Fun with Survey Data: *Likert Scales and Latent Variables*

Paul Harmon

Montana State University

April 10, 2018

Motivation

In institutional research, survey data is a big deal. How many surveys are going on right now at Montana State?

- Diversity Survey
- Student Survey
- Career Destinations Survey
- NESSE
- BESSE
- And these are just the big ones that come through the Office of Planning and Analysis!

Survey Implementation

Did you know that MSU students have access to Qualtrics? If you (or one of your students) has interest in designing and implementing a survey, you can (and probably should) use Qualtrics to set it up.

A handy link:
http://guides.lib.montana.edu/c.php?g=329596p=2213103

More motivation

Study: People are given a survey about their opinions on environment, etc. prior to presentation. Then, they take the same survey again after the presentation (or this could be a series of presentations).
How do we analyze pre/post-treatment differences in scores on this survey?

Likert Scales

Likert Scales refer specifically to scales on which the response is measured on a 5 or 7 point scale. Respondents measure the agree to which they agree or disagree with a statement, with the mean scale being the middle value and the values to the left and right being 'symetric'.
Example: The lack of concrete evidence in favor of bigfoot's existence proves that he does not actually exist.

Response Choice:
Strongly Disagree – Disagree – Neutral – Agree – Strongly Agree

Likert Simulations

It's easy to simulate Likert-Scale data! Let's try simulating a ten question survey with five point scales. Here we simulate two covariates, PRE/POST and GENDER. We simulate no differences between these groups.

Can we use Linear Regression?

Responses take on 5 or 7 values in most cases. Thus, we probably can't get away with using linear regression. (However, it is occasionally used, despite being wrong). They argue that these values can be treated as **interval data**, where the order and differences between values are known.
Better method: Ordinal Multinomial Response!

$$logit(P(y_i \leq j)) = \alpha_i + x_i\beta, j = 1, ..., c-1$$

Multinomial models are GLMS that allow for modeling a response that takes multiple categories. I The response is assumed to be ordinal, not nominal, since likert scales are ordered. We model the odds of being in one category as opposed to a baseline category, and in this case we would have to model single questions.

(Ab)Using Linear Regression vs. Ordinal Responses

Regression Results

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	15.8462	0.3182	49.81	0.0000
PrePost1	-0.0714	0.3814	-0.19	0.8516
GENDER2	-0.8550	0.3837	-2.23	0.0270

Ordinal Response Results

regression results.png

Multinomial Regression Results:

```
Level  2 vs. Level  1
               rrr    se  zStat  pVal ci95.lo ci95.hi
(Intercept) 0.442 0.637 -1.280 0.200   0.127   1.542
PrePost1    2.359 0.696  1.233 0.218   0.603   9.237
GENDER2     0.349 0.671 -1.568 0.117   0.094   1.301

Level  3 vs. Level  1
               rrr    se  zStat  pVal ci95.lo ci95.hi
(Intercept) 6.665 0.374  5.071 0.000   3.202  13.875
PrePost1    0.781 0.410 -0.604 0.546   0.350   1.743
GENDER2     0.410 0.415 -2.150 0.032   0.182   0.924

Level  4 vs. Level  1
               rrr    se  zStat  pVal ci95.lo ci95.hi
(Intercept) 0.533 0.634 -0.993 0.321   0.154   1.846
PrePost1    1.184 0.770  0.219 0.827   0.262   5.350
GENDER2     0.180 0.883 -1.941 0.052   0.032   1.017
>
```

Some Conclusions

Latent Variables:

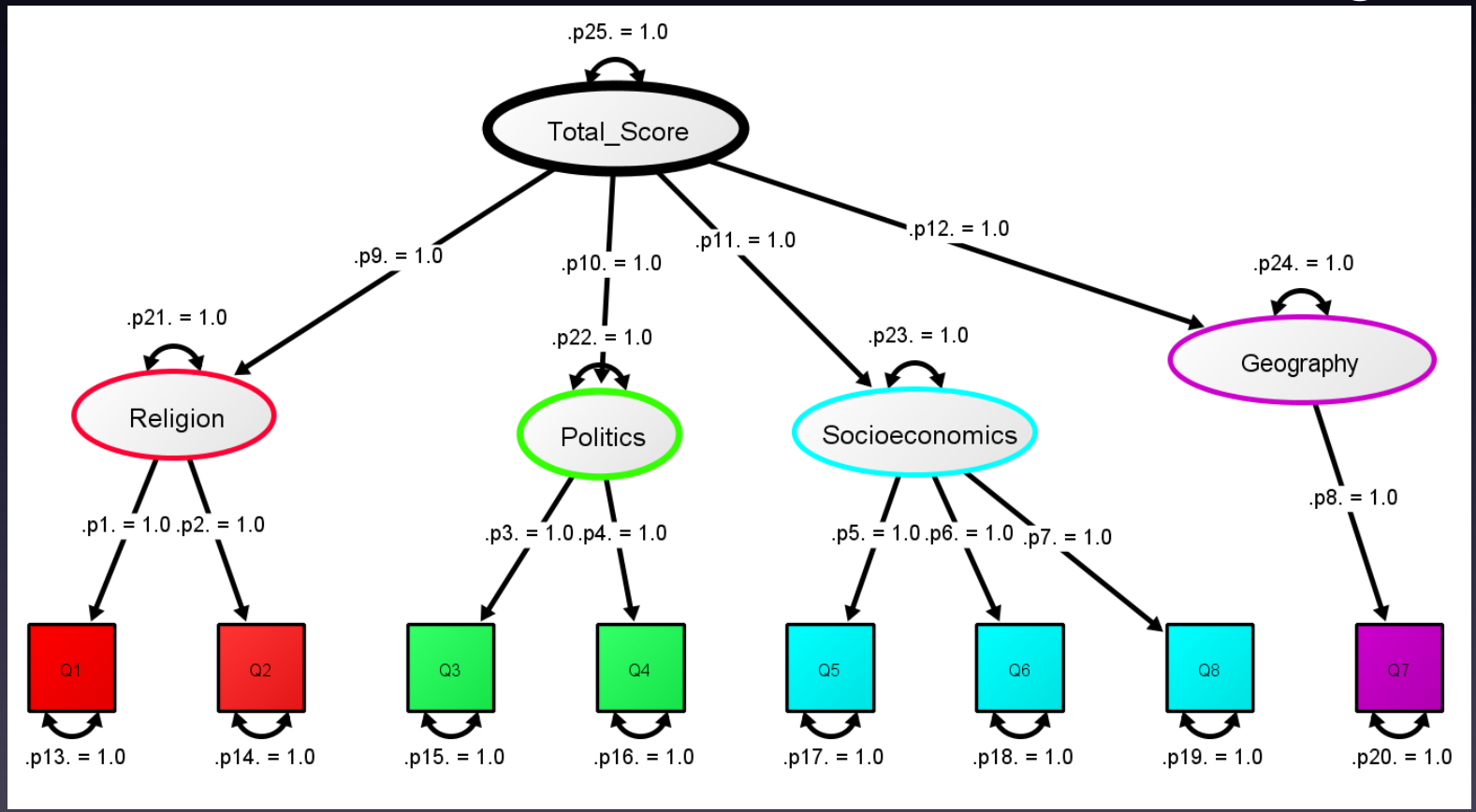
Some surveys are designed to analyze 'latent variables' that cannot directly be measured. A good example of this is the **Chapman University Survey of American Fears**.
Latent variables analyzed:

- Religious beliefs of respondent
- Political beliefs of respondent
- Geographic characteristics of respondent
- Socioeconomic characteristics of respondent

In this survey, you cannot directly measure the factors that impact the things that people fear the most (nor can fear itself be measured directly). But you can ask questions designed to measure pre-specified latent traits and use those in your analysis.

Latent Variable Models: Path Diagrams

We can visualize latent variable models with **Path Diagrams**:



Onyx Package: A Fun Way to Make Path Diagrams

Path diagrams are hard to make look good in R. The lavaan package (used for SEM models) has some code to generate them, but they are ugly!

A Better Way to Do Path Diagrams:

The onyx package is installed via Github and uses Java on your machine to allow you to create path diagrams by hand.

- Onyx works with lavaan
- Onyx can even generate code for a model!

Factor Analysis: Exploratory Analysis

We use exploratory factor analysis if we have measured a bunch of variables (or questions) and we want to see which latent traits they load onto.
We use confirmatory factor analysis if we measure variables expecting them to load onto certain latent factors, and then see if they do what we expect.
In a Structural Equation Model, we are interested in using observed manifest variables to load onto latent traits and then input them into some kind of regression model.

Factor Analysis: Star Wars Survey

The Star Wars survey! Data available on FiveThirtyEight's GitHub page.
See HTML:

Future Work: Structural Equation Models

I'm wondering if you could use this survey's response to predict a binary response. Say you were interested in whether or not fans of the Star Wars prequels are more likely to respond "Greedo shot first" as opposed to "Han shot first".
You could use this survey response to fit latent factors and then regress it on the binary response, Y.
However, you'd have to use Structural Equation Models (see my talk on Friday!) with a GLM binary response. I'm not sure how those work just yet...

Questions

Thanks for coming!