UNIT–8

INTERPRETING TEST SCORES

Written By:
Muhammad Azeem

Reviewed By:
Dr. Muhammad Tanveer Afzal

CONTENT

INTRODUCTION

Raw scores are considering as points scored in test when the test is scored according to the set procedure or rubric of marking. These points are not meaningful without interpretation or further information. Criterion referenced interpretation of test scores describes students' scores with respect to certain criteria while norm referenced interpretation of test scores describes students' score relative to the test takers. Test results are generally reported to parents as a feedback of their young one's learning achievements. Parents have different academic backgrounds so results should be presented them in understandable and usable way. Among various objectives three of the fundamental purposes for testing are (1) to portray each student's developmental level within a test area, (2) to identify a student's relative strength and weakness in subject areas, and (3) to monitor time-to-time learning of the basic skills. To achieve any one of these purposes, it is important to select the type of score from among those reported that will permit the proper interpretation. Scores such as percentile ranks, grade equivalents, and percentage scores differ from one another in the purposes they can serve, the precision with which they describe achievement, and the kind of information they provide. A closer look at various types of scores will help differentiate the functions they can serve and the interpretations or sense they can convey.

OBJECTIVES

After completing this unit, the students will be able to:

- understand what are the test score?
- understand what are the measurement scales used for test scores?
- ways of interpreting test score
- clarifying the accuracy of the test scores
- explain the meaning of test scores
- interpret test scores
- usability of test scores
- learn basic and significant concepts of statistics
- understand and usage of central tendency in educational measurements
- understand and usage of measure of variation in educational measurements
- planning and administration of test

8.1 Introduction of Measurement Scales and Interpretation of Test Scores

Interpreting Test Scores

All types of research data, test result data, survey data, etc is called raw data and collected using four basic scales. Nominal, ordinal, interval and ratio are four basic scales for data collection. Ratio is more sophisticated than interval, interval is more sophisticated than ordinal, and ordinal is more sophisticated than nominal. A variable measured on a "nominal" scale is a variable that does not really have any evaluative distinction. One value is really not any greater than another. A good example of a nominal variable is gender. With nominal variables, there is a qualitative difference between values, not a quantitative one. Something measured on an "ordinal" scale does have an evaluative connotation. One value is greater or larger or better than the other. With ordinal scales, we only know that one value is better than other or 10 is better than 9. A variable measured on interval or ration scale has maximum evaluative distinction. After the collection of data, there are three basic ways to compare and interpret results obtained by responses. Students' performance can be compare and interpreted with an absolute standard, with a criterion-referenced standard, or with a norm-referenced standard. Some examples from daily life and educational context may make this clear:

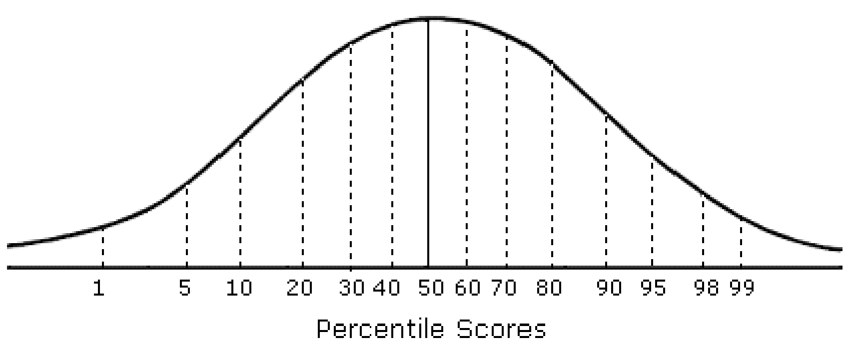
Sr. No.	Standard	Characteristics	daily life	educational context
1	Absolute	simply state the observed outcome	He is 6' and 2" tall	He spelled correctly 45 out of 50 English words
2	criterion-referenced	compare the person's performance with a standard, or criterion.	He is tall enough to catch the branch of this tree.	His score of 40 out of 50 is greater than minimum cutoff point 33. So he must promoted to the next class.
3	norm-referenced	compare a person's performance with that of other people in the same context.	He is the third fastest ballar in the pakistani squad 15.	His score of 37 out of 50 was not very good; 65% of his class fellows did better.

All three types of scores interpretation are useful, depending on the purpose for which comparisons made.

An absolute score merely describes a measure of performance or achievement without comparing it with any set or specified standard. Scores are not particularly useful without any kind of comparison. Criterion-referenced scores compare test performance with a specific standard; such a comparison enables the test interpreter to decide whether the scores are satisfactory according to established standards. Norm-referenced tests compare test performance with that of others who were measured by the same procedure. Teachers are usually more interested in knowing how children compare with a useful standard than how they compare with other children; but norm-referenced comparisons may also provide useful insights.

8.2 Interpreting Test Scores by Percentiles

The students' scores in terms of criterion-referenced scores are most easy to understand and interpret because they are straightforward and usually represented in percentages or raw scores while norm-referenced scores are often converted to derive standard scores or converted in to percentiles. Derived standard scores are usually based on the normal curve having an arbitrary mean to compare respondents who took the same test. The conversion of students' score into student's percentile score on a test indicates what percentage of other students are fell below that student's score who took the same test. Percentiles are most often used for determining the relative standing position of any student in a population. Percentile ranks are an easy way to convey a student's standing at test relative to other same test takers.



For example, a score at the 60th percentile means that the individual's score is the same as or higher than the scores of 60% of those who took the test. The 50th percentile is known as the median and represents the middle score of the distribution.

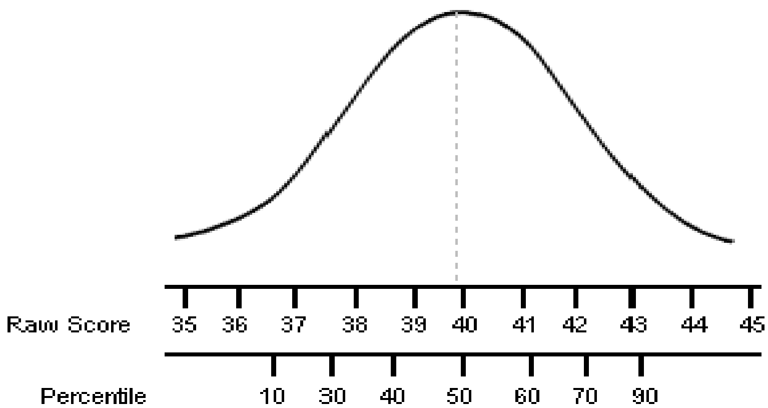
Percentiles have the disadvantage that they are not equal units of measurement. For instance, a difference of 5 percentile points between two individual's scores will have a different meaning depending on its position on the percentile scale, as the scale tends to exaggerate differences near the mean and collapse differences at the extremes.

Percentiles cannot be averaged nor treated in any other way mathematically. However, they do have the advantage of being easily understood and can be very useful when giving feedback to candidates or reporting results to managers.

If you know your percentile score then you know how it compares with others in the norm group. For example, if you scored at the 70th percentile, then this means that you scored the same or better than 70% of the individuals in the norm group.

Percentile score is easily understood when tend to bunch up around the average of the group i.e. when most of the student are the same ability and have score with very small rang.

To illustrate this point, consider a typical subject test consisting of 50 questions. Most of the students, who are a fairly similar group in terms of their ability, will score around 40. Some will score a few less and some a few more. It is very unlikely that any of them will score less than 35 or more than 45.



These results in terms of achievement scores are a very poor way of analyzing them. However, percentile score can interpret results very clearly.

Definition

A **percentile** is a measure that tells us what percent of the total frequency scored at or below that measure. A percentile rank is the percentage of scores that fall at or below a given score. OR

A **percentile** is a measure that tells us what percent of the total frequency scored below that measure. A percentile rank is the percentage of scores that fall below a given score.

Both definitions are seams to same but statistically not same. For Example

Example No.1

If Aslam stand 25^{th} out of a class of 150 students, then 125 students were ranked below Aslam.

Formula:

To find the percentile rank of a score, x, out of a set of n scores, where x is included: $\frac{(B + 0.5E)}{n}.100 = \text{percentilerank}$

Where B = number of scores below x

E = number of scores equal to x

n = number of scores

using this formula Aslam's percentile rank would be:

$$\frac{125+0.5(1)}{150}=\frac{125.5}{150}=.83\bar{6}=84^{th}\text{ percentile}$$

Formula:
To find the percentile rank of a score, x, out of a set of n scores, where x is not included:

$$\frac{\text{number of sources below } x}{n}.100=\text{percentile rank}$$

using this formula Aslam's percentile rank would be:

$$\frac{125}{150}.8\bar{3}=83^{rd}\text{ percentile}$$

Therefore both definition yields different percentile rank. This difference is significant only for small data. If we have raw data then we can find unique percentile rank using both formulae.

Example No.2

The science test scores are: 50, 65, 70, 72, 72, 78, 80, 82, 84, 84, 85, 86, 88, 88, 90, 94, 96, 98, 98, 99 Find the percentile rank for a score of 84 on this test.

Solution:

First rank the scores in ascending or descending order

50, 65, 70, 72, 72, 78, 80, 82, 84, |84, 85, 86, 88, 88, 90, 94, 96, 98, 98, 99

Since there are 2 values equal to 84, assign one to the group "above 84" and the other to the group "below 84".

Solution Using Formula:

$$\frac{(B+0.5E)}{n}.100=\text{percentile rank}$$

$$\frac{8+0.5(2)}{20}.100=\frac{9}{20}.100=45^{th}\text{ percentile}$$

Solution Using Formula:

$$\frac{\text{number of sources below } x}{n}.100 = \text{percentile rank}$$

$$\frac{9}{100}.100 = 45^{th} \text{ percentile}$$

Therefore score of 84 is at the 45^{th} percentile for this test.

Example No.3

The science test scores are: 50, 65, 70, 72, 72, 78, 80, 82, 84, 84, 85, 86, 88, 88, 90, 94, 96, 98, 98, 99. Find the percentile rank for a score of 86 on this test.

Solution:

First rank the scores in ascending or descending order

Since there is only one value equal to 86, it will be counted as "half" of a data value for the group "above 86" as well as the group "below 86".

Solution Using Formula:

$$\frac{(\mathrm{B} + 0.5\mathrm{E})}{n}.100 = \text{percentile rank}$$

$$\frac{11 + 0.5(1)}{20}.100 = \frac{11.5}{20}.100 = 58^{th} \text{ percentile}$$

Solution Using Formula:

$$\frac{\text{number of sources below } x}{n}.100 = \text{percentile rank}$$

$$\frac{11.5}{20}.100 = 57.5 = 58^{th} \text{ percentile}$$

The score of 86 is at the 58^{th} percentile for this test.

Keep in Mind:

- Percentile rank is a number between 0 and 100 indicating the percent of cases falling at or below that score.
- Percentile ranks are usually written to the nearest whole percent: 64.5% = 65% = 65^{th} percentile

- Scores are divided into 100 equally sized groups.
- Scores are arranged in rank order from lowest to highest.
- There is no 0 percentile rank - the lowest score is at the first percentile.
- There is no 100th percentile - the highest score is at the 99th percentile.
- Percentiles have the disadvantage that they are not equal units of measurement.
- Percentiles cannot be averaged nor treated in any other way mathematically.
- You cannot perform the same mathematical operations on percentiles that you can on raw scores. You cannot, for example, compute the mean of percentile scores, as the results may be misleading.
- Quartiles can be thought of as percentile measure. Remember that quartiles break the data set into 4 equal parts. If 100% is broken into four equal parts, we have subdivisions at 25%, 50%, and 75% .creating the:

First quartile (lower quartile) to be at the 25th percentile.
Median (or second quartile) to be at the 50th percentile.
Third quartile (upper quartile) to be a the 75th percentile.

8.3 Interpreting Test Scores by Percentages

The number of questions a student gets right on a test is the student's raw score (assuming each question is worth one point). By itself, a raw score has little or no meaning. For example if teacher says that Fatima has scored 8 marks. This information (8 marks) regarding Fatima's result does not convey any meaning. The meaning depends on how many questions are on the test and how hard or easy the questions are. For example, if Umair got 10 right on both a math test and a science test, it would not be reasonable to conclude that his level of achievement in the two areas is the same. This illustrates, why raw scores are usually converted to other types of scores for interpretation purposes. The conversion of raw score into percentage convey students' achievements in understanding and meaningful way. For example if Sadia got 8 questions right out of ten questions then we can say that Sadia is able to solve

$\frac{8}{10} \times 100$=80% questions. If each question carries equal marks then we can say that Sadia has scored 80% marks. If different questions carry different marks then first count marks obtained and total marks the test. Use the following formula to compute % of marks.

$$\frac{\text{Marks Otained}}{\text{Total Marks}} \times 100 = \%\ \text{marks}$$

Example:

The marks detail of Hussan's math test is shown. Find the percentage marks of Hussan.

Question	Q1	Q2	Q3	Q4	Q5	Total
Marks	10	10	5	5	20	50
Marks obtained	8	5	2	3	10	28

Solution:

Hussan' s marks = 28

Total marks =50

Hussan got $= \frac{\text{Marks Otained}}{\text{Total Marks}} \times 100 = \frac{28}{50} \times 100 = 56\ \%$

For example, a number can be used merely to label or categorize a response. This sort of number (nominal scale) has a low level of meaning. A higher level of meaning comes with numbers that order responses (ordinal data). An even higher level of meaning (interval or ratio data) is present when numbers attempt to present exact scores, such as when we state that a person got 17 correct out of 20. Although even the lowest scale is useful, higher level scales give more precise information and are more easily adapted to many statistical procedures.

Scores can be summarized by using either the mode (most frequent score), the median (midpoint of the scores), or the mean (arithmetic average) to indicate typical performance. When reporting data, you should choose the measure of central tendency that gives the most accurate picture of what is typical in a set of scores. In addition, it is possible to report the standard deviation to indicate the spread of the scores around the mean.

Scores from measurement processes can be either absolute, criterion referenced, or norm referenced. An absolute score simply states a measure of performance without comparing it with any standard. However, scores are not particularly useful unless they are compared with something. Criterion-referenced scores compare test performance with a specific standard; such a comparison enables the test interpreter to decide whether the scores are satisfactory according to established standards. Norm-referenced tests compare test performance with that of others who were measured by the same procedure. Teachers are usually more interested in knowing how children compare with a useful standard than how they compare with other children; but normreferenced comparisons may also provide useful insights.

Criterion-referenced scores are easy to understand because they are usually straightforward raw scores or percentages. Norm-referenced scores are often converted to percentiles or other derived standard scores. A student's percentile score on a test indicates what percentage of other students who took the same test fell below that student's score. Derived scores are often based on the normal curve. They use an arbitrary mean to make comparisons showing how respondents compare with other persons who took the same test.

8.4 Interpreting Test Scores by ordering and ranking

Organizing and reporting of students' scores start with placing the scores in ascending or descending order. Teacher can find the smallest, largest, rang, and some other facts like variability of scores associated with scores from ranked scores. Teacher may use ranked scoes to see the relative position of each student within the class but ranked scores does not yield any significant numerical value for result interpretation or reporting.

8.4.1 Measurement Scales

Measurement is the assignment of numbers to objects or events in a systematic fashion. Measurement scales are critical because they relate to the types of statistics you can use to analyze your data. An easy way to have a paper rejected is to have used either an incorrect scale/statistic combination or to have used a low powered statistic on a high powered set of data. Following four levels of measurement scales are commonly distinguished so that the proper analysis can be used on the data a number can be used merely to label or categorize a response.

8.4.1.1 Nominal Scale.

Nominal scales are the lowest scales of measurement. A nominal scale, as the name implies, is simply some placing of data into categories, without any order or structure. You are only allowed to examine if a nominal scale datum is equal to some particular value or to count the number of occurrences of each value. For example, categorization of blood groups of classmates into A, B. AB, O etc. In The only mathematical operation we can perform with nominal data is to count. Variables assessed on a nominal scale are called **categorical variables;** Categorical data are measured on nominal scales which merely assign labels to distinguish categories. For example, gender is a nominal scale variable. Classifying people according to gender is a common application of a **nominal** scale.

Nominal Data

- classification or gatagorization of data, e.g. male or female
- no ordering, e.g. it makes no sense to state that male is greater than female (M > F) etc
- arbitrary labels, e.g., pass=1 and fail=2 etc

8.4.1.2 Ordinal Scale.

Something measured on an "ordinal" scale does have an evaluative connotation. You are also allowed to examine if an ordinal scale datum is less than or greater than another value. For example rating of job satisfaction on a scale from 1 to 10, with 10 representing complete satisfaction. With ordinal scales, we only know that 2 is better than 1 or 10 is better than 9; we do not know by how much. It may vary. Hence, you can 'rank' ordinal data, but you cannot 'quantify' differences between two ordinal values. Nominal scale properties are included in ordinal scale.

Ordinal Data

- ordered but differences between values are not important. Difference between values may or may not same or equal.
- e.g., political parties on left to right spectrum given labels 0, 1, 2
- e.g., Likert scales, rank on a scale of 1..5 your degree of satisfaction
- e.g., restaurant ratings

8.4.1.3 Interval Scale

An ordinal scale has quantifiable difference between values become interval scale. You are allowed to quantify the difference between two interval scale values but there is no natural zero. A variable measured on an interval scale gives information about more or better as ordinal scales do, but interval variables have an equal distance between each value. The distance between 1 and 2 is equal to the distance between 9 and 10. For example, temperature scales are interval data with 25C warmer than 20C and a 5C difference has some physical meaning. Note that 0C is arbitrary, so that it does not make sense to say that 20C is twice as hot as 10C but there is the exact same difference between 100C and 90C as there is between 42C and 32C. Students' achievement scores are measured on interval scale

Interval Data

- ordered, constant scale, but no natural zero
- differences make sense, but ratios do not (e.g., 30°-20°=20°-10°, but 20°/10° is not twice as hot!
- e.g., temperature (C,F), dates

8.4.1.4 Ratio Scale

Something measured on a ratio scale has the same properties that an interval scale has except, with a ratio scaling, there is an absolute zero point. Temperature measured in Kelvin is an example. There is no value possible below 0 degrees Kelvin, it is absolute zero. Physical measurements of height, weight, length are typically ratio variables. Weight is another example, 0 lbs. is a meaningful absence of weight. This ratio hold true regardless of which scale the object is being measured in (e.g. meters or yards). This is because there is a natural zero.

Ratio Data

- ordered, constant scale, natural zero
- e.g., height, weight, age, length

One can think of nominal, ordinal, interval, and ratio as being ranked in their relation to one another. Ratio is more sophisticated than interval, interval is more sophisticated than ordinal, and ordinal is more sophisticated than nominal.

8.5 Frequency Distribution

Frequency is how often something occurs. The frequency (**f**) of a particular observation is the number of times the observation occurs in the data.

Distribution

The *distribution* of a variable is the pattern of frequencies of the observation.

Frequency Distribution

It is a representation, either in a graphical or tabular format, which displays the number of observations within a given interval. Frequency distributions are usually used within a statistical context.

8.5.1 Frequency Distribution Tables

A frequency distribution table is one way you can organize data so that it makes more sense. Frequency distributions are also portrayed as frequency tables, histograms, orpolygons. Frequency distribution tables can be used for both categorical and numeric variables. The intervals of frequency table must be mutually exclusive and exhaustive. Continuous variables should only be used with class intervals. By counting frequencies, we can make a frequency distribution table. Following examples will figure out procedure of construction of frequency distribution table.

Example 1

For example, let's say you have a list of IQ scores for a gifted classroom in a particular elementary school. The IQ scores are: 118, 123, 124, 125, 127, 128, 129, 130, 130, 133, 136, 138, 141, 142, 149, 150, 154. That list doesn't tell you much about anything. You could draw a frequency distribution table, which will give a better picture of your data than a simple list.