this technique is that nearly all scores will be within a range of five standard deviations from the mean. Since each *SD* is divided into 10 units, *T*-score is based upon a scale of 100 units, thus avoiding negative scores and fractions.

It is important, therefore, to understand that *T*-score, found for any individual, is relevant only to the distribution of scores of the group from which the values have been derived and with which his score is being compared.

Stanine

This term, coined by psychologists in the Army & Air Force during World War II, is yet another variant of the standard score technique. In this method, the standard

population is divided into nine groups, that is "standard nine", termed as *stanine*. Except the ranks of stanine 1 (lowest) and 9 (highest), each unit is equal to one half of a standard deviation. A score of 5 represents the median group, defined as those whose scores are within ± 0.25 *SD* the mean, that is, a range of a half-sigma at the centre of the distribution. Similarly, a rank of stanine 6 represents the group whose scores fall between + 0.25 sigma (σ) and + 0.75 sigma (SD). The meanings of the other stanine rankings can be determined likewise except 1 and 9, since the former represents all scores below, −1.75 sigma, and the latter includes those above +1.75 sigma.

This single digit system of scores has certain advantages for machine computations; and it does eliminate plus and minus signs. Other than these considerations, the stanine method has not much advantage in preference to other methods.

6 Age Norms

Sometimes it is desirable to express norms in terms of children's age. For instance we found by testing the reasoning of children at all ages from five to fifteen years. The average score can be calculated for each group separately. This can be plotted as shown below.

Grade Norms

Grade norms are similar to age norms. In this case the grade levels are taken in place of age on the *X*-axis on the graph. The grades are plotted with regard to the scores on the *Y*-axis as shown in Fig. 10.8.

These norms are very useful for the teachers to understand how well the students are progressing in the grade level.

8. IQ / Intelligence Quotient

Test norms

Test norms consist of data that make it possible to determine the relative standing of an individual who has taken a test. By itself, a subject's raw score (*e.g.*, the number of answers that agree with the scoring key) has little meaning. Almost always, a test score must be interpreted as indicating the subject's position relative to others in some group. Norms provide a basis for comparing the individual with a group.

Norms and It's Importance

Norms refer to information regarding the group performance of a particular reference on a particular measure for which a person can be compared to.

Norms mean standardized scores. Scores on the psychological tests are most commonly interpreted by reference to the norm that represents the test performance on the standardization sample. Norms always represent the best performance.

Basically, there are two purposes of norms:

1. Norms indicate the individual's relative standing in the normative sample and thus permit evaluation of his/her performance in reference to other persons.
2. Norms provide compared measures that permitted a direct comparison of the individual performance on a difference test.

Sir Francis Galton at the first time developed the logic for norm-based testing in the 18^{th} century.

Statistical concept:

1. **Frequency distribution:** A major object of the statistical method is to organize and summarize quantitative data in order to facilitate their understanding. A list of 1000 test scores can be an overwhelming sight. In that form, it conveys little meaning. A first step In bringing order into such a chaos of raw data is to tabulate the scores into a frequency distribution. A distribution is prepared by grouping the scores into convenient class intervals and tallying each score in the appropriate interval. When all scores have been entered the tallies are counted to find the frequency, or a number of cases, in each class interval. The sum of these frequencies will equal N, the total number of cases in the group.
2. **Graphical representation:** The information provided by a frequency distribution can also be presented graphically in the form of a distribution curve. On the baselines, or horizontal axis, are the scores grouped into class intervals; on the vertical axis are the frequencies or number of cases falling within each class interval. The graph has been plotted in two ways. In the histogram, the height of the column erected over each class interval corresponds to the number of persons scoring in that interval. In the frequency polygon, the number of persons in each interval is indicated by a point in the center of the class interval and across from the appropriate frequency. The successive points are then joined by straight lines.

3. **Central Tendency:** A group of scores can also be described in terms of some measure of central tendency. The most familiar of these measures is the average, more technically known as the mean (M), and it is found by adding all sores and dividing the sum by the number of cases (N). Another measure is the mode or most frequent score. In a frequency distribution, the mode is the midpoint of the class interval with the highest frequency. The third measure of central tendency is the median or middlemost score when all scores have been arranged in order of size. The medial is the point that bisects the distribution, half the cases falling above it and half below.
4. **Variability:** Further description of a set of test scores is given by measures of variability, or the extent of individual differences around the central tendency. The most obvious and familiar way for reporting variability is in terms of range between the highest and lowest score. T range, however, is extremely crude and unstable, for it is determined by only two scores. A single unusually high or low score would thus markedly affect its size. A more precise method of measuring variability is based on the difference between each individual's score and the mean of the group.

A much more serviceable measure of variability is the standard deviation (symbolized by either SD) in which the negative signs are legitimately eliminated by squaring each deviation. The sum of this column divided by the number of cases is known as the variance, or mean square deviation. The variance has proved extremely useful in sorting out the contributions of different factors lo individual differences in test performance. The SD also provides the basis for expressing an individual's scores on different tests in terms of norms.

Developmental Norms

One way in which meaning can be attached to test scores is to indicate how far along the normal developmental path the individual has progressed. Developmental systems utilize more highly qualitative descriptions of behavior in specific functions, such as sensorimotor activities or concept formation.

Mental Age: The term “mental age’ was widely popularized through the various translations and adaptations of the Binet-Simon scales, although Binet himself had employed the more neutral term “mental Level”. In age scales such as the Bind and its revisions (prior to 1986), items were grouped into year levels. For example, those items passed by the majority of 7-years olds in the standardization sample were placed in the 7-year level, and so forth. A child’s score on the test would then correspond to the highest year level that he or she could successfully complete In actual practice, the individual’s tests below their mental age and passed some above it. For this reason, it was customary to compute the basal age, that is, the highest age at and below which all tests were passed. Partial credits, in months, were then added to this basal age for all tests passed at higher year levels. Mental age norms have also been employed with tests that are not divided into year levels. In such a case, the child’s raw score is first determined. The mean raw scores obtained by the children in each year group within the standardization sample constitute the age norms for such a test. The mean raw score of the e8-year old children, for example, would represent the 8-year old raw score then her or his mental age on the test is 8 years. All raw scores on such a test can be transformed in a similar manner by reference to the age norms.

Grade Equivalents: Scores on educational achievement tests are often interpreted in terms of grade equivalents. Grade norms are found by computing the mean raw score obtained by children in each grade. Thus, if the average number of problems solved correctly on an arithmetic test by the fourth graders in the **standardization** sample is 23, then a raw score of 23 corresponds to grade equivalents of 4. Intermediate grade equivalents, representing fractions of a grade, are usually found by interpolation, although they can also be obtained directly by testing children at different times within the school years. For example, 4.0 refers to average performance at the beginning of the fourth grade. Grade norms are also subject to misinterpretation unless the test user keeps firmly in mind the manner in which they were derived.

Ordinal Scales: Ordinal scales are designed to identify the stage reached by the child in the development of specific behavior functions. Although scores may be reported in terms of approximate age levels, such scores are secondary to qualitative **description** of the child's characteristics behavior. The ordinarily of such scales refers to the uniform progression of development through successive stages. in so far as these scales typically provide information about what the child is actually able to do(e.g. climbs stairs without assistance; recognizes identity in quantity of liquid when poured into differently shaped containers), they share important features with the domain-referenced tests.