SENRA Academic Publishers, Burnaby, British Columbia
Vol. 6, No. 3, pp. 2191-2194, October 2012
Online ISSN: 1920-3853; Print ISSN: 1715-9997

Canadian Journal of pure&applied sciences

Short Communication

REGRESSION MODELS FOR PREDICTING THE NUMBERS OF TWO STATE OF DELIVERY IN WOMEN

*Ebuh GU[1], Agbor RA[1] and Ebuh AC[2]
[1]Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka,
[2]Department of Computer Science, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria

ABSTRACT

In this study, expressions for the joint probability distributions of the number of normal deliveries and number of cesarean section deliveries in hospitals were obtained when the number of deliveries is assumed to be a random variable. A goodness of fit test was carried out on them using Kolmogorov-Sminov test. Regression equations for predicting the number of normal deliveries for a given number of cesarean deliveries and vice-versa as well as expressions for the correlation coefficients are also obtained. Number of deliveries are assumed to follow the binomial and Poisson distributions. The models are fitted to the data collected from Regina Caeli Hospital (between 2006 and 2010) in Awka, Anambra State, Nigeria. It was discovered that the Poisson distribution provided a better fit. This model gave a correlation coefficient value of 0 between the number of normal deliveries and that of cesarean deliveries with a conditional variance of 0.4034. The variation between the number of deliveries of different modes (normal or cesarean) widens as the number of deliveries increases.

Keywords: Kolmogorov-Sminov, poisson and binomial, correlation, conditional variance, cesarean and normal deliveries.

INTRODUCTION

Researchers who have studied similar things in this regard include Janardan and Rao (1984). In their research, they gave two classes of probability models for the distribution of the number of boys in a family. Their distributions were based on the Polya urn scheme as well as some genetic considerations. They established several interrelationships and indicated some special cases of their results. Sharma (1985) proposed some probability models for the same situation and discussed their adequacy with the same data used by Janardan and Rao (1984). The value of the correlation coefficient between the number of boys and the number of girls in a family were calculated from the same data in each model. His work was based essentially on the general expression for the correlation coefficient between the number of boys and the number of girls in a family, when the family size follows a modified power series distribution obtained by Gupta (1976). Others are Feller (1968), Janardan (1975), Stene (1978), Onankpa and Ekele (2009). This paper proposes to obtain the expressions for the joint probability distributions of the number of normal deliveries and the number of Cesarean deliveries, the marginal distributions of each of them, the conditional distributions, the correlation coefficients, the regression equations, and the conditional variances of the number of normal deliveries for a given number of cesarean deliveries and vice-versa as well as testing the goodness of fit of the probability models. These are obtained when the delivery size is assumed to follow binomial and Poisson. It is expected that the outcome of the paper will be engaged in predicting the number of cesarean deliveries for a given number of normal deliveries.

*Corresponding author email: ablegod007@yahoo.com

MATERIALS AND METHODS

Kolmogorov-smirnov

In line with the stated objective, statistical analysis such as Kolmogorov-smirnov to determine whether or not the observed frequencies are compatible with a hypothesized population distribution that is binomial and Poisson distribution.

Kolmogorov-smirnov test is used in place of Chi-square to avoid pooling of categories and it does not impose a lower limit on those categorized frequencies.

The Kolmogorov - Smirnov test is based on the measurement of the distance between two cummulative relative frequency distributions. The observed cummulative frequency and the expected cumulative frequency are denoted by O_x and E_x respectively.
The test statistic can be computed as:

$$D(N) = Max\left(E_{(x)} - O_{(x)}\right) \tag{1}$$

While the critical value is gotten from the D (N) table at the appropriate significance level

Probability Distribution

A probability distribution shows the possible outcomes of an experiment and the probability of each outcome. It deals mostly on how the future event is.

The Model

Let Y_i and Y_j be the outcome of each delivery.

$$Y_i = \left.\begin{array}{l} \text{I if the } i^{th} \text{ delivery is through a normal birth} \\ \quad \text{with probability P, i=1, 2,…,n} \\ = 0 \text{ otherwise} \end{array}\right\} \quad (2)$$

$$\text{and } Y_j = \left.\begin{array}{l} 1 \text{ if the } j^{th} \text{ delivery is through cesarean} \\ \quad \text{with probability q j=1, 2,…,n} \\ = 0 \text{ otherwise} \end{array}\right\} \quad (3)$$

The probability generating function of Y_i is given as pt_1

The probability generating function of Y_j is given as qt_2

The joint probability generating function of Y_i and Y_j is given as

$$Py_i y_j(t_1,t_2) = pt_1 + qt_2, p,q: fixed, q = 1 - p \quad (4)$$

For n deliveries, the number of normal deliveries (N) and the number of Cesarean deliveries (C) is given as

$$N = \sum_{i=1}^{n} Y_i \quad (5)$$

$$\text{and } C = \sum_{j=1}^{n} Y_j \quad (6)$$

Let $p_r(N = n)$ denote the probability that the delivery times N= n with pgf

$$G(t) = \sum_{n=1}^{\infty} t^n p_r(N = n) \quad (7)$$

So that the joint pgf of the number of normal deliveries and the number of cesarean deliveries in a delivery of time n is given as

$$H(t_1,t_2;n) = G(P(t_1,t_2) \quad (8)$$

(i) Let N be a binomial random variable with parameters (n,θ), where 1-θ is the probability of no delivery.

$$H(t_1,t_2;n) = (1-\theta + p\theta t_1 + q\theta t_2)^n \quad (9)$$

The joint probability distribution N and C is given as

$$p(N = i, C = j) = \frac{n!}{(n-i-j)!\,i!\,j!} p^i q^j \theta^{i+j}(1-\theta)^{n-i-j}; i, j = 1,2,..., n \quad (10)$$

The marginal distributions are given as

$$p(N = i) = \binom{n}{i}(p\theta)^i(1-p\theta)^{n-i}; i = 1,2,..., n \quad (11)$$

$$p(C = j) = \binom{n}{j}(p\theta)^j(1-p\theta)^{n-j}; j = 1,2,..., n \quad (12)$$

The mean number of normal deliveries is given as

$$\mathrm{E}(N) = np\theta \quad (13)$$

$$\text{and } Var(N) = np\theta(1-p\theta) \quad (14)$$

$\mathrm{E}(C)$ and $Var(C)$ are the same as in equations (13) and (14) with p replaced by q.

The conditional distributions are given as

$$p(N = i / C = j) = \binom{n-j}{i}\left(\frac{p\theta}{1-\theta}\right)^i\left(\frac{1-\theta}{1-q\theta}\right)^{n-j}; i = 1,2,..., n-j \quad (15)$$

$$p(C = j / N = i) = \binom{n-i}{j}\left(\frac{q\theta}{1-\theta}\right)^j\left(\frac{1-\theta}{1-p\theta}\right)^{n-i}; j = 1,2,..., n-i \quad (16)$$

The regression equation of the number of normal deliveries on the number of cesarean deliveries is given as

$$\mathrm{E}(N / C) = (n-j)(p\theta)(1-q\theta)^{-1} \quad (17)$$

Table 1. Data on Cesarean Section Delivery and Virginal Delivery in Women

	2006			2007			2008			2009			2010		
	C/S	V/D	Total	C/S	V/D	Total	C/S	V/D	Total	C/S	V/D	Total	C/S	V/D	Total
Jan	2	11	13	2	8	10	2	1	3	2	11	13	2	9	11
Feb	3	10	13	4	15	19	0	4	4	0	6	6	0	9	9
Mar	1	9	10	5	13	18	0	3	3	3	13	16	4	6	10
Apr	4	8	12	5	10	15	3	6	9	3	8	11	1	4	5
May	1	14	15	1	14	15	0	11	11	3	10	13	1	9	10
Jun	3	12	15	3	8	11	1	8	9	2	10	12	2	9	11
July	2	6	8	0	4	4	4	9	13	3	4	7	1	8	9
Aug	0	11	11	0	5	5	2	8	10	2	8	10	1	8	9
Sep	1	10	11	1	6	7	3	3	9	2	4	6	2	7	9
Oct	2	6	8	0	8	8	5	16	21	0	6	6	1	8	9
Nov	9	12	21	3	4	7	5	11	16	2	7	9	3	4	7
Dec	2	6	8	1	4	5	0	14	14	0	6	6	2	11	13
Tot	30	115	145	25	99	124	25	94	120	22	93	115	20	92	112

(Source: Regina Caeli Hospital Awka, Anambra State. 'Labour Unit')

And the conditional variance is given as

$$Var\,(N/C) = (n-j)(1-\theta)p\theta\left(1-q\theta\right)^{-2} \qquad (18)$$

Also $E(C/N) = (n-i)(q\theta)\left(1-p\theta\right)^{-1}$ (19)

$$Var\,(C/N) = (n-i)(1-\theta)q\theta\left(1-p\theta\right)^{-2} \qquad (20)$$

Expression for the product moment correlation coefficient is given as

$$\rho(N,C) = \frac{-pq\theta}{\sqrt{pq(1-p\theta)(1-q\theta)}} \qquad (21)$$

(ii) Let N be a Poisson random variable with parameter $\lambda = n\theta$.

Then

$$H(t_1,t_2;n) = \exp(\lambda p t_1 + \lambda q t_2 - \lambda) \qquad (22)$$

Hence

$$p(N=i, C=j) = (\lambda)^{i+j}(p)^i(q)^j(e)^{-\lambda};.$$
$$i, j = 1,2,... \qquad (23)$$

$$p(N=i) = \frac{(\lambda p)^i (e)^{-\lambda p}}{i!} \qquad (24)$$

$$p(C=j) = \frac{(\lambda q)^j (e)^{-\lambda q}}{j!} \qquad (25)$$

Observe that $p(N=i, C=j) = p(N=i)\,p(C=j)$ which implies independence and that N and C are both Poisson random variables.
Therefore

$$E(N) = \lambda p \qquad (26)$$

$$Var(N) = \lambda p \qquad (27)$$

Usually for Poisson distribution, $E(N) = \lambda p$ is equal to $Var(N) = \lambda p$. Similarly, $E(C) = \lambda q$ is equal to $Var(C) = \lambda q$ as in equations (26) and (27) with p replaced with q.

$$p(N=i/C=j) = \frac{(\lambda p)^i (e)^{-\lambda p}}{i!} \qquad (28)$$

$$p(C=j/N=i) = \frac{(\lambda q)^j (e)^{-\lambda q}}{j!} \qquad (29)$$

So that $E(N/C) = \lambda p$ (30)

and $Var(N/C) = \lambda p$ (31)

Thus $Var(N/C) = Var(N)$ (32)

$E(C/N)$ and $Var(C/N) = Var(C)$ are obtained by replacing p with q in equations (31) and (32).

Hence, $\rho(N,C) = 0$ (33)

RESULTS AND DISCUSSION

The results obtained are illustrated herein using data collected from delivery ward of Regina Caeli Hospital (between 2006 and 2010) in Awka, Anambra State, Nigeria.

The data used in this application is presented in table 1.

Table 2. Goodness of Fit and Other Estimates for Binomial and Poisson Models.

S/N	Binomial		Poisson	
	E (N/C)	Var (N/C)	E (N/C)	Var (N/C)
j = 0	11.4229	2.7241	0.4034	0.4034
j = 1	10.6614	2.5425	0.4034	0.4034
j = 2	9.8999	2.3608	0.4034	0.4034
j = 3	9.1383	2.1792	0.4034	0.4034
j = 4	8.3768	1.9976	0.4034	0.4034
j = 5	7.6153	1.8160	0.4034	0.4034
j = 6	6.8537	1.6344	0.4034	0.4034
j = 7	6.0922	1.4528	0.4034	0.4034
j = 8	5.3307	1.2712	0.4034	0.4034
j = 9	4.5692	1.0896	0.4034	0.4034
j = 10	3.8076	0.9080	0.4034	0.4034
j = 11	3.0461	0.7264	0.4034	0.4034
j = 12	2.2846	0.5448	0.4034	0.4034
j = 13	1.5231	0.3632	0.4034	0.4034
j = 14	0.7615	0.1816	0.4034	0.4034
j = 15	0.0000	0.0000	0.4034	0.4034
D(N)	0.3575		0.0691**	
ρ(N,C)	-0.8407		0	
$\hat{p}$	0.1984		0.1984	
$\hat{q}$	0.8016		0.8016	
$\hat{\theta}$	0.9415		-	
$\hat{\lambda}$	-		2.0333	

** Significant at 0.1%

From table 2 shown above, using the Kolmogorov-Sminov test for goodness of fit test, it was observed that the Poisson distribution provides a better fit atthe 1% significance levell. Even though the binomial performs better than the Poisson distribution, but the model does not adequately fit the data. For the Poisson distribution model, the correlation between the number of normal deliveries and the number of cesarean deliveries in a hospital is *0*.The Poisson model also showed that the expected number of normal deliveries for a given number of cesarean deliveries is constant. The binomial model also showed that the expected number of normal deliveries for a given number of cesarean deliveries tend to decrease slowly as the amount of cesarean deliveries increase. For the binomial model, the correlation between the number of normal deliveries and the number of cesarean deliveries in a hospital is negative with a value of *-0.8407*. It is however worthy of note to observe that the predictions are reasonable for up to *7* normal deliveries or *7* cesarean deliveries, this may not be unconnected with the negative skewness of the distribution of deliveries according to the number of children.

REFERENCES

Feller, W. 1968. An Introduction to Probability Theory and its Applications. (vol. 1). Wiley, New York, USA.

Gupta, RC. 1976. Some Applications of Modified Power Series Distributions in Genetics. Sankhya. 38(13):187-191.

Janardan, KG. 1975. Markov-Polya Urn Models with Predetermined Strategies-1. Gujarat Statistical Review. 2(2):17-32.

Janardan, KG. and Rao, R.1984. Certain inter-relationships between two classes of Probability Models for Ascertainment of Families in Segregation Analysis. Journal of Indian Statistical Association. 22:1-12.

Onankpa, B. and Ekele, B. 2009. Fetal Outcome following Cesarean Section in a University Teaching Hospital. J Natl Med Assoc. 101(6):578-81.

Sharma, HL.1985. Some Probability Models for ascertainment of Families in Segregation Analysis Janasamkhya. 3(2):97-102.

Stene, J. 1978. Ascertainment Models for Human Families selected through Several Affected Children. Biometrics. 34:255-257.

Received: Jan 30, 2012; Revised: Feb 8, 2012;
Accepted: May 12, 2012