Response Formats & Scales

PSY 600K
Dr. Alyssa Gibbons
February 1, 2018

Agenda

0 Items with correct answers.
0 Continuous items:
  0 Guttman & Thurstone scales.
  0 Likert scales.
  0 Choices involved in Likert scales.
0 A few more variations.
0 Good process for item development.

Response Formats & Item Scores

0 Not *necessarily* linked!
  0 You can dichotomize a continuous-looking item or even (sometimes) make what looks like a dichotomous item into a continuous score.
  0 Ultimately, it's the scores that matter.
0 But, depending on your construct, certain kinds of response formats may be more or less likely to get you the type of scores you want.

Items with Correct Answers

0 Completion
  0 Need to specify set of possible correct answers.
0 Multiple choice
  0 Correct answer(s) = *keyed* answers
  0 Incorrect choices = distractors or foils
  0 Need to worry about guessing... more on this later.
0 Two-choice
0 Count items – did the respondent correctly identify all the "right" answers?
0 These types of items can be used for non-cognitive measures too!
  0 "Correct" answer becomes "keyed" answer – indicating the construct we are looking for.
  0 E.g., depression

Continuous Response Scales

0 We want scales that are:
  0 Unidimensional
  0 At least interval-level data
  0 As precise as possible

0 So how can we achieve these goals?

Guttman Scales

0 Idea: in a really unidimensional scale, we expect that responses should be consistent.
  0 If you agree to a "difficult" or more extreme item, you should also agree to an "easy" or less extreme item.
0 Guttman scales are ordered logically:

Not a Guttman scale:	Guttman scale:
"I enjoy parties." "Having lots of friends is important to me."	"My friends are important to me." "I am comfortable in most social settings." "I am only happy when I am around other people."

ERCHANT

Thurstone Scales

- If you really, really want to be sure you have interval data…
- Thurstone scales are made up of items carefully selected to represent all levels of the construct, equally spaced apart.
- Start with lots of potential items, representing a wide range of the construct.
- Ask judges to rate how extreme each item is.
  - Judges are taught to use the rating scale as if the intervals are equal.
- Choose final items:
  - That judges agree on!
  - So that all levels of the construct are represented.

Utility of Guttman & Thurstone Scales

- As it turns out… Guttman & Thurstone scales may not be worth the trouble.
- Likert did not invent the basic ordered-category rating scale…
  - E.g., 1 = strongly disagree – 5 = strongly agree
- … but he did show that scoring ordered-category responses as simple integers performed just as well as more complex strategies.
  - Likert's actual scales were the more complex ones… it's a little ironic that we now call the simple ones "Likert scales."
- Moral: "Likert-type" scales generally work just as well as the Guttman and Thurstone versions and are much easier to write.

Likert Scales

0 A Likert-type scale is any scale where the response choices are *ordered categories:*
  0 Never, occasionally, often, regularly, always
  0 Strongly disagree, disagree, agree, strongly agree
  0 Etc.
0 *And* they are scored by assigning integers to each response.
  0 1 = never, 2 = occasionally, etc.
0 Whether the test-taker sees the integers or not doesn't matter.
0 Looks easy, right?

Choices in Likert Scales

0 What should the response options be?
  0 Frequency? (always, sometimes…)
  0 Degree? (not at all, a little, some, a lot…)
  0 Agreement? (strongly agree, agree…)
  0 Choose response options that make sense for your items, respondents, and purpose.
  0 Keep them consistent as best you can.
0 Remember that we are trying to approximate interval data, so choose response options that are cognitively "evenly spaced."
  0 Pohl (1981) identified response choices that respondents perceived as being equally spaced for different types of scales.

Choices in Likert Scales

0 How many response options?
- 0 Tradeoff: variance vs. accuracy
- 0 Good rule of thumb: 7 +/- 2
  - 0 How many categories can a respondent really meaningfully differentiate?
- 0 Think ahead to potential biases:
  - 0 E.g., leniency in performance appraisal.

0 Using familiar numbers is good.
- 0 Positive integers, arranged with lowest #s on the left and highest on the right.
- 0 You can rescore them however you want later.

Choices in Likert Scales

0 Should there be a middle response?
- 0 Even #s of response options force the respondent to state at least a slight preference.
  - 0 But this may frustrate respondents who either prefer not to say or who really don't have an opinion.
- 0 Odd #s allow an in-between or neutral option.
  - 0 Respondents tend to like this…
  - 0 … but it may not be clear what it means. Is the respondent really in between the extremes? Or do they have no opinion? Or do they just prefer not to say?
  - 0 Empirical research suggests that "middle" responses don't always fall on the middle of the latent trait continuum.

0 Choice depends on your construct – is it possible to be neutral? Is it valuable to force people to state a preference?

Item and Response Content

0 Your choice of scale should be influenced by how you write your items... and vice versa.

0 If using a "strongly agree" rating scale, items should be strongly worded.

  0 What does it mean to strongly agree with a weak statement?

  0 *Degree* is captured in the response items.

0 Less extreme items may work better with a simpler rating scale (e.g., yes / no).

  0 Degree is captured by the "difficulty" or extremeness of the item.

Variations

0 *Semantic differential*: responses are made on a continuum between two opposite words:

  0 *Happy - - - - - - - - - - - - - - - - - Unhappy*

  0 *Successful - - - - - - - - - - - - - - - - - Unsuccessful*

0 Usually scored as Likert scales (integers).

0 Intuitive, makes sense for some item stems... but finding true unambiguous opposites can be difficult.

0 *Graphic rating scale* or *visual analog scale*: mark the point on the continuum that represents your response.

  0 Less reading... but much harder to score and interpret.

Forced Choice

- A particular type of two-choice item – respondent must pick one and only one of two (theoretically) equally desirable (or undesirable) alternatives.
  - I prefer....   Chocolate cake   OR   Crème brulee
  - I am almost always on time for work OR I am willing to put in extra effort to make a customer happy.
- Why might you use these types of items?
- How would you score them?
- What problems can you foresee?

Questions?

Lab Friday: Item Writing
Don't forget Project Plan due tonight by midnight!

For next time:
A Few Critical Statistical Concepts
Read: R & M 2.1 – 2.7
3rd Reading Response