Title: Choosing a statistical model amidst ceiling and floor effects

Rao, V.N.V., Running, K.E., Codding, R.S.

Poster strand: Recent Developments in Assessment and Research

Practice Model Domains:

- Domain 1 (Data based decision making and accountability)
- Domain 9 (Research and Program Evaluation)

Skill Level: Introductory

Primary Index Term: Research Methods

Learner Objectives: This session will help participants

- Identify ceiling and floor effects in their data
- Determine which statistical models to use when estimating group differences
- Learn how to model group differences in the presence of ceiling and floor effects

Abstract: School psychology research often entails group comparisons, commonly modeled by Analysis of Variance (ANOVA) based models. However, these models underestimate true differences between groups when ceiling or floor effects (CFE) are present. To guide researchers, we conducted a simulation study to analyze bias in ANOVA and censored regression models when estimating mean differences between groups. Researchers will learn when and how to use each method through an illustrated example based on a classwide academic intervention.

Choosing a Statistical Model Amidst Ceiling and Floor Effects

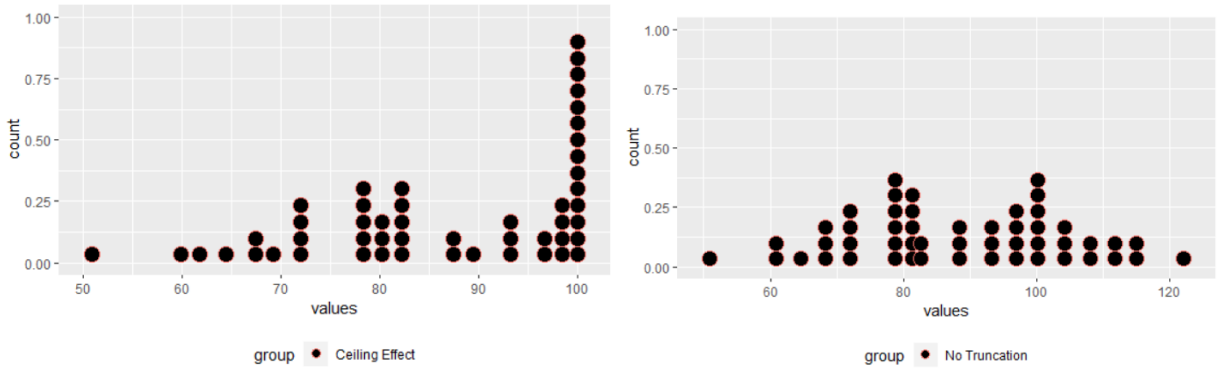
The efficacies of educational interventions are frequently evaluated using ANOVA-based models to compare instructional groups (e.g., Braithwaite & Siegler, 2020; Fuchs et al., 2016; Rittle-Johnson & Koedinger, 2009). However, when school psychology assessment data have ceiling or floor effects (CFE), ANOVA-based models produce biased estimates of instructional group differences. A CFE is when test scores are truncated at either a minimum or a maximum value. Treating scores for a response variable at the ceiling or floor as true values, instead of what it might have been with an unlimited range, results in biased estimates of group means and an underestimation of group differences when comparing scores across groups. Therefore, it is important to identify a CFE and appropriately account for it in statistical models of group comparisons when conducting school psychology research. In this poster, we explain and demonstrate these methods in a non-technical manner using data collected from a classwide academic intervention.

Noticing and Acknowledging a CFE

Analyzing a CFE begins with noticing and acknowledging it (Wild & Pfannkuch, 1999). The dot plot or histogram is a useful visualization to do so. A dot plot for a normally distributed variable with a CFE will look like part of a full normal curve but with extra dots at the ceiling (or floor), as shown in Figure 1. It is important to note the percent of all observations with scores at the ceiling (or floor). Figure 1 shows a dataset with 28% of observations with values at the ceiling of 100, or 14 of 50 observations.

After identifying a CFE, the methods for handling it will vary based on whether it affects a response variable, an explanatory variable, or a covariate. Here, we only focus on CFEs in a response variable and specifically focus on a ceiling effect.

Figure 1
Dot plots with and without a ceiling effect



Censored Regression Models

Censored regression models explicitly aim to account for data truncated at either a floor or a ceiling. The most common censored regression model is known as the Tobit model (Tobin, 1958). Tobit regression assumes an individual's true score is not limited to a bounded range. It then accounts for the possibility of a lower limit to the reported scores, such as 0, and an upper limit to the reported scores, beyond which values are reported as the limits themselves. To produce estimates for regression coefficients and estimate true group means, the Tobit model utilizes cumulative frequency distributions within the range of reported scores.

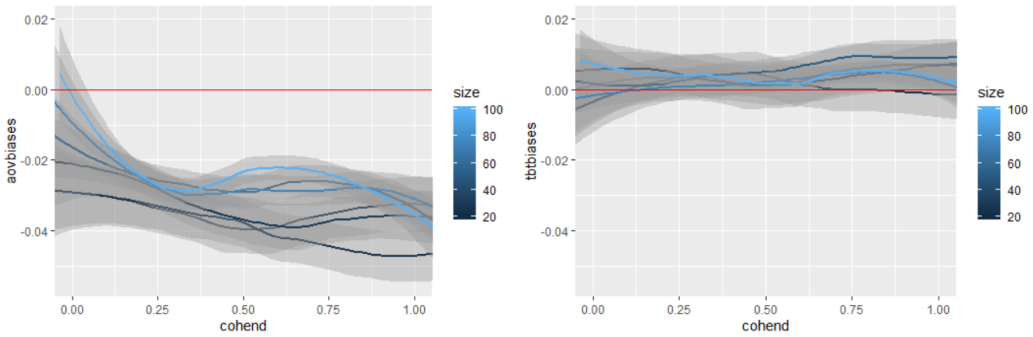
Bias Estimates

To determine when an ANOVA or Tobit model may be appropriate, we conducted a simulation study to estimate the bias of each model's estimate of differences in group means. We

varied three parameters in the simulation study: (a) the size of each group varied between 20 and 100, (b) the effect size varied between 0.2 and 0.8, and (c) the location of the ceiling varied from -1 to 3 standard deviations above the group mean.

Generally speaking, ANOVA will produce underestimates of the true difference between groups. When no more than 30% of observations in either group are at a ceiling, *and* the difference in the percentage of observations at the ceiling between the two groups is no more than 20%, ANOVA will underestimate the effect size by no more than 0.05. Therefore, in such cases, it may be acceptable to use an ANOVA model, especially when attempting to identify moderate and large effects. Notably, the Tobit model produces less bias in those same cases. Therefore, when possible, researchers should utilize Tobit regression models instead of ANOVA models.

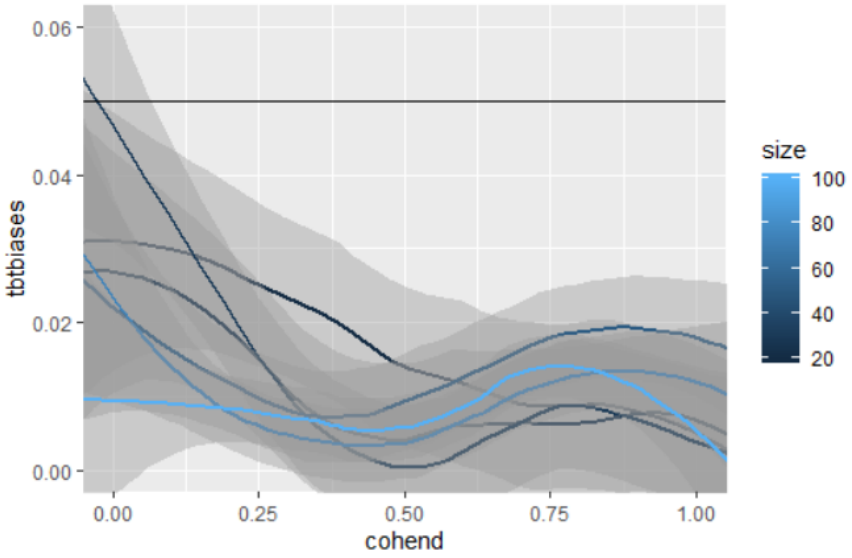
Figure 2
ANOVA and Tobit model bias with less than 30% of observations at the ceiling and difference of the ceiling frequency less than 20% (in effect size units)



Tobit regression models can overestimate the true difference between groups, especially when the effect sizes are small. However, so long as no more than 70% of observations in either group are at a ceiling, Tobit regression will overestimate the effect size by no more than 0.05.

This overestimation is much smaller when group sizes are greater than 30, the effect size is greater than 0.2, or the percentage of observations at ceiling is less than 50%.

Figure 3
Tobit model bias with less than 70% of observations at the ceiling (in effect size units)



Math Intervention Case Example

Using real-life data with a ceiling effect from a classwide math intervention, we will demonstrate the process of identifying and modeling a CFE, and discuss the different interpretations that arise from using an ANCOVA model compared to a Tobit model. Implications for intervention evaluation will be discussed.

Summary

While researchers would ideally utilize assessments with a difficulty targeted to their study population, it is not always possible to guarantee that students' scores will not have CFEs. A CFE, even with as much as two-thirds of a group with maximum (or minimum) scores, does not doom the estimation of group differences. Tobit regression, a censored regression model, can

be employed to compare group differences in such cases. Furthermore, ANOVA-based models may still be used when there is only a slight CFE. Using an example based on a classwide academic intervention, this poster will guide researchers through the process of noticing and acknowledging a CFE, determining which model to use to estimate group differences, and fitting and interpreting model results, using real data relevant to school psychologists.

References

Barros, M., Galea, M., Leiva, V., & Santos-Neto, M. (2018). Generalized Tobit models: diagnostics and application in econometrics. *Journal of Applied Statistics, 45*(1), 145-167.

Braithwaite, D. W., & Siegler, R. S. (2020). Putting fractions together. *Journal of Educational Psychology*.

Fuchs, L. S., Malone, A. S., Schumacher, R. F., Namkung, J., Hamlett, C. L., Jordan, N. C., ... & Changas, P. (2016). Supported self-explaining during fraction intervention. *Journal of Educational Psychology*, *108*(4), 493.

McDonald, J. F., & Moffitt, R. A. (1980). The uses of Tobit analysis. *The review of economics and statistics, 62*(2), 318-321.

Rittle-Johnson, B., & Koedinger, K. (2009). Iterating between lessons on concepts and procedures can improve mathematics knowledge. British Journal of Educational Psychology, 79(3), 483-500.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica, 26(1)*, 24-36.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International statistical review, 67*(3), 223-248.