Research article

An application of ordinal regression to extract social dysfunction levels through behavioral problems

Alka Sabharwal[1], Babita Goyal[2,*] and Lalit Mohan Joshi[3]

[1] Department of Statistics, Kirori Mal College, University of Delhi, Delhi–91–110007, India
[2] Department of Statistics, Ramjas College, University of Delhi, Delhi–91–110007, India
[3] Department of Statistics, University of Delhi, Delhi–91–110007, India

* **Correspondence:** Email: goyalbabita@gmail.com.

Abstract: Psychological problems are complex in nature and accurate identification of these problems is important. For the identification of psychological problems, one of the preliminary tools is the use of interviews/questionnaires. Questionnaires are preferred over interviews if the group under study is large. A strengths and difficulties questionnaire (SDQ) is one of the most widely used and powerful questionnaires to identify behavioral problems and distresses being faced by the respondents, affecting their day-to-day lives (responsible for social dysfunction). This study was held on college/university students in India, with the objective of examining if the extent of social dysfunction as measured by an impact score can be extracted from behavioral problems which are the components of the difficulty score of SDQ. Two surveys were conducted during the COVID-19 pandemic period, between the months of May–June 2020 and October 2020–February 2021 for the study. Only those responses were considered who felt distressed ("yes" to item 26 of SDQ). The numbers of such responses were 772/1020 and 584/743, respectively, in the two surveys. Distress levels were treated as ordered variables and three categories of distress level, viz., "Normal", "Borderline", and "Abnormal" were estimated through behavioral problems using ordinal regression (OR) methods with a negative log-log link function. The fitting of OR models was tested and accepted using Cox and Snell, Nagelkerke, and McFadden test. Hyperactivity-inattention and emotional symptoms were significant contributors to estimating levels of distress among respondents in survey 1 ($p < 0.05$). In addition to these components, in survey 2, peer problems were also significant. OR models were good at estimating the extreme categories; however, the "Borderline" category was not estimated well. One of the reasons was the use of qualitative and complex data with the least wide "Borderline" category, both for the "Difficulty" and the "Impact" scores.