A Double Truncated Binomial Model to Assess Psychiatric Health through Brief Psychiatric Rating Scale: When is Intervention Useful?

Alka Sabharwal[1], Babita Goyal[2,*] and Vinit Singh[3]

[1]Department of Statistics, Kirori Mal College, University of Delhi, Delhi, India

[2]Department of Statistics, Ramjas College, University of Delhi, Delhi, India

[3]Research Scholar, Department of Statistics, University of Delhi, Delhi, India

Abstract: A double truncated binomial distribution model with 'u' classes truncated on left and 'v' classes truncated on right is introduced. Its characteristics, namely, generating functions; and the measures of skewness and kurtosis have been obtained. The unknown parameter has been estimated using the method of maximum likelihood and the method of moments. The confidence interval of the estimate has been obtained through Fisher's information matrix.

The model is applied on cross sectional data obtained through Brief Psychiatric Rating Scale (BPRS) administered on a group of school going adolescent students; and the above-mentioned characteristics have been evaluated. An expert, on the basis of the BPRS score values, suggested an intervention program. The BPRS scores of the students who could be administered the intervention program lied in a range (which was above the lowest and below the highest possible values) suggested by the expert. Whereas the complete data suggested the average number of problem areas is four (which was not in consonance with the observations given by the expert), the double truncated model suggested the number of such areas as five which was consistent with the observations made by the expert. This establishes the usefulness of double truncated models in such scenarios.

Keywords: Brief Psychiatric Rating Scale, Left-right truncated Binomial model, Psychiatric health, Maximum likelihood estimate, Information matrix.

1. INTRODUCTION

In classical probability distributions, often, the theoretical values of random variables are distributed on infinite intervals. However, the corresponding sampling values are distributed over finite intervals. Then the so-called "tails" of the distribution of the general totality are required to be taken into account in some way. One such method is the use of truncated distributions. The variety of truncation methods serves as a source of modelling the new probability distributions. The truncated distributions provide a significant statistical phenomenon applied in numerous domains, viz medicine, reliability theory, industry, queuing systems, and many others. When a range of probability values (either from left or right or both or mid values) for the variables is either ignored or unobservable, probability models are truncated.

Double truncation probability distributions are applied when variable values on both extremes are truncated. Suzuki, M *et al*. (1983) introduced a binomial error model that has a truncated beta distribution (population is assumed to be restricted) as its latent trait distribution for the analysis of the test scores in a university entrance examination. They estimated the truncation points on left and right, when the parameters were either known or estimated [1]. Bilker, W. B., & Wang, M. C. (1996) modified the Mann-Whitney statistic under truncated data by assuming that the parametric form of the truncation distribution to be known. They observed that the reversed Weibull distribution (reversed in time) was the most appropriate form of truncation distribution to study the time from HIV infection to AIDS diagnosis [2]. Efron, B. & Petrosian, V. (1998) used double truncation on astronomical data to develop nonparametric methods for estimation and testing of the model as the quasars can only be observed if their luminosity occurs within a certain finite interval, bounded on both the ends [3]. Tokmachev, M. S. (2018) discussed different methods for modelling either two-sided truncated distributions or one-sided truncated distributions [4]. Aydin, D. (2018) studied certain statistical properties of the doubly-truncated exponentiated inverse Weibull distribution. Unknown parameters were obtained using different estimation methods and the performances of obtained estimators were compared [5].

In the present study, we developed a Left-Right Truncated Binomial Model with parameters $\left(n_{ij}, p\right)$ (*LRTBD*), n_{ij} known, but variable. Left truncation is on the 'u' left most classes of the random variable and right truncation is on 'v' right most classes. A particular case when $\left(n_{ij} = n\right)$ (*n* is a known constant) has been obtained. The unknown parameter *p* has been estimated using the method of maximum likelihood estimation and the method of moments. The 95% confidence interval for the unknown parameter *p* has

*Address correspondence to this author at the Department of Statistics, Ramjas College, University of Delhi, Delhi, India;
E-mail: goyalbabita@gmail.com

been obtained. All the characteristics of the models viz. the moment generating unction, probability generating function, and the characteristic function have been obtained along with the first four moments and the coefficients of skewness and kurtosis.

These models have been applied on real cross sectional data collected through an 18- items Brief Psychiatric rating Scale (BPRS). The data was collected by an expert on 93 students, all in the age group 15-17 years, studying in an NDMC school, New Delhi, India. The models developed in this study have been compared among themselves as well as with the non-truncated binomial model for the cases when

(i) When $n_{ij} = n$ (known)

(ii) When n_{ij} are known but variable

Although truncated models (left or right) are frequently used in medical data but application of a double truncated model is rare. For the best of our knowledge, this is the first study using a double truncated model on psychiatric data. A novelty of this study is the development of a data based LRTBD model with unequal number of classes truncated on left and right. Besides introduction, this paper includes three more sections. Section 2 is about material and methods, followed by results and discussion in Section 3 and a concluding remark in Section 4.

2. MATERIAL AND METHODS

2.1. Material

The cross-sectional data used in this study was collected on 93 adolescents, between 15-17 years of age, studying in a New Delhi Municipal Corporation (NDMC) School, New Delhi, India, by administering an 18-item Brief Psychiatric Rating Scale (BPRS) by an expert.

Developed by Dr. John Overall and Dr. Donald Gorham, BPRS is a test used to assess the positive, negative and affective symptoms of individuals with psychiatric disorders. This instrument is particularly useful for assessing the efficacy of treatment in case of patients with moderate to severe psychiatric disorder(s). It consists of 18 questions which measure an array of traits viz. Somatic Concern, Anxiety, Emotional Withdrawal, Conceptual Disorganization, Guilt Feelings, Tension, Mannerisms and Posturing, Grandiosity, Depressive Mood, Hostility, Suspiciousness, Hallucinatory Behavior, Motor Retardation, Uncooperativeness, Unusual Thought Content, Blunted Affect, Excitement and Disorientation. Each item is rated on a scale of 0-7, where 0 measures 'not assessed', 1 measures 'not present' and 7 measures 'extremely severe'. This scale is administered by an expert who interviews a respondent by a battery of questions which are aimed at measuring the above-mentioned traits. The scores are given on the basis of the current response of the respondent although the incidents affecting the response might have occurred in the past [6-9].

2.2. Methods

2.2.1. Model

Let X_{ij} denote the score of the j^{th} subject in the i^{th} item of BPRS and $X_{ij} \sim Ber(p)$ where p is the probability of getting a success (the item score is greater than or equal to 4, i.e. moderate or higher). Assuming X_{ij}'s to be i.i.d. for every *j*, $\sum_{i=1}^{n_{ij}} X_{ij} \sim Bin(n_{ij}, p); n_{ij}$ is the number of items assessed by an expert for the j^{th} participant and is known.

Case 1: All the items were assessed for a participant, i.e. $n_{ij} = n = 18$, and $\sum_{i=1}^{n} X_{ij} \sim Bin(n, p)$.

Case 2: The number of items assessed for every participant are not necessarily equal.

Depending on the number of successes (i.e. the item scores moderate or higher) an intervention program for the participants was suggested by the expert. The suggested program

i. cannot be recommended if the participant is healthy, i.e. all item scores but at most one were at mild level or below

ii. is not beneficial if more than ten item scores are moderate or at higher levels.

All the participants falling in the above two categories (not recommended/not beneficial) were truncated. The participants were left truncated if healthy (not recommended) and right truncated if the program is not beneficial. The '*u*' classes from left and '*v*' classes from right were truncated.

Let $X_{i1}, X_{i2}, ..., X_{ij}, ..(i = 1,2...,18; j = 1,2,...,93)$ are independently distributed random variables following LRTBD (n_{ij}, p). Then the probability mass function of the random variable X_{ij} for the j^{th} respondent, is given by

$$P\left(X_{ij}=x_{ij}\right)=\frac{\binom{n_{ij}}{x_{ij}}p^{x_{ij}}\left(1-p\right)^{n_{ij}-x_{ij}}}{1-\sum_{k=0}^{u-1}\binom{n_{ij}}{k}p^{k}\left(1-p\right)^{n_{ij}-k}-\sum_{k=v}^{n_{ij}}\binom{n_{ij}}{k}p^{k}\left(1-p\right)^{n_{ij}-k}};x_{ij}=u,...v-1 \tag{1}$$

2.2.2. Estimation of the Parameter of LRTBD

Method of Maximum Likelihood Estimation (MLE) is used to estimate the unknown parameter *p*. The likelihood function for estimating *p* is given by

$$L=\prod_{j=1}^{k}P(X_{ij}=x_{ij})=\prod_{j=1}^{k}\left(\frac{\binom{n_{ij}}{x_{ij}}p^{x_{ij}}\left(1-p\right)^{n_{ij}-x_{ij}}}{\sum_{x_{ij}=u}^{v-1}\binom{n_{ij}}{x_{ij}}p^{x_{ij}}\left(1-p\right)^{n_{ij}-x_{ij}}}\right) \tag{2}$$

(*k* = number of cases)

The log-likelihood above equation is

$$\log(L)=\sum_{j=1}^{k}\log\binom{n_{ij}}{x_{ij}}+\log(p)\sum_{j=1}^{k}x_{ij}+\log(1-p)\sum_{j=1}^{k}\left(n_{ij}-x_{ij}\right)-\sum_{j=1}^{k}\log\left(\sum_{x_{ij}=u}^{v-1}\binom{n_{ij}}{x_{ij}}p^{x_{ij}}(1-p)^{n_{ij}-x_{ij}}\right)$$

Upon solving $\frac{\partial\log\left(L\right)}{\partial p}=0$, we have

$$\frac{\sum_{j=1}^{k}x_{ij}-p\sum_{j=1}^{k}n_{ij}}{p(1-p)}=\sum_{j=1}^{k}\left(\frac{\sum_{x_{ij}=u}^{v-1}c_1\left(x_{ij}(1-p)^{n_{ij}-x_{ij}}p^{x_{ij}-1}-(n_{ij}-x_{ij})(1-p)^{n_{ij}-x_{ij}-1}p^{x_{ij}}\right)}{\sum_{x_{ij}=u}^{v-1}c_1p^{x_{ij}}(1-p)^{n_{ij}-x_{ij}}}\right) \quad \text{where } c_1=\binom{n_{ij}}{x_{ij}} \tag{3}$$

The maximum likelihood estimate $\hat{p}$ of p is obtained by solving the above equation by themethod of iteration.

2.2.3. The 95% Confidence Interval Based on MLE of p

95% confidence interval (C.I.) of *p* is

$\left(\hat{p}\mp1.96\sqrt{Var(\hat{p})}\right)$ where $Var(\hat{p})=\frac{1}{E\left(-\frac{\partial^2\log(L)}{\partial p^2}\right)}$

and $I(\theta)=-E\left(\frac{\partial^2\log(L)}{\partial p^2}\right)$ is the Fisher's information matrix .

Again, differentiating equation (3) with respect to p , we have

$$\frac{\partial^2 \log(L)}{\partial p^2} = \frac{-p^2 \sum_{j=1}^{k} n_{ij} - (1-2p)\sum_{j=1}^{k} x_{ij}}{p^2 (1-p)^2}$$

$$-\sum_{j=1}^{k} \frac{\left(\left(\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{n_{ij}-x_{ij}} \right) \left(\sum_{x_{ij}=u}^{v-1} c_1 \left(\left(x_{ij}(1-p)^{n_{ij}-x_{ij}} (x_{ij}-1) p^{x_{ij}-2} - x_{ij} p^{x_{ij}-1} (n_{ij}-x_{ij})(1-p)^{n_{ij}-x_{ij}-1} \right) - \left((n_{ij}-x_{ij})(1-p)^{n_{ij}-x_{ij}-1} x_{ij} p^{x_{ij}-1} - (n_{ij}-x_{ij}) p^{x_{ij}} (n_{ij}-x_{ij}-1)(1-p)^{n_{ij}-x_{ij}-2} \right) \right) \right) \right)}{\left(\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{n_{ij}-x_{ij}} \right)^2}$$

$$+\sum_{j=1}^{k} \frac{\left(\sum_{x_{ij}=u}^{v-1} c_1 \left(x_{ij}(1-p)^{n_{ij}-x_{ij}}\, p^{x_{ij}-1} - (n_{ij}-x_{ij})(1-p)^{n_{ij}-x_{ij}-1}\, p^{x_{ij}} \right) \right)^2}{\left(\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{n_{ij}-x_{ij}} \right)^2}$$

$$\Rightarrow E\left(-\frac{\partial^2 \log(L)}{\partial p^2} \right) = \frac{p^2 \sum_{j=1}^{k} n_{ij} + (1-2p)\sum_{j=1}^{k} E(x_{ij})}{p^2 (1-p)^2}$$

$$+\sum_{j=1}^{k} \frac{\left(\left(\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{n_{ij}-x_{ij}} \right) \left(\sum_{x_{ij}=u}^{v-1} c_1 \left(\left(x_{ij}(1-p)^{n_{ij}-x_{ij}} (x_{ij}-1) p^{x_{ij}-2} - x_{ij} p^{x_{ij}-1} (n_{ij}-x_{ij})(1-p)^{n_{ij}-x_{ij}-1} \right) - \left((n_{ij}-x_{ij})(1-p)^{n_{ij}-x_{ij}-1} x_{ij} p^{x_{ij}-1} - (n_{ij}-x_{ij}) p^{x_{ij}} (n_{ij}-x_{ij}-1)(1-p)^{n_{ij}-x_{ij}-2} \right) \right) \right) \right)}{\left(\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{n_{ij}-x_{ij}} \right)^2}$$

$$+\sum_{j=1}^{k} \frac{\left(\sum_{x_{ij}=u}^{v-1} c_1 \left(x_{ij}(1-p)^{n_{ij}-x_{ij}}\, p^{x_{ij}-1} - (n_{ij}-x_{ij})(1-p)^{n_{ij}-x_{ij}-1}\, p^{x_{ij}} \right) \right)^2}{\left(\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{n_{ij}-x_{ij}} \right)^2} \tag{4}$$

and

$$E(X_{ij}) = \sum_{x_{ij}=u}^{v-1} \frac{\left(x_{ij} c_1 p^{x_{ij}} (1-p)^{x_{ij}} \right)}{\left(\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{x_{ij}} \right)} \tag{5}$$

Hence

$$E\left(-\frac{\partial^2 \log(L)}{\partial p^2} \right) = \frac{p^2 \sum_{j=1}^{k} n_{ij} + (1-2p)\sum_{j=1}^{k} \left(\sum_{x_{ij}=u}^{v-1} \frac{\left(x_{ij} c_1 p^{x_{ij}} (1-p)^{x_{ij}} \right)}{\left(\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{x_{ij}} \right)} \right)}{p^2 (1-p)^2}$$

$$+\sum_{j=1}^{k}\frac{\left(\left(\sum_{x_{ij}=u}^{v-1}c_1 p^{x_{ij}}(1-p)^{n_{ij}-x_{ij}}\right)\left(\sum_{x_{ij}=u}^{v-1}c_1\left(\left(x_{ij}(1-p)^{n_{ij}-x_{ij}}(x_{ij}-1)p^{x_{ij}-2}-x_{ij}p^{x_{ij}-1}(n_{ij}-x_{ij})(1-p)^{n_{ij}-x_{ij}-1}\right)-\left((n_{ij}-x_{ij})(1-p)^{n_{ij}-x_{ij}-1}x_{ij}p^{x_{ij}-1}-(n_{ij}-x_{ij})p^{x_{ij}}(n_{ij}-x_{ij}-1)(1-p)^{n_{ij}-x_{ij}-2}\right)\right)\right)\right)}{\left(\sum_{x_{ij}=u}^{v-1}c_1 p^{x_{ij}}(1-p)^{n_{ij}-x_{ij}}\right)^2}$$

$$+\sum_{j=1}^{k}\frac{\left(\sum_{x_{ij}=u}^{v-1}c_1\left(x_{ij}(1-p)^{n_{ij}-x_{ij}}\,p^{x_{ij}-1}-(n_{ij}-x_{ij})(1-p)^{n_{ij}-x_{ij}-1}\,p^{x_{ij}}\right)\right)^2}{\left(\sum_{x_{ij}=u}^{v-1}c_1 p^{x_{ij}}(1-p)^{n_{ij}-x_{ij}}\right)^2} \tag{6}$$

Then, 95% confidence interval (C.I.) for p is

$$\left(\hat{p}\,\mathrm{m}1.96\left(1/\sqrt{E\left(-\frac{\partial^2\log(L)}{\partial p^2}\right)}\right)\right)$$

2.2.4. Moment Estimator of p

Here number of unknown parameters is 2. Therefore, we solve the two equations given by

$$\mu_1' = m_1' \Rightarrow \mu_1' = \frac{1}{k}\sum_{j=1}^{k}x_{ij} \;\Rightarrow\; \mu_1' = \bar{x} \tag{7}$$

$$\mu_2' = m_2' \Rightarrow \;\mu_2' = \frac{1}{k}\sum_{j=1}^{k}x_{ij}^2 \tag{8}$$

$\bar{x} = \mu_1' = E(X_{ij})$ where $E(X_{ij})$ is given by equation (5). On solving (7) and (8) for p, the moment estimator is obtained.

2.2.5. Characteristics of LRTBD

(a) Moment Generating Function: Moment generating function of LRTBD is given by

$$M_{x_{ij}}(t) = E(e^{tx_{ij}}) = \sum_{x_{ij}=u}^{v-1} e^{tx_{ij}}P(X_{ij}=x_{ij})$$

which, upon using equation (1) is given by

$$M_{x_{ij}}(t) \;= \sum_{x_{ij}=u}^{v-1} e^{tx_{ij}}\left(\frac{c_1 p^{x_{ij}}(1-p)^{x_{ij}-n_{ij}}}{\sum_{x_{ij}=u}^{v-1}c_1 p^{x_{ij}}(1-p)^{x_{ij}-n_{ij}}}\right) \tag{9}$$

(b) Characteristic Function: Characteristic function is given by

$$\phi_{x_{ij}}(t) = E(x^{itx_{ij}}) = \sum_{x_{ij}=u}^{v-1} e^{itx_{ij}}P(X_{ij}=x_{ij})$$

$$= \sum_{x_{ij}=u}^{v-1} e^{itx_{ij}} \left(\frac{c_1 p^{x_{ij}} (1-p)^{x_{ij}-n_{ij}}}{\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{x_{ij}-n_{ij}}} \right) \tag{10}$$

(c) Probability Generating Function: Probability generating function is given by

$$Z_{x_{ij}}(t) = E(z^{x_{ij}}) = \sum_{x_{ij}=u}^{v-1} z^{x_{ij}} P(X_{ij} = x_{ij})$$

$$= \sum_{x_{ij}=u}^{v-1} z^{x_{ij}} \left(\frac{c_1 p^{x_{ij}} (1-p)^{x_{ij}-n_{ij}}}{\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{x_{ij}-n_{ij}}} \right) \tag{11}$$

2.2.6. The First Four Moments of LRTBD

Using the moment generating function

$$\mu_1' = E(X_{ij}) = \frac{d}{dt}\left[M_{x_{ij}}(t)\right]\Big|_{t=0}$$

$$Var(x_{ij}) = \mu_2 = \mu_2' - (\mu_1')^2$$

Hence

$$Mean = E(X_{ij}) = \sum_{x_{ij}=u}^{v-1} \left(\frac{x_{ij} c_1 p^{x_{ij}} (1-p)^{x_{ij}}}{\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{x_{ij}}} \right) \tag{12}$$

and

$$Var(x_{ij}) = E(x_{ij}^2) - \left(E(x_{ij})\right)^2$$

where

$$E(X_{ij}^2) = \mu_2' = \frac{d^2}{dt^2}\left[M_{x_{ij}}(t)\right]\Big|_{t=0} = \frac{d}{dt}\left[\frac{d}{dt} M_{x_{ij}}(t)\right]\Big|_{t=0}$$

$$\Rightarrow \mu_2' = \sum_{x_{ij}=u}^{v-1} \left(\frac{x_{ij}^2 c_1 p^{x_{ij}} (1-p)^{x_{ij}}}{\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}} (1-p)^{x_{ij}}} \right)$$

Hence

$$Var(x_{ij}) = \sum_{x_{ij}=u}^{v-1}\left(\frac{x_{ij}^2 c_1 p^{x_{ij}}(1-p)^{x_{ij}}}{\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}}(1-p)^{x_{ij}}}\right) - \left(\sum_{x_{ij}=u}^{v-1}\left(\frac{x_{ij} c_1 p^{x_{ij}}(1-p)^{x_{ij}}}{\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}}(1-p)^{x_{ij}}}\right)\right)^2 \tag{13}$$

Similarly

$$E(X_{ij}^3) = \mu_3' = \frac{d^3}{dt^3}\left[M_{x_{ij}}(t)\right]\Big|_{t=0} = \frac{d}{dt}\left[\frac{d^2}{dt^2}M_{x_{ij}}(t)\right]\Big|_{t=0}$$

$$\mu_3' = \sum_{x_{ij}=u}^{v-1}\left(\frac{x_{ij}^3 c_1 p^{x_{ij}}(1-p)^{x_{ij}}}{\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}}(1-p)^{x_{ij}}}\right)$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3 \tag{14}$$

and

$$E(X_{ij}^4) = \mu_4' = \frac{d^4}{dt^4}\left[M_{x_{ij}}(t)\right]\Big|_{t=0} = \frac{d}{dt}\left[\frac{d^3}{dt^3}M_{x_{ij}}(t)\right]\Big|_{t=0}$$

$$\mu_4' = \sum_{x_{ij}=u}^{v-1}\left(\frac{x_{ij}^4 c_1 p^{x_{ij}}(1-p)^{x_{ij}}}{\sum_{x_{ij}=u}^{v-1} c_1 p^{x_{ij}}(1-p)^{x_{ij}}}\right)$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 \tag{15}$$

The skewness and kurtosis can be calculated by $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$ and $\beta_2 = \frac{\mu_4}{(\mu_2)^2}$ which can be calculated by using μ_2, μ_3 and μ_4 given in equation (13), (14) & (15) respectively.

2.2.7. Data Reliability Tests

Cronbach's alpha: It is a measure of reliability of data, especially psychiatric data. Mathematically, Cronbach's alpha is given by

$$\alpha = \frac{N.\bar{c}}{\bar{v} + (N-1).\bar{c}}$$

where, N is the number of items, $\bar{c}$ is the mean inter-item covariance among the items and $\bar{v}$ is the mean variance. A high value of Cronbach's alpha indicates the high internal consistency of the data, i.e. how closely related are the items of the data as a group [10-11].

Guttman's lambda: It is a reliability quotient used to estimate score correlation for parallel measures. Mathematically, Guttman's lambda is calculated as

$$\lambda_2 = 1 - \sum \frac{\sigma_j^2}{\sigma_x^2} + \sqrt{\left(\frac{n \sum \sum \sigma_{ij}^2}{n-1} \right) / \sigma_x^2} ; i \neq j \text{ where } \sigma_{ij} = Cov\left(i^{\text{th}} \text{ item}, j^{\text{th}} \text{ item} \right)$$

The flow diagram (Figure **1**) below presents the development and application of the model.

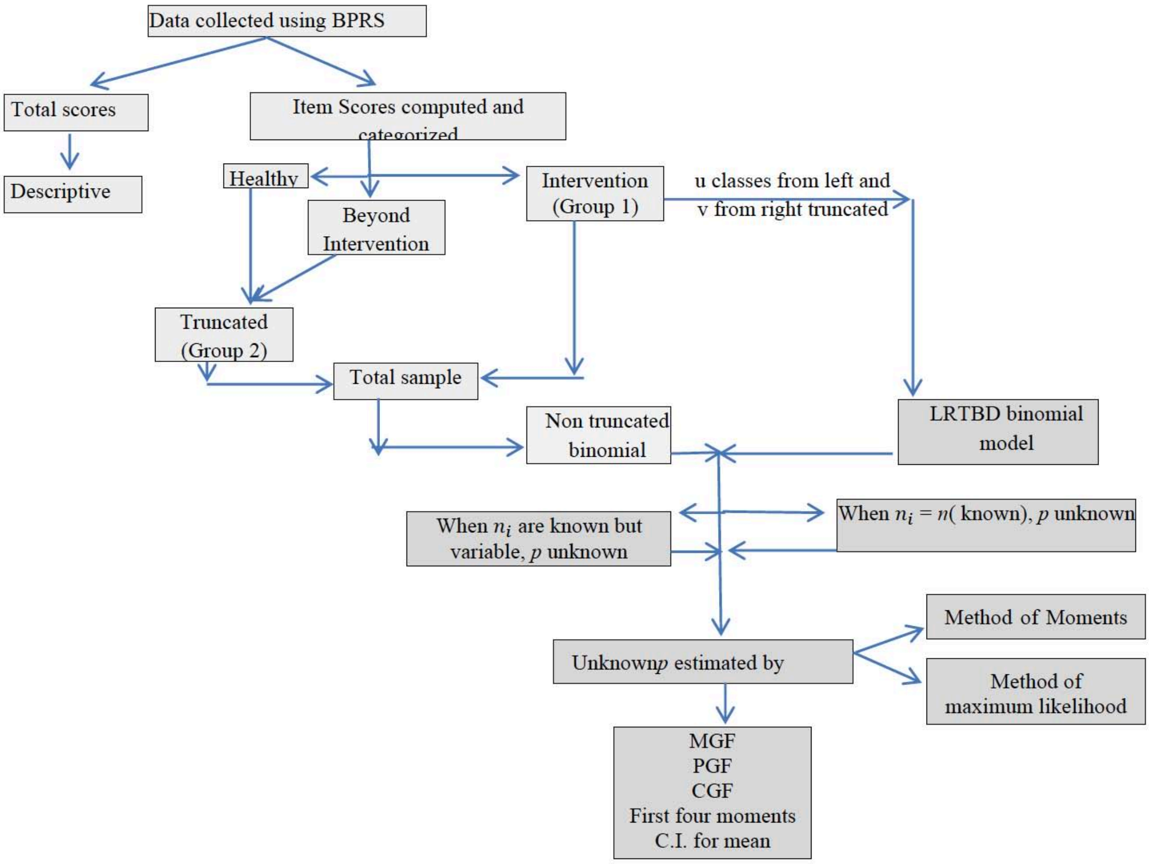

Figure 1: Flow chart of the model developed and executed for BPRS data.

3. RESULTS AND DISCUSSION

The 18-items BPRS instrument was administrated on 93 participants. **Cronbach's alpha** and **Guttman Lambda** were respectively, 0.86 and 0.91.

The LRTBD model developed above was applied to determine the range of the BPRS scores within which an intervention program can be suggested. Thus, participants were divided into two groups, formed on the basis of the item scores of the participants. For every participant, total number of those items was counted for which the item scores were moderately severe/ severe /extremely severe (score≥4). If for a participant the number of such items was zero or one (participant considered healthy); or more than ten (participant was beyond intervention as per experts), the participant was placed in Group 2 otherwise the participant was placed in Group 1. Thus the first group (Group 1) was of those participants for which intervention was suggested. Out of 93 participants, 72 were in Group1. The participants in the second group (Group 2) were either healthy or were beyond intervention.

The item wise descriptive statistics is given below in Table **1** for complete data (N = 93) and for the participants belonging to Group1 (N1 = 72). It is evident from the table that Anxiety, Tension and Excitement were the main problem areas with their mean values being close to moderate category, and in other areas mean scores were close to the mild category [12].

The total BPRS scores were calculated for the complete data and for the participants in Group 1. The descriptive statistics for total scores are given below in Table **2**. The table shows that the minimum score for the two data sets are 18 and 26 respectively; and, the maximum scores are 100 and 69 respectively. The differences in the two scores for the two data sets portray the effect of truncation; minimum score highlighting the effect of left truncation and the maximum score highlighting the effect of right truncation. This justifies the use of double truncation for suggesting the intervention program for Group 1 participants. The third quartile value at 53 indicates that at least 75% of the participants in both the data sets do not belong to the severe category of scores (Figure **2**).

Table 1: Item-Wise Descriptive Statistics of Two Groups; All the Participants and Group 1 Participants

Attribute	Item	Mean		Minimum		Maximum		Standard Deviation	
		N=93	N1=72	N=93	N1=72	N=93	N1=72	N=93	N1=72
Somatic concern	Q1	2.94	2.96	0	0	7	7	1.81	1.87
Anxiety	Q2	4.08	4.26	1	1	7	7	1.67	1.67
Emotional Withdrawal	Q3	2.17	2.36	1	1	7	7	1.36	1.39
Conceptual Disorganization	Q4	2.65	2.78	1	1	6	6	1.25	1.21
Guilt feelings	Q5	2.42	2.54	1	1	5	5	1.25	1.2
Tension	Q6	3.16	3.44	1	1	7	7	1.71	1.64
Mannerism and Posturing	Q7	1.8	1.86	1	1	5	5	1.13	1.15
Grandiosity	Q8	2.69	2.85	0	0	7	7	1.61	1.57
Depressive Mood	Q9	2.33	2.47	1	1	7	7	1.53	1.46
Hostility	Q10	2.77	3	1	1	7	7	1.83	1.82
Suspiciousness	Q11	2.54	2.74	0	0	6	6	1.68	1.68
Hallucinatory Behavior	Q12	2.01	2.1	0	0	7	6	1.57	1.49
Motor Retardation	Q13	2.44	2.64	0	0	7	7	1.59	1.56
Uncooperativeness	Q14	1.78	1.83	0	0	6	6	1.33	1.32
Unusual thought content	Q15	2.18	2.26	0	0	5	5	1.45	1.43
Blunted Effect	Q16	1.71	1.6	0	0	7	5	1.4	1.17
Excitement	Q17	3.34	3.57	0	0	7	7	1.86	1.7
Disorientation	Q18	1.76	1.85	0	0	6	6	1.27	1.31

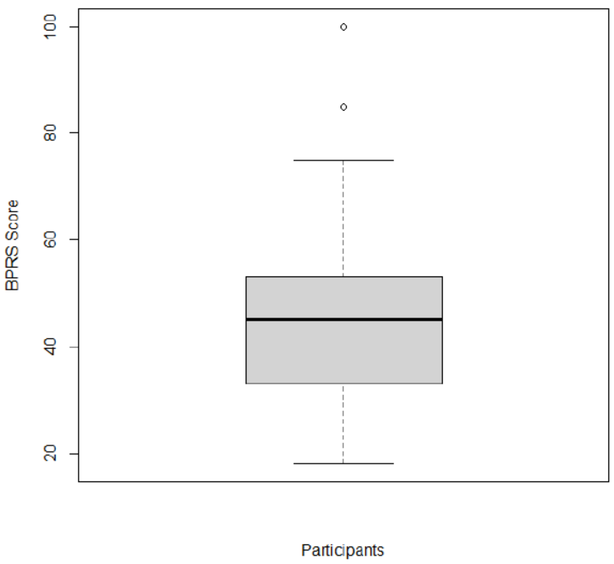

Figure 2: The distribution of total BPRS scores for the complete group.

Table 2: Score Wise-Descriptive Statistics of the Respondents without and with Truncation

Descriptive Statistics	Complete data	After truncation
Mean	44.77	47.11
Standard Deviation	14.95	10.48
Minimum	18	26
Maximum	100	69
Third quartile	53	53

3.1. BPRS Categories Based on Total Scores

The Figure **3** below is based on the total scores of all the 18 items for every participant. The bifurcation of various categories of scores is the standard bifurcation of BPRS instrument [11]. According to our grouping criteria discussed above (based on Likert scale), the proportions in mild and moderate categories belong to the intervention group (Group 1). The proportions in very mild categories are those of the healthy participants. The proportion of the participant who are 'beyond intervention' is a subset the severe category as shown in the figure. According to the BPRS criterion, a participant with total score of 53 or

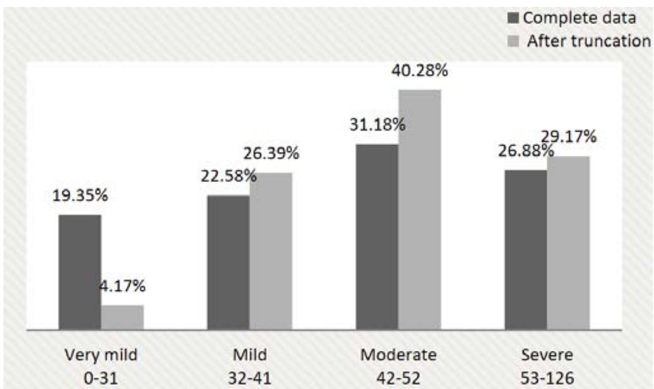

Figure 3: The bar charts representing the proportions of respondents in the four severity categories of BPRS instrument.

Table 3: Frequency Table for the BPRS Total Scores of the Complete and Group 1 Data

S. No.	Problem Area (X_{ij})	Frequency (*fi*)	
		Complete	Group 1
1	0	8	-
2	1	9	-
3	2	7	7
4	3	14	14
5	4	10	10
6	5	14	14
7	6	7	7
8	7	7	7
9	8	6	6
10	9	6	6
11	10	1	1
12	11	1	-
13	12	1	-
14	13	1	-
15	18	1	-

above will lie in the severe category. For complete data, 26.88%; and for Group 1 data, 29.17% (21 out of 72) of the participants belong to this category.

The Table **3** above represents the frequency distribution according to the problem areas faced by the participants for the 'Complete' and Group 1 respectively.

As is clear from Figure **4**, the truncated classes, and particularly on right (11-15 items/problem areas) have very small frequencies in case of complete data also, justifying truncation on account of rare events.

3.2. Truncation Criterion for the Binomial Distribution

For the application of $Bin(n_{ij}, p)$ distribution for the data explained above, n_{ij} represents the number of items assessed by an expert for j^{th} participant, j = 1,2,,,93. The p is the probability of a success which has been defined as "getting a score greater than or equal to 4" in an item of the BPRS instrument. The random variable X_{ij} is the number of successes in the i items for j^{th} participant. According to the truncation criterion discussed above, 21 participants were truncated out of which17 are left truncated which is in accordance with

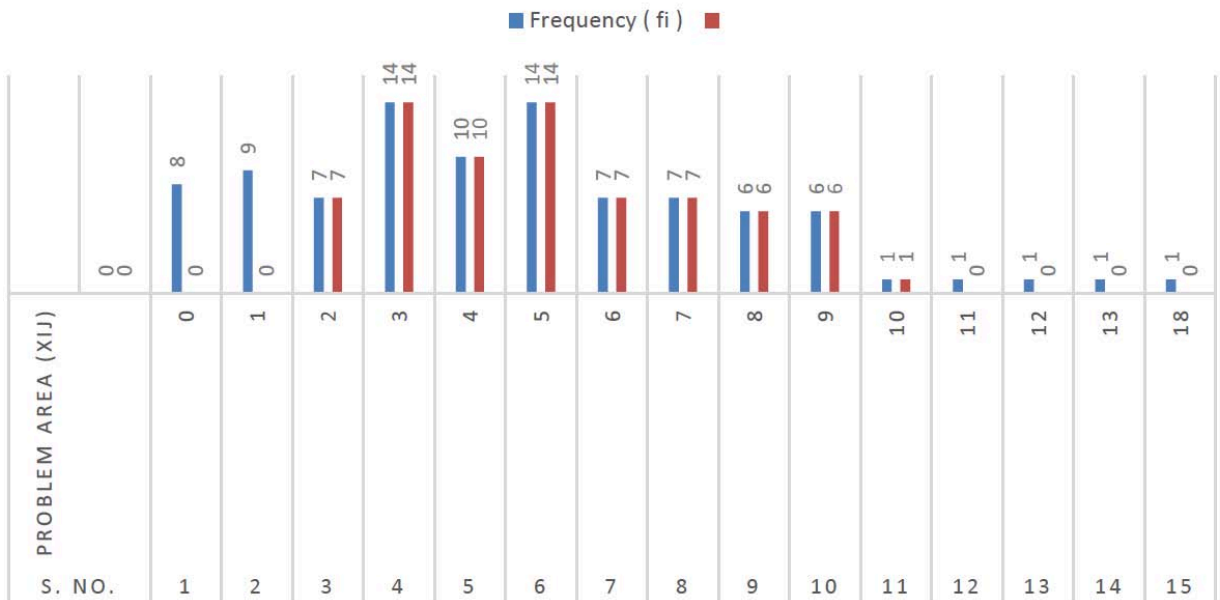

Figure 4: Frequency distribution for the BPRS total scores of the complete and the truncated group.

the item score criterion also. However, according to our criterion, only 4 participants are right truncated ('beyond intervention') whereas according to the item score criterion (item score should be either 6 or 7), 11 participants are in 'beyond intervention' category so there is a mismatch of 7 participants in the right truncation.

3.2.1. Estimating the unknown Parameter p

The model developed in section 2.2 is applied on the BPRS data. The unknown parameter *p* has been estimated under three different models:

i. $Bin\left(n_{ij}, p\right)$; the number of items assessed for every participant is the same and $n_{ij} = n = 18; j = 1,2,...,93$ i.e. for the complete data.

ii. $Bin\left(n_{ij}, p\right)$; the number of items assessed for every participant is not the same and $n_{ij} = 2,...,10$; $j = 1,2,...,93$ i.e. for the complete data.

iii. LRTBD $\left(n_{ij}, p\right)$ - truncation applies but under the assumption of all the 18 items are being assessed and $n_{ij} = n = 18 \forall j$

iv. LRTBD$\left(n_{ij}, p\right)$- truncation applies but with actual number of items assessed by the expert and $n_{ij} = 2,...,10\ \forall\, j$.

The parameter has been estimated for all the four models using the method of maximum likelihood and the method of moments. The estimates from both the methods have been found to be same. The variances and 95% confidence intervals have been obtained. The probability of success is highest for the fourth model which signifies the role of actual number of items assessed (in place of assuming assessment of all the 18 items). The results are presented in Table **4** below:

The Table **5** below shows first four moments about origin and mean for the numbers of problem areas (as identified by the values that the random variables X_{ij} assume;) for the 'Complete' and the Group 1 data respectively. The value of first moment about origin (i.e. Mean) of the 'Complete' group is 4.451, and 5.386 for Group 1 under model 3. This means that the mean score of 'Complete' group is less than that of 'Truncated' group This further implies on an average number of problems faced by respondents is high in the 'Truncated' group (around five) which is consistent with the experts' opinion. In contrast, the mean number of problem areas in the complete group (without truncation) is around four. The value of second moment about mean (i.e. variance) of the 'Complete' group is 10.638, which is higher than the variance (4.692) of the 'Truncated' group. The number of problems faced by respondents (i.e. Score) in 'Complete' group is more spread out or dispersed than the 'Truncated' group.

Table 4: Estimates, Variances of the Estimates and the 95% Confidence Interval for the 'Complete Data' and the 'Group 1' Data

Model	$\hat{p}$	$Var\left(\hat{p}\right)$	95% Confidence Interval	
			LCL	UCL
MLE (without Truncation)	0.2581	0.00011437	0.2578	0.2583
MLE (without Truncation, and actual number of items assessed)	0.2675	0.00011345	0.2466	0.2884
MLE (with Truncation, and equal number of items assessed)	0.2796	0.00016935	0.27921	0.2799
MLE (with Truncation, and actual number of items assessed	0.2985	0.00017494	0.2982	0.2988

Table 5: First Four Moments about Origin and Mean for all the Three Models

Moments	Complete data, equal number of items assessed		Complete data, actual number of items assessed		Group 1, equal number of items assessed		Group 1, actual number of items assessed	
	About Origin	About Mean	About Origin	About Mean	About Origin	About Mean	About Origin	About Mean
1st	4.451	0	4.353	0	5.352	0	5.386	0
2nd	32.215	10.638	25.060	3.765	30.958	4.692	32.501	3.359
3rd	282.580	34.105	150.824	0.441	210.875	4.118	215.389	3.671
4th	3024.731	548.249	987.684	45.071	1560.792	46.992	1539.897	24.022
Skewness		0.965		0.003		0.164		0.355
Kurtosis		4.843		3.179		2.134		2.128

This establishes the usefulness of double truncated models in such scenarios.

4. CONCLUDING REMARK

In all the parametric distributions, the characteristics such as the variance, confidence intervals, skewness and kurtosis, are generally mean based which are affected by the outliers /extreme values of the data. The usual ways of handling the issue of outliers is (i) to retain them in the study as they are; and (ii) to discard them treating them as nuisance. In both the cases, the results obtained may be misleading or not representative of the true values. However outliers are not always nuisance but may be the outcomes of rare events. For example, in our case, the right most values correspond to the rare events for healthy adolescents (these are not nuisance values but the conditions of those adolescents went unnoticed until this study). Truncated models in such situations provide a better solution by eliminating the direct effect and retaining the indirect effect of extreme values by incorporating them in their probability mass functions.

REFERENCES

[1] Suzuki M. A Binomial Error Model with a Truncated Latent Distribution and its Application. Behaviormetrika 1983; 10: 49-58. https://doi.org/10.2333/bhmk.10.14_49

[2] Bilker WB, Wang MC. A semiparametric extension of the Mann-Whitney test for randomly truncated data. Biometrics 1996; 10-20. https://doi.org/10.2307/2533140

[3] Efron B, Petrosian V. Nonparametric methods for doubly truncated data. Journal of the American Statistical Association 1999; 824-34. https://doi.org/10.1080/01621459.1999.10474187

[4] Tokmachev MS. Modeling of truncated probability distributions. InIOP Conference Series: Materials Science and Engineering 2018; 441(1): 012056). IOP Publishing. https://doi.org/10.1088/1757-899X/441/1/012056

[5] Aydin D. The doubly-truncated exponentiated inverse Weibull distribution. Anadolu University Journal of Science and Technology B-Theoretical Sciences 2018; 6(1): 55-74. https://doi.org/10.20290/aubtdb.326782

[6] Fulford D, Pearson R, Stuart BK, Fisher M, Mathalon DH, Vinogradov S, Loewy RL. Symptom assessment in early psychosis: the use of well-established rating scales in clinical high-risk and recent-onset populations. Psychiatry Research 2014; 220(3): 1077-83. https://doi.org/10.1016/j.psychres.2014.07.047

[7] Overall JE, Gorham DR. The Brief Psychiatric Rating Scale (BPRS): recent developments in ascertainment and scaling. Psychopharmacology Bulletin 1988.

[8] Thompson PA, Buckley PF, Meltzer HY. The Brief Psychiatric Rating Scale: Effect of Scaling: System on Clinical Response Assessment. Journal of Clinical Psychopharmacology 1994; 14(5): 344-6. https://doi.org/10.1097/00004714-199410000-00010

[9] Unni KE, Srivastav R, Sabharwal A, Goyal B. Assessing the effect of Parental Psychiatric disorder on Adolescents using Brief Psychiatric Rating Scale-A Case Study. International Journal of Research in Social Sciences 2019; 9(7): 401-16.

[10] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16(3): 297-334. https://doi.org/10.1007/BF02310555

[11] Tavakol M, Dennick R. Making sense of Cronbach's alpha. International Journal of Medical Education 2011; 2: 53. https://doi.org/10.5116/ijme.4dfb.8dfd

[12] Leucht S, Kane JM, Kissling W, Hamann J, Etschel EV, Engel R. Clinical implications of brief psychiatric rating scale scores. The British Journal of Psychiatry 2005; 187(4): 366-71. https://doi.org/10.1192/bjp.187.4.366

Received on 20-11-2023 Accepted on 14-12-2023 Published on 11-01-2024

https://doi.org/10.6000/1929-6029.2024.13.01