Video 9 - Validity and reliability

Steve Simon

Measurement quotes (1 of 2)

– "The government is extremely fond of amassing great quantities of statistics. These are raised to the Nth degree, the cube roots are extracted, and the results are arranged into elaborate and impressive displays. What must be kept ever in mind, however, is that in every case, the figures are first put down by a village watchman, and he puts down anything he damn well pleases."

• Sir Josiah Stamp, as quoted on Quotetab.

Measurement quotes (2 of 2)

– "only scientists are arrogant enough to think that they always observe with rigorous and objective scrutiny"

- Stephen Jay Gould, The Mismeasure of Man, page 36.

Measurements that warrant closer scrutiny

– Patient reported outcomes

- Participant report

– Researcher evaluations

- Only when concerned about subjectivity

– Psychological constructs

– Composite scores

Composite measure - PHQ-9

PHQ questionnaire

Composite measure - NES

NES questionnaire

Three types of validity

– Internal validity
  • "The extent to which we can infer that the independent variable caused the dependent variable."
– External validity
  • "The extent to which the findings will generalize to other populations, settings, measures, and treatments."
– Measurement validity
  • "The quality of accuracy of individual measures or scores. The extent to which a score measures what it was intended to measure."

Measurement Reliability

– Synoynms: consistency, precision, stability
– Classical test theory
  • Observed value = True value + Measurement error
  • This is a purely hypothetical model
– Reliability coefficient
  • Variance of true values / Variance of measured values
– No measurement is perfectly reliable
  • Strive for 0.7 or higher in research
  • 0.6 is "borderline".
  • Might require 0.9 or higher for individual decisions

3/18/2019

Take a break here

– What you have learned.
  • Measurements that require special scrutiny
  • Reliability coefficient
– What's coming next
  • Indirect measures of the reliability coefficient

Indirect measures of the reliability coefficient

– Test-retest
– Interrater
– Parallel forms
– Internal consistency
  • Split-half
  • Kuder-Richardson 20
  • Cronbach's alpha

Test-retest reliability

– Also called repeatability
– Correlation of two measurements separated by time
– Length of time interval is critical
  • No carry-over
  • No changes in the true score

Inter-rater reliability

– Used for researcher evaluations only
– Simplest case
  • Two independent raters
  • Ratings for every patient
– Analysis
  • Intraclass correlation
  • Cohen's Kappa
– Extensions
  • Rate random subsets
  • More than two raters

Take a second break

– What have you learned so far.
  • Test-retest measures of reliability
  • Inter-rater reliability
– What is coming next
  • Measures of internal consistency

Parallel forms

– "No man ever steps in the same river twice, for it's not the same river and he's not the same man."
  • Heraclitus
– Used when you can't run the same measurement twice.
– How to develop parallel forms
  • Change the question order
  • Minor changes to the wording
– Difficult to develop two parallel forms of the same measurement.

Split half reliability

– Only used for composite measurements
– Split into halves, correlated
  • Odd-even split
  • Random split
– Brown-Spearman adjustement

Kuder-Richardson 20

– Only for composite measures with binary items
– Book's formula is confusing
  • S^2 and σ^2 used interchangably
  • Σpq is a theoretical minimum variation
  • S^2 is observed variation
  • $S^2 = \Sigma pq$ implies randomness
  • $S^2 > \Sigma pq$ implies internal consistency

Cronbach's alpha

- – Used for composite measurements with continuous items
- – Book's formula is confusing
  - ΣS^2 should be ΣS_i^2
  - ΣS_i^2 is a theoretical minimum variation
  - S^2 is observed variation
  - $S^2 = \Sigma S_i^2$ implies randomness
  - $S^2 > \Sigma S_i^2$ implies internal consistency
- – Cronbach's alpha is NOT a measure of unidimensionality

Practical guidance on reliability

- – Is there previous literature?
  - Report their reliability coefficients
- – Is your setting similar?
  - Different demographics?
  - Different cultural norms?
  - Different literacy?
  - Different language?
- – Compare to reliability in your sample
  - Test-retest and inter-rater reliability preferred.
  - 0.7 or higher

Time for a third break

– What have you learned so far.
  • Measures of internal consistency
  • Practical advice about reliability
– What is coming next
  • Measurement validity

Measurement Validity

– Reliability by itself is not enough.
  • Consistent measures of the "wrong thing" is bad
– Examples of the wrong thing
  • Measuring anxiety instead of stress
  • Measuring transient changes in a patient's mood rather than chronic depression
– Validity
  • "Degree to which a measure … measures that which it was intended to measure"
– Reliability is a pre-requisite for validity
– Validity is a journey and not a destination

Types of measurement validity

– Face validity
– Content validity
– Criterion validity
– Construct validity

Face validity and content validity

– Only used for composite measures
– Face validity
  • Opinions from your patients
  • Subjective and unquantifiable
– Content validity
  • Opinions from experts
  • Also subjective and unquantifiable

Response process evidence

- Observe the process
  - Watch as patients fill out the form
  - Ask questions along the way
  - Monitor response times
  - Encourage them to think aloud
- Supplement with interview
- Goal is to identify problematic elements
  - Confusion, misunderstandings, language issues

Take fourth break here

- What have you learned
  - General concept of validity
  - Face and content validity
  - Response process evidence
- What's coming up
  - Criterion validity
  - Construct validity

Criterion validity

– Comparison to external criterion
  - Represents "truth"
  - Not always available
– Predictive evidence
  - Measurement in the future
  - Be careful about dropouts
– Concurrent evidence
  - Measured at the same time

Construct validity

– Used for a psychological construct
– No direct measure of the truth exists
– Define associations consistent with your constuct
  - Does your measurement show the expected association?
  - Known as convergent evidence
– Define non-associations with your construct
  - Does your measurement also show non-association?
  - Known as discriminant or divergent evidence

Alternative framework for validity

- Content
- Response processes
- Internal structure
- Relations to other variables
- Consequences

Validity of diagnostic tests

- Sensitivity
  - A test's ability to obtain a positive result when the target condition is really present
- Specificity
  - A test's ability to obtain a negative result when the target condition is really absent

Stop here

– What have you learned so far?
  • Criterion validity
  • Construct validity
– Next week
  • Spring break
  • After that, data collection issues

–