Estimating Length of Stay and Duration of Illness for Psychiatric Inpatients using Multivariate Modelling

Gurprit Grover[1], Alka Sabharwal[2,*], Sakshi Kaushik[1]

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

Abstract   Mental and behavioural disorders is a significant contributor of global burden of disease. As per WHO estimates, this burden is likely to increase by 15 percent by 2020, significantly impacting health and major social, human rights and economic consequences in all countries of the world. This paper provides a procedure for estimation of length of stay (LOS) in the hospital, total duration of illness (TDI) and recent duration of illness (RDI) using multivariate normal (MVN) and bivariate (BVN) normal distributions. To accomplish this, a retrospective data of 146 patients with complete record history, diagnosed with mental and behavioural disorders (as per APA's DSM-V as well as the WHO's ICD-10) is collected from the Department of Psychiatry, Lady Hardinge Medical College & Smt. S.K, Hospital, New Delhi, India for the calendar year 2013-2014. The estimated values of the above mentioned variables are found to be consistent with the observed values. Finally MVN distribution is applied to estimate the variables LOS and TDI for the patients for whom the information on these variables is not known. The model derived in this paper will facilitate the medical fraternity to not only guide the patients about their approximate length of stay in the hospital at the time of admission, but also assist them in development and management of appropriate interventions for patients.

Keywords   Duration of illness, Length of stay, Mardia test, Mental disorder, Multivariate normal distribution, Psychiatric rating scale

1. Introduction

Mental disorders are common non-communicable diseases rising with epidemic rates globally with over one third of people in most countries reporting sufficient criteria to be diagnosed at some point in their life [1]. Mental illness is associated with a significant burden of morbidity and disability. According to DSM-IV each mental disorder can be conceptualized as "a clinically significant behavioural or psychological syndrome or pattern that occurs in a person and that is associated with present distress (a painful symptom) or disability (impairment in one or more important areas of functioning) or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom". Diagnostic and Statistical Manual of Mental Disorders (DSM), a publication of American Psychiatric Association (APA), is a standard classification of mental disorders used by mental health professionals worldwide for diagnosing mental disorders. Another widely used diagnostic manual for mental disorders is chapter V of International Classification of Diseases (ICD), a publication of World Health Organization (WHO).

Mental and behavioural disorders account for 12% of the global burden of disease. According to WHO 2001 report, nearly 450 million of world population is suffering from mental disorders [2]. In 2005, the National Commission on Macroeconomics and Health, India reported that 10-20 million (1-2 per cent) of population suffered from severe mental disorders such as schizophrenia and bipolar disorder and nearly 50 million (5 per cent of population) suffered from common mental disorders like depression and anxiety, yielding an overall estimate of 6.5 per cent [4]. The burden of these disorders is likely to increase by 15 percent by 2020 according to WHO estimates [2]. The burden of mental disorders continues to grow with significant impacts on health and major social, human rights and economic consequences in all countries of the world [5].

Severe mental illness is often defined by its duration and the disability it produces. These illnesses include disorders that produce psychotic symptoms, such as schizophrenia and schizoaffective disorder, and severe forms of other disorders, such as major depression and bipolar disorder [6]. Schizophrenia is often described in terms of positive and negative (or deficit) symptoms [7]. Positive symptoms are those that most individuals do not normally experience but are present in people with schizophrenia. Negative symptoms are deficits of normal emotional responses or of

other thought processes, and are less responsive to medication [9]. Bipolar disorder causes extreme mood swings ranging from mania or extreme happiness, grandiosity, euphoria or irritability, or decreased need for sleep. Typically a person with bipolar disorder cycles from one extreme to the other while experiencing periods with few or no symptoms in between [10].

Acute and transient psychotic disorder (ATPD) has certain key features, such as acute onset (within 2 weeks) and rapidly changing, variable polymorphic picture, which are accepted as required criteria for diagnosis [12]. Depression is a common mental disorder, characterized by sadness, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, feelings of tiredness and poor concentration. It is associated with poorer lifelong health outcomes for the individual which may, in part, be related to suboptimal brain maturational processes [13].

Previous studies suggest that duration of illness, length of stay in the hospital and severity measured by psychiatric rating scales are the three major contributors of mental health of a patient [14]. The duration of mental illness not only reflects the chronic condition of the disorder but also plays a vital role for its diagnosis. According to ICD-10, the diagnosis of schizophrenia depends upon the presence of typical delusions, hallucinations or other symptoms, clearly present for a minimum duration of one month while the duration requirement for the diagnosis of delusional disorder is at least three months. Clearly the efficiency of diagnosis of a mental disorder is dependent on its duration indicating that the initiation of effective treatment is also contingent on the duration of the disease. Length of stay (LOS) in hospital is generally used as an indirect indicator of effectiveness and efficiency of treatment programs [15]. While LOS is not only a final indicator of effective and efficient care, it is a variable of considerable interest in hospital psychiatry as an indicator of resource utilization [16]. Many different questionnaires, interviews, checklists, outcome assessments and other instruments are used by psychiatrists and mental health professionals to aid in treatment planning by helping to establish a diagnosis, identify co morbid conditions, and assess levels of functioning. These are collectively called psychiatric rating scales or rating instruments [17]. The decision whether to hospitalize a patient, and, if so, in what type of setting is one of the most basic, yet, tricky decisions about level of care. Thus, various psychiatric rating scales are used to differentiate among psychiatric patients who require hospitalization from those who are suitable for outpatient care [18].

Previous work reveals that a lot of research has been done for predicting LOS. Huntley ET. al (1998) [19] and Ithman et al (2014) [21] applied regression modelling to predict LOS in hospital with independent predictors: a primary diagnosis of schizophrenia, the number of previous admissions, a primary diagnosis of mood disorder, age and a secondary diagnosis of alcohol- or other drug related disorder; and age, marital status, involuntary admission and diagnosis of an affective disorder or a psychotic disorder, respectively.

Anderson et. al (2004) examined whether assessment data from administration of the extended version of the Brief Psychiatric Rating Scale (BPRS-E) to state hospital patients within 72 hours of their admission could be used to predict LOS. Factor analysis extracted the four factors: patient's resistance to treatment, positive symptoms, mood, and negative symptoms while Discriminant analysis showed that the negative symptoms factor correctly predicted that LOS for 94 percent of cases exceeded the mean LOS. These findings suggest that the severity of negative symptoms can be a useful predictor of LOS among patients with severe and persistent mental illness [20].

Multivariate normal (MVN) distribution is one of the most frequently made distributional assumptions for the analysis of multivariate data. The assumption of multivariate normality is often checked by individually examining the univariate normality through various P-P plots or some other plots but it can be very subjective at times. Further, the marginal normality of each random variable does not guarantee the multivariate normality. Thus, one of the relatively simpler and mathematically tractable ways to find a support for the assumption of multivariate normality is by using the tests based on Mardia's multivariate skewness and kurtosis measures [25, 26].

Grover et. al (2014) applied MVN distribution to a dataset of type-2 diabetic patients for estimating the duration of diabetes on the basis of their latest renal health status. They fitted the MVN distribution model on a dataset of diabetic nephropathy patients with complete information (from the time of diagnosis till termination of study), validated the model by using simulation and then used it for estimating the duration of diabetes for patients with incomplete record history. The results are also derived for bivariate normal (BVN) distribution, as a special case of MVN distribution. Their work has proved that MVN and BVN distributions can contribute meaningfully to approximate the duration of disease and hence evaluate the severity of disease [27].

Despite the massive toll and the powerful impact mental disorders can have on individual and society at large, they do not receive nearly the amount of attention paid to other diseases. While many communicable diseases have been shrinking in terms of the number of people they affect over the last two decades, the global burden of mental illness has largely remained the same. Not much statistical modeling has been applied to mental disorder data. This motivates us to build a model and test its validity for the same.

Previous studies reveal that the variables LOS and duration of illness are important contributors for progression of disease while the psychiatric rating scale quantifies the intensity and severity of mental disorder. The information on above mentioned variables is generally available, if not for all, at least for some. Previous studies have presented the important predictors on the basis of qualitative information available, for e.g. Ithman et. al (2014) [21] showed marital status, age, involuntary admission and diagnosis of an affective disorder or a psychotic disorder to be independent variables that predicted length of stay, while in this study we

have attempted to use the maximum available quantitative information on variables LOS and duration of illness to estimate the unknown variable. This builds the background of our study.

The aim of present study is to derive a model which could be implemented immediately for estimating the duration of illness and LOS in hospital for patients with complete information and also for whom the information on all variables of interest is not known. This model not only estimates the current values of these variables but can also predict the future values for the same.

The four variables considered appropriate for this study are length of stay (LOS) in the hospital, recent duration of illness (RDI), rating scale score (RSS) and total duration of illness (TDI). LOS is defined as the duration for which a patient is admitted to a hospital following an episode. TDI is the total length of duration when the patient was symptomatic (as per ICD-10 and DSM-5) for the first time till the current episode whereas RDI is the recent duration for which the patient is symptomatic for the latest (recorded) episode as observed by the patient and/or his acquaintances. RSS (scaled down to 100) is the score of the psychiatric rating scales depicting the status of illness of the patient.

This paper attempts to provide a procedure for estimation of LOS in the hospital, TDI and RDI using multivariate and bivariate normal distributions. The joint distribution of four variables: LOS, RDI, RSS and TDI have been analyzed to estimate the LOS, RDI and TDI for patients with mental illness. Prerequisites for application of MVN distribution are checked as follows: Firstly various distributions are fitted to the four variables LOS, RDI, TDI and RSS, and appropriate distributions are selected on the basis of Akaike Information Criterion (AIC) for each random variable [28]. Secondly, univariate normality is checked for each distribution. It is observed that except RSS, univariate distributions of other variables are Gamma and lognormal, on applying log and square root transformations, all the four variables tend to normal distribution. Lastly, multivariate normality has been verified for the variables LOS, RDI, TDI and RSS by applying Mardia's test [26]. It is observed that the above mentioned four variables are marginally as well as jointly normally distributed after transformation. This supported the usage of four dimensional MVN distribution model to the data.

The first dataset used for conducting this study is retrospective data of 146 patients with complete record history, diagnosed with mental and behavioural disorders (as per American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) as well as the World Health Organization (WHO) International Classification of Diseases (ICD-10)), collected from the Department of Psychiatry, Lady Harding Medical College & Smt. S.K, Hospital, New Delhi, India for the calendar year 2013-2014. Data set 2 consists of incomplete data of 21 patients with mental disorders, collected from the same hospital. On the basis of the known variables, the unknown variables are estimated using the validated model derived in this paper.

Using dataset 1, LOS is estimated by applying the knowledge of conditional expectation under MVN distribution for given values of RDI, RSS and TDI. Further, LOS is obtained by applying conditional expectation under MVN distribution for given values of RDI and RSS. In the similar manner, RDI is obtained conditionally for given values of LOS and RSS. The above procedure is firstly applied to estimate variables for the first dataset for which these variables are already known. The estimated values of the above mentioned variables are found to be very close to their observed values, which validates the above mentioned procedure to be an appropriate method of estimation. Further, the validity of the model is confirmed by using simulation. Finally, MVN distribution is applied to estimate the variables LOS and TDI for the patients belonging to dataset 2 for whom information on these variables is not known. Three dimensional MVN and BVN distributions are also applied as special cases of four dimensional MVN distribution.

Not much statistical and clinical work has been done for estimation of total as well as recent duration of mental illness. Also, the application of four dimensional MVN distribution to mental disorders is a novel concept, which, to the best of our knowledge has not been explored previously. This paper not only exhibits a practical approach towards estimation of variables LOS and duration of mental illness but also imparts a theoretically plausible derivation of conditional expectation under four dimensional MVN distribution. The model derived in this paper will facilitate the medical fraternity to not only guide the patients about their approximate length of stay in the hospital at the time of admission, but also assist them in development and management of appropriate interventions for patients. This will further aid in monitoring and evaluating the severity of mental disorders. The course of this paper along with introduction is as follows. In section 2, development of model is discussed. In Section 3, the model is applied to the data. Section 4 discusses the results with previous studies and the paper is concluded in section 5.

2. Methodology

2.1. Multivariate Normal (MVN) Distribution

The random vector $\underset{\sim}{X}$ is said to have a p-dimensional multivariate normal distribution (MVN) with a mean vector $\underset{\sim}{\mu}$ and variance-covariance matrix Σ if it has the joint probability density function of the form [29],

$$f(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \left(\underset{\sim}{x} - \underset{\sim}{\mu} \right)^T \Sigma^{-1} \left(\underset{\sim}{x} - \underset{\sim}{\mu} \right) \right\};$$

2.1.1. Conditional Density of a 4-Dimensional Random Vector with MVN Distribution

The 4-dimensional random vector with MVN distribution is a special case of MVN distribution with p=4. Assume a

4-dimensional random vector $\underset{\sim}{X}$ has multivariate normal distribution (MVN) with mean vector $\underset{\sim}{\mu}$ and variance-covariance matrix Σ, i.e.,

$$\underset{\sim}{X} \sim MVN\left(\underset{\sim}{\mu}, \Sigma\right), \text{ where } X_{4X1} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix}, \quad \mu_{4X1} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} \text{ and } \Sigma_{4X4} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix} \text{ with the joint probability}$$

density function [29],

$$f(\underset{\sim}{x}) = \frac{\exp\left\{-\frac{1}{2}\left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)^T \Sigma^{-1}\left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)\right\}}{(2\pi)^2 |\Sigma|^{\frac{1}{2}}}; \tag{1}$$

Where, $-\infty < (\underset{\sim}{x}, \mu) < \infty, \sigma_{ii} > 0$ and $-\infty < \sigma_{ij} < \infty$

Since X_1, X_2, X_3 and X_4 are jointly normal, the conditional density of X_1 given X_2, X_3 and X_4 is

$$f\left(x_1 \mid x_2, x_3, x_4\right) = \frac{f\left(x_1, x_2, x_3, x_4\right)}{f\left(x_2, x_3, x_4\right)} \tag{2}$$

Let $A = \Sigma^{-1} = \left[a_{ij}\right]_{4X4}$ and $y_i = x_i - \mu_i$; i, j = 1, 2, 3, 4 $\tag{3}$

Considering the exponent part in (1), let

$$Q = \left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)^T \Sigma^{-1} \left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)$$

$$= (y_1 \ y_2 \ y_3 \ y_4) \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

$$= a_{11} y_1^2 + \sum_{i,j=2}^{4} y_i a_{ij} y_j + y_1 \sum_{j=2}^{4} a_{1j} y_j + y_1 \sum_{i=2}^{4} a_{i1} y_i$$

Since $A = \Sigma^{-1}$ is a symmetric matrix, $\sum_{j=2}^{4} a_{1j} y_j = \sum_{i=2}^{4} a_{i1} y_i$. Thus,

$$Q = a_{11} y_1^2 + \sum_{i,j=2}^{4} y_i a_{ij} y_j + 2 y_1 \sum_{j=2}^{4} a_{1j} y_j \tag{4}$$

$$f(x_1, x_2, x_3, x_4) = \frac{\exp\left\{-\frac{1}{2}\left(a_{11} y_1^2 + \sum_{i,j=2}^{4} y_i a_{ij} y_j + 2 y_1 \sum_{j=2}^{4} a_{1j} y_j\right)\right\}}{(2\pi)^2 |\Sigma|^{\frac{1}{2}}}$$

$$= \frac{\exp\left\{-\frac{1}{2} \sum_{i,j=2}^{4} y_i a_{ij} y_j\right\} \cdot \exp\left\{-\frac{1}{2}\left(a_{11} y_1^2 + 2 y_1 \sum_{j=2}^{4} a_{1j} y_j\right)\right\}}{(2\pi)^2 |\Sigma|^{\frac{1}{2}}} \tag{5}$$

Let $U = \sum_{j=2}^{4} a_{1j} y_j$, then

$$\exp\left\{-\frac{1}{2}\left(a_{11}y_1^2 + 2y_1\sum_{j=2}^{4}a_{1j}y_j\right)\right\} = \exp\left\{-\frac{a_{11}}{2}\left(y_1^2 + \frac{2y_1 U}{a_{11}}\right)\right\}$$

$$= \exp\left\{-\frac{a_{11}}{2}\left[\left(y_1 + \frac{U}{a_{11}}\right)^2 - \left(\frac{U}{a_{11}}\right)^2\right]\right\}$$

$$= \exp\left\{-\frac{U^2}{2a_{11}}\right\}\cdot\exp\left\{-\frac{a_{11}}{2}\left(y_1 + \frac{U}{a_{11}}\right)^2\right\}$$

Thus,
$$f(\underset{\sim}{x}) = \frac{\exp\left\{-\frac{1}{2}\sum_{i,j=2}^{4}y_i a_{ij}y_j\right\}\cdot\exp\left\{-\frac{U^2}{2a_{11}}\right\}\cdot\exp\left\{-\frac{a_{11}}{2}\left(y_1 + \frac{U}{a_{11}}\right)^2\right\}}{(2\pi)^2|\Sigma|^{\frac{1}{2}}}$$

$$f(\underset{\sim}{x}) = D(y_2, y_3, y_4)\cdot\exp\left\{-C\left(y_1 + \frac{U}{2C}\right)^2\right\} \tag{6}$$

where, $C = \dfrac{a_{11}}{2}$ and,

$D(y_2, y_3, y_4) =$ a function of y_2, y_3, y_4 not involving y_1

$$= \frac{1}{(2\pi)^2|\Sigma|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\sum_{i,j=2}^{4}y_i a_{ij}y_j\right\}\cdot\exp\left\{-\frac{U^2}{2a_{11}}\right\}$$

Further,

$$f(x_2, x_3, x_4) = \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4)\, dx_1$$

$f(x_2, x_3, x_4) =$ a function involving y_2, y_3, y_4 and independent of y_1.

$\qquad = B(y_2, y_3, y_4)$, (say)

Then the conditional density of X_1 given X_2, X_3 and X_4 has the form:

$$f(x_1 \mid x_2, x_3, x_4) = \frac{D(y_2, y_3, y_4)}{B(y_2, y_3, y_4)}\cdot\exp\left\{-C\left(y_1 + \frac{U}{2C}\right)^2\right\} \tag{7}$$

From the form of the density it can be concluded that $X_1 \mid X_2, X_3, X_4$ is normal.

The conditional variance of $X_1 \mid X_2, X_3, X_4$ (same as the conditional variance of $(Y_1 = X_1 - \mu_1)$) is

$$\mathrm{var}(X_1 \mid X_2, X_3, X_4) = \frac{1}{a_{11}} \quad\text{and the conditional mean of } Y_1 \text{ is } -\frac{U}{2C} = -\frac{1}{a_{11}}\sum_{j=2}^{4}a_{1j}y_j$$

Thus, the conditional mean and variance of X_1 are [37]

$$E(X_1 \mid X_2, X_3, X_4) = \mu_1 - \frac{1}{a_{11}}\sum_{j=2}^{4}a_{1j}(X_j - \mu_j) \quad\text{and}\quad \mathrm{var}(X_1 \mid X_2, X_3, X_4) = \frac{1}{a_{11}} \tag{8}$$

2.1.2. Conditional Density of a 3-Dimensional Random Vector with MVN Distribution

The 3-dimensional random vector with MVN distribution is a special case of MVN distribution with p=3. Assume a 3-dimensional random vector $\underset{\sim}{X}$ has multivariate normal distribution (MVN) with mean vector $\underset{\sim}{\mu}$ and variance-covariance matrix Σ, i.e.,

$\underset{\sim}{X} \sim MVN(\underset{\sim}{\mu}, \Sigma)$, where,

$$\underset{\sim}{X}_{3X1} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}, \quad \underset{\sim}{\mu}_{3X1} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} \quad \text{and} \quad \Sigma_{3X3} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \quad \text{with the joint probability density function [29],}$$

$$f(\underset{\sim}{x}) = \frac{\exp\left\{-\frac{1}{2}\left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)^T \Sigma^{-1} \left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)\right\}}{(2\pi)^{3/2} |\Sigma|^{\frac{1}{2}}}; \quad -\infty < \underset{\sim}{x} < \infty \tag{9}$$

Let $A = \Sigma^{-1} = \left[a_{ij}\right]_{3X3}$; i, j=1, 2, 3

Proceeding in the similar manner as mentioned in the previous section, the conditional mean and variance of $X_1 / X_2, X_3$ can be obtained as

$$E(X_1 \mid X_2, X_3) = \mu_1 - \frac{1}{a_{11}} \sum_{j=2}^{3} a_{1j}(X_j - \mu_j) \quad \text{and} \quad \text{var}\left(X_1 \mid X_2, X_3\right) = \frac{1}{a_{11}} \tag{10}$$

2.1.3. Conditional Density under Bivariate Normal (BVN) Distribution

The bivariate normal distribution is a special case of MVN distribution with p=2. If x and y are two related variables following normal distributions, $x \sim N\left(\mu_x, \sigma_x^2\right)$ and $y \sim N\left(\mu_y, \sigma_y^2\right)$ with correlation coefficient ρ , then the joint probability density function of x and y is defined as [29],

$$f(x,y) = \frac{\exp\left\{-\frac{1}{2\left(1-\rho^2\right)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

Then the conditional expectation under BVN distribution is given as

$$E(x \mid y) = \mu_x - \frac{1}{\sigma_x}\sigma_{xy}(y - \mu_y) \tag{11}$$

2.2. Distribution of Variables

The response variables under study are assumed to follow some statistical distributions. Among the different distributions applied to the variables under study, Gamma, lognormal and normal distribution fitted those best. Gamma distribution is one of the most widely used distributions in medical research. The lognormal distribution (with two parameters) may be defined as the distribution of a random variable whose logarithm is normally distributed. Thus a normal variate can be obtained from a lognormal variate by applying a logarithmic transformation on it. The resulting transformed variable will follow approximately normal distribution [33]. The normal distributions are a very important class of statistical distributions. They form the base for application of MVN distribution to the joint random variables under study.

The probability density functions of aforementioned distributions are defined as follows:

2.2.1. Gamma Distribution

$$f(x / \lambda, \gamma) = \frac{e^{-x/\lambda} x^{\gamma-1}}{\lambda^{\gamma} \overline{|\gamma}} \quad ; \lambda > 0, \gamma > 0 \ \& \ x > 0$$

2.2.2. Lognormal Distribution

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\} \quad x > 0$$

2.2.3. Normal Distribution

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right); \quad -\infty < (x, \mu) < \infty, \sigma > 0$$

2.3. Transformations

Many statistical methods require the data to fit assumptions of normal distribution and uniform variance. When data do not fit these, one approach is to make them do so by a mathematical transformation. The most frequently used transformations are the square root, the logarithm, and the reciprocal. The square root can be used for variables which are greater than or equal to zero, the log and the reciprocal can only be used for variables which are strictly greater than zero [30]. In this article square root and logarithm transformations have been applied to the variables under study. Square root and logarithm transformations are a special case of Box-Cox or Power transforms. Box-Cox or Power transforms are a family of functions for transforming data. The various forms of the transforms utilize a single parameter, λ, which is either specified or is computed from the data using some form of optimization procedure on the assumption that the transformed sample is to be normally distributed. The simplest form of power transformation is [31]:

$$z = (x+c)^{\lambda}, \lambda > 0, x > 0, c > 0, \text{and}$$
$$z = \ln(x+c), \lambda = 0$$

where c is an optional constant. Ahrens and Dieter showed that a suitable square root transformation of a gamma random variable with mean $a \geq 1$ yields a probability density close to the standard normal density [32].

2.4. Mardia Test for Multivariate Normality

Classical methods of multivariate analysis are mostly based on the assumption that the data are coming from a multivariate normal population. Thus, it becomes crucial to test the assumption of multivariate normality of the data at hand. Several statistical and graphical approaches, including the Q-Q plot, P-P plot, box-plot, stem and leaf plot, Shapiro-Wilk and Kolmogorov-Smirnov tests to evaluate the univariate normality, contour and perspective plots for assessing bivariate normality, and the chi-square Q-Q plot to check the multivariate normality, are available [34]. One of the relatively convenient, most widely used and accurate method for testing multivariate normality is by using the tests based on Mardia multivariate skewness and kurtosis measures. The sample measures of multivariate skewness and kurtosis are,

$$\hat{\gamma}_{1,p} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij}^3 \quad \text{and} \quad \hat{\gamma}_{2,p} = \frac{1}{n^2} \sum_{i=1}^{n} m_{ii}^2$$

where, $m_{ij} = (x_i - \bar{x})'S^{-1}(x_j - \bar{x})$,

$$S = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'\right] \& \bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$$

and, p is the number of variables. The test statistic for skewness, $\kappa_1 = \frac{n\hat{\gamma}_{1,p}}{6}$, is approximately χ^2 distributed

with p(p+1)(p+2)/6 degrees of freedom. The test statistic for kurtosis $\kappa_2 = \dfrac{\left(\hat{\gamma}_{2,p} - p(p+2)\right)}{\sqrt{\left(8p(p+2)/n\right)}}$ tends to standard normal distribution [26].

2.5. Criteria for Model Checking

The distributional assumptions made about the data are a matter of special relevance in medical studies as it critically influences the results obtained. The selection of an appropriate approximating model is critical to statistical inference from many types of empirical data. Application of Akaike Information Criterion (AIC) is one of the most used information criterion for model selection and is defined as:

$$AIC = -2\log(L) + 2p$$

Where, L refers to the likelihood under the fitted model and p is the number of parameters in the model. One should select the model that yields the smallest value of AIC because this model is estimated to be "closest" to the unknown reality that generated the data, from among the candidate models considered [28].

3. Application

The medical records for the datasets used in this study have been collected from the Department of Psychiatry, Lady Hardinge Medical College & Smt. S.K, Hospital, New Delhi, India for the calendar year 2013-2014. The patients enrolled in the study are diagnosed with mental and behavioural disorders as per American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) as well as the World Health Organization (WHO) International Classification of Diseases (ICD-10). The patients included in the study belonged to the age group 20-60 years. An attempt is made to include almost equal number of patient from both the genders. The data consisted of patients diagnosed with Bipolar Affective Disorder (BPAD), Schizophrenia, Depression and other disorders. The others category includes disorders viz. Acute Transient Psychotic Disorder (ATPD), Dementia, Psychotic Disorder: Not otherwise Specified (NOS) and Dissociative Motor Disorder. The patients diagnosed with substance abuse, pregnant women and suffering from any other chronic medical illness along with mental and behavioural disorders are excluded from the study.

Out of two datasets, the first dataset is a retrospective data of 146 patients diagnosed with mental and behavioural disorders with complete record history. It consisted of the admission and discharge records of all the patients along with the complete details of their recent and previous episode history. Complete demographic and clinical information regarding the subjects is available in their admission records. In the present data, Psychopathology is

assessed by the Positive and Negative Syndrome Scale (PANSS), Bush-Francis Catatonia Rating Scale (BFCRS), Brief Psychiatric Rating Scale (BPRS), The Hamilton Rating Scale for Depression (HAMD), Young Mania Rating Scale (YMRS), Hamilton Anxiety Rating Scale (HAM-A) [35-40]. The Complete Blood Count (CBC), Electrocardiogram (ECG) and Magnetic Resonance Imaging (MRI) test reports of all patients are provided. All patients were receiving antipsychotic medications. The entire eating, sleeping and hygienic pattern of patients are noted. 46.58 percent males and 53.42 percent females are included in the study. The percentage of patients observed to be diagnosed with BPAD, Schizophrenia, Depression and others are 21.23, 50, 8.90 and 19.86 percent, respectively. Table 1 represents the descriptive statistics of 146 patients of data set 1 diagnosed with mental and behavioural disorders giving total number, minimum, maximum, range, mean and standard deviation of the variables: age, number of episodes, LOS, RDI, TDI and RSS.

Table 1. Disease-wise Descriptive Statistics of 146 Patients of data set 1 with Mental Disorders giving Total number of cases in each category, Minimum, Maximum, Range, Mean and Standard Deviation of Age at Diagnosis, Number of Episodes, Length of Stay(LOS), Duration of Recent Illness(RDI), Total Duration of Illness(TDI) and Rating Scale Score(RSS)

Mental Disorders	Statistic	Age (years)	Number of Episodes	LOS (days)	RDI (days)	TDI (months)	RSS (out of 100)
BPAD	N	31	31	31	31	31	31
	Minimum	24	0	2	1.7260	0.1644	6.6667
	Maximum	70	8	70	600	420	80
	Range	46	8	68	598.2739	419.8356	73.3333
	Mean	45.06	3.0645	14.1935	65.7653	116.2311	41.4163
	Standard Deviation	12.920	1.6519	12.5044	115.9347	100.9325	17.8629
SCHIZOPHRENIA	N	73	73	73	73	73	73
	Minimum	20	0	1	1.7260	0.1644	4.4776
	Maximum	63	16	70	2880	420	80
	Range	43	16	69	2878.2739	419.8356	75.5223
	Mean	41.71	2.6438	17.2739	207.6195	93.3584	34.4201
	Standard Deviation	9.139	2.0908	11.1184	444.5852	92.8724	12.0678
DEPRESSION	N	13	13	13	13	13	13
	Minimum	30	1	3	1.7260	1	14.9254
	Maximum	62	5	58	165	420	48.4848
	Range	32	4	55	163.2739	419	33.5595
	Mean	42.62	2.0769	18.0769	37.9789	93.5384	36.1694
	Standard Deviation	11.177	1.1875	13.9908	45.5501	114.2465	11.7681
OTHERS	N	29	29	29	29	29	29
	Minimum	20	0	2	1.7260	0.1644	4.4776
	Maximum	56	4	68	900	420	80
	Range	36	4	66	898.2739	419.8356	75.5224
	Mean	38	1.7241	13.7241	78.4733	92.4083	29.6789
	Standard Deviation	9.301	1.0315	12.5780	173.2734	100.5348	18.2752
TOTAL	N	146	146	146	146	146	146
	Minimum	20	0	1	1.7260	0.1644	4.4776
	Maximum	70	16	70	2880	420	80
	Range	50	16	69	2878.2739	419.8356	75.5224
	Mean	41.77	2.50	15.9863	136.7424	98.0422	35.1196
	Standard Deviation	10.406	1.8089	11.9758	334.7482	97.5321	15.1337

Data of patients with mental disorders

Dataset 1 (146 patients)

Dataset 2 (21 patients)

Variables Known:
➢ Length of Stay (LOS)
➢ Recent Duration of Illness (RDI)
➢ Rating Scale Score (RSS)
➢ Total Duration of Illness (TDI)

For 9 patients
➢ Variables Known:
  ✓ RDI
  ✓ RSS
  ✓ TDI
➢ Variable Unknown: LOS

For 12 patients
➢ Variables Known:
  ✓ RDI
  ✓ RSS
➢ Variable Unknown: TDI

Information from dataset 1 is used to estimate unknown variables in dataset 2

AIC is used to estimate the most appropriate distribution of LOS, RDI, RSS & TDI

➢ LOS ~ Gamma ($\hat{\gamma}$=1.7084 , $\hat{\lambda}$=9.8205) ,
➢ RDI ~ Lognormal ($\hat{\mu}$=3.5016, $\hat{\sigma}$=1.4747)
➢ RSS ~ Normal ($\hat{\mu}$=34.8035,$\hat{\sigma}$=14.8783) ,
➢ TDI~Gamma ($\hat{\gamma}$=0.8692, $\hat{\lambda}$=117.8032)

Transformations are applied to convert random variables to normal variates

Transformations Used:
➢ *LOS = $\sqrt{LOS}$
➢ *RDI = ln (RDI)
➢ *RSS = RSS
➢ *TDI = $\sqrt{TDI}$

Using Mardia's test multivariate normality is established

Application in Dataset 1

*Estimation in Dataset 2

1. LOS given RDI, RSS, TDI by Four variate Normal distribution model
2. LOS given RDI, RSS by Three variate Normal distribution model
3. RDI given LOS, RSS by Three variate Normal distribution model
4. TDI given RDI by Bivariate Normal distribution model

1. LOS given RDI, RSS, TDI by Four variate Normal distribution model
2. TDI given RDI, RSS by Three variate Normal distribution model
3. TDI given RDI by Bivariate Normal distribution model
*Estimation of unknown variables in dataset 2 is done after validation of models on dataset 1with simulation study

Figure 1.   Algorithm depicting modelling and its application

Data set 2 consists of incomplete data of 21 patients with mental disorders, collected from the Department of Psychiatry, Lady Hardinge Medical College & Smt. S.K, Hospital, New Delhi. It includes patients whose assessment reports are incomplete in some respect, either the LOS in the hospital of the patient is unknown or the total duration (TDI) of disease is not available. Since some patients discontinued the treatment, only partial information is available for them and thus, LOS for such patients is unknown. The reason for unspecified TDI could be attributed to non availability of previous medical records or accompaniment of patient with strangers or a family member who had no information about the previous history of patient. The patients who went LAMA (Left against medical advice) or absconded the hospital are also included in this category. Out of 21 patients, the data for 12 patients did not consist of total duration of disease while for 9 patients length of stay in the hospital is found to be missing.

For both the cases the other variables are known. Out of 21 patients, 38.1 percent (8) are males while 69.1 percent (13) are female patients. Out of these 8 male patients, 62.5 percent (5) patients are observed to be diagnosed with schizophrenia while the rest 37.5 percent (3) patients are diagnosed with mental disorders falling in others category.

No male patient is found to be diagnosed with BPAD and depression in data set 2. Out of 13 female patients, 30.77 percent (4) patients are observed to be diagnosed with schizophrenia, BPAD and disorders falling in the others category while only one patient is observed to be diagnosed with depression. Table 2 represents the descriptive statistics of 21 patients belonging to data set 2 diagnosed with mental and behavioural disorders giving total number, minimum, maximum, range and mean of the variables: age, number of episodes, LOS, RDI, TDI and RSS. Figure 1 represents a brief algorithm of the modelling applied for estimation of unknown variables.

Table 2. Disease-wise Descriptive Statistics of 21 Patients of data set 2 with Mental Disorders giving Total number of cases in each category, Minimum, Maximum, Range and Mean of Age at Diagnosis, Number of Episodes, Length of Stay(LOS), Duration of Recent Illness(RDI), Total Duration of Illness(TDI) and Rating Scale Score(RSS)

Mental Disorders	Statistic	Age (years)	Number Of Episodes	LOS (days)	RDI (days)	TDI (months)	RSS (out of 100)
BPAD	N	4	4	2	4	2	4
	Minimum	29	1	7	15	95	38.8889
	Maximum	39	6	70	250	180	65
	Range	10	5	63	235	85	26.1111
	Mean	34	3.5	38.50	112.722	137.5000	51.9445
SCHIZOPHRENIA	N	9	9	5	9	4	9
	Minimum	24	1	1	7	6	15.9420
	Maximum	47	3	38	60	120	57.0833
	Range	23	2	37	53	114	41.1413
	Mean	42.33	2	22.40	29	46	34.0916
DEPRESSION	N	1	1	1	1	1	1
	Minimum	30	1	Estimated	11	12	45.4545
	Maximum	30	1		11	12	45.4545
	Range	0	0		0	0	0
	Mean	30	1		11	12	45.4545
OTHERS	N	7	7	5	7	2	7
	Minimum	37	1	2	3	48	14.9254
	Maximum	54	4	20	165	54	47.50
	Range	17	3	18	162	6	32.5746
	Mean	40	2.29	10.80	67.2857	51	35.9123
TOTAL	N	21	21	12	21	9	21
	Minimum	24	1	1	3	0.1644	14.9254
	Maximum	54	6	70	250	180	65
	Range	30	5	69	247	179.8356	50.0746
	Mean	39.38	2.33	20.25	50.8571	53.1294	38.6401

3.1. Distribution Selection for LOS, RDI, RSS and TDI

For testing the assumption of underlying distribution of a random variable most commonly used techniques include use of P-P plots and AIC. While plots give a rough estimate of the distribution of random variable, AIC, on the other hand provides more specific comparative results for model selection. In this article we have utilized AIC to determine the most appropriate distribution for the random variable. From among the various distributions applied to the variable LOS, Gamma distribution is observed to have minimum AIC value of 529.202, suggesting that Gamma distribution with parameters $\hat{\gamma} = 1.7084$ and $\hat{\lambda} = 9.8205$ is the most appropriate distribution for LOS. For RDI, lognormal distribution is found to have the minimum AIC value of 767.502. Thus, lognormal distribution with parameters $\hat{\mu} = 3.5016$ and $\hat{\sigma} = 1.4747$ is most suitable distribution for RDI. Since normal distribution gives minimum AIC value of 595.12 for the variable RSS, normal distribution with parameters $\hat{\mu} = 34.8035$ and $\hat{\sigma} = 14.8783$ is the most accurate distribution for RSS. And for TDI, Gamma distribution has minimum AIC value of 803.578,

indicating that Gamma distribution with parameters $\hat{\gamma} = 0.8692$ and $\hat{\lambda} = 117.8032$ is the most appropriate distribution for TDI. Table 3 presents the AIC values for different distribution tested for variables LOS, RDI, RSS and TDI along with the maximum likelihood estimations of distributions selected.

3.2. Transformations Applied on LOS, RDI and TDI

Since the distributions selected for LOS, RDI and TDI are not normal, power transformations are applied to them to follow marginal normal distribution approximately. Square root transformation is applied on LOS and TDI as mean values of LOS and TDI are greater than 1 while RDI is transformed using logarithm transformation. The resulting transformed variables *LOS, *RDI and *TDI approximately follow normal distribution as is apparent from graphs of *LOS, *RDI, RSS and *TDI presented in Figure 2, 3, 4 and 5. Hence from Figure 2, 3, 4 and 5 it can be concluded that all the four random variables viz. *LOS, *RDI, RSS and *TDI marginally tend to normal distribution.

Table 3. Akaike Information Criterion (AIC) Values of Different Distributions for Length of Stay (LOS), Duration of Recent Illness (RDI), Rating Scale Score (RSS) and Total Duration of Illness (TDI)

Variables	Distribution	AIC Values	Selected Distribution	MLE of Parameters
Length of Hospital Stay (LOS)	Normal	574.854	GAMMA	$(\hat{\gamma} = 1.7084, \hat{\lambda} = 9.8205)$
	Lognormal	533.018		
	Gamma	529.202		
	Exponential	552.208		
	Weibull	537.438		
	Rayleigh	551.978		
Duration of Recent Illness (RDI)	Normal	914.356	LOGNORMAL	$(\hat{\mu} = 3.5016, \hat{\sigma} = 1.4747)$
	Lognormal	767.502		
	Gamma	780.296		
	Exponential	790.718		
	Weibull	775.706		
	Rayleigh	1005.178		
Rating Scale Score (RSS)	Normal	595.12	NORMAL	$(\hat{\mu} = 34.8035, \hat{\sigma} = 14.8783)$
	Lognormal	624.288		
	Gamma	607.336		
	Exponential	657.16		
	Weibull	599.94		
	Rayleigh	600.406		
Total Duration of Illness (TDI)	Normal	864.364	GAMMA	$(\hat{\gamma} = 0.8692, \hat{\lambda} = 117.8032)$
	Lognormal	842.534		
	Gamma	803.578		
	Exponential	816.55		
	Weibull	814.302		
	Rayleigh	899.69		

Probability Distribution Plot of *Length Of Stay in hospital (LOS)
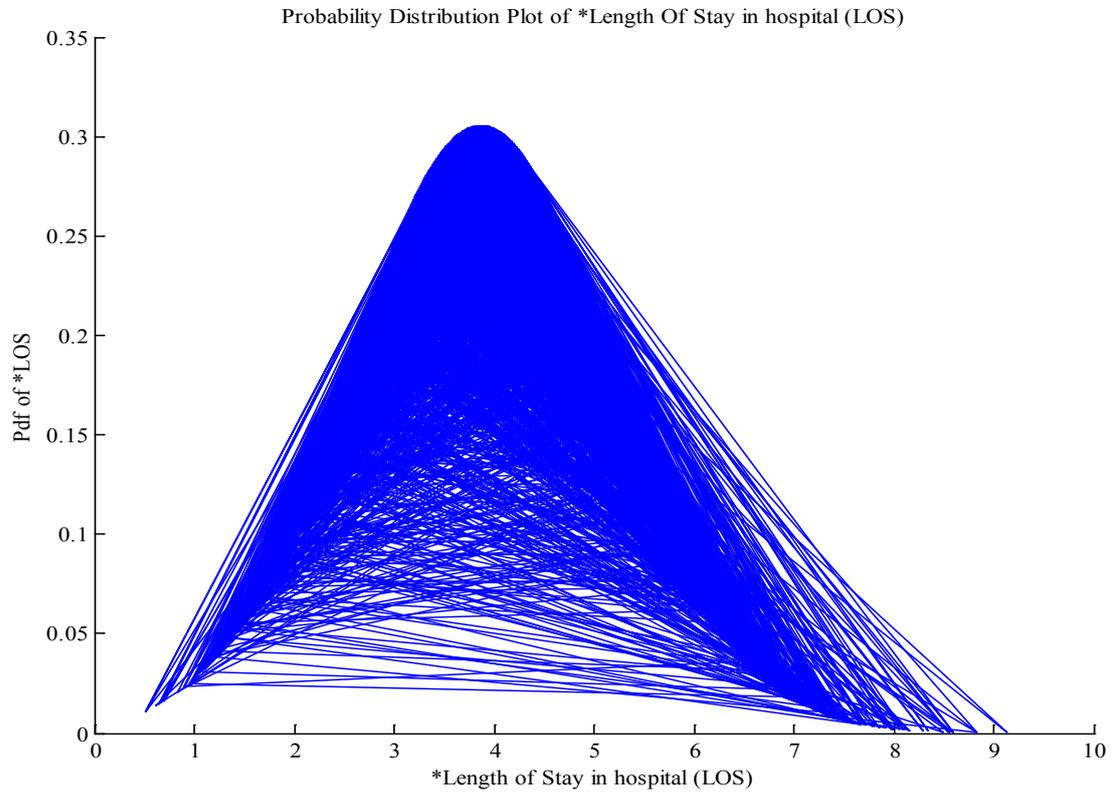
Figure 2. Normal approximation of * Length of Stay in hospital (LOS)

Probability Distribution Plot of *Duration of Recent Illness (RDI)
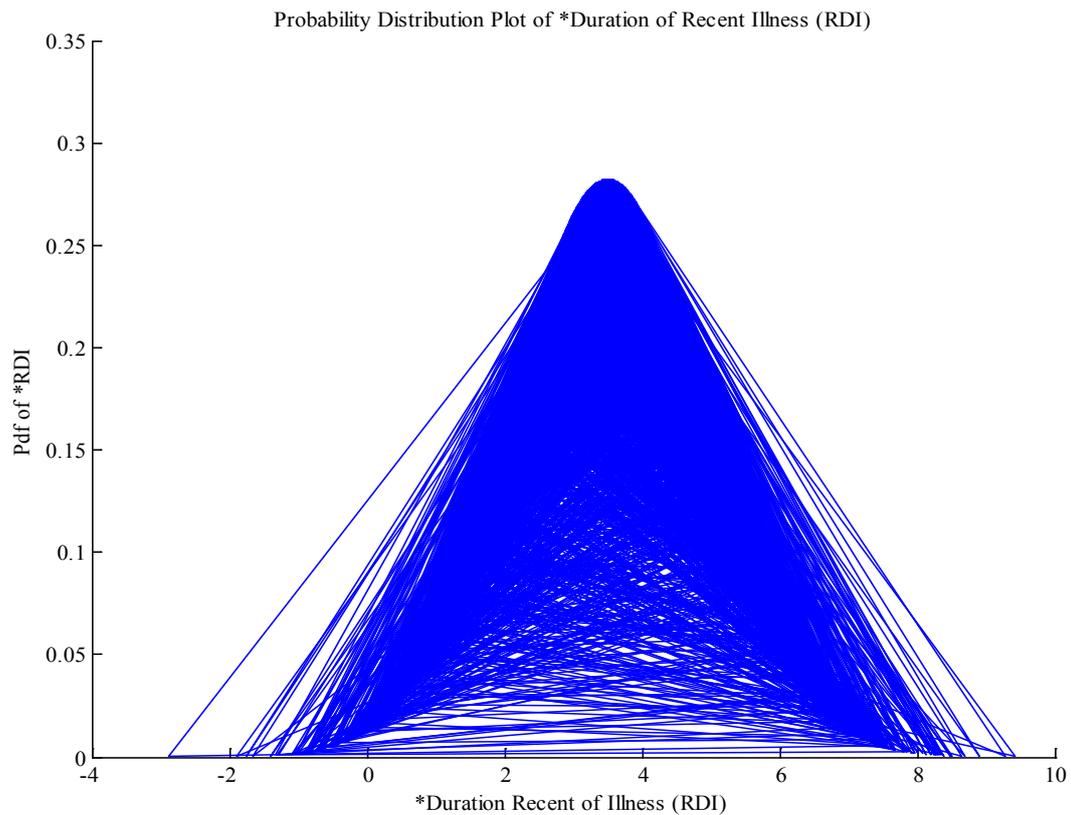
Figure 3. Normal approximation of * Recent Duration of Illness (RDI)

Probability Distribution Plot of Rating Scale Score (RSS)
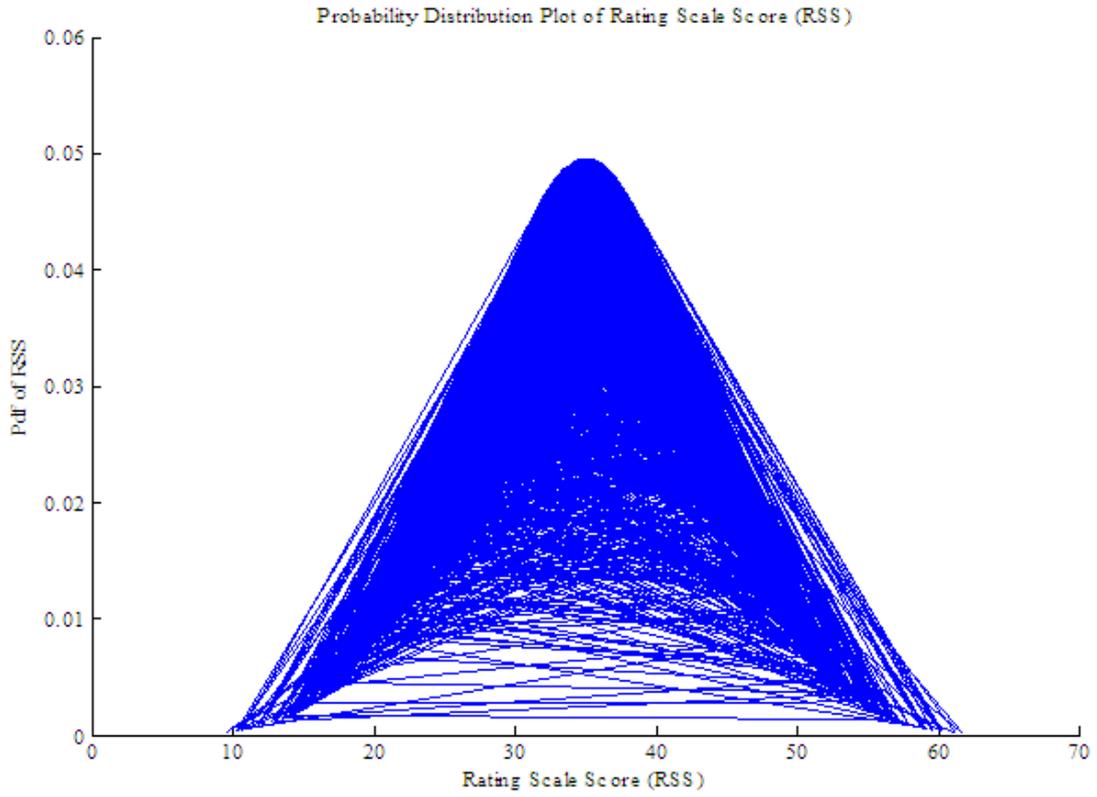
Figure 4. Normal approximation of Rating Scale Score (RSS)

Probability Distribution Plot of *Total Duration of Illness (TDI)
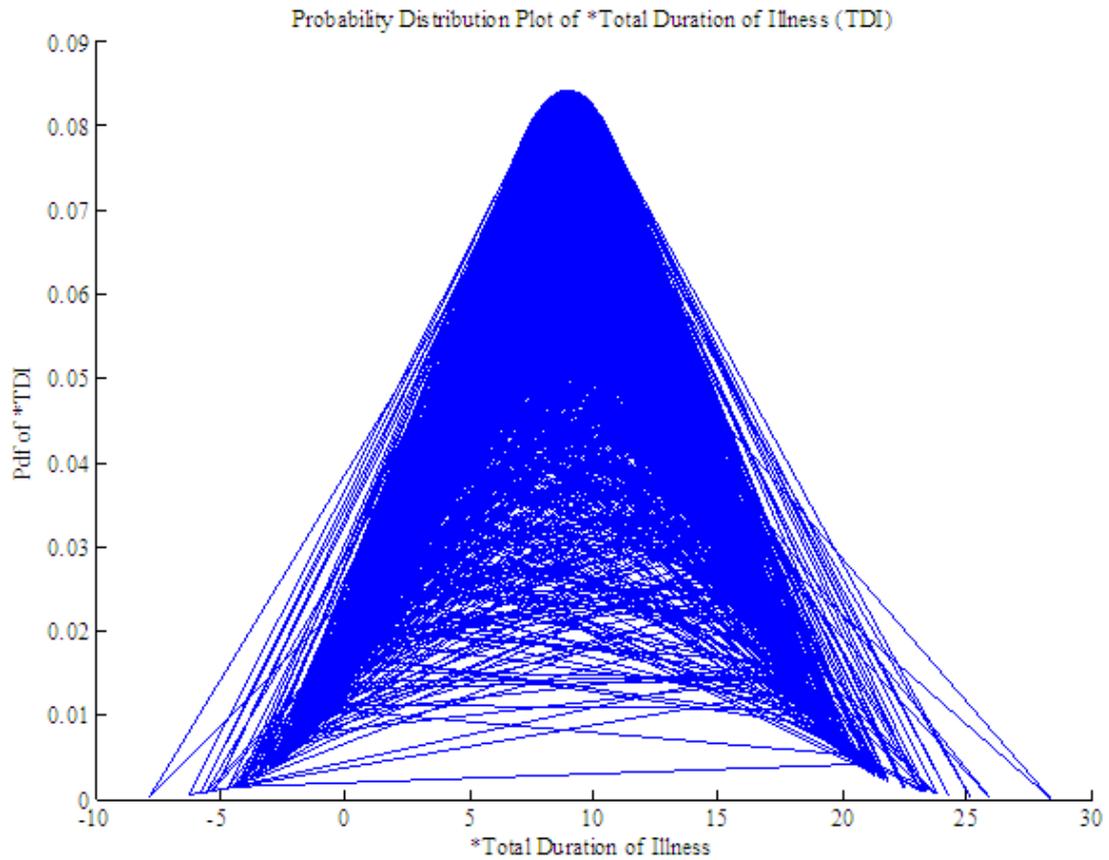
Figure 5. Normal approximation of *Total Duration of Illness (TDI)

3.3. Checking Multivariate Normality of *LOS, *RDI, RSS and *TDI

Next, we have verified multivariate normality of *LOS, *RDI, RSS and *TDI by applying Mardia's test. Mardia's measure for skewness and kurtosis are calculated to be $\kappa_1 = 7.8287$ (at 20 degrees of freedom, p-value = $0.6455 > 0.05$) and $\kappa_2 = -0.8776$ (p-value = 0.3802 > 0.05), respectively, connoting that *LOS, *RDI, RSS and *TDI are jointly normally distributed. Thus, MVN distribution can be applied to the variables *LOS, *RDI, RSS and *TDI since they satisfy the basic assumption of multivariate normality. The resulting transformed variables are denoted as X_1=*LOS , X_2=*RDI , X_3=RSS and X_4=*TDI .

3.4. Application of Four Variate Normal Distribution

This section highlights the application of MVN distribution. The joint distribution of variables X_1, X_2, X_3 and X_4 is observed to be MVN distribution, defined as,

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} \sim MVN \left(\underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix} \right) \quad (12)$$

Where, $-\infty < (\underset{\sim}{x}, \underset{\sim}{\mu}) < \infty, \sigma_{ii} > 0$ and $-\infty < \sigma_{ij} < \infty$

3.4.1. Estimation of Mean LOS for 146 Patients of Dataset 1

In this section, we have estimated mean LOS for 146 Patients of dataset 1 by applying four variate normal distribution.

The first dataset consists of complete data of 146 patients with mental disorders for whom the variables LOS, RDI, RSS and TDI are known. The primary objective of the study is to estimate the LOS, RDI and TDI for these patients by applying MVN distribution. To accomplish this, firstly a random sample of size 5000 is generated from MVN distribution defined in equation (12) and then used to estimate the LOS for different intervals of X_1 (transformed length of stay in the hospital) defined as $X_1 \le 2$, $2 < X_1 \le 3$,...., $X_1 > 8$. The mean LOS for each interval is estimated by applying the following equation (derived in methodology section):

$$E(X_1 \mid X_2, X_3, X_4) = \mu_1 - \frac{1}{a_{11}} \sum_{j=2}^{4} a_{1j}(X_j - \mu_j) \quad (13)$$

Where, μ_1 represents the mean value of X_1 calculated from simulated data corresponding to a specific interval of X_1 , X_j (j=2, 3, 4) is the observed value of variables X_2, X_3 and X_4 from the data obtained after applying the respective transformations on variables RDI, RSS and TDI,

corresponding to a specific interval while μ_j (j=2,3,4) is the mean value of X_2, X_3 and X_4 calculated from the generated data corresponding to the specific interval. a_{11}, a_{1j} (j=2,3,4) are elements of first row of matrix $A = \Sigma^{-1}$ calculated from the generated data corresponding to the specific interval, where, $\Sigma = \left[\sigma_{ij} \right]_{4X4}$, σ_{ij} being covariance between X_i and $X_j (i, j = 1, 2, 3, 4)$. The procedure of calculation for the first interval $X_1 \le 2$ is illustrated below:

1. Firstly, the observed data of 146 patients on variables LOS, RDI, RSS and TDI is transformed to X_1, X_2, X_3 and X_4 , where X_1=*LOS $= \sqrt{LOS}$, X_2=*RDI=ln (RDI), X_3=RSS and X_4=*TDI$= \sqrt{TDI}$.

2. The mean value of X_1 is calculated from the data of 146 patients for whom X_1 is less than or equal to 2 and is found to be 1.2866, while for the same range the mean value of X_1 computed from simulated data was found to be 1.6147.

3. The values of variables X_2, X_3 and X_4 for patients with $X_1 \le 2$ lied in the ranges $1.0986 \le X_2 \le 2.7081$ with mean value 2.0766, $14.9254 \le X_3 \le 38.3333$ with mean value 25.2712 and $4.8990 \le X_4 \le 15.4919$ with mean value 9.8211.

4. The mean vector $\underset{\sim}{\mu}$, variance-covariance matrix Σ and matrix $A = \Sigma^{-1} = \left[a_{ij} \right]_{4X4}$ of X_1 , X_2 , X_3 and X_4 calculated from the simulated data corresponding to the ranges of observed X_1, X_2, X_3 and X_4 are

$$\underset{\sim}{\mu} = \begin{pmatrix} 1.6147 \\ 3.3426 \\ 35.0839 \\ 9.4375 \end{pmatrix} \Sigma = \begin{bmatrix} 0.0906 & 0.0127 & -0.0405 & 0.1047 \\ 0.0127 & 2.0220 & -0.5478 & -0.3946 \\ -0.0405 & -0.5478 & 66.5119 & 0.6867 \\ 0.1047 & -0.3946 & 0.6867 & 23.5494 \end{bmatrix}$$

and

$$A = \Sigma^{-1} = \left[a_{ij} \right]_{4X4} = \begin{bmatrix} 11.1143 & -0.0781 & 0.0066 & -0.0509 \\ -0.0781 & 0.4978 & 0.0040 & 0.0086 \\ 0.0066 & 0.0040 & 0.0151 & -0.0004 \\ -0.0509 & 0.0086 & -0.0004 & 0.0428 \end{bmatrix}.$$

5. Conditional expectation of $X_1 \mid X_2, X_3, X_4$ is obtained by substituting the above values in equation (13), which is calculated to be 1.5022.

6. The mean LOS (untransformed) in hospital for patients with $X_1 \le 2$ is then obtained by squaring the value of $E(X_1 \mid X_2, X_3, X_4)$ which comes out to be 2.2566 days.

Following the above procedure, the mean LOS for all intervals are estimated and presented in Table 4. Figure 5 shows the comparison of observed and estimated LOS given RDI, RSS and TDI of 146 patients of dataset 1 for model derived in this section.

Table 4. Estimated Mean Length of Stay (LOS) given Recent Duration of Illness (RDI), Rating Scale Score (RSS) & Total Duration of Illness (TDI) for 146 Patients with Mental Disorders for Different intervals of *LOS using simulated data of Size 5000 from MVN Distribution (X_1=*LOS=$\sqrt{LOS}$, X_2=*RDI=ln(RDI), X_3=RSS and X_4=*TDI=$\sqrt{TDI}$)

Interval	Real Data				Simulated Data				$E[X_1 / X_2, X_3, X_4]$	Observed LOS (days)	Estimated LOS (days)
	Mean X_1	Mean X_2	Mean X_3	Mean X_4	Mean X_1	Mean X_2	Mean X_3	Mean X_4			
$X_1 \leq 2$	1.2866	2.0766	25.2712	9.8211	1.6148	3.3426	35.0840	9.4375	1.5022	1.6553	2.2566
$2 < X_1 \leq 3$	2.6003	3.3144	30.9549	7.7400	2.5748	3.5320	34.6111	9.1376	2.5542	6.7617	6.5241
$3 < X_1 \leq 4$	3.5360	3.8995	28.4699	9.0170	3.5029	3.4785	35.2394	9.0114	3.4732	12.5030	12.0633
$4 < X_1 \leq 5$	4.3991	3.4582	42.8994	10.0268	4.4458	3.5022	34.6387	9.0719	4.4859	19.3517	20.1231
$5 < X_1 \leq 6$	5.3254	3.4878	43.5761	8.1941	5.4192	3.5201	34.7482	8.7558	5.4883	28.3596	30.1212
$6 < X_1 \leq 7$	6.1644	1.9459	41.2698	7.7460	6.4124	3.5186	35.4025	9.1668	6.6728	38	44.5261
$7 < X_1 \leq 8$	7.6158	5.1059	40.9091	7.7460	7.3821	3.3906	34.0826	8.8284	7.2333	58	52.3203
$X_1 > 8$	8.3064	2.5053	37.2321	11.2218	8.4552	3.6432	34.8570	9.1571	8.5643	68.9964	73.3471

Comparison of observed and estimated length of stay for four variate model
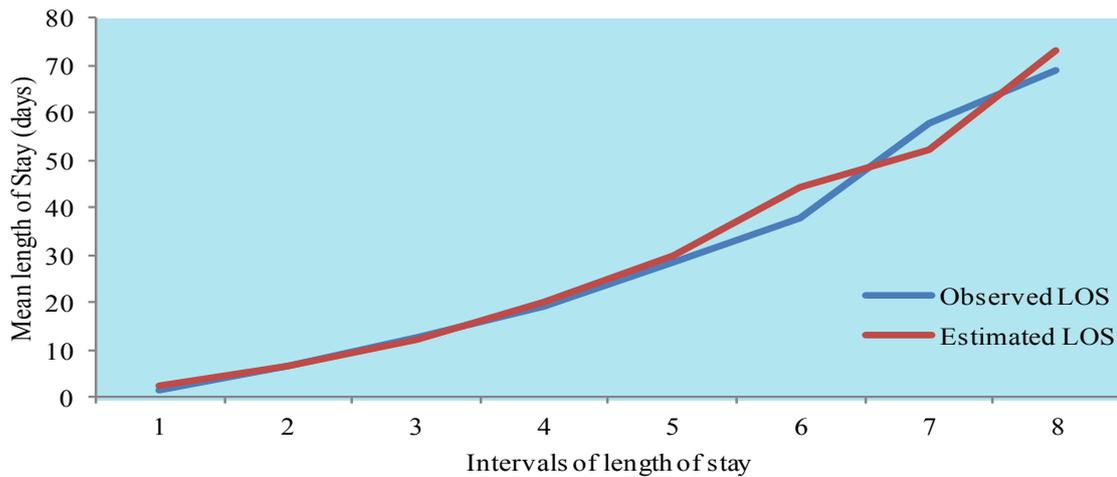
Figure 6. Comparison of observed and estimated Length of Stay (LOS) of 146 patients of dataset 1 for Four Variate Multivariate Normal Distribution Model

3.4.2. Estimation of Mean LOS for 9 Patients of Dataset 2

The LOS in hospital for 9 patients (with missing LOS) with mental disorders of dataset 2 is estimated on the basis of known values of RDI, RSS and TDI. The procedure of calculation of mean LOS for first patient with given mean values of RDI, RSS and TDI is as follows:

1. The observed values of RDI, RSS and TDI for first patient are 10 days, 15.9420 (out of 100) and 6 months, respectively, which are treated as mean values for the purpose of calculations.

2. These values of RDI, RSS and TDI are transformed as X_2=ln (RDI) , X_3=RSS and X_4=$\sqrt{TDI}$ for application of MVN distribution model. Thus, for first patient, X_2=2.3026 , X_3=15.9420 and $X_4 = 2.4495$.

3. The mean vector μ , variance-covariance matrix Σ and matrix $A = \Sigma^{-1} = \begin{bmatrix} a_{ij} \end{bmatrix}_{4X4}$ of X_1 , X_2 , X_3 and X_4 are calculated from the simulated data corresponding to the ranges of X_2, X_3 and X_4 , i.e., from simulated data, we have taken means of those observations for

which $X_2 \cong 2.3026$, $X_3 \cong 15.9420$ and $X_4 \cong 2.4495$.

4. Conditional expectation of $X_1 \mid X_2, X_3, X_4$ is obtained by substituting the above values in equation (13).

5. The mean LOS in hospital for first patient is then obtained by squaring the value of $E(X_1 \mid X_2, X_3, X_4)$ which is found to be 6.5128 days. This gives the mean length of stay in the hospital for the first patient with known RDI, RSS and TDI.

The mean LOS for all 9 patients with known RDI, RSS and TDI are estimated by applying the above procedure and are presented in Table 5.

Table 5. Mean Length of Stay (LOS) given Recent Duration of illness (RDI), Rating Scale Score (RSS) & Total Duration of Illness (TDI) for 9 Patients with Mental Disorders using a Generated Sample of Size 5000 from MVN Distribution

S. No.	RDI (days)	RSS (out of 100)	TDI (months)	Estimated LOS (days)	Disorder
1	10	15.9420	6	6.5128	Schizophrenia
2	3	34.9206	48	8.0809	Others
3	30	65.0000	180	19.1071	BPAD
4	60	37.5000	8	19.2093	Schizophrenia
5	11	45.4545	12	12.3423	Depression
6	45	27.7778	50	14.5300	Schizophrenia
7	7	41.2698	120	17.1724	Schizophrenia
8	165	40.9091	54	15.2420	Others
9	250	38.8889	95	19.9518	BPAD

3.5. Application of Three Variate Normal Distribution

In this section we have shown the application of MVN distribution with three random variables. The joint distribution of variables X_1, X_2 and X_3 is observed to be MVN distribution, defined as,

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim MVN \left(\underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \right) \quad (14)$$

Firstly we estimated LOS given RDI and RSS and RDI given LOS and RSS for 146 patients of dataset 1, both in absence of TDI and then once the model is verified, it is used to estimate LOS given RDI and RSS and TDI given RDI and RSS for 9 patients (with unknown LOS) and 12 patients (with unknown TDI) of dataset 2, respectively.

3.5.1. Estimation of Mean LOS for 146 Patients of Dataset 1

In this section, only three variable viz. LOS, RDI and RSS have been considered. To estimate LOS, firstly a random sample of size 5000 is generated from three variate distribution defined in equation (14) and then used to estimate the LOS for different intervals of X_1 (transformed length of stay in the hospital) defined as

$X_1 \le 2$, $2 < X_1 \le 3$,....,$X_1 > 8$. The mean LOS for each interval is estimated by applying the following equation (derived in methodology section):

$$E(X_1 \mid X_2, X_3) = \mu_1 - \frac{1}{a_{11}} \sum_{j=2}^{3} a_{1j}(X_j - \mu_j) \quad (15)$$

Where, μ_1 represents the mean value of X_1 calculated from simulated data corresponding to a specific interval of X_1, X_j (j=2, 3) is the observed value of variables X_2 and X_3 from the data obtained after applying the respective transformations on variables RDI and RSS, corresponding to a specific interval while μ_j (j=2,3) is the mean value of X_2 and X_3 calculated from the generated sample corresponding to the specific interval. a_{11}, a_{1j} (j=2,3) are elements of first row of matrix $A = \Sigma^{-1}$ calculated from the generated sample corresponding to the specific interval, where, $\Sigma = \left[\sigma_{ij} \right]_{3X3}$, σ_{ij} being covariance between X_i and $X_j (i, j = 1, 2, 3)$. The procedure for calculation is similar to the procedure followed for four variate case except that here three variables are considered. The steps for calculating mean LOS the interval $3 < X_1 \le 4$ are as follows:

1. Firstly, the observed data of 146 patients on variables LOS, RDI and RSS is transformed to X_1, X_2 and X_3, where X_1=*LOS $= \sqrt{LOS}$, X_2=*RDI=ln (RDI) and X_3=RSS.

2. The mean value of X_1 is calculated from the data of 146 patients for whom $3 < X_1 \le 4$ and is found to be 3.5359 while for the same range the mean value of X_1 computed from simulated data is found to be 3.5029.

3. The values of variables X_2 and X_3 for patients with $3 < X_1 \le 4$ lied in the ranges $0.6931 \le X_2 \le 6.2916$ with mean value 3.8995 and $4.4776 \le X_3 \le 55$ with mean value 28.4699.

4. The mean vector $\underset{\sim}{\mu}$, variance-covariance matrix Σ and matrix $A = \Sigma^{-1} = \left[a_{ij} \right]_{3X3}$ of X_1, X_2 and X_3 calculated from the simulated data corresponding to the ranges of observed X_1, X_2 and X_3 are

$$\underset{\sim}{\mu} = \begin{pmatrix} 3.5029 \\ 3.4785 \\ 35.2394 \end{pmatrix} \quad \Sigma = \begin{bmatrix} 0.0875 & -0.0108 & -0.0528 \\ -0.0042 & 2.0797 & 0.6366 \\ -0.0533 & 0.5333 & 62.2512 \end{bmatrix} \quad \text{and}$$

$$A = \Sigma^{-1} = \left[a_{ij} \right]_{3X3} = \begin{bmatrix} 11.4312 & 0.0569 & 0.0091 \\ 0.0201 & 0.4822 & -0.0049 \\ 0.0096 & -0.0041 & 0.0161 \end{bmatrix}.$$

5. Conditional expectation of $X_1 \mid X_2, X_3$ is obtained by substituting the above values in equation (15) which is

calculated to be 3.5080.

6. The mean LOS (untransformed) in hospital for patients with $3 < X_1 \le 4$ is then obtained by squaring the value of $E(X_1 \mid X_2, X_3)$ which comes out to be 12.3061 days.

3.5.2. Estimation of Mean RDI for 146 Patients of Dataset 1

For estimating mean RDI given LOS and RSS for 146 patients belonging to dataset 1, procedure similar to the estimation of mean LOS for 146 Patients of dataset 1, described in section 3.5.1 has been followed.

Following the above procedure, the mean LOS for all intervals are estimated and presented in Table 6.

For estimation of RDI, the variable viz. LOS and RSS have been considered. To accomplish this, firstly a random sample of size 5000 is generated from three variate distribution defined in equation (14) and then used to estimate the RDI for different intervals of X_2 (transformed recent duration of illness) defined as $X_2 \le 1$, $1 < X_2 \le 1.5$,...., $X_2 > 6$. The mean RDI for each interval is estimated by applying the following equation (derived in methodology section):

$$E(X_2 \mid X_1, X_3) = \mu_2 - \frac{1}{a_{22}}\left(a_{21}(X_1 - \mu_1) + a_{23}(X_3 - \mu_3)\right) \text{ (16)}$$

Where, μ_2 represents the mean value of X_2 calculated from simulated data corresponding to a specific interval of X_2, X_j (j=1, 3) is the observed value of variables X_1 and X_3 from the data obtained after applying the respective transformations on variables LOS and RSS, corresponding to a specific interval while μ_j (j=1,3) is the mean value of X_1 and X_3 calculated from the generated sample corresponding to the specific interval. a_{22}, a_{2j} (j=1,3) are elements of second row of matrix $A = \Sigma^{-1}$ calculated from the generated sample corresponding to the specific interval, where, $\Sigma = \left[\sigma_{ij}\right]_{3X3}$, σ_{ij} being covariance between X_i and $X_j (i, j = 1, 2, 3)$. The procedure for calculation is similar to the procedure described in section 3.5.1 for estimation of mean LOS for 146 Patients of dataset 1. The steps for calculating mean RDI for the first interval $X_2 \le 1$ are as follows:

1. Firstly, the observed data of 146 patients on variables LOS, RDI and RSS is transformed to X_1, X_2 and X_3, where X_1=*LOS=$\sqrt{LOS}$, X_2=*RDI=ln (RDI) and X_3=RSS.

2. The mean value of X_2 is calculated from the data of 146 patients for whom $X_2 \le 1$ and is 0.7121 while for the same range the mean value of X_2 computed from simulated data is 0.4565.

3. The values of variables X_1 and X_3 for patients with $X_2 \le 1$ lied in the ranges $2.2361 \le X_1 \le 5.3852$ with mean value 3.9247 and $17.9104 \le X_3 \le 56.6667$ with mean value 37.9625.

4. The mean vector μ, variance-covariance matrix Σ and matrix $A = \Sigma^{-1} = \left[a_{ij}\right]_{3X3}$ of X_1, X_2 and X_3 calculated from the simulated data corresponding to the ranges of observed X_1, X_2 and X_3 are

$$\mu = \begin{pmatrix} 3.8526 \\ 0.4565 \\ 34.6876 \end{pmatrix} \quad \Sigma = \begin{bmatrix} 1.8734 & 0.0021 & -0.3756 \\ 0.0021 & 0.2424 & 0.1494 \\ -0.3756 & 0.14946 & 0.8407 \end{bmatrix} \quad \text{and}$$

$$A = \Sigma^{-1} = \left[a_{ij}\right]_{3X3} = \begin{bmatrix} 0.5345 & -0.0066 & 0.0033 \\ -0.0066 & 4.1315 & -0.0102 \\ 0.0033 & -0.0102 & 0.0165 \end{bmatrix}.$$

5. Conditional expectation of $X_2 \mid X_1, X_3$ is obtained by substituting the above values in equation (16) which is calculated to be 0.4647.

6. The mean RDI (untransformed) for patients with $X_2 \le 1$ is then obtained by squaring the value of $E(X_2 \mid X_2, X_3)$ which comes out to be 0.2159 days.

Following the above procedure, the mean RDI for all intervals are estimated and presented in Table 7.

3.5.3. Estimation of Mean TDI for 12 Patients of Dataset 2

Section 3.5.1 and 3.5.2 highlight the application of three variate normal distribution model and from Tables 6 and 7 and Figure 7 it has been verified that three variate normal distribution model provides estimates very close to the observed values. Thus, it can now be applied to estimate TDI in hospital for 12 patients (with unknown TDI) with mental disorders of dataset 2 on the basis of known values of RDI and RSS. The joint distribution of variables X_2, X_3 and X_4 is observed to be MVN distribution, defined as,

$$\begin{pmatrix} X_2 \\ X_3 \\ X_4 \end{pmatrix} \sim MVN \left(\mu = \begin{pmatrix} \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}, \Sigma = \begin{bmatrix} \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix} \right) \text{ (17)}$$

The mean TDI for each interval is estimated by applying the following equation (derived in methodology section):

$$E(X_4 \mid X_2, X_3) = \mu_4 - \frac{1}{a_{44}} \sum_{j=2}^{3} a_{4j}(X_j - \mu_j) \text{ (18)}$$

Where, μ_4 represents the mean value of X_4 calculated from simulated data corresponding to a specific interval of X_4, X_j (j=2, 3) is the observed value of variables X_2 and X_3 from the data obtained after applying the respective transformations on variables RDI, RSS and TDI, corresponding to a specific interval while μ_j (j=2,3) is the

mean value of X_2 and X_3 calculated from the generated sample corresponding to the specific interval. a_{44}, a_{4j} (j=2,3) are elements of last row of matrix $A = \Sigma^{-1}$ calculated from the generated sample corresponding to the specific interval, where, $\Sigma = \left[\sigma_{ij} \right]_{3X3}$, σ_{ij} being covariance between X_i and $X_j (i,j = 2,3,4)$.

Table 6. Estimated Mean Length of Stay (LOS) given Recent Duration of Illness (RDI) and Rating Scale Score (RSS) for 146 Patients with Mental Disorders for Different intervals of X_1 using a Generated Sample of Size 5000 from MVN Distribution (X_1=*LOS=$\sqrt{LOS}$, X_2=*RDI=ln(RDI) and X_3=RSS)

| Interval	Original Data			Simulated Data			$E[X_1 / X_2, X_3]$	Observed LOS (days)	Estimated LOS (days)
	Mean X_1	Mean X_2	Mean X_3	Mean X_1	Mean X_2	Mean X_3			
$X_1 \leq 2$	1.2866	2.0766	25.2712	1.6148	3.3426	35.0840	1.6125	1.6553	2.6001
$2 < X_1 \leq 3$	2.6003	3.3143	30.9549	2.5748	3.5320	34.6111	2.5750	6.7617	6.6306
$3 < X_1 \leq 4$	3.5360	3.8995	28.4699	3.5029	3.4785	35.2394	3.5080	12.5030	12.3059
$4 < X_1 \leq 5$	4.3991	3.4582	42.8994	4.4458	3.5022	34.6387	4.4416	19.3517	19.7279
$5 < X_1 \leq 6$	5.3254	3.4878	43.5761	5.4192	3.5201	34.7482	5.3916	28.3596	29.0693
$6 < X_1 \leq 7$	6.1644	1.9459	41.2698	6.4124	3.5186	35.4025	6.4046	38	41.0189
$7 < X_1 \leq 8$	7.6158	5.1059	40.9091	7.3821	3.3906	34.0823	7.4292	58	55.1924
$X_1 > 8$	8.3064	2.5053	37.2321	8.4552	3.6432	34.8570	8.3835	68.9964	70.2830
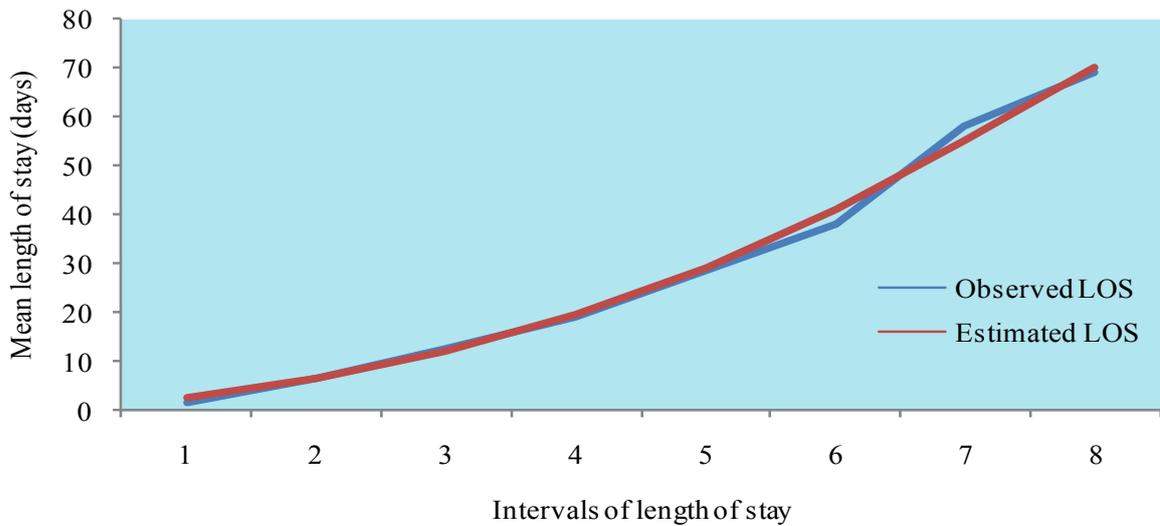
Figure 7. Comparison of observed and estimated Length of Stay (LOS) of 146 patients of dataset 1 for Three Variate Normal Distribution Model

Table 7. Estimated Mean Recent Duration of Illness (RDI) given Length of Stay (LOS) and Rating scale Score (RSS) for 146 Patients with Mental Disorders for Different intervals of X_2 using a Generated Sample of Size 5000 from MVN Distribution (X_1=*LOS=$\sqrt{LOS}$, X_2=*RDI=ln(RDI) and X_3=RSS)

Interval	Original Data			Simulated Data			$E[X_2/X_1,X_3]$	Observed RDI (days)	Estimated RDI (days)
	Mean X_2	Mean X_1	Mean X_3	Mean X_2	Mean X_1	Mean X_3			
$X_2 \leq 1$	0.7121	3.9247	37.9624	0.4565	3.8526	34.6876	0.46467	2.0383	1.5915
$1 < X_2 \leq 1.5$	1.2424	3.5118	47.6087	1.2752	3.7289	35.2718	1.2731	3.4641	3.5719
$1.5 < X_2 \leq 2$	1.8005	4.4346	36.9784	1.7692	3.9141	35.3633	1.7712	6.0524	5.8781
$2 < X_2 \leq 2.5$	2.2158	2.8099	39.9311	2.2671	3.9096	34.7950	2.2641	9.1688	9.6225
$2.5 < X_2 \leq 3$	2.6716	2.6043	30.0000	2.7636	3.7998	34.4196	2.7666	14.4629	15.9047
$3 < X_2 \leq 3.5$	3.3735	4.0236	32.8323	3.2655	3.8901	34.8065	3.2647	29.1819	26.1718
$3.5 < X_2 \leq 4$	3.8067	4.2320	35.3472	3.7383	3.9190	34.7870	3.7404	44.9999	42.1139
$4 < X_2 \leq 4.5$	4.2261	4.1376	35.6695	4.2359	3.9142	35.0296	4.2365	68.4518	69.1631
$4.5 < X_2 \leq 5$	4.7875	3.5765	27.1710	4.7307	3.7973	35.1801	4.7304	119.9998	113.3424
$5 < X_2 \leq 5.5$	5.1456	4.8227	38.3568	5.2329	3.7750	34.4945	5.2285	171.6796	186.5059
$5.5 < X_2 \leq 6$	5.7135	3.6988	32.5754	5.7313	3.8934	35.8172	5.7300	302.9264	307.9816
$X_2 > 6$	6.2723	3.6961	28.6096	6.6060	3.9642	34.4956	6.6015	529.7102	736.2060

The procedure of calculation of TDI for first patient with given mean values of RDI and RSS is as follows:

1. The observed mean values of RDI and RSS for first patient are 15 days and 38.8889 (out of 100) respectively, which are treated as mean values for the purpose of calculations.
2. These values of RDI and RSS are transformed as X_2=ln (RDI) and X_3=RSS for application of MVN distribution model, i.e., for first patient, X_2=2.7081 and X_3=38.8889.
3. The mean vector $\underset{\sim}{\mu}$, variance-covariance matrix Σ and matrix $A = \Sigma^{-1} = \left[a_{ij} \right]_{3X3}$ of X_2, X_3 and X_4 are calculated from the simulated data corresponding to the ranges of observed X_2, X_3 and X_4, i.e., from simulated data, we have taken means of those observations for which $X_2 \cong 2.7081$ and $X_3 \cong 38.8889$.
4. Conditional expectation of $X_4 \mid X_2, X_3$ is obtained by substituting the above values in equation (18). The mean TDI in hospital for first patients is then obtained by squaring the value of $E(X_4 \mid X_2, X_3)$ which comes out to be 77.6514 months. This gives the mean total duration of illness for the first patient with known RDI and RSS.

The mean TDI for 12 patients with known RDI and RSS are estimated by applying the above procedure and are presented in Table 8.

Table 8. Estimated Mean Total Duration of Illness (TDI) given Recent Duration of Illness (RDI) and Rating scale Score (RSS) for 12 Patients with Mental Disorders using a Generated Sample of Size 5000 from MVN Distribution (X_2=*RDI=ln(RDI), X_3=RSS and X_4=*TDI=$\sqrt{TDI}$)

S. No.	RDI (days)	RSS (Out of 100)	Estimated TDI (months)
1	15	38.8889	77.6514
2	30	37.3016	82.5666
3	10	15.9420	76.6515
4	60	37.5000	66.2824
5	30	57.0833	103.8282
6	12	14.9254	75.2127
7	11	34.9206	76.4758
8	160	40.9091	89.9886
9	90	47.5000	102.6534
10	9	41.2698	84.1770
11	30	32.5397	88.2094
12	30	65.0000	110.3265

3.6. Application of BVN Distribution

This section highlights the application of BVN distribution. In this section, the application of BVN distribution has been examined for two variables viz. TDI and RDI. The joint distribution of variables X_2 and X_4 is

observed to be BVN distribution, defined as,

$$\begin{pmatrix} X_2 \\ X_4 \end{pmatrix} \sim MVN \left(\underset{\sim}{\mu} = \begin{pmatrix} \mu_2 \\ \mu_4 \end{pmatrix}, \Sigma = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{42} & \sigma_{44} \end{bmatrix} \right) \quad (19)$$

Firstly, TDI given RDI is estimated for 146 patients of dataset 1 with complete record history, in absence of LOS and RSS. Once the model is verified, it is used to estimate TDI given RDI for 12 patients (with unknown TDI) of dataset 2, respectively.

3.6.1. Estimation of Mean TDI for 146 Patients of Dataset 1

In this section, only two variable viz. RDI and TDI have been considered. For estimating mean TDI, firstly a random sample of size 5000 is generated from BVN distribution defined in equation (19) and then used to estimate the TDI for different intervals of X_4 (transformed total duration of illness) defined as $X_4 \le 2.5$, $2.5 < X_4 \le 4.5$,...., $X_4 > 18.5$. The mean TDI for each interval is estimated by applying the following equation (derived in methodology section):

$$E(X_4 \mid X_2) = \mu_4 - \frac{1}{\sigma_{44}} \sigma_{24} (X_2 - \mu_2) \quad (20)$$

Where, X_2 is the observed value of variable X_2 from the data obtained after applying the respective transformations on variables RDI and TDI corresponding to a specific interval while $\mu_i (i = 2, 4)$, the mean value of X_i (i=2,4), σ_{44}, variance of X_4 and σ_{24}, the covariance

between X_2 and X_4 are calculated from simulated data corresponding to a specific interval of X_4. The method for calculating mean TDI for the last interval $X_4 > 18.5$ is as follows:

1. Firstly, the observed data of 146 patients on variables RDI and TDI is transformed to X_2 and X_4, by using the transformation X_2=*RDI=ln (RDI) and X_4=*TDI=$\sqrt{\text{TDI}}$.

2. The mean value of X_4 calculated from the data of 146 patients for whom X_4 is greater than 18.5 is found to be 20.4939.

3. The mean vector $\underset{\sim}{\mu}$ and variance-covariance matrix Σ of X_2 and X_4 calculated from the simulated data corresponding to the ranges of observed X_2 and X_4 are found to be $\underset{\sim}{\mu} = \begin{pmatrix} 3.5345 \\ 20.4724 \end{pmatrix}$ $\Sigma = \begin{bmatrix} 1.9828 & -0.0878 \\ -0.0878 & 3.0171 \end{bmatrix}$.

4. Conditional expectation of $X_4 \mid X_2$ obtained by substituting the above values in equation (20) is calculated as 20.5137.

5. The mean TDI (untransformed) for patients with $X_4 > 18.5$ obtained by squaring the value of $E(X_4 \mid X_2)$ is 420.8123 months.

Following the above procedure, the mean TDI for all intervals are estimated and presented in Table 9.

Table 9. Estimated Mean Total Duration of Illness (TDI) given Recent Duration of Illness (RDI) for 146 Patients with Mental Disorders for Different intervals of X_4 using a Generated Sample of Size 5000 from MVN Distribution $(X_2$=*RDI=ln(RDI) and X_4=*TDI=$\sqrt{\text{TDI}}$)

Interval	Original Data		Simulated Data		$E[X_4 / X_2]$	Observed TDI (months)	Estimated TDI (months)
	Mean X_4	Mean X_2	Mean X_4	Mean X_2			
$X_4 \le 2.5$	1.3671	3.5842	0.3811	3.5408	0.3819	1.8689	0.1459
$2.5 < X_4 \le 4.5$	3.2768	3.6157	3.5712	3.4617	3.5694	10.7374	12.7407
$4.5 < X_4 \le 6.5$	5.4474	2.8423	5.5515	3.3704	5.5351	29.6747	30.6376
$6.5 < X_4 \le 8.5$	7.5830	3.5072	7.5096	3.5821	7.5092	57.5022	56.3887
$8.5 < X_4 \le 10.5$	9.6697	3.7944	9.4977	3.5083	9.4991	93.5029	90.2325
$10.5 < X_4 \le 12.5$	11.4792	3.3867	11.4342	3.5224	11.4335	131.7722	130.7251
$12.5 < X_4 \le 14.5$	13.1889	2.2499	13.4278	3.4617	13.4145	173.9483	179.9492
$14.5 < X_4 \le 16.5$	15.4919	3.4320	15.3852	3.4016	15.3862	240	236.7355
$16.5 < X_4 \le 18.5$	17.8836	4.0147	17.4035	3.5160	17.3860	319.8233	302.2718
$X_4 > 18.5$	20.4939	4.0943	20.4974	3.5345	20.5137	420	420.8123

Comparison of observed and estimated total duration of illness for bivariate model
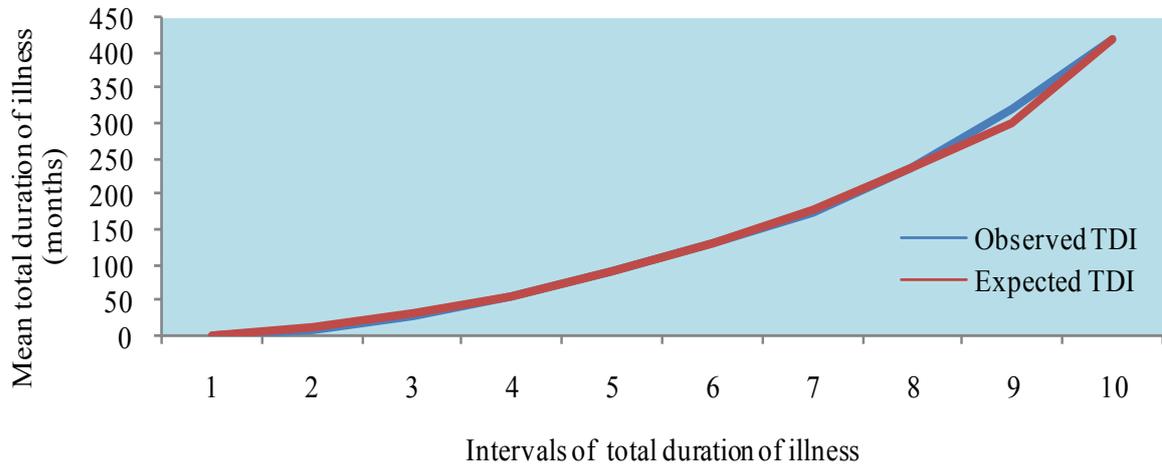

Figure 8.　Comparison of observed and estimated Total Duration of Illness (TDI) of 146 patients of dataset 1 for Bivariate Normal Distribution Model

3.6.2. Estimation of Mean TDI for 12 Patients of Dataset 2

Table 10.　Estimated Mean Total Duration of Illness (TDI) given Recent Duration of Illness (RDI) for 12 Patients with Mental Disorders using a Generated Sample of Size 5000 from MVN Distribution

S. No.	RDI (days)	Estimated TDI (months)
1	15	78.2393
2	30	75.8931
3	10	92.2723
4	60	80.8669
5	30	75.8931
6	12	82.6077
7	11	83.1499
8	160	85.6459
9	90	83.4830
10	9	84.2614
11	30	75.8931
12	30	75.8931

In this section, we applied BVN distribution model to estimate TDI in hospital for 12 patients (with unknown TDI) with mental disorders of dataset 2 on the basis of known values of RDI. The procedure of calculation of TDI for fourth patient with given mean value of RDI is as follows:

1. The observed mean value of RDI for fourth patient is 60 days, which is treated as mean value for the purpose of calculation.

2. This value of RDI is transformed as $X_2 = \ln$ (RDI) for application of BVN distribution model, i.e., for fourth patient, $X_2 = 4.0943$.

3. Mean and variance of X_2 (μ_2, σ_{22}) and X_4 (μ_4,

σ_{44}), and covariance (σ_{24}) between X_2 and X_4 are calculated from the simulated sample corresponding to the ranges of X_2, i.e., from simulated data, we have taken means of those observations for which $X_2 \cong 4.0943$.

4. Conditional expectation of $X_4 \mid X_2$ is obtained by substituting the above values in equation (20). The mean TDI given RDI for fourth patient is then obtained by squaring the value of $E(X_4 \mid X_2)$ which comes out to be 80.8669 months.

The mean TDI for 12 patients with known RDI are estimated by applying the above procedure and are presented in Table 10.

Matlab, version R2013a and R-3.2.1 statistical packages are used for the calculation and analysis.

4. Discussion

Mental and behavioural disorders are a significant contributor of global burden of disease. The global prevalence of mental and behavioural disorders among the adult population is estimated to be 10% and contributed to four of the ten leading causes of disability, with one in four families suffering the burden [4]. Thus, the problems related to mental disorders require immediate attention as early intervention can help control mental illness. Previous studies focussed primarily on predicting LOS in hospital [41, 19, 20, 21] but not much statistical research has been done to estimate the duration of illness.

The aim of this paper is to provide a model which will elevate the treatment procedure of the patients with mental disorders, further improving their quality of life.

The four variables considered appropriate for the study

were LOS, RDI, TDI and RSS. These variables are considered to be of utmost importance since they not only reflect the chronicity of the disease but also play a vital role in its diagnosis. These variables are a major determinant of mental health of a patient. They also account for the response of a patient towards the treatment. As per the previous studies, DSM and ICD are the standard tools used by mental health professionals' world over for classifying and diagnosing patients with mental disorders. In this paper, the data of patients diagnosed with mental disorders is in accordance with DSM-V and ICD-10. According to ICD-10, the diagnostic criteria for different mental disorders change according to the different durations for which a patient was symptomatic, suggesting that the duration of illness plays a cardinal role in diagnosing a patient with a particular mental disorder. Previous studies also reveal the importance of duration of disease in disease progression [22, 23, 24]. In this paper, RDI as well as TDI have been estimated and are also employed for estimating other variable like LOS which is dependent on them. Further, TDI is also estimated for given RDI accentuating the relationship between the two variables. Various psychiatric rating scales have been developed to measure the intensity and severity of mental disorders. For example, the most commonly used rating scale for depression is Hamilton Rating Scale for Depression (HAM-D). In this paper, the various psychiatric rating scales employed are BPRS and PANSS for schizophrenia, HAM-D for depression, YMRS for bipolar disorder, Bush-Francis Catatonia Scale for catatonia and Hamilton Anxiety Rating Scale (HAM-A) for anxiety, which are scaled down to 100 for comparison. Currently, there is a growing interest in the LOS in psychiatric admissions as this may be used as an indicator of efficiency for inpatient care, quality of care and as an important factor in the planning and distribution of hospital resources. LOS is universally recorded with an admission and discharge date, thus making it an easily accessible quantitative measure for statistical purposes [42]. Various studies have been performed to estimate or predict LOS on the basis of some factors. Manuel Martin-Carrasco et. al (2012) predicted LOS on the basis of number of prior episodes and Montgomery-Asberg Depression Rating Scale (MADRS) score for patients with acute mania [43] while Andeson et. al (2004) predicted that extended version of Brief Psychiatric Rating Scale (BPRS-E) can be employed to predict the length of stay (LOS) [41]. In this article, we have predicted LOS on the basis of TDI, RDI and RSS.

In this paper, we firstly obtained appropriate univariate distributions of random variables LOS, RDI, TDI and RSS using Akaike Information Criterion (AIC), transformed those variables which are not normally distributed using appropriate power transformations and then corroborated their joint normality using Mardia's test. AIC is well known method for the identification of an optimum model in a class of competing models. It trades off model's likelihood against its complexity and provides better criterion of model selection as compared to the conventional Chi-square goodness of fit test. The assumption of multivariate

normality of variables forms the basis for the application of MVN distribution. To check this assumption various tests are available but out of them Mardia's test is a relatively convenient, most widely used and accurate method for testing multivariate normality [27, 28, 44]. The Second step consists in application of MVN distribution to these variables. MVN distribution has been applied to solve various problems related to reliability analysis and hydrological engineering design problems [45, 46]. In medical research, Grover et al (2014) have already established the utility of MVN distribution for estimating the duration of diabetes of a patient with type-2 diabetes [27]. This encourages us for the implementation of MVN distribution to our present study. The Department of Psychiatry, Lady Hardinge Medical College & Smt. S.K, Hospital, New Delhi supported us by providing the data on variables of our interest. Thus, MVN distribution could be applied on variables LOS, RDI, RSS and TDI and the validity of the model could be established. Next, we validated our model by using simulation. Simulation is a procedure to test particular hypotheses and assess the appropriateness and accuracy of a variety of statistical methods in relation to the known truth. These techniques provide empirical estimation of the sampling distribution of the parameters of interest that could not be achieved from a single study and enable the estimation of accuracy measures as the truth is known. The comparison of the simulated results with the true values used to simulate the data provides a measure of the performance and associated precision of the simulation process [47]. Finally, we used our model for estimating the LOS and TDI for patients for whom information on these variables is not known.

From our data we observed that the mean LOS in hospital is found to be more for women than for men. The maximum LOS in hospital for a women diagnosed with depression is 58 days while for men it is observed to be 20 days. Maximum mean LOS for patients with schizophrenia is observed to be 70 days. The average TDI is discerned to be 112.39 months for men while for women it is 85.53 months. According to our data, average total duration of Schizophrenia is found to be more for men as compared to women while for depression women are observed to have more total duration than men.

Since information on all the variables is rarely available, thus application of multivariate normal (MVN) distribution for four, three and two variables has been demonstrated so that appropriate model can be chosen according to the information available. Previous studies suggest that when the assumptions are met, the statistical power of parametric distribution is higher than nonparametric and semi parametric models. We found in our study that results derived from these models did not vary significantly from each other. As per our study, expected values of the variables estimated did not deviate much from the observed values. Further, the results obtained are also consistent with simulated study. Thus, according to the availability of records, appropriate model could be directly employed for estimation of unknown variable. The findings of our study

are not only consistent with previous studies but extend them further, thus, adding new evidence to literature on the estimation of length of stay and total as well as recent duration of illness for patients with mental disorders. Multivariate normal distribution model produces good results for our study but application of this model can be extended to other medical research as well.

5. Conclusions

The work done in this paper points to the possibility of immediate implementation of a model which, at the time of admittance, could be used to predict the length of time which a given patient would be likely to stay in the hospital, the total duration of illness and the recent duration of illness for which the patient was symptomatic, further stimulating the treatment procedure. Although length of stay is considered to be one of the most important predictor of mental health of an inpatient, our work manifests recent and total duration of illness to be significant factors as well influencing disease progression.

Multivariate normal distribution model is complicated, yet, powerful tool for estimating variable of interest if the underlying assumptions are met. In this study, strong transformations are applied to variables to follow marginal normal distribution. In this paper, we applied multivariate normal distribution to the four variables under study (after applying log and power transformations) and found that estimated values of these variables are consistent with the respective observed values. The following can be concluded from the study:

> The expected length of stay (LOS) from our model (four variate normal distribution) is estimated to be 14.79 days while it is observed to be 15.99 days for the real data.
> The expected recent duration of illness (RDI) from our model is estimated to be 18.07 weeks; whereas the RDI observed from the real data is 15.50 weeks.
> The expected RSS estimated by our model is found to be 36.64 while from the real data it is observed to be 35.12.
> TDI estimated on the basis of our model is 104.27 months whereas form the data it is observed to be 98.04 months.

On the basis of these results, it can be concluded that estimated values of expected length of stay (LOS), expected recent duration of illness (RDI), expected rating scale score (RSS) and expected total duration of illness (TDI) are consistent with the observed values.

The results depicted in Table 11, based on 146 patients (Lady Hardinge Medical College & Smt. S.K, Hospital, Delhi, India for the calendar year 2013-2014) and simulated data, are comparable to previous study in which the average hospital stay for patients with schizophrenia in the USA is found to be (20.92 +5.42) days and Germany are (37.29 +4.18) days while the average hospital stay for patients with bipolar disorder in the USA are (15.37+6.23) days and in Germany are (40.31 +7.71) days. The above study was given by Auffarth et. al (2008) [55] for large sample data.

Table 11. Observed and Estimated Mean Length of stay (LOS) in the Hospital for Patients with Schizophrenia and Bipolar Affective Disorder (BPAD)

Disorder	Percentage of cases	Observed LOS (days)	Estimated LOS (days)
Schizophrenia	50	29	30
Bipolar Affective Disorder	21.23	13.72	14.21

Table 12. Comparison between Regression Model by Huntley et. al (1998) and Multivariate Normal Distribution (MVN) Model (Four variate) for Estimating Mean Length of stay (LOS)

Regression Model			MVN Model		
Actual LOS	Predicted LOS	Difference between the means (in %)	Actual LOS	Estimated LOS	Difference between the means (in %)
26.30	17.81	32.28	6.7617	6.5241	3.51
16.77	16.31	2.74	12.5030	12.0633	3.52
16.65	15.39	7.57	19.3517	20.1231	3.99
8.37	10.94	30.70	28.3596	30.1212	6.21
15.04	15.64	3.99	38	44.5261	17.17
22.38	17.09	23.64	58	52.3203	9.79
14.44	14.73	2.01	68.9964	73.3471	6.31

The columns (3) and (6) of Table 12 depict the difference (in percentage) between the actual and predicted/estimated length of stay in hospital by regression model given by Huntley et. al (1998) [19] and multivariate normal distribution (four variate) model, presented in this study, respectively. It is evident from the above table that the results obtained by multivariate model are more consistent with the observed values as compared to the regression model.

The multivariate normal distribution may be preferred over bivariate normal distribution as more the information the better would be the estimates as is evident from the results obtained in Table 4, Table 6 and Table 9.

From our data, we also observed that the length of stay in the hospital is more than the average for patients who received electroconvulsive therapy. The patients with length of stay in the hospital exceeding the average value along with patients receiving electroconvulsive therapy can be studied exclusively. The relationship between number of episodes of illness and its chronic condition can be taken up in next study.

Though the cumulative effect of qualitative variables is incorporated in this study through the psychiatric rating scale score and length of stay in hospital but there individual effect on disease progression can also be assessed according to the availability of data. In this section, we have compared results obtained for schizophrenia and

bipolar affective disorder with previous studies but cannot do the same for acute and transient psychotic disorder and depression as the number of cases for these disorders is not enough for the comparison which can be taken as a limitation of our study.

ACKNOWLEDGEMENTS

REFERENCES

[1] Cross-national comparisons of the prevalences and correlates of mental disorders. WHO International Consortium in Psychiatric Epidemiology. Bull. World Health Organ. 78 (4): 413–26. 2000. doi:10.1590/S0042-96862000000400003. PMC 2560724. PMID 10885160.

[2] WHO, World Health Report 2001.

[3] Mental Health conditions, https://www.nami.org/Learn-More/Mental-Health-conditions.

[4] Burden of Disease in India, September 2005, National Commission on Macroeconomics and Health, Ministry of Health & Family Welfare, Government of India.

[5] WHO Fact Sheets, October 2014, http://www.who.int/media centre/factsheets/fs396/en/.

[6] Severe Mental Illness Defined by Duration and Disability: http://www.bhevolution.org/public/severe_mental_illness.page.

[7] Sims, A. Symptoms in the mind: an introduction to descriptive psychopathology. Philadelphia: W. B. Saunders 2002. ISBN 0-7020-2627-1.

[8] Kneisl, C. and Trigoboff, E. Contemporary Psychiatric-Mental Health Nursing. 2nd edition. London: Pearson Prentice Ltd. p. 371. 2009.

[9] Carson, V.B. Mental health nursing: the nurse-patient journey W.B. Saunders. ISBN 978-0-7216-8053-8. p. 638. 2000.

[10] Mood-related Disorders, http://www.bhevolution.org/public/mood_disorders.page.

[11] World Health Organization. The ICD-10 classification of mental and behavioral disorders. Geneva: World Health Organization; 1992.

[12] Mehta, S. Diagnostic Stability of Acute and Transient Psychotic Disorders in Developing Country Settings: An Overview. PMC, Ment Illn. 2015 Feb 24; 7(1): 5640. doi: 10.4081/mi.2015.5640.

[13] Hagan, C. C., Graham, J. M. E., Tait, R., Widmer, B., van Nieuwenhuizen, A. O., Ooi, C., … Suckling, J. (2015). Adolescents with current major depressive disorder show dissimilar patterns of age-related differences in ACC and thalamus. NeuroImage: Clinical, 7, 391–399. http://doi.org/10.1016/j.nicl.2014.12.019.

[14] Schinnar, A.P., Rothbard, A.B., Kanter, R., Jung, Y.S. An empirical literature review of definitions of severe and persistent mental illness. Am J Psychiatry. 1990; 147(12): 1602-8.

[15] Ballesteros, R.J., Martínez, P.L., Martín, C.M., Ibarra G.N., Bulbena, V.A. Assessment of the Life Skills Profile and the Brief Psychiatric Rating Scale as predictors of the length of psychiatric hospitalization. Actas Esp Psiquiatr. NCBI. Pubmed 2002 Jul-Aug;30(4):225-32.

[16] Kiesler, C.A., Simpkins, C., Morton, T. Predicting length of hospital stay for psychiatric inpatients. Hosp Community Psychiatry. 1990;41:149–54.

[17] Kaplan & Sadock. Concise textbook of clinical psychiatry, 3rd edition. Wolters Kluwer Health, Lippincott Williams & Wilkins.

[18] Roy-Byrne, P., Russo, J., Rabin, L., Fuller, K., Jaffe, C., Ries, R., Dagadakis, C., et al. (1998). A Brief medical necessity scale for mental disorders: Reliability, validity, and clinical utility. The Journal of Behavioral Health Services and Research, 25(4), 412-424. doi:10.1007/BF02287511.

[19] Huntley, D.A., Cho, D.W., Christman, J., Csernansky, J.G. Predicting length of stay in an acute psychiatric hospital. Psychiatric Services 49:1049-1053,vol 8, 1998.

[20] Shawanda, W., Anderson, A.J., Crist, and Nanetta, P. Predicting Inpatient Length of Stay With the Expanded Version of the Brief Psychiatric Rating Scale (Version 4.0), Psychiatric Services 2004 55:1, 77-79.

[21] Ithman, M.H., Goplarkrishna G., Beck, N.C., Das, .J, Petroski, G. Predictors of Length of Stay in an Acute Psychiatric Hospital. J Biosafety Health Educ 2: 119. doi:10.4172/2332-0893.1000119. 2014.

[22] Green, M., Marzo-Ortega,H., Mcgonagle, D., Wakefield, R., Proudman, S., Conaghan, P., Gooi, J. , and Emery, P. (1999). Persistence of mild, early inflammatory arthritis, The Importance of Disease Duration, Rheumatoid Factor, and the Shared Epitope. ARTHRITIS & RHEUMATISM Vol. 42, No. 10, October 1999, pp 2184–2188.

[23] Anderson, J. J., Wells, G., Verhoeven, A.C., Felson, D.T. (2000). Factors predicting response to treatment in rheumatoid arthritis: the importance of disease duration. Arthritis Rheum. 2000 January; 43(1): 22–29. doi: 10.1002/1529-0131(200001).

[24] Grover, G., Sabharwal, A. A parametric approach to estimate survival time of diabetic nephropathy with left truncated and right censored data. International Journal of Statistics and Probability, Vol. 1, No. 1 (2012). DOI: 10.5539/ijsp.v1n1p128.

[25] Multivariate analysis concepts, http://support.sas.com/publishing/pubcat/chaps/56903.pdf.

[26] Mardia, K.V. Measures of multivariate skewness and kurtosis with applications. Biometrika, 57(3):519–530, 1970.

[27] Grover, G., Sabharwal, A., Mittal, J. Application of Multivariate and Bivariate Normal Distributions to Estimate Duration of Diabetes, International Journal of Statistics and

Applications, 2014; 4(1): 46-57. doi:10.5923/j.statistics.2014 0401.05.

[28] Bradman, M.J., Clark, T.G., Love, S.B., Altman, D.G., 2003, survival analysis part III: Multivariate data analysis-choosing a model and assessing its adequacy and fit, British Journal of Cancer, 89, 605-11, doi: 10.1038/sj.bjc.6601120.

[29] Johnson, N. L., and Kotz, S., Distributions in statistics: Continuous multivariate distributions, John Wiley & Sons, New York, 1972.

[30] Transformations,http://www-users.york.ac.uk/~mb55/msc/cl inbio/week5/transfm_gif.pdf.

[31] Box-Cox and Power transforms, http://www.statsref.com/HT ML/index.html?freeman-tukey.html.

[32] Ahrens, J.H., Dieter, U. Generating gamma variates by a modified rejection technique. Communications of the ACM, Vol. 25, Issue 1, Jan 1982, Pages 47-54. doi>10.1145/358315.358390.

[33] Crow, E.L., and Shimizu, K. Lognormal Distributions: Theory and Applications, Statistics, textbooks and monographs, vol 88, Marcel Dekker, Inc, New York, 1988.

[34] Burdenski, T. Evaluating univariate, bivariate, and multivariate normality using graphical and statistical procedures. Multiple Linear Regression Viewpoints, 26(2): 15–28, 2000.

[35] Kay, S, R.., Fizbein, A., Opler, A. (1987) The positive and Negative Syndrome Scale (PANSS) for schizophrenia. Schizophrenia Bulletin, 2, 261-276.

[36] Hamilton, M. J. (1960). A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960 Feb; 23(1): 56–62.

[37] Overall, J.E, Gorham, D.R. (1962). The brief psychiatric rating scale. Psychological Reports 1962 vol. 10, pp799-812.

[38] Hamilton, M. The assessment of anxiety states by rating. Br J Med Psychol 1959; 32:50–55.

[39] Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978; 133:429–35.

[40] Bush, G., Fink, M., Petrides, G., Dowling, F. and Francis, A.. Catatonia. I. Rating scale and standardized examination. Acta Psychiatrica Scandinavica, Volume 93, Issue 2, pages 129–136, February 1996.

[41] Anderson, S.W., Crist, A.J., Payne, N. Predicting Inpatient Length of Stay With the Expanded Version of the Brief Psychiatric Rating Scale (Version 4.0). Psychiatric Services 55, 2004.

[42] Adegunloye, O.A., Yussuf, A.D., Ajiboye, P.O., Issa, B.A. and Buhari, O.I.N. (2009). Correlates of Length of Stay among Psychiatric In-Patients in a Tertiary Health Institution

in Nigeria. Research Journal of Medical Sciences, 3(2), 56-61.

[43] Martin-Carrasco et al.: Number of prior episodes and the presence of depressive symptoms are associated with longer length of stay for patients with acute manic episodes. Annals of General Psychiatry 2012 11:7.

[44] Harada,T., Ariyoshi, N., Shimura, H., Sato, Y., Yokoyama, I., Takahashi, K., Yamagata,S., Imamaki, M., Kobayashi, Y., Ishii, I., Miyazaki, M., Kitada, M. Application of Akaike information criterion to evaluate warfarin dosing algorithm. Thromb Res. 2010 Sep; 126(3): 183-90. doi: 10.1016/j.thromres.2010.05.016.

[45] Lipow, M., and Eidemiller, R. L., 1964. Application of the bivariate normal distribution to a stress vs strength problem in reliability analysis, Technometrics, 6(3), 325-328, http://www.jstor.org/stable/1266043.

[46] Yue, S., 1999. Applying bivariate normal distribution to flood frequency analysis, Water International, 24(3), 248-254. http://dx.doi.org/10.1080/02508069908692168.

[47] Burton, A., Altman, D.G., Royston, P..and Holder, R.L. The design of simulation studies in medical statistics. STATISTICS IN MEDICINE Statist. Med. 2006; 25:4279–4292. www.interscience.wiley.com. DOI: 10.1002/sim.2673.

[48] The Multivariate Normal Distribution, http://www.math.uiuc .edu/~r-ash/Stat/StatLec21-25.pdf.

[49] Mental Health prevalence in U. S., http://www.nimh.nih.gov /health/statistics/prevalence/any-mental-illness-ami-among-a dults.shtml.

[50] WHO, Fact files, http://www.who.int/features/factfiles/ment al_health/mental_health_facts/en/index4.html.

[51] Hopko, D.R., Lachar, D., Bailley, S.E., et al: Assessing predictive factors for extended hospitalization at acute psychiatric admission. Psychiatric Services 52:1367–1373, 2001.

[52] Woerner, M.G., Mannuzza, S., Kane, J.M. Anchoring the BPRS: an aid to improved reliability. Psychopharmacology Bulletin 24: 112–117, 1998.

[53] Rudorfer, M.V., Henry, M.E., Sackeim, H.A. (2003). Electroconvulsive therapy. In A Tasman, J Kay, JA Lieberman (eds) Psychiatry, Second Edition. Chichester: John Wiley & Sons Ltd, 1865–1901.

[54] Ect, https://www.sane.org/information/factsheets-podcasts/4 45-electroconvulsive-therapy-ect.

[55] Auffarth I., Busse R., Dietrich D., Emrich H. (2008). Length of Psychiatric Inpatient Stay: Comparison of Mental Health Care Outlining a Case Mix From a Hospital in Germany and the United States of America. German Journal of Psychiatry • http://www.gjpsy.uni-goettingen.de • ISSN 1433-1055.