Reliability and Validity

Marilyn K Simon
What Makes a Good Instrument?

- **Validity**
  - Does the instrument measure what it’s supposed to?

- **Reliability**
  - Does the instrument produce the same results when completed under the same conditions?
  - Are the measurements consistent?
Validity/Reliability

- We want to hit the bull’s eye

---

Validity

- Validity denotes the extent to which an instrument is measuring what it is supposed to measure.
- It indicates extent of relationship between a scale and the measure of independent criterion variable.
- For an instrument (and thus research) to be trustworthy we need to be able to assume that the data collected are both reliable and valid.
Content Validity

- is concerned with the relevance of contents of items, individually and as a whole.

Methods:
- Correlating experts’ judgment
- Item-item or item-total correlation
Criterion - Related Validity

- Correlating the test scores obtained on a criterion variable with the variable being predicted – usually the dependent variable.

- Methods:
  - Correlating the test with criteria during data collection (Concurrent validity)
Construct Validity

- is concerned with the extent to which an instrument may be said to measure a theoretical construct or trait. Construct is similar to a concept, which is formally proposed with a definition and is related to empirical data.

- **Methods:**
  - Factor analysis (Factorial validity)
  - Correlating with other theoretical measure with which the developing instrument should correlate (Convergent validity)
  - Correlating with other theoretical measure with which the developing instrument should not correlate (Discriminant validity)
Predictive Validity

- Predictive validity involves testing a group of participants for a certain construct, and then comparing the outcome with results obtained at some point in the future.

- There is a time gap in collection of data from the test to the criterion measure.
Reliability Vs Validity

- Reliability has to do with the quality of measurement.
- Reliability is the "consistency" or "repeatability" of measures.
- Reliability does not imply validity.
- In terms of accuracy and precision, reliability is analogous to precision, while validity is analogous to accuracy.
Types of Reliability

- **Inter-rater reliability**: is assessed by having two or more independent judges score the test.

- **Test-retest reliability**: is measured by administering a test twice at two different points in time. This type of reliability assumes that there will be no change in the quality or construct being measured. In most cases, reliability will be higher when little time has passed between tests.

- **Parallel-forms reliability**: is gauged by comparing to different tests that were created using the same content.

- **Internal Consistency reliability**: is used to judge the consistency of results across items on the same test.
Cronbach’s alpha

- Cronbach's alpha ($\alpha$) is a measure of internal consistency, that is, how closely related a set of items are as a group.
- A "high" value of $\alpha$ is often used (along with substantive arguments and possibly other statistical measures) as evidence that the items measure an underlying (or latent) construct.
- We generally look for $\alpha$ values between 0.7 and 0.8.
Steps to computing Cronbach’s $\alpha$ with SPSS

- Enter the data into SPSS
- Select "Analyze"
- Select "Scale"
- Select "Reliability Analysis"
- A pop-up window will appear for reliability analysis. In this window are two boxes, one to the left and one to the right. The left contains the variables, or items, entered in SPSS, the box to the right, which is labeled "Items," is where one moves those scale items for which Cronbach's $\alpha$ is desired.
- Click on the "Statistics" button
- check mark next to "Scale" and "Scale if item deleted."
- Click Continue and Run statistics.
Computing Cronbach's Alpha
Selecting the Statistics for the Output

 reli Reliability Analysis: Statistics

 Descriptives for
  ✔ Scale
  ✔ Scale if item deleted

 Inter-Item
  ✔ Correlations
  ✔ Covariances

 Summaries
  ✔ Means
  ✔ Variances
  ✔ Covariances
  ✔ Correlations

 ANOVA Table
  ✔ None
  ✔ F test
  ○ Friedman chi-square
  ○ Cochran chi-square

 □ Hotelling’s T-square
 □ Tukey’s test of additivity

 Intra-class correlation coefficient
  Model: Two-Way Mixed
  Type: Consistency
  Confidence interval: 95%
  Test value: 0
Completing the Specifications
Cronbach’s $\alpha$

- The result will indicate if the two factors that are being compared are correlated.
- The results will indicate if respondents who rate high (low) scores for one item also tended to select high (low) scores for the other;
- If $\alpha$ is low, the ability to predict scores from one item would not be possible.
Alternate-Form Method

- Construction of two tests that supposedly sample the same material.
- Parallel-form or equivalence method. They are given to the same participants. Correlation coefficient (Pearson r) > 0.70
Split-Half Method

- Splits the questionnaire into two random halves, calculate scores, and correlates them Odd and even numbers of the scale.

- Correlation coefficient (Pearson r) Spearman-Brown prophecy formula is often used.
Concluding Remarks

- The importance of an instrument achieving a reasonable level of reliability and validity cannot be overemphasized.
- If an instrument lacks validity or reliability, the meaning of individual scores becomes otiose.
- A score of 90 on an invalid or unreliable test would be no different from a score of 50.
