JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

An Application of the Symmetrical Bivariate Negative Binomial Distribution to the Study of Absence Proneness.

Ms. Moon,
Department of Statistics, Patna University, Patna

Dr. Kumar Satyendra Yadav
Assistant Professor, Department of Statistics
Patna University, Patna
and
Dr. Dilip Kumar
Associate Professor, Population Research Centre
Patna University, Patna

Abstract

In this paper an attempt has been made to use the Symmetrical Bivariate Negative Binomial Distribution (SBNBD) to study and predict the absenteeism behavior of the shift workers. Apart from Providing good fit to the observed bivariate distributions of the shift works, it helps in the determination of the correlation coefficient, the line of regression and the operating characteristics for the absenteeism occurred in two equaling periods. These, in fact, help in making good predictions of the absenteeism of the shift workers on the basis of previous year data.

Keywords: Absence proneness; Bivariate distribution; Maximum likelihood, operating characteristics.

1. Introduction :

Arbous and Sichel (1954) have made an attempt to study the absenteeism in an Organization. It was found that shortage of competent man-power became the problem of loss in the productivity in industry which encouraged both managerial and the scientific interest in such problems. These problems became more serious to the management due to limited availability of

man-power. Consequently, the social scientists considered absenteeism as systematic of occupational casualty. In the past the analysis of absence was confined largely to the calculation of a series of 'rates' for innumerable classifications of the data. Traditional method of analyzing absenteeism data by means of frequency, severity and disability rates was found inadequate for the purpose of scientific investigation and of planning a practical approach to the absenteeism problem in industry. In fact, Greenwood and Yule (1920), made an attempt to study that accident data, and claimed thereby to have established the phenomenon of accident proneness. Thus, the term 'proneness' was first used in the study accidents. Following the suggestions of Lundberg (1940) and Maritz (1950) proneness began to be measured by the product moment correlation coefficient existing between absences occurring to a group in successive non-overlapping periods of exposure, irrespective of the specific type of underlying bivariate distribution. Here, the model of Symmetrical Bivariate Negative Binomial Distribution (SBNBD) has been fitted to enable both management and social scientists to analyze the absenteeism data more precisely for the better planning and management. The technique of SBNBD clearly predicts absence behavior so that one can identify in a group of those who would be absence offenders in future, and so attention should be paid to them.

2. The Suggested Model

Arbous and Sichel (1954) pointed out that the bivariate negative binomial distribution (BNBD) was first given by Lundberg (1940) and has again derived by Arbous and Kerrich (1951) to analyse accident statistics Arbous and Kerrich (1951) developed BNBD which they called a bivariate compound Poisson distribution this is given as follows:

$$P(X = x, Y = y)\ \frac{(k)^k \alpha^{x+y} \sqrt{k+x+y}}{(k+\alpha d)^{k+x+y} \sqrt{k}} \times \frac{d_0^x}{x!y!} \qquad (1)$$

Where d_0 and d_1 are two sub-intervals of the exposed period such that $d = d_0 + d_1$. x, y and $x + y$ are the number of accidents an individual suffers during the interval d_0, d_1 and d.

While studying the absence proneness Arbou and Sichel (1954) have derived the BNBD but they called it symmetrical bivariate negative binomial distribution (SBNBD). In fact, it is the particular case of (1) and SBNBD suggested by Arbous and Sichel after putting $d_0 = d = 1$ and $(1/2)\alpha = m$. It is given by.

$$P(X=x,Y=y)=\left\{\frac{(k)}{(k+2m)}\right\}^{k}\frac{\sqrt{k+x+y}}{\sqrt{k}\sqrt{x+1}\sqrt{y+1}}\left\{\frac{(m)}{(k+2m)}\right\}^{x+y}$$

This is known as the probability mass function of the SBNBD "M" and "k" are its parameters.

The marginal distribution of (2), their means and variances are given by equations (3), (4), (5) respectively

$$F(X=x) = \left\{\frac{\{k\}^{k}}{\{k+m\}}\right\}\frac{\sqrt{k+x+y}}{\sqrt{k}\sqrt{x+1}}\left\{\frac{\{m\}^{k\to y}}{\{k+m\}}\right\} \quad (3)$$

$$\text{mean} = m \quad (4)$$

$$\text{Variance} = m + \left\{\frac{(m)^2}{(k)^2}\right\} \quad (5)$$

The conditional probability mass function of y is given by

$$P(Y=y:X=x)=\left\{\frac{(k+m)}{(k+2m)}\right\}^{k\to x}\left\{\frac{(m)^{y}}{(k+2m)}\right\}\frac{\sqrt{k+k+y}}{\sqrt{k+x}\sqrt{y+1}} \quad (6)$$

Similarly, $P\{X=x:Y=y\}$ may be defined.

The product moment correlation coefficient of SBNBD is obtained by

$$P = \left\{\frac{(m)}{(k+m)}\right\} \quad (7)$$

The mean and the variance are given by

$$E\{Y \mid X=x\} = P\{k+x\} \quad (8)$$

$$V\{Y \mid X=x\} = P(1+p)(k+x) \quad (9)$$

Similarly,

$$E(X \mid y=y) = p(k+y)$$

$$V(X \mid Y=y) = p(1+p)(k+y)$$

The operating characteristics of absenteeism of the shift workers in the second interval is defined as the probability to have α or more absenteeism when y absenteeism occur in the first interval have already been taken place. It is obtained by.

$$P(x >, \alpha | Y = y) = 1 - \left\{\frac{\{k+m\}^{y+k}}{\{k+2m\}}\right\}.\left\{\frac{\{1\}}{\{\sqrt{y+k}\}}\right\}\sum_{x=0}^{\alpha-1}\left\{\frac{\{m\}^{x}}{\{k+2m\}}\right\}\frac{\sqrt{x+y+k}}{\sqrt{x+1}} \quad (12)$$

Similarly, P (Y>, β|X = x) may be obtained.

In order to estimate the parameters of the SBNBD we obtain first the distribution of x+y (i.e. univariate distribution). The estimator "k" & "p" are computed to the distribution and then the maximum likelihood estimates are obtained as given by Johnson and Kotz (1969), Sinha (1984) and Kumar (1990). Intact, the value of "m" is the half of the mean of the univariate distribution.

Examples and discussion

In order to illustrate the technique, the univariate negative binomial distribution has been fitted to the absenteeism data of 248 shift workers which have been referred to by Arbous and Sichel (1954). The fit is quite good. The adequacy of the model has been examined by the Chi-square test. The maximum likelihood method has been used for the estimation of parameters. Thus, for the fitting of SBNBD we get.

m= 4.59072, k = 1.58542, and p = 0.74329.

Hence, 2m and k are the parameters of the univariate distribution. The SBNBD provides good fit to the bivariate distribution.

Table 1 gives the observed and the expected frequencies. The adequacy of the model has been judged by the method of goodness of fit test of the chi-square.

Y^X	0	1	2	3	4	5	6	7	8+	Total
0	7 (11.90)	7 (8.04)	8 (4.43)	1 (2.26)	1 (1.11)	- (0.53)	- (0.25)	- (0.11)	- (0.09)	24 (28.72)
1	9 (8.04)	6 (8.86)	8 (6.72)	7 (4.42)	1 (2.63)	3 (1.48)	- (0.80)	- (0.42)	- (0.41)	34 (33.84)
2	9 (4.43)	9 (6.78)	6 (6.63)	5 (5.26)	1 (3.69)	3 (2.39)	- (1.46)	2 (0.85)	1 (1.03)	36 (32.52)
3	2 (2.26)	8 (4.42)	9 (5.26)	2 (4.92)	4 (3.98)	3 (2.91)	- (1.98)	- (1.28)	4 (1.88)	32 (28.69)
4	1	7	5	4	3	-	2	1	3	26

	(1.11)	(2.63)	(3.69)	(3.98)	(3.64)	(2.98)	(2.24)	(1.58)	(2.76)	(24.61)
5	2 (0.53)	1 (1.48)	1 (2.39)	2 (2.91)	5 (2.98)	2 (2.69)	1 (2.21)	1 (1.69)	5 (3.56)	20 (20.44)
6	- (0.25)	1 (0.80)	1 (1.46)	2 (1.98)	2 (2.24)	3 (2.21)	1 (1.98)	2 (1.64)	4 (4.11)	16 (16.67)
7	- (0.11)	- (0.42)	- (0.85)	1 (1.28)	1 (1.58)	2 (1.69)	1 (1.64)	- (1.46)	3 (6.04)	8 (13.43)
8+	- (0.09)	- (0.41)	1 (1.03)	1 (1.88)	5 (2.76)	4 (3.56)	4 (4.11)	3 (6.04)	34 (29.20)	52 (49.08)
Total	= 30 (28.72)	39 (33.84)	39 (32.84)	25 (28.69)	23 (24.61)	20 (20.44)	9 (21.23)	9 (13.43)	54 (49.02)	248 (248.00)

Further, the 95 percent confidence limits for the product moment correction coefficient (0.72801) are.

0.66361 (lower) an 0.78167 (upper).

These limits include the value of P (the correlation coefficient) obtained on the basis of equation (7). This illustrates the suitability of the theory with the observed data. The observed and the expected average number of absenteeism of the shift workers in the year 2018 corresponding to the absenteeism in the year 2017 are given in. Table2. The result has been obtained on the basis of the following equation.

$$E(X \mid Y = y) = 1.17844 + 0.74329\, y.$$

Table 2: The observed (x) and the expected (xe) average number of absenteeism of shift workers in 2018 having Y number of absenteeism in 2017.

Y	X	XE
0	1.250 = 1	1.178 = 1
1	1.823 = 2	1.922 = 2
2	2.138 = 2	2.665 = 3
3	2.969 = 3	3.408 = 3
4	3.231 = 3	4.152 = 4
5	4.600 = 5	4.895 = 5
6	5.250 = 5	5.638 = 6
7	8.875 = 9	6.381 = 6
8+	20.807 = 21	17.124 = 17

It is evident that the correspondence between the observed and the expected frequencies is very good. This justifies the application of the SBNBD in the present problem.

Further, on substituting the values of k and m in the (12), we obtain

For each value of y we obtain

F(x >, α | Y = y), and the expected number of shift workers to record α (α =1, 2,..., 6)

or more absenteeism in 2018. These values along with the observed frequencies are mentioned in Table 3.

Table 3. The observed and the expected number of shift workers to have α (α =1, 2, ...,6) or more absenteeism in 2018 when they had Y(Y=0, 1,....., 8+) absenteeism in 2017.

Y	0	1	2	3	4	5	6	7	8+
P(X>, 1|Y)	0.58	0.76	0.86	0.92	0.95	0.97	0.98	0.99	0.99
Obs. No.	17	25	27	30	25	18	16	8	52
Exp. No.	14.1	25.9	31.1	29.5	24.8	19.5	15.7	7.9	51.7
P(x>,2|Y)	0.31	0.49	0.65	0.77	0.85	0.90	0.94	0.96	0.97
Obs. No.	10	19	18	22	18	17	16	8	52
Exp. No.	7.3	16.9	23.6	24.6	22	18	14.9	7.7	50.7
P (X>, 3|Y)	0.15	0.29	0.45	0.59	0.69	0.78	0.85	0.90	0.93
Obs. No.	2	11	12	13	13	16	14	8	51
Exp. No.	3.6	10.2	16.3	18.8	18.1	15.7	13.7	7.2	48.2
P (X>, 4|Y)	0.07	0.16	0.29	0.42	0.53	0.64	0.73	0.81	0.85
Obs. No.	1	4	7	11	9	14	12	7	50
Exp. No.	1.5	5.7	10.5	13.4	13.9	12.8	11.6	6.5	44.3
P (X>, 5|Y)	0.04	0.09	0.18	0.28	0.38	0.49	0.59	0.70	0.75
Obs. No.	1	3	6	7	6	9	10	6	45
Exp. No.	0.8	3.1	6.4	8.9	10.1	9.8	9.4	5.6	38.0
P (X>,	0.02	0.04	0.10	0.18	0.26	0.36	0.45	0.59	0.63

6|Y)									
Obs. No.	1	0	3	4	6	7	7	4	41
Exp. No.	0.4	1.6	3.8	5.8	6.9	7.2	7.3	4.7	32.9

The resemblance between the observed and the expected frequencies is quite good. This in fact, reflects *The adequacy of the SBNBD model to describe the bivariate distribution of the absenteeism data.

Through this approach it is easy to predict the absenteeism of the shift workers on the basis of previous year data.

REFERENCES

Arbous, A.G. & Kerrich, J.C. (1951). Accident statistics and the concept of accident proneness. **Biometics**, 7,340-42.

Arobous, A.G. & Sichel, H.S. (1954). New Techniques for the analysis of absenteeism data. **Biometrika**, 41/77-90.

Greenwood, M.& Yule, G.U. (1920). An inquiry in to the nature of frequency distributions representative of multiple happenings, **J.R. Statist. Soc.**, A83, 255-79.

Johnson, N.L.& Kotz, S. (1969). Discrete distributions. John Wiley & Sons Ince., New York, U.S.A.

Kumar, D (1990). The Bivariate Negative Binomial Distribution and its Application to Population Sciences and other areas. Ph.D. thesis, Patna University, Patna.

Lundberg, 0. (1940). On random Processes and their application to sickness and accident statistics. Univ. of Stockholm, Uppsala, Sweden.

Martiz, J.S. (1950). On the validity of inferences drawn from the fitting of Poisson and negative binomial distributions to observed accident data. **Psychological Bulletin**, 47, 434-43.

Sinha, A.K. (1984). The Negative Binomial Distribution and its Role in Operations Research and population studies, Ph.D. Thesis, Patna University, Patna.