I

Imputation-Based Modeling for Outcomes With Ceiling and Floor Effect

Selen Yilmaz Isikhan
https://orcid.org/0000-0002-3725-2987
Hacettepe University, Turkey

Gozde Erturk Zararsiz
Erciyes University, Turkey

INTRODUCTION

Data with ceiling or floor effects frequently occur in psychology and educational researches, and even in the evaluation of treatment effect, well-being, and responses from scales such as health-related quality of life in clinical studies. According to the definitions in some studies (Reissmann et al., 2022), ceiling or floor effects occur when the questionnare is relatively easy or difficult, so that a significant proportion of individuals will achieve maximum or minimum scores (Liu & Wang, 2021). The ceiling effect occurs at the top of a scale range, while the floor effect occurs at the lower end of the scale range. When floor or ceiling effects occur, actual differences between individuals with the lowest (or largest) possible score are indistinguishable.

Limited information is available about the scores of those who actually scored at the bottom or the ceiling. Scores on the floor and ceiling are said to be censored because only the lower and upper limit of the score is known (McBee, 2010). The highest (or lowest) possible value of a measurement tool is called the censoring point (Gustavsen et al., 2022). For instance, scenarios such as the following can be considered to define floor and ceiling effective scores in clinical studies. When patients are tested for pain degrees after surgery, patients report pain degrees between 0-10. In this case, depending on the time passed after surgery, many patients may have a score of "0" or many patients may report the pain they feel distinctly as "10." Another example is that, in the audiology department, a 5-point Likert scale questionnaire can be applied to patients who have a unilateral implant or bilateral implants to measure the success in hearing after a while. In this case, the actual hearing success (development) of those who answer the question as "5" may differ from each other. Therefore, the hearing achievement test has a very low ceiling for distinguishing the true hearing level of patients who score their hearing achievement as "5."

The main purpose of this chapter is to compare the performances of a few commonly used estimation methods in the solution of ceiling and floor effect observations (i.e., excluding ceiling and floor effective observations, Tobit regression, and zero-inflated regression) and the regression-based imputation method that considers these observations as missing data in the missing not at random (MNAR) structure (Loos et al., 2022).

DOI: 10.4018/978-1-7998-9220-5.ch060

BACKGROUND

Ignoring the observations with ceiling or floor effect and analyzing this type of data with classical statistical analysis methods causes problems such as biased, artificial, and nonlinear solutions or the regression coefficients being insignificant (Yenilmez et al., 2018).

As a first method, (Liu & Wang, 2021) applied t-test and ANOVA by ignoring the ceiling observations (i.e., assuming the real value). As a second method, they suggested removing ceiling effective observations. As a third method, they showed censored regression, that is, Tobit regression. As a final method, they proposed an adjusted mean and variance estimation based on the truncated distribution. However, their second approach may not yield good results when the number of observations is insufficient. For this reason, estimating ceiling effective observations as if they were missing data by regression-based imputing according to the MNAR approach, instead of removing problematic observations, formed the basis of this study.

Very few studies have examined and compared the regression-based methods used to estimate the outcome variable with ceiling and floor effect. In some identified studies, methods such as ordinary least square (OLS) regression, ridge regression, Tobit regression, and beta regression were examined (Ogundimu & Collins, 2018; Sayers et al., 2020; Taku et al., 2018).

METHODS

In the following subsections, the authors explain Tobit regression, zero-inflated Poisson regression, which are the main estimation methods the authors considered in the study, and the regression-based imputation method proposed for ceiling and floor effective observations, respectively.

Tobit Regression

The Tobit model, also known as the censored regression model, was developed to predict linear relationships between variables when there is left or right censoring in the dependent variable. The Tobit model is defined as a latent variable model. The Tobit regression model basically improved as a left-censored model as follows:

$$y_i^* = x_i^{'}\beta + u_i \text{ , } \mathrm{u_i} \sim \mathrm{N}(0, \sigma^2) \tag{1}$$

a defines the censoring point:

$$y_i = y_i^* \; y_i^* > a \tag{2}$$

$$y_i = a \; y_i^* \leq a \tag{3}$$

Here, y_i is an observed value that is equal to y_i^* when $y_i^* > a$, and $y_i = a$ when $y_i^* \leq a$. On the other hand, the right-censored Tobit model can be defined as follows:

$$y_i^* = x_i^{'}\beta + u_i \text{ , } u_i \sim N(0, \sigma^2) \tag{4}$$