

[mux-lab] Statistical methods for Likert-type data

1 message

Brian Gleeson <gleesonb@cs.ubc.ca>
To: mux-lab@cs.ubc.ca, caris-lab@interchange.ubc.ca

Fri, Oct 25, 2013 at 11:19 AM

Hi all,

I've been trying to figure out how to analyze some Likert-type data. I've asked around the lab, without getting any definitive answers, so I did a bit of my own research and found a bottomless rabbit hole full of contentious opinions from belligerent social scientists. Since I'm sure I won't be the last person in the lab to face these issues, I thought I'd share what I found.

Disclaimer: I'm no expert. You may want to do your own reading to before using these tests. If any of you already knows more about this than I do, feel free to enlighten me.

The question is, "Can I use parametric statistics (ANOVA, t-test, etc.) to analyze Likert data?" Depending on who you ask, the you will find that parametric stats are either perfectly alright[1][2] or absolutely forbidden[3]. I'm choosing to join the parametric camp, but parametric stats can only be used if you design you study correctly.

Using Parametric Stats on Likert Scales.

You probably shouldn't use parametric stats on **Likert formatted** items. This is a Likert formatted item:

UBC has a smoothly functioning bureaucracy:

strongly disagree – disagree – neither agree nor disagree – agree - strongly agree

You may, however, use parametric stats on **Likert Scales**. A Likert Scale is a group of 4 or more questions that are all addressing the same underlying belief. E.g., rate on a 5-point, strongly disagree -strongly agree scale:

UBC's bureaucracy is fast and efficient

The UBC bureaucracy never impedes my work

UBC's bureaucracy always resolves my requests quickly and satisfactorily

UBC's bureaucratic processes are fine as they are

Taken together, these four questions could be a Likert scale. Sum or average the responses to all four questions, within subjects (each subject produces a single number, a UBC bureaucracy score). These scores can be thrown into your standard parametric test. Reviewers might grouse about these methods, but you have citations to back you up[1][2]. This is the method used by Likert himself.

Apparently, there is a real art to designing the scale questions. More on what makes a good Likert scale:

<http://www.john-uebersax.com/stat/likert.htm>

<http://www.joe.org/joe/2012april/tt2.php>

Also, see [2].

When you Can't Use Parametric Stats:

If you have single, Likert-formatted items, you're probably stuck with non-parametric stats.

Although, I spoke with a social science faculty member who told me that she uses parametric stats on single, Likert-formatted questions, particularly when using a 7 or 9 point response scale. I'm a little nervous about this

approach, so I'm still recommending non-parametric stats for single Likert-formatted questions.

To summarize results in 1 Number:

Use the Mode.

To compare results of 2, independent questions (e.g., condition A vs. condition B)

Use a Mann–Whitney U test.

-or-

Lump your results into categories (e.g. 1-2 = disapprove, 3 = neutral, 4-5 = approve), and use Pearson's chi-squared test.

You could also use chi-squared test to compare a single condition against a null distribution (e.g., null: 40% disapprove, 20% neutral, 40% approve).

To compare results of more than 2 questions (e.g., conditions A, B, C):

Use a Kruskal–Wallis test. Not as robust as Mann-Whitney, because it assumes each group has the same distribution.

When your questions are not independent, or you have multiple independent variables (e.g., within subjects design)

Use an Ordered Logit test (proportional odds ordinal logistic regression) model, with Condition and Subject as factors (plus any other factors you may have).

This doesn't always work. Sometimes you just can't fit a good model to your data (look at the Model Likelihood Ratio result). I'm not sure what you do in that case. Maybe you just cry ☹

-or-

Cheat: Just assume that your conditions are independent. Depending the design of your experiment, this may not be that bad of an assumption. The experts remind us that "The F-Test is not made of glass"[2] and that you don't always have to worry about violating assumptions, so long as you are actually thinking about your results and just blindly copying down p values.

Some notes on how to interpret the results of logit analysis:<http://www.princeton.edu/~otorres/Logit.pdf>

I hope this saves some people some suffering,

Brian

[1] G. Norman, "Likert scales, levels of measurement and the 'laws' of statistics.," *Advances in health sciences education : theory and practice*, vol. 15, no. 5, pp. 625–32, Dec. 2010.

[2] J. Carifio and R. J. Perla, "Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes," *Journal of Social Sciences*, vol. 3, no. 3, pp. 106–116, Mar. 2007.

[3] S. Jamieson, "Likert scales: how to (ab)use them.," *Medical education*, vol. 38, no. 12, pp. 1217–8, Dec. 2004.