Needs of Scaling

Scales are such which can be used in quantifying every measurable property of objects or the variables. To measure the physical phenomena different kinds of scales are to be needed. The scaling techniques are applied in ordering a series of items along some sort of continuum. The method of scaling is the technique of turning a series of qualitative facts into a quantitative series. The main needs of scaling techniques are:

(1) **For attaining scientific maturity**: The fundamental form of the movement in the direction of the greater precision is measuring graduation of the objects for a scientific study. Hence, the scaling technique is for the scientific maturity.

(2) **For objective measurement**: The scaling techniques are very useful in technical and social studies and the reliable inference about the technical/social phenomenon can be drawn by the use of the objective measurement.

(3) **For the improvement of more precise measuring device**: The existing measuring instrument and techniques can be improved by the use of different scales. More precise measuring devices are developed through the scientific use and practice of the existing scales.

Characteristics of a Scale

The essential characteristics of a good scale are:

a**) Continuum** **b) Reliability** **c) Validity** **d) Practicability**

Continuum

It is the characteristics of scale that it should be in the form of continuous series and the factors to be measured interrelated. The continuum to be defined depends upon the nature of the phenomenon and the nature of the factors to be defined.

Reliability

To obtain consistent result the measuring instrument should be steadfast. The reliability of the measuring scale must be stable, consistent and the error function should be correctly defined. It means the stability, consistency, dependability, predictability and accuracy of the scale used. According to Ebil 'The term reliability means the consistency with which set of test scores measure whatever they do measure'. Reliability is necessary for validity but not sufficient. The reliability of a test suffers to the consistency of score obtained of the some item/individual on different occasions or with different sets of equivalent.

Definition

a) If we get the same measuring result for the same object / observation in many times then the measure (scale) may be considered as reliable measure (scale).

b) The absence of measurement error in measuring instrument is the reliability of the measure.

c) The true measure of the property measure is called the reliable measure.

Validity

The most critical criterion of the scaling technique is the validity. A scale is said to be valid when it measures correctly what is expected to be measure. In other words, validity is the extent (degree) to which differences found with a measuring instrument reflect true difference among those being tested. It is thought as utility of the measure.

In experimentation, **reliability** is the extent to which the measurements resulting from a test reflect characteristics of the subject of measurement. An experiment is reliable if we are getting consistent results from the same measure. It is unreliable if repeat measurements of the same thing give different results.

In statistics a valid measure is one which is measuring what it is supposed to measure. **Validity** implies reliability (accuracy). A valid measure must be reliable, but a reliable measure may not be valid. Validity refers to getting results that accurately reflect the concept being measured.

Practicability

It is the characteristics of the measure that must be practicable to use. It should be reasonable economically, conveniently and interpretably. It should be easy to administer. The scales used must be supplemented by i) detailed instruction of handling ii) scoring techniques (keys) iii) guides for utilize and iv) evidence about reliability.

Reliability

Definitions of Reliability:

A measuring scale is said to be reliable if it provides consistent result. Prof. Goode and Hatt say 'a scale is reliable when it will consistently produce the same results when applied to the same sample'. 'Just as a ruler which shrank or expand materially when exposed to temperature, changes would be useless, so would it be a scale which yielded a different result upon each application?

To test the reliability of a measure (test) statistical method of analyzing (i) the scores obtaining in a single measure (test) or (ii) the correlation of the scores obtained in the more than one measure (tests) are used.

Methods of Estimating Reliability (Tests of Reliability)

To obtain the reliability of the different scores, the following four method of computing reliability, coefficient are used.

a) Test- retests Method

b) Parallel/Alternate/Equivalent forms Method

c) Split Half Method

d) Rational Equivalence Method (Kuder Richardson Method)

(a) **Test-retest Method**

In this method the same set of objects/ items is measured (tested) again and again by using the same or the comparable measuring instrument. The results so obtain are compared by computing correlation coefficient between the scores of the different tests (measures). If it is impossible to use such method due to the long space of time, it is considered whether the effects of causative factor in the period of two tests are present or not. To find the result related to this problem technique of control group (a team checking by trained and motivated persons) is applied.

(b) **Parallel Forms Method**

In this method two analogous forms of test-scales are constructed and alternatively applied to the same samples/items selected/objects selected. The analogous forms mean the alternate or parallel forms of the tests having of the same kind difficulty category and design. To obtain the reliability coefficient the correlation coefficient between the test-results is computed. If the results are in high degree of connection then the scale may be said to be reliable. The correlation of the forms is called self-correlation and it becomes an index of equivalence of two sets. The parallel forms are mostly useful for the standard psychological and educational achievement tests.

(c) Split half Method

In this method the scores are divided in two equal parts randomly. For example, the scores on odd no. of test-set as a first half and the scores on even no. of test-set as the other half. Considering a part as complete, scaling procedure is applied for these two half-parts. If there exists a high degree of correlation between the scores of these halves, then the scaling is considered to be reliable. The split half method is generally used when it is difficult to construct the parallel form test. The coefficient of internal consistency or the self-correlation coefficient of the whole test is called Stepped up reliability and estimated by using Spearman-Brown formula. The general formula is

$$Rw = \frac{n * r_p}{1+(n-1)r_p}$$

Where, R_w is stepped up reliability coefficient; n the number of parts

r_p = the correlation coefficient between two parts

Spearman- Brown formula for computing test reliability having two halves is

$R_{xy} = \frac{2 \times r_{xy}}{1 + r_{xy}}$

Where,

R_{xy} = Stepped up reliability coefficient of first and second half

2 is the number of parts and

r_{xy} = The correlation coefficient between two parts X and Y.

If the score are expressed in the ranks then, correlation coefficient is calculated by

$r_{xy} = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$, Which is called Spearman's Rank correlation Coefficient?

If the scores are in the numeric scale then, correlation coefficient is calculated by,

$$r_{xy} = \frac{n\Sigma XY - \Sigma X \times \Sigma Y}{\sqrt{nY^2 - (\Sigma Y)^2}\sqrt{n\Sigma X^2 - (\Sigma X)^2}}$$

Which is called Karl Pearson's coefficient of correlation, where, X represent the scores of the first half and Y the scores on the second set.

Example

A test-score is divided in two halves as the scores on the odd numbered questions and the scores on the even numbered questions. The correlation coefficient between them is obtained as 0.72, what is the reliability coefficient of the whole test.

Solution

Here, Correlation coefficient (r_{xy}) = 0.72

n = 2

Reliability coefficient, $R_{xy} = \frac{2 \times r_{xy}}{1 + r_{xy}} = \frac{2 \times 0.72}{1 + 0.72} = 0.8272$

The reliability coefficient is 82.72%; the dependability of the whole score seems to be very good.

d) Rational Equivalence Method (Kuder-Richardson Method)

Two forms of a test are defined as equivalence when corresponding items are interchangeable and inter item correlation is same for both forms. Kuder-Richardson method is the method of obtaining reliability by using the internal consistency between the measures (questions) of the same scaling (test). The reliability coefficient for this method is obtained by the following two formulae. Kuder-Richardson's first formula for reliability is denoted by KR_1, is computed as

$$KR_1 = R_w = \frac{n}{n-1}\left[1 - \frac{\Sigma pq}{\sigma^2}\right]$$

Where, R_w = The reliability coefficient of the whole test

n - The number of items in the test

σ -Standard deviation of the test score

p-the proportion of the answering test item correctly

q = (1 - p)

Note 1: If the values of p for each test is equal then Σpq = npq

2: If the values of p for 'n' tests are $p_1, p_2 \ldots\ldots p_n$, then $\Sigma pq = p_1q_1 + p_2q_2 + \ldots\ldots\ldots\ldots + p_nq_n$

Example

In a test there are 60 questions. The proportion of answering each question correctly is 70%, if the standard deviation is 10 what is the reliability coefficient?

Solution

Number of questions (n) = 60,

Proportion of correct answers (p) = 0.7 q = 1 - p = 1 - 0.7 = 0.3

Σ pq = npq

S.D. (s) = 10, therefore, the reliability coefficient is given by

$$KR_1 = R_w = \frac{n}{n-1}\left[1 - \frac{\Sigma pq}{\sigma^2}\right]$$

$$KR_1 = \frac{60}{60-1}\left[1 - \frac{60 \times 0.7 \times 0.3}{10 * 10}\right]$$

Hence reliability coefficient is 88.88%