The Korean Communications
in Statistics Vol. 10 No. 3, 2003
pp. 997-1005

The Confidence Intervals for Logistic Model in Contingency Table[1]

Tae Kyoung Cho[2]

Abstract

We can use the logistic model for categorical data when the response variables are binary data. In this paper we consider the problem of constructing the confidence intervals for logistic model in $I \times J \times 2$ contingency table. These constructions are simplified by applying logit transformation. This transforms the problem to consider linear form which called the logit model. After obtaining the confidence intervals for the logit model, the reverse transform is applied to obtain the confidence intervals for the logistic model.

Keywords : logistic model, contingency table, logit model

1. Introduction

A categorical variable is one for which the measurement scale consists of a set of categories. Categorical scales are common in the social and biomedical science. For instance, demographic characters may be measured such as gender(male, female), race(white, black, yellow, others), and social class(upper, middle, lower); smoking status might be measured using categories "never smoked," "former smoker," and "current smoker". Contingency tables are used to summaries the number of observations with the corresponding values of the categorical values. Goodman and Kruskal (1979) summarized the historical development of measures of association for contingency tables.

We consider the $I \times J \times 2$ contingency table. Suppose there are two factors, A and B, for binary response. Let I denote the number of levels of A, J denote the number of levels of B, and 2 denote the number of responses. The cells of the table represent the IJ possible outcomes.

Let π_{ij} denote the probability that (A, B) falls in the cell in row i and column j. Let the binary response variable, Y_{ij}, be values 0 or 1. The logistic model specifies the

1) This work was supported by the research fund of Dongguk University
2) Assistant Professor, Dept. of Statistics and Information Science, Dongguk University, Kyongju, 780-714, Korea
E-mail : tkcho@mail.dongguk.ac.kr

probability of response, when $Y_{ij}=1$, as

$$\pi_{1|ij}=\frac{\exp(\mu+\alpha_i+\beta_j)}{1+\exp(\mu+\alpha_i+\beta_j)}$$

where μ is the effect of the general mean,
α_i is the effect of the factor A,
β_j is the effect of the factor B.

We assume that $\sum_{i=1}^{I}\alpha_i=0$ and $\sum_{j=1}^{J}\beta_j=0$. The logit model is defined as

$$\ln\left(\frac{\pi_{1|ij}}{1-\pi_{1|ij}}\right)=\mu+\alpha_i+\beta_j$$

When more than one observation on Y_{ij} occurs, it is sufficient to record the number of observations m_{ij}. Let $Y_{1|ij}$ denote the number of times response "1" occurs when the factor A is at level i and the factor B is at level j.

In this paper we consider the problem of constructing the confidence intervals on $\pi_{1|ij}$ for all i and j.

Hodges (1958), Cox (1970) and Albert and Anderson (1984) discussed maximum likelihood estimation for logistic model. Silvapull(1981) made necessary and sufficient conditions for the maximum likelihood estimators for logit model.

When binary responses are independent Bernoulli random variables, $\{Y_{1|ij}\}$ are independent binomial random variables with parameter $\{\pi_{1|ij}\}$. Thus, the likelihood function is

$$l(\mu,\alpha_1,\cdots,\alpha_I,\beta_1,\cdots,\beta_J)=\prod_{i=1}^{I}\prod_{j=1}^{J}\binom{m_{ij}}{y_{1|ij}}(\pi_{1|ij})^{y_{1|ij}}(1-\pi_{1|ij})^{m_{ij}-y_{1|ij}}$$

The log likelihood function is defined as

$$\begin{aligned}L&=\ln[l(\mu,\alpha_1,\cdots,\alpha_I,\beta_1,\cdots,\beta_J)]\\&=\sum_{i=1}^{I}\sum_{j=1}^{J}\left[\ln\binom{m_{ij}}{y_{1|ij}}+y_{1|ij}(\mu+\alpha_i+\beta_j)-m_{ij}\ln(1+\exp(\mu+\alpha_i+\beta_j))\right]\end{aligned}$$

To find the values of $\mu, \alpha_1, \cdots, \alpha_I, \beta_1, \cdots, \beta_J$ that maximize L we differentiate L with respect to $\mu, \alpha_1, \cdots, \alpha_I, \beta_1, \cdots, \beta_J$, respectively, and set the resulting expression equal to zero. The likelihood equations are as follows:

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{I}\sum_{j=1}^{J}(y_{1|ij} - m_{ij}\ \hat{\pi}_{1|ij}) = 0$$

$$\frac{\partial L}{\partial \alpha_i} = \sum_{j=1}^{J}(y_{1|ij} - m_{ij}\ \hat{\pi}_{1|ij}) = 0 \quad \text{for} \quad i = 1,\ 2,\ \cdots,\ I$$

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{I}(y_{1|ij} - m_{ij}\ \hat{\pi}_{1|ij}) = 0 \quad \text{for} \quad j = 1,\ 2,\ \cdots,\ J$$

where $\hat{\pi}_{1|ij} = \dfrac{\exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)}{1 + \exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)}$ denotes the maximum likelihood estimate of $\pi_{1|ij}$.

Because the likelihood equations are nonlinear functions of maximum likelihood estimates $\hat{\mu}, \hat{\alpha}_1, \cdots, \hat{\alpha}_I, \hat{\beta}_1, \cdots, \hat{\beta}_J$, they require an iterative solution. We can use the Newton-Raphson method to solve the likelihood equations. For details on the Newton-Raphson method, see Bard (1974) and Haberman (1978). The estimate of α_I can be calculated by $\hat{\alpha}_I = -\hat{\alpha}_1 - \hat{\alpha}_2 - \cdots - \hat{\alpha}_{I-1}$ and similarly $\hat{\beta}_J = -\hat{\beta}_1 - \hat{\beta}_2 - \cdots - \hat{\beta}_{J-1}$.

2. The Confidence Intervals on $\pi_{1|ij}$ for all i and j

We consider the problem of constructing the confidence intervals on $\pi_{1|ij}$ for all i and j. These constructions are simplified by applying the logit transformation. This transforms the problem to consideration of the linear form which is called the logit model. After getting the confidence intervals for the logit model, the reverse transform is applied to obtain the confidence intervals for the logistic model.

Let $\lambda' = (\mu\ \alpha_1\ \alpha_2 \cdots \alpha_{I-1}\ \beta_1\ \beta_2 \cdots \beta_{J-1})$ be the $1 \times (I+J-1)$ vector, and $\mathbf{F}' = (\ f(\pi_{1|11})\ \ f(\pi_{1|12}) \cdots f(\pi_{1|1J}) \cdots\cdots f(\pi_{1|I1})\ f(\pi_{1|I2}) \cdots f(\pi_{1|IJ})\ )$ be the $1 \times IJ$ vector where $f(\pi_{1|ij}) = \ln\left(\dfrac{\pi_{1|ij}}{1 - \pi_{1|ij}}\right) = \mu + \alpha_i + \beta_j$ for $i = 1, 2, \cdots, I,\quad j = 1, 2, \cdots, J$.

We have an $IJ \times (I+J-1)$ design matrix such as

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & \cdots & 0 & -1 & -1 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & -1 & -1 & \cdots & -1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & -1 & -1 & \cdots & -1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & -1 & -1 & -1 & \cdots & -1 & -1 & -1 & -1 & \cdots & -1 \end{bmatrix}$$

Then $\mathbf{F} = \mathbf{X}\boldsymbol{\lambda}$ and $\pi_{1|ij} = \dfrac{\exp(\mathbf{x}_h'\boldsymbol{\lambda})}{1+\exp(\mathbf{x}_h'\boldsymbol{\lambda})}$ where $\mathbf{x}_h'$ is h_th row vector of a design matrix $\mathbf{X}$.

Denote the maximum likelihood estimators of $\boldsymbol{\lambda}$ by $\hat{\boldsymbol{\lambda}}$. The information matrix is the negative expected value of the matrix of second partial derivative of the log likelihood. The maximum likelihood estimators of parameters have a large-sample normal distribution with a covariance matrix equal to the inverse of the information matrix.

We have, asymptotically,

$$\hat{\boldsymbol{\lambda}} \sim N_{I+J-1}(\boldsymbol{\lambda},\ \boldsymbol{\Sigma}^{-1})$$

The second partial derivatives of the log likelihood functions are as follows :

$$\frac{\partial^2 L}{\partial \mu^2} = -\sum_{j=1}^{J}\sum_{i=1}^{I}\{m_{ij}\,\pi_{1|ij}\,(1-\pi_{1|ij})\}$$

$$\frac{\partial^2 L}{\partial \mu\,\partial \alpha_i} = \frac{\partial^2 L}{\partial \alpha_i^2} = -\sum_{j=1}^{J}\{m_{ij}\,\pi_{1|ij}\,(1-\pi_{1|ij})\} \text{ for } i=1,\ 2,\ \cdots,\ I$$

$$\frac{\partial^2 L}{\partial \mu\,\partial \beta_j} = \frac{\partial^2 L}{\partial \beta_j^2} = -\sum_{i=1}^{I}\{m_{ij}\,\pi_{1|ij}\,(1-\pi_{1|ij})\} \text{ for } j=1,\ 2,\ \cdots,\ J$$

$$\frac{\partial^2 L}{\partial \alpha_i\,\partial \alpha_h} = \frac{\partial^2 L}{\partial \beta_j\,\partial \beta_k} = 0 \text{ for } i \neq h \text{ and } j \neq k$$

$$\frac{\partial^2 L}{\partial \alpha_i \, \partial \beta_j} = -\{m_{ij}\, \pi_{1|ij}\,(1-\pi_{1|ij})\} \text{ for } i=1,\ 2,\ \cdots,\ I,\quad j=1,\ 2,\ \cdots,\ J$$

Since the second partial derivatives of the log likelihood functions are not a function of $\{Y_{1|ij}\}$, the expected second derivative matrix are identical. Thus, the asymptotic estimated covariance matrix is given by

$$\hat{\boldsymbol{\Sigma}}^{-1} = [\ \mathbf{X}'\mathrm{diag}\{m_{ij}\, \hat{\pi}_{1|ij}\,(1-\hat{\pi}_{1|ij})\}\ \mathbf{X}\ ]^{-1}$$

where $\mathrm{diag}\{m_{ij}\ \hat{\pi}_{1|ij}\ (1-\ \hat{\pi}_{1|ij})\}$ is an $IJ\times IJ$ diagonal matrix.

Let $\mathbf{D}=\mathrm{diag}(\delta_s)$ be the diagonal matrix of the eigenvalues of $\boldsymbol{\Sigma}$ and $\mathbf{U}$ be the matrix of corresponding orthogonal eigenvectors. Then

$$\boldsymbol{\Sigma}^{-1} = (\ \mathbf{UDU'})^{-1} = \mathbf{UD}^{-1}\mathbf{U'}$$

We define the notations such as :

$\mathbf{d}_h = \mathbf{D}^{-1/2}\mathbf{U}'\ \mathbf{x}_h = (\ d_{h1}\ d_{h2}\ \cdots\ d_{hs})'$

$\boldsymbol{\eta} = \mathbf{D}^{1/2}\ \mathbf{U}'\boldsymbol{\lambda} = (\ \eta_1\ \eta_2\ \cdots\ \eta_s)'$

$\mathbf{D}^{1/2} = \mathrm{diag}(\sqrt{\delta_s}\ )$

$\hat{\boldsymbol{\eta}} = \mathbf{D}^{1/2}\ \mathbf{U}'\hat{\boldsymbol{\lambda}}$ is maximum likelihood estimator of $\boldsymbol{\eta}$.

Then we have

$$E(\hat{\boldsymbol{\eta}}) = \boldsymbol{\eta} \text{ and } V(\hat{\boldsymbol{\eta}}) = \mathbf{I}$$

where $\mathbf{I}$ is an $s\times s$ identity matrix.

Thus, asymptotically,

$$\hat{\boldsymbol{\eta}} \sim N_s(\boldsymbol{\eta},\ \mathbf{I})$$

Let $\boldsymbol{\theta} = (\theta_1\ \theta_2\ \cdots\ \theta_s)'$, where $\theta_k = \hat{\eta}_k - \eta_k$, $k=1,\ 2,\ \cdots,s$. Then we have s multivariate asymptotic standard normal distribution,

$$\boldsymbol{\theta} \sim N_s(\mathbf{0},\ \mathbf{I})$$

Since θ_k, $k=1, 2, \cdots, s$ are independent, we have

$$\begin{aligned} &\Pr\{-c_{\alpha/2} \le \theta_1 \le c_{\alpha/2},\ -c_{\alpha/2} \le \theta_2 \le c_{\alpha/2},\ \cdots,\ -c_{\alpha/2} \le \theta_s \le c_{\alpha/2}\} \\ &\quad = \Pr\{-c_{\alpha/2} \le \theta_1 \le c_{\alpha/2}\} \Pr\{-c_{\alpha/2} \le \theta_2 \le c_{\alpha/2}\} \cdots \Pr\{-c_{\alpha/2} \le \theta_s \le c_{\alpha/2}\} \\ &\quad = 1-\alpha \end{aligned} \tag{1}$$

where $c_{\alpha/2}$ is a number such that $\int_{-c_{\alpha/2}}^{c_{\alpha/2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta_k^2}{2}\right) d\theta_k = (1-\alpha)^{1/s}$.

Substituting $\theta_k = \widehat{\eta}_k - \eta_k$ into (1), we have the following a rectangular confidence set on $\boldsymbol{\eta}$ with the confidence coefficient of $1-\alpha$:

$$\begin{cases} \hat{\eta}_1 - c_{\alpha/2} \le \eta_1 \le \hat{\eta}_1 + c_{\alpha/2} \\ \qquad \vdots \\ \hat{\eta}_s - c_{\alpha/2} \le \eta_s \le \hat{\eta}_s + c_{\alpha/2} \end{cases} \tag{2}$$

From a confidence set on $\boldsymbol{\eta}$ in (2), we have an inequality

$$\begin{aligned} \hat{\eta}_1 d_{h1} + \cdots + \hat{\eta}_s d_{hs} - c_{\alpha/2}(|d_{h1}| + \cdots + |d_{hs}|) &\le \eta_1 d_{h1} + \cdots + \eta_s d_{hs} \\ &\le \hat{\eta}_1 d_{h1} + \cdots + \hat{\eta}_s d_{hs} + c_{\alpha/2}(|d_{h1}| + \cdots + |d_{hs}|) \end{aligned}$$

for all $d_{h1}, d_{h2}, \cdots, d_{hs}$.

This is equivalent to the following matrix form :

$$\mathbf{d}_h{}'\hat{\boldsymbol{\eta}} - c_{\alpha/2}\left(\sum_{k=1}^{s} |d_{hk}|\right) \le \mathbf{d}_h{}'\boldsymbol{\eta} \le \mathbf{d}_h{}'\hat{\boldsymbol{\eta}} + c_{\alpha/2}\left(\sum_{k=1}^{s} |d_{hk}|\right) \text{ for all } \mathbf{d}_h \tag{3}$$

Substituting $\mathbf{d}_h = \mathbf{D}^{-1/2}\mathbf{U}'\mathbf{x}_h$ and $\boldsymbol{\eta} = \mathbf{D}^{1/2}\mathbf{U}'\boldsymbol{\lambda}$ into (3), we have

$$\mathbf{x}_h{}'\hat{\boldsymbol{\lambda}} - c_{\alpha/2}\left(\sum_{k=1}^{s} |d_{hk}|\right) \le \mathbf{x}_h{}'\boldsymbol{\lambda} \le \mathbf{x}_h{}'\hat{\boldsymbol{\lambda}} + c_{\alpha/2}\left(\sum_{k=1}^{s} |d_{hk}|\right) \text{ for all } \mathbf{x}_h \tag{4}$$

Therefore, 100(1- α)% approximate confidence bands for $\mu + \alpha_i + \beta_j$ over all i and j are given by

$$[L_b,\ U_b] = \left[\mathbf{x}_h{}'\hat{\boldsymbol{\lambda}} - c_{\alpha/2}\left(\sum_{k=1}^{s} |d_{hk}|\right),\ \mathbf{x}_h{}'\hat{\boldsymbol{\lambda}} + c_{\alpha/2}\left(\sum_{k=1}^{s} |d_{hk}|\right)\right]$$

The corresponding 100(1- α)% confidence intervals on $\pi_{1|ij}$ over all i and j are given by taking the inverse logit transform of inequality in (4) :

$$\frac{\exp(L_b)}{1+\exp(L_b)} \le \pi_{1|ij} \le \frac{\exp(U_b)}{1+\exp(U_b)}$$

3. Numerical Example

A 2×3×2 table is examined. The data in Table 1 for illustrating these confidence intervals calculation came from National Opinion Research Center, 1975 General Social Survey.

Table 1. Subject in 1975 General Social Survey, Cross-Classified by Attitude Toward Women Staying at Home, Sex of Respondent, and Education of Respondent

Sex of Respondent	Education of Respondent yrs	Response				Total Number
		Agree		Disagree		
		Number	Probability	Number	Probability	
Male	≤8	72	0.605	47	0.395	119
	9-12	110	0.359	196	0.641	306
	≥13	44	0.197	179	0.803	223
Female	≤8	86	0.694	38	0.306	124
	9-12	173	0.379	283	0.621	456
	≥13	28	0.130	187	0.870	215

Denote the probability of response Agree, when factor sex is at level i and factor education is level
j by

$$\pi_{1|ij} = \frac{\exp(\mu+\alpha_i+\beta_j)}{1+\exp(\mu+\alpha_i+\beta_j)}, \quad i=1,\ 2,\ j=1,\ 2,\ 3.$$

SAS/IML(1990) is used for computation. The maximum likelihood estimates of μ, α_1, α_2, β_1, β_2 and β_3 are $\hat{\mu}=-0.511551$, $\hat{\alpha}_1=-0.01172$, $\hat{\alpha}_2=0.01172$, $\hat{\beta}_1=1.1312745$, $\hat{\beta}_2=-0.017027$ and $\hat{\beta}_3=-1.1142475$. The estimated asymptotic covariance matrix of maximum likelihood estimators is given by

$$\hat{\Sigma}^{-1} = \begin{bmatrix} 0.0044982 & 0.0002404 & 0.0015373 & -0.002576 \\ 0.0002404 & 0.003494 & -0.000181 & 0.0004576 \\ 0.0015373 & -0.000181 & 0.0105225 & -0.003447 \\ -0.002576 & 0.0004576 & -0.003447 & 0.0064154 \end{bmatrix}$$

The estimated numbers of response and the estimated probabilities of response are in Table 2 and 95% approximate confidence region on π_{ij} for all i and j are in Table 3.

Table 2. The Estimated Numbers of Response and the Estimated Probabilities of Response

Sex of Respondent	Education of Respondent yrs	Response				Total Number
		Agree		Disagree		
		Number	Probability	Number	Probability	
Male	≤8	77.05	0.647	41.95	0.353	119
	9-12	112.64	0.368	193.36	0.632	306
	≥13	36.31	0.163	186.69	0.837	223
Female	≤8	80.95	0.653	43.05	0.347	124
	9-12	170.36	0.374	285.64	0.626	456
	≥13	35.69	0.166	179.31	0.834	215

Table 3. 95% Confidence Intervals on π_{ij} for all i and j

Sex of Respondent	Education of Respondent yrs	Response			
		Agree		Disagree	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound
Male	≤8	0.513	0.762	0.238	0.487
	9-12	0.280	0.467	0.533	0.720
	≥13	0.104	0.246	0.754	0.896
Female	≤8	0.519	0.766	0.234	0.481
	9-12	0.302	0.451	0.549	0.698
	≥13	0.106	0.250	0.750	0.894

References

[1] Albert, A. and Anderson, J. A. (1984). On the Existence Maximum Likelihood Estimates in Logistic Models, *Biometrika*, 71, 1-10.

[2] Bard, Y. (1974). *Nonlinear Parameter Estimates*, New York: Academic Press.

[3] Cox, D. R. (1970). *The Analysis of Binary Data*, London: Methuen.

[4] Goodman, L. A. and Kruskal, W. H. (1979). *Measures of Association for Cross classifications*, New York: Spring-Verlag.

[5] Haberman, S. J. (1978), *Analysis of Qualitative Data*, Vol 1, New York : Academic Press.

[6] Hodges, J. L., Jr. (1958). Fitting the Logistic by Maximum Likelihood, *Biometrics*, 14, 453-461.

[7] SAS Institute Inc. (1990). *SAS/IML*.

[8] Silvapull, M. J. (1981). On the existence of maximum likelihood estimators for the binomial response models, *J. Roy. Statist. Soc. Ser.* B, 43, 310-313

[Received June 2003, Accepted September 2003]