



Reliability and Validity



P The concepts of reliability and validity of measures

▶ Formal definitions

- Reliability: The extent to which you can expect to get the same answer if you apply it twice to the same object.

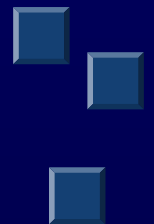
- Repeatability

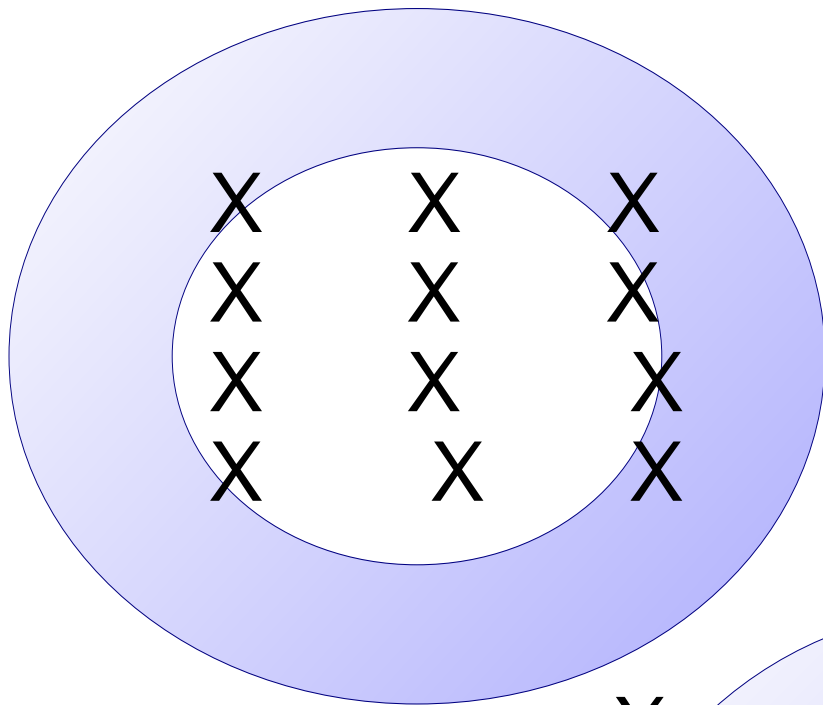
- Book says “measuring something consistently and dependably”

- Validity: The extent to which an operational definition represents the conceptual definition

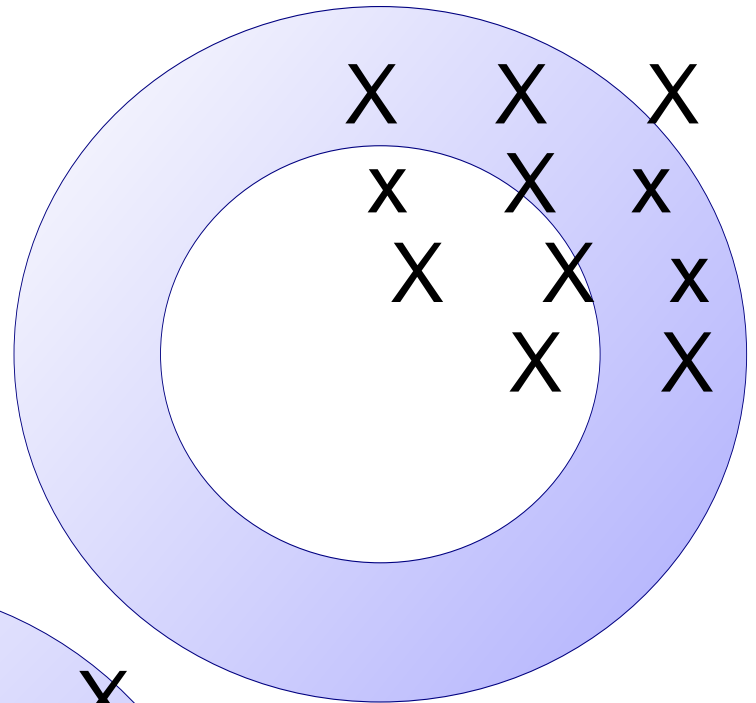
- Are you measuring what you intended to measure?

▶ Target analogy

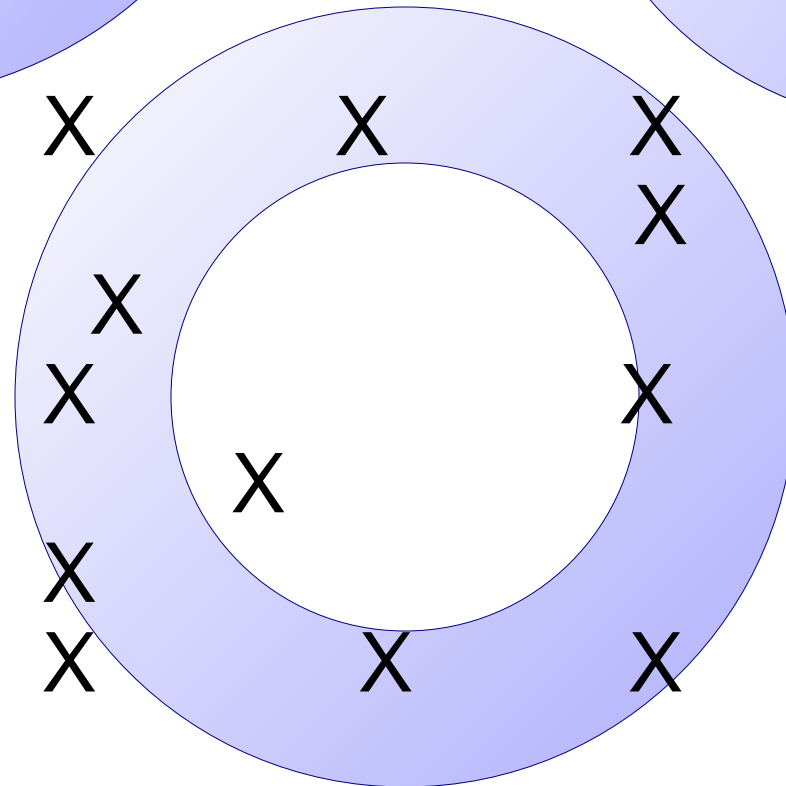




Reliable and valid



Reliable but not valid



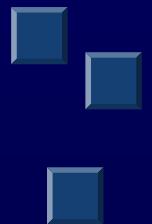
Not reliable, but somewhat valid



Reliability and Validity

P Reliability


- Factors in whether it is high or low
 - In observations, what were your problems?
 - Different point of view
 - Lack of clarity in operational definition
- What about an open-end code? How improve reliability?
 - Have clear definitions, with examples
 - Training of coders
 - Practice
- Ways of assessing it, for indexes such as yours
 - Test-retest
 - Alternate forms
 - Internal consistency





Reliability and Validity

P Validity: the concept of a criterion

- ▶ What is a criterion? A separate measure by which we can assess the validity of a measure we have developed.
 - ▶ Example of football player
 - Referee calls his catch out of bounds. What is criterion to assess validity of that measure?
 - Right, instant replay.
 - He claims to be self-confident. What is criterion to assess validity of that statement?
 - Ask his friends, coach, other team members
 - Observe his behavior
 - Take physiological measures
- 

Reliability and Validity

P Kinds of validity

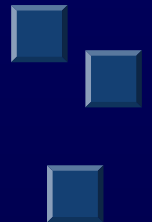
- ▶ face
 - Just what it says: on its face, it looks like a measure
- ▶ convergent/concurrent
 - A second similar kind of measure is correlated with it
 - E.g., open-ended vs. Index in your questionnaires
- ▶ criterion
 - Known groups differ in expected ways on measure
- ▶ construct/predictive
 - Relates as it should, theoretically, to other variables
- ▶ discriminant
 - It measures *differently* from a measure known or thought to measure a different concept: e.g. love and liking



Reliability and Validity

P Scales and Indexes

- ▶ Why do we use scales and indexes instead of single item measures?
 - Only need them when measuring complex concepts with multiple dimensions
 - They increase both reliability and validity
 - No single item is very reliable, so having multiple items increases reliability of measure. Notion of equivalence. Errors cancel out
 - No single item captures all of a complex concept. So having multiple items allows for coverage of entire concept.
- ▶ Two kinds of scaling
 - Rational
 - Empirical

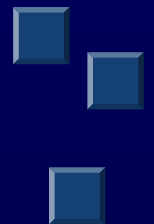




Reliability and Validity

P Rubin article

- ▶ Purpose
 - Develop and validate a measure of romantic love
- ▶ Method
 - Rational or empirical?
 - Where did he get the items?
 - Steps after developing initial list of items
 - Judging and sorting into love and liking
 - Factor analysis --> two sets of items
 - Couples in romantic relationships answer questions
 - Gazing experiment
- ▶ Your assessment?
- ▶ Walster scale

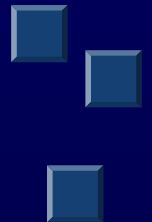




Reliability and Validity

P The questionnaire study. What is it for?

- ▶ Think of it as a pretest for a future study
 - Trying to develop a reliable and valid measure of your central concept
 - Concept of equivalence: assume each of items is equally good at measuring concept, although may measure different aspects or dimensions
 - Don't want to use more items than needed. Look at items for
 - Confusion – here is where you look at “process”
 - Range – power to move people around - for this you look at frequencies





Reliability and Validity

- Finally, reliability – how they correlate with other items – for this you look at item analysis
 - Correlation matrix
 - Item to total correlations
- ▶ Then check index for validity
 - Two tests of validity
 - Concurrent/convergent validity: open-ended vs index test
 - Construct/predictive validity: Obvious hypothesis test

