

Assignment 2: Cumulative Scaling

The General Social Survey tolerance items, originally developed by Stouffer (1955), are usually summed to create a summated rating scale of tolerance attitudes in the U.S. The assumption, of course, is that larger values on the resultant scale reflect greater levels of intolerance toward “fringe” groups. However, no study has ever considered whether this assumption is met by the data.

The Stata dataset “Tolerance, 2016 GSS.dta” contains the responses to 18 tolerance questions asked on the 2016 GSS. The stem question for each of the groups reads: “There are always some people whose ideas are considered bad or dangerous by other people. For instance, somebody who _____.” The groups asked about include: atheists, communists, racists, militarists, homosexuals, and muslims. For each group, respondents are asked whether they would (not) grant one of three civil liberties:

- “If such a person wanted to make a speech in your (city/town/community), should he be allowed to speak, or not?” (variables labeled with **s**)
- “Should such a person be allowed to teach in a college or university, or not?” (variables labeled with **t**)
- “If some people in your community suggested that a book he wrote should be taken out of your public library, would you favor removing this book, or not?” (variables labeled with **b**)

For each of the 18 items (6 groups \times 3 liberties), denying the civil liberty is coded 1 and granting it is coded 0. As such, we might label the scale “intolerance,” rather than “tolerance.”

For this assignment, you must assess whether the tolerance battery fits the cumulative scaling model developed by Mokken (1971). Do all items fit the cumulative scaling model? If not, which should be removed? Are the assumptions of the monotone homogeneity model met? What about the double monotonicity model? What substantive inferences can we make about the nature of tolerance in America in 2016 using the tolerance scale (hint: consider distribution, item ordering)?

Next, estimate Rasch and two-parameter IRT models. Are there any differences in the inferences you can make with the Rasch and 2PL models than with the nonparametric cumulative scaling model? Which model is “best?”