Introduction to Psychometrics

SDS 390 Structural Equation Modeling Monday Feb 4, 2019

First, let's make some data!





- DataFest 2019 Info session on Wednesday 7-8p in SR 301
- Brittney Bailey lunchtime talk on Thursday 12:10-1:10p in McConnell B15



- Scales of measurement
- Reliability and validity
- O Download the data!

What is measurement?



One unit is "one feels"!

Types of Measurement

- 1. Behavioral Measures
 - Objective/Physiological measures (e.g., Bodily activity; nervous system)
 - Directly observing of participants' behavior (think creepy two-may mirror)

2. Self-report

• Participants provide information about themselves

Basic Levels of Measurement

• nominal: categorical variable whose levels have no ordering.

Ogender, race, etc.

• ordinal: categorical variable whose levels have an order.

Oeducation level

 interval: numerical variable where we assume the distance between points is equal. No true zero.

Oscores on a "self-esteem" scale, measured from 1 to 7

• **ratio:** numerical variable that has a true zero point.

Ostudents' times to complete cognitive task

Measure Quality



Making Useful Measurements

• Measurement **reliability**—the <u>consistency</u> of a measurement

- Reliable measures provide <u>consistent</u> measurement from occasion to occasion, or item to item
- O Luckily, we have a statistic to measure sample reliability!
- Measurement **validity**—how <u>close</u> a measurement is to what it is really trying to measure
 - O Also called construct validity
 - Assessed by using your head, peer review, expert judgement
 - O And also with statistics...

Assessing Measurement Reliability

- Test-Retest reliability (across time)
- O Inter-rater reliability (across raters)
- Internal consistency/reliability (across items)
- Split-half reliability (across-halves)

Internal Consistency/Reliability

- Relevant for psychological measures that consist of more than 1 item (e.g., average scores on scale items, or when several behavioral observations are used to obtain a single score)
- Internal consistency refers to inter-item reliability, and assesses the degree of consistency among items
- Cronbach's Alpha:
 - conceptually, it is the average consistency across all items with all other items

$$\alpha = \frac{n_i \bar{r}_{ij}}{1 + (n_i - 1)\bar{r}_{ij}}$$

where n_i is the number of items and \bar{r}_{ij} is the average inter-item correlation.

High Reliability

Participant	Q1	Q2	Q3	Q4	Q5
Rachel	1	1	1	1	1
Emma	5	7	6	6	7
James	2	3	2	1	2
Avery	4	4	4	4	4
Charles	2	1	1	1	1
George	1	2	1	2	2
Seana	7	7	7	7	7

Low Reliability

Participant	Q1	Q2	Q3	Q4	Q5
Betsy	1	7	1	6	1
Seamus	5	1	7	4	1
Eli	1	7	1	7	1
Paul	1	6	2	7	4
Obama	1	7	4	1	7
Mitt	1	7	2	5	1
Hilary	1	7	2	3	6

Validity of a Measurement

- A good measure must not only be reliable, but also valid
- Construct validity the extent to which a measure actually measures what it is intended to measure
- A measure cannot be valid unless it is reliable, but a reliable measure may not be valid

Reliability vs. Validity







Reliable, but not valid

Neither reliable nor valid

Reliable and valid

No validity without reliability

Measurement (Construct) Validity

• Types of measurement validity:

- Face Validity (cannot be statistically tested)
- Convergent and Discriminant Validity
- O Predictive Validity
- Like reliability, validity is not absolute

What is this scale trying to measure?

- 1. On the whole, I am satisfied with myself.
- At times, I think I am no good at all.
- I feel that I have a number of good qualities.
- I am able to do things as well as most other people.
- 5.* I feel I do not have much to be proud of.
- 6.* I certainly feel useless at times.
- I feel that I'm a person of worth, at least on an equal plane with others.
- 8.* I wish I could have more respect for myself.
- 9.* All in all, I am inclined to feel that I am a failure.
- 10. I take a positive attitude toward myself.

Rosenberg Self-Esteem Scale (1965)

- On the whole, I am satisfied with myself.
- At times, I think I am no good at all.
- I feel that I have a number of good qualities.
- I am able to do things as well as most other people.
- 5.* I feel I do not have much to be proud of.
- 6.* I certainly feel useless at times.
- I feel that I'm a person of worth, at least on an equal plane with others.
- 8.* I wish I could have more respect for myself.
- 9.* All in all, I am inclined to feel that I am a failure.
- 10. I take a positive attitude toward myself.

What is this scale trying to measure?

- 1. I often find myself performing tasks that I had intended to do days before.
- 2.* I do not do assignments until just before they are to be handed in.
- 3.* When I am finished with a library book, I return it right away regardless of the date it is due.
- When it is time to get up in the morning, I most often get right out of bed.
- A letter may sit for days after I write it before mailing it.
- 6. I generally return phone calls promptly.
- Even with jobs that require little else except sitting down and doing them, I find they seldom get done for days.
- I usually make decisions as soon as possible.
- I generally delay before starting on work I have to do.
- 10.* I usually have to rush to complete a task on time.

Procrastination Scale (Lay, 1986)

- 1. I often find myself performing tasks that I had intended to do days before.
- I do not do assignments until just before they are to be handed in.
- 3.* When I am finished with a library book, I return it right away regardless of the date it is due.
- When it is time to get up in the morning, I most often get right out of bed.
- A letter may sit for days after I write it before mailing it.
- I generally return phone calls promptly.
- Even with jobs that require little else except sitting down and doing them, I find they seldom get done for days.
- I usually make decisions as soon as possible.
- I generally delay before starting on work I have to do.
- 10.* I usually have to rush to complete a task on time.

What is this scale trying to measure?

- No matter how accomplished his is, a man is not truly complete as a person unless he has the love of a woman.
- 2. In a disaster, women ought not necessarily to be rescued before men.
- People are often truly happy in life without being romantically involved with a member of the other sex.
- 4. Many women have a quality of purity that few men possess.
- 5. Women should be cherished and protected by men.
- 6. Every man ought to have a woman whom he adores.
- 7. A good woman should be set on a pedestal by her man.
- 8. Women, compared to men, tend to have a superior moral sensibility.
- Men should be willing to sacrifice their own well being in order to provide financially for the women in their lives.
- 10. Women, as compared to men, tend to have a more refined sense of culture and good taste.

Benevolent Sexism (Glick & Fiske, 1996)

- No matter how accomplished his is, a man is not truly complete as a person unless he has the love of a woman.
- 2. In a disaster, women ought not necessarily to be rescued before men.
- People are often truly happy in life without being romantically involved with a member of the other sex.
- 4. Many women have a quality of purity that few men possess.
- 5. Women should be cherished and protected by men.
- 6. Every man ought to have a woman whom he adores.
- 7. A good woman should be set on a pedestal by her man.
- 8. Women, compared to men, tend to have a superior moral sensibility.
- Men should be willing to sacrifice their own well being in order to provide financially for the women in their lives.
- 10. Women, as compared to men, tend to have a more refined sense of culture and good taste.



- Face validity refers to the extent to which a measure '<u>appears</u>' to measure what it is supposed to measure
- Not statistical—involves the judgment of the researcher (and the participants)

Convergent and Discriminant Validity

- To have construct validity, a measure should both:
- <u>Correlate</u> with other measures that it should be related to (convergent validity)
 - Body dissatisfaction should be positively related to body shame
 - Body dissatisfaction should be negatively related to body esteem
- And, <u>not correlate</u> with measures that it should not correlate with (discriminant validity)
 - Body dissatisfaction should not be related (or should have a weak relationship) to job dissatisfaction

Predictive Validity

 Refers to the extent to which a measure correlates with or <u>predicts</u> outcomes that it should predict

- A valid measure of marital conflict should correlate with behavioral observations (e.g., number of fights)
- The Scholastic Aptitude Test (SAT) is valid to the extent that it distinguishes between students that do well in college versus those that do not