RESEARCH PAPER | Statistics | Volume : 5 | Issue : 9 | September 2015 | ISSN - 2249-555X

Logistic Regression: A Tool for Estimating Still Birth

KEYWORDS: Logistic regression Model, Still Births

Dipti J Bhavsar	Dr. C D Bhavsar
C.C.Sheth Colege of Commerce Navgujarat Campus Ashram Road Ahmedabad	Gujarat University, Dept. of Statistics

ABSTRACT *Integrated programme of health and family planning of Government of India could be made more realistic if it could be based on an adequate knowledge of the factors associated with foetal wastage and infant mortality rates. Shah K. S. (1982) considered the study of factor influencing on foetal wastage in Anandtown. DiptiBhavsaretc (2012) considered the study for Gandhinagar districtfor factor influencing onfoetal wastage using binary multiple regression technique. In this paper, Logistic regression model have been applied for the detailed study of factor influencing on still birth. An attempt has also been made to know the factors which affect the still births and the appropriate reasons to decrease the rate of foetal wastage. Since the data for abortions is not available, it has been obtained from the results for still births.*

1. Introduction:

The causes of still birth and abortion are multiple, the importance of each of those causes is not implicitly determined. One can examine the crude impact of the factors: Sex ,Type of delivery and Maternal age on still birth rate, on one hand and the relative effect of each of these factors individually, in the absence of the influence of the other factors on the other hand. Child which survives longer grows into a full man and becomes a citizen and makes his/her country progress further. Unfortunately, In India, it has been observed that there are number of foetal wastages and early neonatal deaths. There are various factors to this natural calamity, in the problem of human survival. Fetal wastage means, the total number of still births and abortions. If it can be decreased, the mothers have a good health. The causes of still birth are as follows:

1. The non-availability of medical treatment during the pregnancy period.
2. The lack of nutritious food.
3. To save the pregnant mother's life.
4. Pregnant mother's ignorance about the changes in the condition of her health.

Two types of abortions take place in society.

1. Abortions resulting from illegitimate relations.
2. Abortions with approval of law, society or religion.

1.1Abortions resulting from illegitimate relations:

Sometimes illegitimate sexual relations take place in society. Supposing a lonely woman is raped or seduced in a lonely place by an unknown man. She has not entered the bond of wedlock with him. And because of the forced sexual act she becomes pregnant. She is required to get abortion. If she does not get abortion, she will be rejected by her family and society. Her whole family has to suffer because of this scandal. This abortion resulting out of illegitimate relation is resorted to not only by unmarried woman but by married woman and widows also. Sometimes it also happens that a woman unwillingly sells her body just to earn her live hood. Sometimes women with unfulfilled desires like to sleep with other men. If such women are either ignorant about the contraceptives or careless in some cases, they become pregnant. Such women are forced to get abortions. It is not possible to avoid abortions of this kind because the social evils are too strong and powerful. It is not possible for a researcher to get the details about the exact number of such abortions as the women in question do not get their names registered in hospitals records. They keep the abortion a secret in order to escape public fury. So it is not possible to find out effective means to avoid such abortions.

1.2Abortions with approval of law, society or religion:

Women get their names registered in the hospital records without any fear because these abortions are legal. The occasions on which women get the legal abortions are as follows:

- When the child in the womb of mother is weak because of malnutrition, the doctor advices such mothers to get abortions.
- When a mother has a large number of children, the doctor advises her to take recourse to abortion.
- When a pregnant mother's life is in danger, the doctor asks her to get abortion.
- Some couples are not willing to take up the responsibility of bringing up children in their early married life. So in order to avoid this responsibility they resort to abortions.

It is not possible to stop abortions resulting from illegitimate relations. But one can decrease the number of abortions which are approved by the law, society or religion, by giving the nutritious food to pregnant mother, knowledge of family planning and medical treatment during her pregnancy period.

It is extremely difficult to collect complete and accurate information about pregnancy losses in any retrospective study. The problems of detecting early miscarriages, memory lapses, unwillingness to divulge information on abortions and still births, lack of knowledge in distinguishing a still birth from the death of a new born infants etc. are well known and make accurate assessment of pregnancy wastage difficult. To overcome such difficulties data are collected from the case card records of mothers registered for delivery, during the period 2007-2009 in Gandhinagar-CivilHospital, Gandhinagar. During that period 4313 pregnancies are reported with complete information as per Performa. And out of these 137 are found to have ended into still births.

2. Methodology:
To study the influence of each of the factors under investigation on still birth rate a Logistic regression method is adopted. In this study, still birth rate is assumed to have the impact of the following three factors: (1) Maternal Age (2) Types of Delivery. (3) Sex of the infant. Each factor is divided into following subclasses:

Maternal Age: Mother included in this study is divided into r_1 = 5 subgroups with regards to their age: (1) less than 20 years (2) 20-24 years (3) 25-29 years (4) 30-34 years and (5) above 35 year. Four binary variables X_1, X_2, X_3, and X_4 are used to denote these subclasses. (0, 0, 0, 0,) means that mother belong to the class (5) i.e., they are of age 35 years and above.

Types of Delivery: Different types of deliveries are divided in to 4 subclasses. Three binary variables X_5, X_6 and X7areused to denote Normal, L.S.D.S and Breech type of delivery respectively for these subclasses. (0, 0, 0) means that mother belong to the class (4) i.e., the type of delivery of a mother is Forceps.

Sex: Still births are divided into r_3 = 2 categories according to their sex: (1) Male (2) Female. One binary variable X_8 is used to denote these subclasses. (0) means that the live birth is female.

For the purpose of multiple regression analysis, all the variables are considered as binary variables taking the value 0 or 1. And each subclass of the variable is considered here as separate regressor (Feldstein, 1966). The regression equation of the still births on the independent variables mentioned above is given as follows:

$$y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + + \beta_7 X_7 + \beta_8 X_8 + \varepsilon_t \qquad (2.1)$$

Where y= 1 if birth results in a still births.

= 0 if birth results is not in a still births.

X_0 = Dummy variable (always).

(X_1, X_2, X_3, X_4)

= (1, 0, 0, 0) If the birth belong to mother with age less than 20 years.

= (0, 1, 0, 0) If the birth belong to mother with age 20-24 years.

= (0, 0, 1, 0) If the birth belong to mother with age 25-29 years.

= (0, 0, 0, 1) If the birth belong to mother with age 30-34 years.

= (0, 0, 0, 0) If the birth belong to mother with age 35 years and above.

(X_5, X_6, X_7)

= (1, 0, 0) If the type of delivery is Normal.

= (0, 1, 0) If the type of delivery is L.S.D.S

= (0, 0, 1) If the type of delivery is Breech.

= (0, 0, 0) If the type of delivery is Forceps.

X_8

= 1 if live birth is male.

= 0 if live birth is female.

ε_t

= Error component such that $E(\varepsilon_t) = 0$ and $E(X_{it}, \varepsilon_{t+\theta}) = 0$ for all i, t and θ.

Since the interest in this study is to measure the average effect of each factor on the still birth mortality rather than its prediction, it has not assumed interactions between the effects of the four different factors in the above model. The above regression model can be rewritten as follows:

$$E(y) = X\beta \qquad (2.2)$$

Where y is the vector of N observations, X is the **N * 9** matrix (X_0, X_1, X_2, X_3, X_4, X_5, X_6,X_7, X_8) of zeros and ones and β is the 9*1 vector of regression coefficients, (β_0 , β_1 , β_2 , β_3 , β_4 , β_5 , β_6 , β_7, β_8).

3. Logistic regression Model:
When the usual assumptions of normality and constant variance are not satisfied the generalized linear model (GLM) is useful to handle the non constant variance problem. In a GLM, the response variable distribution must be a member of the exponential family, which includes the normal, Poisson, binomial, exponential and gamma distributions as members. Logistic regeression model is just a special case of the general linear model. If the response variable has only two possible outcomes, generally called "Success" and "failure" and denoted by 0 and 1 then the logistic regression model is applied. Note that here the response is essentially qualitative, since the designation "success" or "failure" is entirely arbitrary.

3.1Models with a Binary Response Variable:
Consider the situation where the response variable in a regression problem takes on only two possible values, 0 and 1. Suppose thatthe model has the form

$$y_i = x_i' \beta + \epsilon_i \qquad (3.1)$$

Where $x_i' = (1, x_{i1}, x_{i2}, \ldots, x_{i8})$, $\beta' = (\beta_0, \beta_1, \ldots, \beta_8)$

and the response variable y_i takes the value 0 or 1(i=1,2,…,N). If the response variable y_i is a Bernoulli random variable with probability distribution function as follows:

y_i	Probability
1	$P(y_i = 1)$
0	$P(y_i = 0)$

Now since E(€i) = 0, the expected value of the response variable is

$$E(y_i) = 1\pi_i + 0(1 - \pi_i) = \pi_i$$

This implies that

$$E(y_i) = x_i' \beta = \pi_i$$

This means that the expected response given by the response function

$E(y_i) = \underline{x}_i' \underline{\beta}$ Is just the probability that the response variable takes on the value 1. Note that if the response is binary, then the error terms ε_i can take on only two values, namely

$$\epsilon_i = 1 - \underline{x}_i' \underline{\beta} \quad \text{when } y_i = 1$$
$$\epsilon_i = -\underline{x}_i' \underline{\beta} \quad \text{when } y_i = 0$$

Consequently, the error in the model is not normal and the error variance is not constant, since

$$\sigma^2_{y_i} = (E\{y_i - E(y_i)\}^2$$
$$= (1-\Pi_i)^2 \Pi_i + (1-\Pi_i)^2 (1-\Pi_i)$$
$$= \Pi_i (1-\Pi_i)$$

Note that the last expression is just

$$\sigma^2_{y_i} = E(y_i)[1 - E(y_i)]$$
$$E(y_i) = \underline{x}_i' \underline{\beta} = \Pi_i$$

This indicates that the variance of the observations is a function of the mean, and hence there is a constraint on the response function, because $0 \le E(y_i) = \Pi_i \le 1$.

This restriction can cause serious problems with the choice of a linear response function, it has been assumed in equation (3.1). It would be possible to fit a model to the data for which the predicted values of the response lie outside the 0, 1 interval. Generally when the response variable is binary, there is considerable empirical evidence indicating that the shape of the response function should be nonlinear. This non linear function is called logistic regression function, and has the form

$$E(y) = \frac{\exp(x'\beta)}{1+\exp(x'\beta)} \quad (3.2)$$

Or equivalently

$$E(y) = \frac{1}{1+\exp(-x'\beta)} \quad (3.3)$$

The logistic response function can be easily lineralized. Let

$$\eta = x'\beta$$

Be the linear predictor where ε is defined by the transformation

$$\eta = \ln \frac{\pi}{1-\kappa}$$

This transformation is often called the logit transformation of the probability π, and the ratio π / (1- π) in the transformation is called the odds. Sometimes the logit transformation is called the log odds.

3.2 Estimation of the Parameters:

The general form of the logistic regression model is

$$y_i = \underline{x}_i' \underline{\beta} + \epsilon_i$$

Where the observations y_i are independent Bernoulli random variables with expected values

$$E(y_i) = \Pi_i$$
$$= \frac{\exp(x'\beta)}{1+\exp(x'\beta)}$$

If $\underline{\hat{\beta}}$ is the maximum likelihood estimate of $\underline{\beta}$ then asymptotically, one can have

$$E(\underline{\hat{\beta}}) = \underline{\beta} \text{ and } \operatorname{var}(\underline{\hat{\beta}}) = (x' v^{-1} x)^{-1}$$

Provided the model assumptions are correct. Hence, the estimated value of the linear predictor is, $\hat{\eta}_i = \underline{x}_i' \underline{\hat{\beta}}$ and the fittedvalue of the logistic regression model is

$$\hat{y}_i = \hat{\pi}_i = \frac{\exp(\hat{\eta}_i)}{1+\exp(\hat{\eta}_i)}$$
$$= \frac{\exp(\underline{x}_i' \underline{\hat{\beta}})}{1+\exp(\underline{x}_i' \underline{\hat{\beta}})}$$
$$= \frac{1}{1+\exp(\underline{x}_i' \underline{\hat{\beta}})}$$

4. Interpretation:

It is relatively easy to interpret the parameters in a logistic regression model. If the linear predictor has only a single regressor then the fitted value of the model at a particular value of x, say $\hat{x}_i$, is

$$\hat{\eta}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

The fitted value at $(x_i + 1)$ is

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \hat{\beta}_1$$

Now, $\hat{\eta}(x_i)$ are just the log odds when the regressor variable is equal to x_i and $\hat{\eta}(x_i + 1)$ is just the log odds when the regressor is equal to $x_i + 1$

If one takes antilog, the odd ratio is obtained

$$\hat{O}_R = \frac{odds_{x_i+1}}{odds_{x_i}} = e^{\hat{\beta}_1} \quad (3.4)$$

The odds ratio is interpreted as the estimated increase in the probability of success associated with a unit change in the value of the predictor variable. In general, the estimated increase in the odds ratio associated with a change of d units in the predictor variable in $\exp(d\hat{\beta}_1)$.

Note that the interpretation of the regression coefficients in the multiple logistic regression models is similar to that for the case where the linear predictor contains only one regressor. That is, the quantity $\exp(\hat{\beta}_j)$ is the odds ratio for regressor x_j, assuming that all other predictor variables are constant.

In this case, the fitted logistic regression model for sex is

$$\hat{y} = \frac{1}{1+e^{\beta_0+\beta_1 x}} = \frac{1}{1+e^{-3.556+0.191x}}$$

to the still birth data. Since the linear predictor contains only one regressor variable (sex) and $\hat{\beta}_1$ **= .191,** one can compute the odds ratio from equation (3.4) as

$$\hat{O}_R = e^{\hat{\beta}_1} = e^{\square} = 1.21046$$

This implies that every additional number of male increase the odds of still birth by 20 %. If the number of male increase by 5 (10) then the odds ratio becomes $\exp(5\hat{\beta}_1)$ = 2.60 (6.75). This indicates that the odds increase more than twice (five times) with five (ten) number of male.

The fitted logistic regression model for type of delivery is

$$\hat{y} = \frac{1}{1 + e^{\beta 0 + \beta 1\, x1 + \beta 2\, x2 + \beta 3\, x3}}$$

$$= \frac{1}{1 + e^{-21.203 + 17.815 x_5 + 15.504 x_6 + 19.091 x_7}}$$

$$O_{R\hat{\beta}1} = e^{\hat{\beta}1} = e^{17.82} = 54,843,816.3$$

$$O_{R\hat{\beta}2} = e^{\hat{\beta}2} = 5,389,698.5$$

Similarly all odds are too high sothat one can't conclude finallythat there is no effect of type of delivery on still birth

$$\hat{y} = \frac{1}{1 + e^{\beta 0 + \beta 1\, x1 + \beta 2\, x2 + \beta 3\, x3 + \beta 4\, x4}}$$

$$= \frac{1}{1 + e^{-3.496 - 0.270 x_1 - 0.296 x_2 + 0.194 x_3 + 0.51}}$$

We can observe that

$$e^{\hat{\beta}3} = e^{.194} = 1.21$$

$$e^{\hat{\beta}4} = e^{.512} = 1.66$$

$O_{R\hat{\beta}3} = 1.21$ (25-29 age group)

$O_{R\hat{\beta}4} = 1.66$ (30-34 age group)

The values of odds ratio $O_{R\hat{\beta}3}$ and $O_{R\hat{\beta}4}$ indicates that the number of female in the age group 25-29 and 30-34 are affect on the number of still birth significantly.

REFERENCE 1. Douglas C. Montgomery, Elizabeth A. Peck and G. Geoffrey Vining (2001), "Introduction to Linear Regression Analysis", Third edition, John Wiley and Sons, Inc. New York. | 2. DiptiBhavsar and C.D.bhavsar (2012) Factor Influencing Foatel Wastage. Indian Journal of Applied Research; Vol-1, Issue-7, April 2012, pp. 185-188. (ISSN-2249-555x) | 3. Feldstein, M. S. (1966), 'A Binary Variable Multiple Regression Method of Analyzing Factors Affecting Perinatal Mortality and other outcomes of pregnancy. Journal of the Royal Statistical Society, Series , 129, Part I, pp. 61-73. | 4. Gandotra M. M., Das N and Bhatt R. V. (1975) Foetal Wastage in Indian Women Monograph Series No. 29, Demographic Research Center Baroda. | 5. Potter,R. G., Wyon, J. B., New, N. and Corden, J. E., (1965): Foetal Wastage in Eleven Punjab Villages', Ruman Biology, 37, September, pp. 262-273. | 6. Shah, P and Hallen Abbey, (1971): Effects of some factors on neo natal and Post neo natal mortality. Analysis by a Binary Variable Multiple Regression Method', Mill Bank Memorial Fund Quarterly. Vol 49, No. 1, January,pp. 33 – 55. | 7. Shah K. S. (1982) "Population explosion" Subline publications Jaipur 302006 (India) 1995. | 8. Shapiro, S. N. Jones, W., and Densan, P. M., (1962): A life table of Pregnancy termination and correlates of Foetal losses. Mill Bank Memorial Fund Quarterly, Vol 40, January, pp. 7 – 45. |