Paper – Business Research Methods Dr. James Hussain
Paper Code-MB 202, Unit-III Assistant Professor (Guest Faculty)
MBA- Sem-II Email.-mbajames123@gmail.com

TOPIC- RELIABILITY & VALIDITY

RELIABILITY

Reliability refers to the consistency or repeatability of an operationalized measure. A reliable measure will yield the same results over and over again when applied to the same thing. It is the degree to which a test consistently measures whatever it measures. If you have a survey question that can be interpreted several different ways, it is going to be unreliable. One person may interpret it one way and another may interpret it another way. You do not know which interpretation people are taking. Even answers to questions that are clear may be unreliable, depending on how they are interpreted.

Reliability refers to the consistency of scores obtained by the same persons when they are re-examined with the same tests on different occasions, or with different sets of equivalent items, or under other variable examining conditions. Research requires dependable measurement. Measurements are reliable to the extent that they are repeatable and that any random influence which tends to make measurements different from occasion to occasion or circumstance to circumstance is a source of measurement error. Errors of measurement that affect reliability are random errors and errors of measurement that affect validity are systematic or constant errors. Reliability of any research is the degree to which it gives an accurate score across a range of measurement. It can thus be viewed as being 'repeatability' or 'consistency'.

TYPES OF RELIABILITY

Different methods are used to measure the reliability of Score

1.Test-Retest Reliability: (Same people, different times) The most obvious method for finding the reliability of test scores is by repeating the identical test on a second occasion. Test-retest reliability is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals. The scores from 'Time 1' and 'Time 2' can then be correlated in order to evaluate the test for stability over time. The reliability coefficient in this case is simply the correlation between the scores obtained by the same persons on the two administrations of the test. The error variance corresponds to the random fluctuations of performance from one test session to the other.

Advantage: Simple and straightforward.

Disadvantage: More or less the disadvantages of this reliability are - Practice effect: Practice will probably produce varying amounts of improvement in the retest scores of different individuals. Interval effect: If the interval between retest is fairly short, the test takers may recall many of their former response.

2. Split-Half Reliability: Split-half reliability is a subtype of internal consistency reliability. In split-half reliability we randomly divide all items that purport to measure the same construct into two sets. We administer the entire instrument to a sample of people and calculate the total score for each randomly divided half. The most commonly used method to split the test into two is using the odd-even strategy. The split-half reliability estimate,

From a single administration of one form of a test; it is possible to arrive at a measure of reliability by various split-half procedures. In such a way, two scores are obtained for each person by dividing the test into equivalent halves. Split-half reliability provides a measure of consistency with regard to content sampling.

The steps that are followed for this reliability -

Step 1: Divide the test into equivalent halves.

Step 2: Compute a ‘Pearson r’ between scores on the two halves of the test.

Step 3: Adjust the half-test reliability using the ‘Spearman-Brown’ formula. The Spearman-Brown formula is widely used in determining reliability by the spilt-half method, many test manuals reporting reliability in this form. When applied to split-half reliability, the formula always involves doubling the length of the test. Under these conditions, it can be simplified as follows -

$$\text{Reliability} = \frac{2 \times r_{half\text{-}test}}{1 + r_{half\text{-}test}}$$

3.Parallel-Forms Reliability (Different people, same time, different test)

Parallel forms reliability is a measure of reliability obtained by administering different versions of an assessment tool (both versions must contain items that probe the same construct, skill, knowledge base, etc.) to the same group of individuals. The scores from the two versions can then be correlated in order to evaluate the consistency of results across alternate versions. In parallel forms reliability you first have to create two parallel forms. One way to accomplish this is to create a large set of questions that address the same construct and then randomly divide the questions into two sets. You administer both instruments to the same sample of people. The correlation between the two parallel forms is the estimate of reliability. For example if you wanted to evaluate the reliability of a critical thinking assessment, you might create a large set of items that all pertain to critical thinking and then randomly split the questions up into two sets, which would represent the parallel forms. One major problem with this approach is that you have to be able to generate lots of items that reflect the same construct. Furthermore, this approach makes the assumption that the randomly divided halves are parallel or equivalent. Even by chance this will sometimes not be the case.

Cronbach's Alpha (α) Reliability- Cronbach's alpha (often symbolized by the lower case Greek letter α) is a measure of internal consistency, that is, how closely related a set of items are as a group. It is considered to be a measure of scale reliability. A 'high' value for alpha does not imply that the measure is unidimensional. If, in addition to measuring internal consistency, you wish to provide evidence that the scale in question is unidimensional, additional analyses can be performed. Exploratory factor analysis is one method of checking dimensionality. Technically speaking, Cronbach's alpha is not a statistical test - it is a coefficient of reliability (or consistency). Cronbach's alpha is the most common measure of internal consistency (reliability). It is most commonly used when you have multiple Likert questions in a survey/questionnaire that form a scale and you wish to determine if the scale is reliable. For example a researcher has devised a nine-question questionnaire to measure how safe people feel at work at an industrial complex. Each question was a 5-point Likert item from 'strongly disagree' to 'strongly agree'. In order to understand whether the questions in this questionnaire all reliably measure the same latent variable (feeling of safety) [so a Likert scale could be constructed], a Cronbach's alpha was run on a sample size of 15 workers. The alpha coefficient for the items is .839, suggesting that the items have relatively high internal consistency. Note that a reliability coefficient of .70 or higher is considered 'acceptable' in most social science research situations

VALIDITY

Validity refers to whether the measure actually measures what it is supposed to measure. If a measure is unreliable, it is also invalid. That is, if you do not know what it is measuring, it certainly cannot be said to be measuring what it is supposed to be measuring. On the other hand, you can have a consistently unreliable measure. For example, if we measure income level by asking someone how many years of formal education they have completed, we will

get consistent results, but education is not income (although they are positively related). If the 'trade dress' of a product refers to the total image of a product, then measuring how people perceive the product's color and shape by themselves falls far short of measuring the product's 'trade dress'. It is an invalid measure. In general, validity is an indication of how sound your research is. More specifically, validity applies to both the design and the methods of your research. Validity in data collection means that your findings truly represent the phenomenon you are claiming to measure. Valid claims are solid claims

Validity is described as the degree to which a research study measures what it intends to measure. There are two main types of validity, internal and external. Internal validity refers to the validity of the measurement and test itself, whereas external validity refers to the ability to generalize the findings to the target population.

TYPES OF VALIDITY

1.Face Validity: Face validity refers to the degree to which a test appears to measure what it purports to measure. The stakeholders can easily assess face validity. Although this is not a very 'scientific' type of validity, it may be an essential component in enlisting motivation of stakeholders.

2. Predictive Validity: Predictive validity refers to whether a new measure of something has the same predictive relationship with something else that the old measure had. In predictive validity, we assess the operationalization's ability to predict something it should theoretically be able to predict.

3. Content Validity: In content validity, you essentially check the operationalization against the relevant content domain for the construct. This approach assumes that you have a good detailed description of the content domain, something that's not always true. In content validity, the criteria are the construct definition itself - it is a direct comparison.

4. Convergent Validity: Convergent validity refers to whether two different measures of presumably the same thing are consistent with each other - whether they converge to give the same measurement. In convergent validity, we examine the degree to which the operationalization is similar to (converges on) other operationalizations that it theoretically should be similar to.

5. Concurrent Validity: Concurrent validity is the degree to which the scores on a test are related to the scores on another already established, test administered at the same time or to some other valid criterion available at the same time.

6. Construct Validity: Construct validity is used to ensure that the measure is actually measure what it is intended to measure (i.e. the construct), and not other variables. Using a panel of 'experts' familiar with the construct is a way in which this type of validity can be assessed. The experts can examine the items and decide what that specific item is intended to measure