CORNERSTONE
MINNESOTA STATE UNIVERSITY MANKATO

Journal of Undergraduate Research at Minnesota State University, Mankato

Volume 3 Article 2

2003

Inferences in Log-Rate Models

Herbert C. Heien
Minnesota State University, Mankato

William A. Baumann
Minnesota State University, Mankato

Inferences in Log-Rate Models

Herbert C. Heien
(herbert.heien@mnsu.edu)

William A. Baumann
(william.baumann@mnsu.edu)

and

Mezbahur Rahman
(mezbahur.rahman@mnsu.edu)

Department of Mathematics and Statistics

Minnesota State University, Mankato, MN 56002, USA

Abstract

Log-Rate models are used in analyzing rates of individuals who are exposed to a risk of having a certain characteristic. The explanatory variables could be categorical or in a continuous scale. In finding a Log-Rate Model, parameters are estimated and goodness-of-fit are studied to carefully extract the best model to fit our data. Here we revisit three aspects of Log-Rate Models using the data set give at the end of the paper. The three aspects are parameter estimation, goodness-of-fit of the model, and marginal effect of the factors.

1. Introduction

In the categorical data analysis literatures, the survival models and/or the hazard rate models are treated differently than standard logit models. In general, these models are termed as Rate Models or Log-Rate Models. In its simplest form, a rate is defined as the number of individuals or observations possessing a particular characteristic divided by the total amount of exposure to the risk of having such a characteristic. The Rate Models can easily be connected to the standard Poisson Models. Then the Poisson Models are directly related to the Exponential Models by making conversion of rates per unit interval with the waiting time until the first occurrence. Here we use a Log-Rate Model to determine the likelihood of premarital births in adolescent populations.

Let $t_1, t_2, \ldots, t_n$ be the waiting times of n individuals, and assume the distribution function to be $F(t) = \Pr(T < t)$ with probability density function $f(t)$. The hazard rate is denoted by $\Lambda(t)$, and can be viewed as the instantaneous probability of an event in the interval $[t, t + \Delta t]$, given the event has not occurred before time t. Formally, the hazard rate is defined by the following limit:

$$\Lambda(t) = \lim_{\substack{\Delta t \to 0 \\ \Delta t > 0}} \frac{1}{\Delta t} \Pr[t \le T < t + \Delta t \mid T \ge t]. \tag{1}$$

The probability of an event not occurring up to time t is given by the function

$$S(t) = \Pr[T > t] = 1 - F(t) = \int_t^{\infty} f(u)du. \tag{2}$$

Assuming the waiting times are exponentially distributed, equation (2) may be written as

$$S(t) = \exp(-\Lambda t_i). \tag{3}$$

The hazard rate is defined by the ratio

$$\Lambda(t_i) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} = \Lambda, \tag{4}$$

and the general hazard rate model may be written as

$$\Lambda(\mathbf{x}_i^T \boldsymbol{\beta}_i) = \exp(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{pi} x_{pi}), \tag{5}$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are unknown constants as the rate is determined by several regressors. This exponential hazard rate model can be estimated using a Poisson regression model for counts. The interested reader may see Powers and Xie (2000) pp. 154-156 for a brief explanation. In a time interval of length t, the probability of d events is given by

$$\Pr(d \mid \Lambda, t) = \frac{(t\Lambda)^d \exp(-t\Lambda)}{d!}. \tag{6}$$

Because the mean number of events in the time interval is $\mu = t\Lambda$, for the i^{th} individual, the expected number of events in the time interval is

$$\mu_i = t_i \Lambda_i = t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}_i). \tag{7}$$

Taking the log (log stands for natural logarithm) of the Poisson means results in the log-linear regression model

$$\log(\mu_i) - \log(t_i) = \mathbf{x}_i^T \boldsymbol{\beta}_i, \tag{8}$$

where $\mathbf{x}^T = [1, x_1, x_2, x_3, ..., x_p]$.

The plan of the paper is as follows: In sections 2 and 3 we discuss the estimation procedures and goodness-of-fit of the models. Marginal effects and their inferences are discussed in section 4. In section 5, we use a real life data to demonstrate the use of a log-rate model, and we write a brief conclusion and give future research potentials in section 6.

2. Estimation of Parameters

We wish to use the method of maximum likelihood to estimate the parameters of this model. Generally, the estimators are obtained by maximizing the logarithm of the likelihood function. The likelihood is defined as

$$\mathrm{L} = \prod_{i=1}^{n} \{t_i \Lambda(\mathbf{x}_i^T \boldsymbol{\beta}_i)\}^{d_i} \exp\{-t_i \Lambda(\mathbf{x}_i^T \boldsymbol{\beta}_i)\}/(d_i!), \tag{9}$$

where d_i is the number of occurrence of the event of interest in the time interval t_i and n is the sample size. The log-likelihood (log stands for natural logarithm) function is

$$\log \mathrm{L} = \sum_i \{d_i[\log(t_i) + \mathbf{x}_i^T \boldsymbol{\beta}_i] - t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}_i) - \log(d_i!)\}. \tag{10}$$

The system of likelihood equations

$$\frac{\delta \log(L(\boldsymbol{\beta}))}{\delta \boldsymbol{\beta}} = \mathbf{U}(\boldsymbol{\beta}) = \sum_i \{d_i - t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}_i)\}\mathbf{x}_i = \mathbf{0} \tag{11}$$

are nonlinear and hence we use the standard Newton-Raphson numerical solution method (see Scarborough (1979), pp. 201-203) to solve them. The second derivatives, which are used in iteration to find maximum likelihood estimates, are given by

$$-\left[\frac{\delta^2 \log(L(\boldsymbol{\beta}))}{\delta \boldsymbol{\beta} \boldsymbol{\beta}^T}\right] = \mathbf{I}(\boldsymbol{\beta}) = \sum_i \{t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}_i)\mathbf{x}_i \mathbf{x}_i^T\}. \tag{12}$$

At the $k^{\underline{th}}$ iteration, the estimates are obtained using the equation

$$\hat{\boldsymbol{\beta}}^{(k)} = \hat{\boldsymbol{\beta}}^{(k-1)} + \left[\mathbf{I}\left(\hat{\boldsymbol{\beta}}^{(k-1)}\right)\right]^{-1}\mathbf{U}\left(\hat{\boldsymbol{\beta}}^{(k-1)}\right), \tag{13}$$

where $\hat{\boldsymbol{\beta}}^{(0)}$ is obtained using the least squares method by regressing the *y* on the *x*'s in a linear regression set up as the link function is defined in (8). The iteration is stopped when the consecutive iteration values are close and/or the log-likelihood values are maximized (see Powers and Xie (2000) pp. 61-63for details).

3. Goodness-of-fit

Log L cannot be used alone as an index of fit because it is dependent on the size of the sample. Different values of log L result when competing models, models that differ in the number of parameters, are fit to the same data. The number of parameters, in general, should be more than one, and significantly less than the number of observations. To assess model fit, we need to know how one model fits relative to another. An indicator of model fit which measures the extent to which the current model deviates from a more generalized model is given by the likelihood-ratio statistic:

$$\mathrm{G}^2 = -2\log\left(\frac{L_c}{L_f}\right) = -2(\log L_c - \log L_f), \tag{14}$$

where $\log L_c$ is the log-likelihood of the current model, and $\log L_f$ is the log-likelihood of the more generalized model. The likelihood ratio statistic has a Chi-Square distribution with $K_2 - K_1$ degrees of freedom, where K_2 and K_1 denote the number of parameters in the more generalized model and the current model, respectively. A comparative study of different choices of general models can be seen in Simonoff (1998).

4. Marginal Effects

For the log-rate model, the marginal effects can be thought of as the relative risk associated with a certain variable. The overall mean effect in (7) is

$$\Lambda(\mathbf{x}^T\boldsymbol{\beta}) = \exp(\mathbf{x}^T\boldsymbol{\beta}). \tag{15}$$

Then the marginal effect due to the $\mathrm{k}^{\underline{\mathrm{th}}}$ factor can be considered as

$$\theta_k = \frac{\delta\Lambda(\mathbf{x}^T\boldsymbol{\beta})(\mathbf{x})}{\delta\beta_k} = \mathrm{x}_k\exp(\mathbf{x}^T\boldsymbol{\beta}). \tag{16}$$

The estimate of θ_k can be computed as

$$\hat{\theta}_k = \bar{x}_k \exp(\bar{\mathbf{x}}^T \hat{\boldsymbol{\beta}}) \tag{17}$$

where $\bar{x}_k$ is the mean of the $k^{\underline{th}}$ factor values in the sample and $\bar{\mathbf{x}}^T$ is the vector of the means of the factor values in the sample. The estimate of the variance for $\hat{\theta}_k$ can be obtained using the delta method (see Ramsey and Schafer (2002) pp. 328-329 for details) as follows: the first derivative of $\hat{\theta}_k$ with respect to $\hat{\beta}_i$'s are

$$\hat{\theta}'_{ki} = \frac{\delta \hat{\theta}_k}{\delta \hat{\beta}_i} x_k x_i \exp(\bar{\mathbf{x}}^T \hat{\boldsymbol{\beta}}), \text{ for } i = 0, 1, 2, \ldots, p. \tag{18}$$

Then the approximate variance of $\hat{\theta}_k$ can be written as

$$V(\hat{\theta}_k) = \sum_{i=0}^{p} \sum_{j=0}^{p} \theta'_{ki} \theta'_{kj} \text{Cov}(\hat{\beta}_i, \hat{\beta}_j). \tag{19}$$

The estimate of the variance can be obtained as

$$\hat{V}(\hat{\theta}_k) = \sum_{i=0}^{p} \sum_{j=0}^{p} \hat{\theta}'_{ki} \hat{\theta}'_{kj} \left[\mathbf{I}\left(\hat{\boldsymbol{\beta}}\right)\right]^{-1}_{lj}, \tag{20}$$

where $\left[\mathbf{I}\left(\hat{\boldsymbol{\beta}}\right)\right]^{-1}_{lj}$ is the lj^{th} element of the matrix $\left[\mathbf{I}\left(\hat{\boldsymbol{\beta}}\right)\right]^{-1}$ in (13).

5. Application

We consider data analyzed by Powers and Xie (2000). The data studied was occurrence-exposure data on premarital births to young women participating in the National Longitudinal Survey of Youth from 1979 to 1988. It provided retrospective information on the dates and occurrence of first birth and first marriage, and was categorized using three age intervals, race, and family structure

Information on those experiencing premarital births was recorded and is displayed in Table 1. In Table 1, “E” represents the number of person-months from age 14 to age at premarital birth, and “D” represents the cell-specific number of events.

The approach taken here is that which is presented in Simonoff (1998). Different log-rate models are compared to find the best model. A saturated model and its nested models are defined. We then use the deviations between the maximized log-likelihood from each model to perform a series of Chi-square tests so as to ascertain which model fits best as described in the Section 3.

The models used were:

$$\ln(\mu_i) - \ln(\mathrm{E_i}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 \quad (21)$$

$$\ln(\mu_i) - \ln(\mathrm{E_i}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \quad (22)$$

$$\ln(\mu_i) - \ln(\mathrm{E_i}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (23)$$

$$\ln(\mu_i) - \ln(\mathrm{E_i}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_4 x_4 \quad (24)$$

$$\ln(\mu_i) - \ln(\mathrm{E_i}) = \beta_0 + \beta_3 x_3 + \beta_4 x_4 \quad (25)$$

where x_1 is an indicator variable representing the age interval 16 - 18, x_2 is an indicator variable representing the age interval 18 - 20, x_3 is an indicator variable representing a non - intact family structure, x_4 is an indicator variable representing a non - white person, $x_5 = x_1 x_3$, $x_6 = x_2 x_3$, $x_7 = x_1 x_4$, $x_8 = x_2 x_4$ and $x_9 = x_3 x_4$.

Using SPSS, MATLAB, and Mathematica to perform the iterations necessary for the maximum likelihood method, the following results were obtained:

Model	Fitted Model	Maximum Log-Likelihood	Iterations Needed
(21)	$\ln(\mu_i) - \ln(\mathrm{E}_i) = -6.9593 + 1.3927x_1 + 2.6380x_2 + 0.7286x_3 + 1.0466x_4 - 0.00278x_5 - 0.0305x_6 - 0.1517x_7 - 0.1558x_8 - 0.1232x_9$	-4164.86	5
(22)	$\ln(\mu_i) - \ln(\mathrm{E}_i) = -6.80798 + 1.35019x_1 + 2.4327x_2 + 0.561x_3 + 0.904995x_4$	-4165.94	4
(23)	$\ln(\mu_i) - \ln(\mathrm{E}_i) = -6.2460 + 1.3310x_1 + 2.4148x_2 + 0.6647x_3$	-4234.20	4
(24)	$\ln(\mu_i) - \ln(\mathrm{E}_i) = -6.5705 + 1.3415x_1 + 2.4076x_2 + 0.9766x_4$	-4179.19	4
(25)	$\ln(\mu_i) - \ln(\mathrm{E}_i) = -5.57251 + 0.50343x_3 + 0.871393x_4$	-4486.82	3

Using these results, we tested the competing models using the likelihood-ratio statistic as described in section 3 in order to determine goodness-of-fit. To perform the tests, we started by testing the saturated model (21) against the main factors model (22), and then tested the main factors model against its nested counterparts. The results of the Chi-square tests, performed with $\alpha = 0.05$, are as follows:

Test	G^2	d.f.	Critical Value	Conclusion
(22) vs. (21)	-2[(-4165.94)-(-4164.86)] = 2.16	6	12.59	Adequate fit for (20)
(23) vs. (22)	-2[(-4234.20)-(-4165.94)] = 136.52	1	3.84	Adequate fit for (20)
(24) vs. (22)	-2[(-4179.19)-(-4165.94)] = 26.5	1	3.84	Adequate fit for (20)
(25) vs. (22)	-2[(-4486.82)-(-4165.94)] = 641.76	2	5.99	Adequate fit for (20)

The main factors model (22) compared to the saturated model (with all the interactions) (21) had adequate fit. Model (22) had adequate fit compared to all the other models. Thus, we decided to qualify the main factors model (22) as the adequate model for this data.

The marginal effects are computed as described in the Section 4. The marginal effect for the first factor, age interval 16-18, is calculated as $\hat{\theta}_1 = \exp(\hat{\beta}_1) = 3.858$. This means, the target population of age group 16-18 has a 3.858 times higher rate of premarital births than the other age groups. A similar interpretation can be made for the marginal effect for the second factor, age interval 18-20, which is $\hat{\theta}_2 = \exp(\hat{\beta}_2) = 11.390$. The marginal effect for the third factor, non-intact family structure, is $\hat{\theta}_3 = \exp(\hat{\beta}_3) = 1.752$, which means that non-intact families have a 1.752 times higher rate of premarital births than intact families and the marginal effect for the fourth factor, non-white, is $\hat{\theta}_4 = \exp(\hat{\beta}_4) = 2.472$, which means that non-white families have a 2.472 times higher rate of premarital births than white families.

The variances of the marginal effects are computed using the delta method mentioned in section 4. The computed variance-covariance matrix for the estimates of the parameters is,

$$\mathrm{V}\left(\underline{\hat{\beta}}\right) = \left[\mathbf{I}\left(\underline{\hat{\beta}}\right)\right]^{-1} = \begin{bmatrix} 0.0132 & -0.0080 & -0.0080 & -0.0025 & -0.0050 \\ -0.0080 & 0.0107 & 0.0078 & 0.0001 & 0.0001 \\ -0.0080 & 0.0078 & 0.0109 & 0.0002 & 0.0001 \\ -0.0025 & 0.0001 & 0.0002 & 0.0052 & -0.0007 \\ -0.0050 & 0.0001 & 0.0001 & -0.0007 & 0.0070 \end{bmatrix}.$$

The partial derivatives of the marginal effects are computed as

Marginal Effect	$\hat{\theta}'_k$
16-18	3.858
18-20	11.390
Non-Intact	1.752
Non-White	2.472

Then, using (20), the variances are computed as $V(\hat{\theta}_1) = 0.0413$, $V(\hat{\theta}_2) = 0.1241$, $V(\hat{\theta}_3) = 0.0091$, and $V(\hat{\theta}_4) = 0.0173$. These estimates of variances would be used in testing and in finding confidence intervals for the corresponding marginal effects to see the significance of the marginal effect estimates.

6. Conclusion

In conclusion, we see that the older age group has a higher rate of premarital births. Non-white have 2.472 times higher incidence of premarital births. Non-intact families have a 1.752 higher birth rate than intact families.

Such procedures can be applied to any risk exposure data in which are categorical and/or quantitative in nature, such as studies of rare diseases in different cross-sections of the society. For too many iterations in the parameter estimation procedure, estimates might have higher variations and the log-likelihood function is not maximized.

Table 1

	Intact				Non-intact			
	White		Non-white		White		Non-white	
Age	D	E	D	E	D	E	D	E
14-16	17	13220	33	13838	10	7332	68	12827
16-18	39	10266	104	9823	42	5417	160	8516
18-20	43	3552	112	3331	42	1599	128	2594

7. References

Cox, D. R. and Snell, E. J. (1989), *Analysis of Binary Data*, 2nd edition. London: Chapman and Hall.

Efron, Bradley (1983), Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation, *Journal of the American Statistical Association*,Vol.78 (382), pp. 316-331.

Efron, Bradley (1986), How Biased is the Apparent Error Rate of a Prediction Rule?, *Journal of the American Statistical Association*, pp. 461-470.

McCullagh, P. and J. A. Nelder (1989), *Generalized Linear Models*, 2nd edition. New York: Chapman and Hall.

Powers, Daniel A. and Xie, Yu (2000), *Statistical Methods for Categorical Data Analysis*. San Diego, CA: Academic Press.

Rahman, Mezbuhar, Cortes, Judy and Pardis, Cyrus (2001), A Note on Logistic Regression, *Journal of Statistics & Management Systems*, Vol. 4, No. 2, pp. 175-187.

Ramsey, Fred L. and Schafer, Daniel W. (2002), *The Statistical Sleuth: A Course in Methods of Data Analysis*, 2nd edition. Pacific Grove, CA: Duxbury.

Ryan, Thomas P. (1997), *Modern Regression Methods*, New York: Wiley.

Scarborough, James B. (1979), *Numerical Mathematical Analysis.* Baltimore, MD: The Johns Hopkins Press.

Siminoff, Jeffrey S. (1998), Logistic Regression, Categorical Predictors, and Goodness-of-Fit: It Depends on Who You Ask, *The American Statistician*, Vol. 52, No.1, pp. 10-14.

Van Houwelingen, J. C. and S. Le Cessie (1990), Predictive Value of Statistical Models, *Statistics in Medicine*, 1303-1325.