Quadrant II – Transcript and Related Materials

Programme: Bachelor of Arts

Subject: Psychology

Course Code: PSC 106

Course Title: Psychological Testing

Unit: II Norms In Psychological Testing

Module Name: Standard Scores

Name of the Presenter: Ms. Sweta Shyam Matonkar

Notes

Standard Scores

Simply stated, a standard score is a raw score that has been converted from one scale to another scale, where the latter scale has some arbitrarily set mean and standard deviation. Why convert raw scores to standard scores? Raw scores may be converted to standard scores because standard scores are more easily interpretable than raw scores. With a standard score, the position of a testtaker's performance relative to other testtakers is readily apparent. Different systems for standard scores exist, each unique in terms of its respective mean and standard deviations. We will briefly describe z scores, T scores, stanines, and some other standard scores. First for consideration is the type of standard score scale that may be thought of as the zero plus or minus one scale. This is so because it has a mean set at 0 and a standard deviation set at 1. Raw scores converted into standard scores on this scale are more popularly referred to as z scores. Although percentiles are the most popular type of transformed score, standard scores exemplify the most desirable psychometric properties. A standard score uses the standard deviation of the total distribution of raw scores as the fundamental unit of measurement. The standard score expresses the distance from the mean in standard deviation units. For example, a raw score that is exactly one standard deviation above the mean converts to a standard score of +1.00. A raw score that is exactly one-half a standard deviation below the mean converts to a standard score of -.50. Thus, a standard score not only expresses the magnitude of deviation from the mean, but the direction of departure (positive or negative) as well. Computation of an examinee's standard score (also called a z score) is simple: Subtract the mean of the normative group from the examinee's raw score and then divide this difference by the standard deviation of the normative group. Standard scores possess the desirable psychometric property of retaining the relative magnitudes of distances between successive values found in the original raw scores. This is because the distribution of standard scores has exactly the same shape as the distribution of raw scores. As a consequence, the use of standard scores does not distort the underlying measurement scale. This fidelity of the transformed measurement scale is a major advantage of standard scores over percentiles and percentile ranks. As previously noted, percentile scores are very distorting, especially at the extremes. A specific example will serve to illustrate the non-distorting feature of standard scores. Consider four raw scores of 55, 60, 70, and 80 on a test with mean of 50 and standard deviation of 10. The first two scores differ by 5 raw score points, while the last two scores differ by 10 raw score points—twice the difference of the first pair. When the raw scores are converted to standard scores, the results are +0.50, +1.00, +2.00, and +3.00, respectively. The reader will notice that the first two scores differ by 0.50 standard scores, while the last two scores differ by 1.00 standard scores—twice the difference of the first pair. Thus, standard scores always retain the relative magnitude of differences found in the original raw scores. Standard score distributions possess important mathematical properties that do not exist in the raw score distributions. When each of the raw scores in a distribution is transformed to a standard score, the resulting collection of standard scores always has a mean of zero and a variance of 1.00. Because the standard deviation is the square root of the variance, the standard deviation of standard scores (11.00) is necessarily 1.00 as well. One reason for transforming raw scores into standard scores is to depict results on different tests according to a common scale. If two distributions of test scores possess the same form, we can make direct comparisons on raw scores by transforming them to standard scores. Suppose, for example, that a first-year college student earned 125 raw score points on a spatial thinking test for which the normative sample averaged 100 points (with SD of 15 points). Suppose, in addition, he earned 110 raw score points on a vocabulary test for which the normative sample averaged 90 points (with SD of 20 points). In which skill area does he show greater aptitude, spatial thinking or vocabulary?

If the normative samples for both tests produced test score distributions of the same form, we can compare spatial thinking and vocabulary scores by converting each to standard scores. The spatial thinking standard score for our student is (125 - 100)>15 or 1.67, whereas his vocabulary standard score is (110 - 90)>20 or +1.00. Relative to the normative samples, the student has greater aptitude for spatial thinking than vocabulary. But a word of caution is appropriate when comparing standard scores from different distributions. If the distributions do not have the same form, standard score comparisons can be very misleading. When two distributions of test scores do not possess the same form, equivalent standard scores do not signify comparable positions within the respective normative samples.

z Scores

A z score results from the conversion of a raw score into a number indicating how many standard deviation units the raw score is below or above the mean of the distribution. Let's use an example from the normally distributed "National Spelling Test" data in Figure 3–7 to demonstrate how a raw score is converted to a z score. We'll convert a raw score of 65 to a z score by using the formula Standard Score = Z = X – M/SD In essence, a z score is equal to the difference between a particular raw score and the mean divided by the standard deviation. In the preceding example, a raw score of 65 was found to be equal to a z score of +1. Knowing that someone obtained a z score of 1 on a spelling test provides context and meaning for the score. Drawing on our knowledge of areas under the normal curve, for example, we would know that only about 16% of the other testtakers obtained higher scores. By contrast, knowing simply that someone obtained a raw score of 65 on a spelling test conveys virtually no usable information because information about the context of this score is lacking.

In addition to providing a convenient context for comparing scores on the same test, standard scores provide a convenient context for comparing scores on different tests. As an example, consider that Crystal's raw score on the hypothetical Main Street Reading Test was 24 and that her raw score on the (equally hypothetical) Main Street Arithmetic Test was 42. Without knowing anything other than these raw scores, one might conclude that Crystal did better on the arithmetic test than on the reading test. Yet more informative than the two raw scores would be the two z scores. Converting Crystal's raw scores to z scores based on the performance of other students in her class, suppose we find that her z score on the reading test was 1.32 and that her z score on the arithmetic test was −0.75. Thus, although her raw score in arithmetic was higher than in reading, the z scores paint a different picture. The z scores tell us that, relative to the other students in her class (and assuming that the distribution of scores is relatively normal), Crystal performed above average on the reading test and below average on the arithmetic test. An interpretation of exactly how much better she performed could be obtained by reference to tables detailing distances under the normal curve as well as the resulting percentage of cases that could be expected to fall above or below a particular standard deviation point (or z score).

T Scores

If the scale used in the computation of z scores is called a zero plus or minus one scale, then the scale used in the computation of T scores can be called a fifty plus or minus ten scale; that is, a scale with a mean set at 50 and a standard deviation set at 10. Devised by W. A. McCall (1922, 1939) and named a T score in honor of his professor E. L. Thorndike, this standard score system is composed of a scale that ranges from 5 standard deviations below the mean to 5 standard deviations above the mean. Thus, for example, a raw score that fell exactly at 5 standard deviations below the mean would be equal to a T score of 0, a raw score that fell at the mean would be equal to a T of 50, and a raw score 5 standard deviations above the mean would be equal to a T of 100. One advantage in using T scores is that none of the scores is negative. By contrast, in a z score distribution, scores can be positive and negative; this can make further computation cumbersome in some instances. Formula : -

$$T = 10 \frac{(X - M)}{SD} + 50$$

Other Standard Scores

Numerous other standard scoring systems exist. Researchers during World War II developed a standard score with a mean of 5 and a standard deviation of approximately 2. Divided into nine units, the scale was christened a stanine, a term that was a contraction of the words standard and nine.

Stanine scoring may be familiar to many students from achievement tests administered in elementary and secondary school, where test scores are often represented as stanines. Stanines are different from other standard scores in that they take on whole values from 1 to 9, which represent a range of performance that is half of a standard deviation in width.

The 5th stanine indicates performance in the average range, from 1/4 standard deviation below the mean to 1/4 standard deviation above the mean, and captures the middle 20% of the scores in a normal distribution. The 4th and 6th stanines are also 1/2 standard deviation wide and capture the 17% of cases below and above (respectively) the 5th stanine. Another type of standard score is employed on tests such as the Scholastic Aptitude Test (SAT) and the Graduate Record Examination (GRE). Raw scores on those tests are converted to standard scores such that the resulting distribution has a mean of 500 and a standard deviation of 100. If the letter A is used to represent a standard score from a college or graduate school admissions test whose distribution has a mean of 500 and a standard deviation of 100, then the following is true: (A = 600) = (z = 1) = (T = 60) Have you ever heard the term IQ used as a synonym for one's score on an intelligence test? Of course you have. What you may not know is that what is referred to variously as IQ, deviation IQ, or deviation intelligence quotient is yet another kind of standard score. For most IQ tests, the distribution of raw scores is converted to IQ scores, whose distribution typically has a mean set at 100 and a standard deviation set at 15. Let's emphasize typically because there is some variation in standard scoring systems, depending on the test used. The typical mean and standard deviation for IQ tests results in approximately 95% of deviation IQs ranging from 70 to 130, which is 2 standard deviations below and above the mean. In the context of a normal distribution, the relationship of deviation IQ scores to the other standard scores we have discussed so far (z, T, and A scores).

Finally, we give brief mention to three raw score transformations that are mainly of historical interest. The stanine (standard nine) scale was developed by the United States Air Force during World War II. In a stanine scale, all raw scores are converted to a single-digit system of scores ranging from 1 to 9. The mean of stanine scores is always 5, and the standard deviation is approximately 2. The transformation from raw scores to stanines is simple: The scores are ranked from lowest to highest, and the bottom 4 percent of scores convert to a stanine of 1, the next 7 percent convert to a stanine of 2, and soon. The main advantage of stanines is that they are restricted to single-digit numbers. This was a considerable asset in the premodern computer era in which data was keypunched on Hollerith cards that had to be physically carried and stored on shelves. Because a stanine could be keypunched in a single column, far fewer cards were required than if the original raw scores were entered. Statisticians have proposed several variations on the stanine theme. Canfield proposed the 10-unit **sten scale**, with 5 units above and 5 units below the mean. Guilford and Fruchter proposed the **C scale** consisting of 11 units. Although stanines are still in widespread use, variants such as the sten and C scale never roused much interest among test developers.

Standard scores converted from raw scores may involve either linear or nonlinear transformations. A standard score obtained by a linear transformation is one that retains a direct numerical relationship to the original raw score. The magnitude of differences between such standard scores exactly parallels the differences between corresponding raw scores. Sometimes scores may undergo more than one transformation. For example, the creators of the SAT did a second linear transformation on their data to convert z scores into a new scale that has a mean of 500 and a standard deviation of 100. A nonlinear transformation may be required when the data under consideration are not normally distributed yet comparisons with normal distributions need to be made. In a nonlinear transformation, the resulting standard score does not necessarily have a direct numerical relationship to the original, raw score. As the result of a nonlinear transformation, the original distribution is said to have been normalized. Normalized standard scores Many test developers hope that the test they are working on will yield a normal distribution of scores. Yet even after very large samples have been tested with the instrument under development, skewed distributions result. What should be done? One alternative available to the test developer is to normalize the distribution. Conceptually, normalizing a distribution involves “stretching” the skewed curve into the shape of a normal curve and creating a corresponding scale of standard scores, a scale that is technically referred to as a normalized standard score scale.

Normalization of a skewed distribution of scores may also be desirable for purposes of comparability. One of the primary advantages of a standard score on one test is that it can readily be compared with a standard score on another test. However, such comparisons are appropriate only when the distributions from which they derived are the same. In most instances, they are the same because the two distributions are approximately normal. But if, for example, distribution A were normal and distribution B were highly skewed, then z scores in these respective distributions would represent different amounts of area subsumed under the curve. A z score of −1 with respect to normally distributed data tells us, among other things, that about 84% of the scores in this distribution were higher than this score. A z score of −1 with respect to data that were very positively skewed might mean, for example, that only 62% of the scores were higher.

For test developer's intent on creating tests that yield normally distributed measurements, it is generally preferable to fine-tune the test according to difficulty or other relevant variables so that the resulting distribution will approximate the normal curve. That usually is a better bet than attempting to normalize skewed distributions. This is so because there are technical cautions to be observed before attempting normalization. For example, transformations should be made only when there is good reason to believe that the test sample was large enough and representative enough and that the failure to obtain normally distributed scores was due to the measuring instrument.

References

1. Cohen, J. R. & Swerdlik, M. E. (2018). Psychological Testing and Assessment: An Introduction to Tests and Measurement. (9th ed.). New Delhi: McGraw-Hill Education.
2. Gregory, R. J. (2017). Psychological Testing: History, Principles, and Applications (7th Ed.).New Delhi: Pearson (India) Pvt. Ltd.