The Korean Communications
in Statistics Vol. 8, No. 3, 2001
pp. 581-588

The Confidence Band of ED_{100p} for the Simple Logistic Regression Model[1)]

Tae Kyoung Cho[2)] and Mi Young Shin[3)]

Abstract

The ED_{100p} is that value of the dose associated with 100p% response rate in the analysis of quantal response data. Brand, Pinnock, and Jackson (1973) studied the confidence bands of ED_{100p} obtained by solving extremal values algebraically on the ellipsoid confidence region of the parameters in the simple logistic regression model. In this paper, we develope and illustrate a simpler method for obtaining confidence bands for ED_{100p} based on the rectangular confidence region of parameters.

Keywords : Logistic regression, Confidence band, Effective dose

1. Introduction

Consider a quantal assay experiment in which k different doses of preparation are administered to k groups of subjects. Suppose n_i subjects exposed to doses x_i of a substance and y_i subjects exhibit a response $(i = 1,\ 2,\ \cdots,\ k)$.

With the Bernoulli random variable the conditional mean is

$$p_i \ = \Pr(Y_i | X = x_i) = E(Y_i | X = x_i)$$

where p_i is the probability of response corresponding to dose x_i.

Because of the structural problem with linear regression model, curvilinear relationship between p_i and x_i are studied by Cox(1970).

The model usually adopted is

1) This work was supported by the reseach fund of Dongguk Unversity

2) Assistant Professor, Dept. of Statistics and Information Science, Dongguk University, Kyongju, 780-714, Korea
E-mail : tkcho@mail.dongguk.ac.kr

3) Associate Professor, Dept. of Mathematics, The Catholic University of Korea, Puchon City, 420-743, Korea
E-mail : myshin@www.cuk.ac.kr

$$p_i = \int_{-\infty}^{\beta_0+\beta_1 x_i} f(t)dt = F(\beta_0+\beta_1 x_i)$$

where β_0 and β_1 are parameters to be estimated. For logit analysis, $f(t)$ is the logistic density function and the logistic response curve with parameters β_0 and β_1 is given by

$$p_i = e^{\beta_0+\beta_1 x_i} / \left(1+e^{\beta_0+\beta_1 x_i}\right).$$

Confidence bands for the logistic response function were studied by Brand, Pinnock and Jackson (1973), Hauck (1983) and others.

It is often interesting to estimate the value of explanatory variables that correspond to a specified value of the response probability. For example, in bioassay, one is often interested in the concentration of a chemical that is expected to produce a response in 50% of the individuals exposed to it. This dose is termed the median effective dose.

Let the ED_{100p} be that the value of the dose associated with a 100p% response rate. Then

$$p = F(\beta_0 + \beta_1 ED_{100p})$$

and we estimate

$$\widehat{ED_{100p}} = (F^{-1}(p) - \widehat{\beta}_0) / \widehat{\beta}_1.$$

When the simple logistic regression model is used to find the dose-response relationship, we have

$$ED_{100p} = (\ln(p/(1-p)) - \beta_0)/\beta_1 \tag{1}$$

For example, median effective dose is given by $ED_{50} = -\beta_0/\beta_1$.

Brand, Pinnock and Jackson (1973) studied the confidence band of ED_{100p} obtained by solving extrema on the ellipsoid confidence region of the β_0 and β_1. Carter and etc. (1986) developed a method for estimating a large sample confidence region about the ED_{100p} from the logistic curve in the case of multiple explanatory variables.

In this paper, we develop and illustrate a method for obtaining a confidence band for ED_{100p} based on the rectangular confidence region of β_0 and β_1.

2. Confidence Band of ED_{100p}

Let Y be a dichotomous response variable, with possible values 0 and 1. The simple logistic regression model relating the expected value cf Y to a explanatory variable x_i, is given by

$$p_i = e^{\beta_0+\beta_1 x_i} / (1+e^{\beta_0+\beta_1 x_i}).$$

The log-likelihood for β_0 and β_1 is

$$l(\beta_0, \beta_1) = \sum_{i=1}^{k} \left\{ \log \binom{n_i}{y_i} + y_i \log(p_i) + (n_i - y_i) \log(1 - p_i) \right\}.$$

Bradley and Gart (1962) and Cox (1970) discussed the asymptotic properties of maximum likelihood estimators. Under regularity conditions, maximum likelihood estimators have a large-sample normal distribution with a covariance matrix equal to the inverse of the information matrix $\boldsymbol{\Sigma}$.

Suppose $\widehat{\beta_0}$ and $\widehat{\beta_1}$ are maximum likelihood estimates of β_0 and β_1, respectively. Then, $\widehat{\beta_0}$ and $\widehat{\beta_1}$ have a large-sample normal distribution with a covariance matrix $\boldsymbol{\Sigma}^{-1}$. That is, we have asymptotically,

$$\begin{bmatrix} \widehat{\beta_0} \\ \widehat{\beta_1} \end{bmatrix} \sim N_2\left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \boldsymbol{\Sigma}^{-1} \right)$$

Let I_{11}, I_{12}, and I_{22} be estimated component of the information matrix. Then,

$$I_{11} = \sum_{i=1}^{k} n_i \widehat{p_i}(1 - \widehat{p_i}), \quad I_{12} = \sum_{i=1}^{k} n_i x_i \widehat{p_i}(1 - \widehat{p_i}), \quad I_{22} = \sum_{i=1}^{k} n_i x_i^2 \widehat{p_i}(1 - \widehat{p_i}).$$

where $\widehat{p_i}$ is p_i evaluated at the maximum likelihood estimates of β_0 and β_1.

Brand, Pinnock and Jackson(1973) considered the confidence interval of ED_{100p} which is obtained by finding the extremal values for ED_{100p} over the following ellipsoid confidence region of β_0 and β_1.

$$I_{11}(\widehat{\beta_0} - \beta_0)^2 + 2 I_{12}(\widehat{\beta_0} - \beta_0)(\widehat{\beta_1} - \beta_1) + I_{22}(\widehat{\beta_1} - \beta_1)^2 = \chi^2_{(1-\alpha)}$$

where $\chi^2_{(1-\alpha)}$ denotes the $100(1-\alpha)\%$ percentile of the chi-squared distribution with two degrees of freedom. The extremal values occur for those lines in the family which are tangent to the boundary of the elliptical confidence region.

In this paper, we consider the confidence intervals of ED_{100p} for simple logistic regression model based on the rectangular confidence region of β_0 and β_1.

To simplify the covariance matrix, let $\mathbf{D} = \mathrm{diag}(\lambda_i)$, $i = 0, 1$, be the diagonal matrix of the eigenvalues of the information matrix $\boldsymbol{\Sigma}$ and $\mathbf{U}$ be the matrix of corresponding orthonormal eigenvectors. Then the covariance matrix can be written as

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

Let us define the following notations.

$\mathbf{D}^{1/2} = \mathrm{diag}(\sqrt{\lambda_i})$, $i = 0, 1$

$\mathbf{U} = (u_{ij})$, $i = 0, 1$, $j = 0, 1$v

$\boldsymbol{\beta}' = [\beta_0 \quad \beta_1]$

$\boldsymbol{\eta} = \mathbf{D}^{1/2}\mathbf{U}'\boldsymbol{\beta}$

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

Then we have, asymptotically,

$$\hat{\boldsymbol{\eta}} \sim N_2(\boldsymbol{\eta}, \mathbf{I})$$

where $\mathbf{I}$ is an 2×2 identity matrix.

Let $\boldsymbol{\theta}' = (\theta_0\ \theta_1)$, where $\theta_j = \hat{\eta}_j - \eta_j$, $j=0, 1$. Then we have asymptotic bivariate standard normal distribution,

$$\boldsymbol{\theta} \sim N_2(\mathbf{0}, \mathbf{I}).$$

Since θ_j, $j=0, 1$ are independent, we have

$$\begin{aligned}
&\Pr\{-c_{\alpha/2} \le \theta_0 \le c_{\alpha/2},\ -c_{\alpha/2} \le \vartheta_1 \le c_{\alpha/2}\} \\
&= \Pr\{-c_{\alpha/2} \le \theta_0 \le c_{\alpha/2}\}\Pr\{-c_{\alpha/2} \le \theta_1 \le c_{\alpha/2}\} \qquad (2) \\
&= 1-\alpha
\end{aligned}$$

where $c_{\alpha/2}$ is a number such that

$$\int_{-c_{\alpha/2}}^{c_{\alpha/2}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta_j^2}{2}\right) d\theta_j = (1-\alpha)^{1/2}.$$

Substituting $\theta_j = \hat{\eta}_j - \eta_j$ in (2), we have the following a rectangular confidence region of $\boldsymbol{\eta}$ with confidence coefficient of $1-\alpha$:

$$\begin{cases} \hat{\eta}_0 - c_{\alpha/2} \le \eta_0 \le \hat{\eta}_0 + c_{\alpha/2} \\ \hat{\eta}_1 - c_{\alpha/2} \le \eta_1 \le \hat{\eta}_1 + c_{\alpha/2} \end{cases}$$

A rectangular confidence region for β_0 and β_1 with confidence coefficient of $1-\alpha$ obtained by substituting $\boldsymbol{\eta} = \mathbf{D}^{1/2}\mathbf{U}'\boldsymbol{\beta}$ is:

$$\begin{cases} \sqrt{\lambda_0}\, u_{00}\hat{\beta}_0 + \sqrt{\lambda_0}\, u_{10}\hat{\beta}_1 - c_{\alpha/2} \le \sqrt{\lambda_0}\, u_{00}\beta_0 + \sqrt{\lambda_0}\, u_{10}\beta_1 \le \sqrt{\lambda_0}\, u_{00}\hat{\beta}_0 + \sqrt{\lambda_0}\, u_{10}\hat{\beta}_1 + c_{\alpha/2} \\ \sqrt{\lambda_1}\, u_{01}\hat{\beta}_0 + \sqrt{\lambda_1}\, u_{11}\hat{\beta}_1 - c_{\alpha/2} \le \sqrt{\lambda_1}\, u_{01}\beta_0 + \sqrt{\lambda_1}\, u_{11}\beta_1 \le \sqrt{\lambda_1}\, u_{01}\hat{\beta}_0 + \sqrt{\lambda_1}\, u_{11}\hat{\beta}_1 + c_{\alpha/2} \end{cases}$$

The extremal point of pair (β_0, β_1) over the rectangular confidence region are given by

$$\left(\hat{\beta}_0 + \frac{c_{\alpha/2}(\sqrt{\lambda_0}\, u_{10} - \sqrt{\lambda_1}\, u_{11})}{\Delta},\ \hat{\beta}_1 - \frac{c_{\alpha/2}(\sqrt{\lambda_0}\, u_{00} - \sqrt{\lambda_1}\, u_{01})}{\Delta}\right)$$

$$\left(\hat{\beta}_0 - \frac{c_{\alpha/2}(\sqrt{\lambda_0}\, u_{10} + \sqrt{\lambda_1}\, u_{11})}{\Delta},\ \hat{\beta}_1 + \frac{c_{\alpha/2}(\sqrt{\lambda_0}\, u_{00} + \sqrt{\lambda_1}\, u_{01})}{\Delta}\right)$$

$$\left(\hat{\beta}_0 + \frac{c_{\alpha/2}(\sqrt{\lambda_0}\, u_{10} + \sqrt{\lambda_1}\, u_{11})}{\Delta},\ \hat{\beta}_1 - \frac{c_{\alpha/2}(\sqrt{\lambda_0}\, u_{00} + \sqrt{\lambda_1}\, u_{01})}{\Delta}\right)$$

$$\left(\hat{\beta}_0 - \frac{c_{\alpha/2}(\sqrt{\lambda_0}\, u_{10} - \sqrt{\lambda_1}\, u_{11})}{\Delta},\ \hat{\beta}_1 + \frac{c_{\alpha/2}(\sqrt{\lambda_0}\, u_{00} - \sqrt{\lambda_1}\, u_{01})}{\Delta}\right)$$

where $\Delta = \sqrt{\lambda_0}\sqrt{\lambda_1}\,(u_{00}u_{11} - u_{01}u_{10})$.

Equation (1) represents a family of straight lines, $\beta_0 = -\mathrm{ED}_{100p}\,\beta_1 + \ln(p/(1-p))$, which intersect the β_0 axis at $\ln(p/(1-p))$ and have a slop $-\mathrm{ED}_{100p}$. The maximum and minimum slopes occur for the straight lines in the family which are extremal points of pair over the rectangular confidence region of β_0 and β_1. These values provide the confidence interval of ED_{100p}. For the confidence band of ED_{50}, we can use the straight line through the origin such as $\beta_0 = -\mathrm{ED}_{50}\,\beta_1$.

3. Numerical Example

To illustrate the confidence band which is developed in section 2, we use data that come from a study of quantit analysis (Copenhaver and Mielke: 1977) and list them in Table 1. The data set is given in the form : dose (x_i), number responding (y_i), and number in experiment (n_i).

Table 1. Data from quantit analysis (Copenhaver and Mielke: 1977)

x_i	y_i	n_i
0.71	16	49
1.00	18	48
1.31	34	48
1.48	47	49
1.61	47	50
1.70	48	48

The maximum likelihood estimates of β_0, β_1 and estimated components of the information matrix computed from data are $\widehat{\beta}_0 = -4.451$, $\widehat{\beta}_1 = 4.460$, $I_{11} = 37.3505$, $I_{12} = 42.6072$ and $I_{22} = 52.3660$. Table 2 shows the estimated value of ED_{100p} and 95% confidence intervals for ED_{100p}.

Figure 1 represents the rectangular confidence region for the parameter pair (β_0, β_1). The extremal values for β_0 and β_1 over the set provide confidence intervals for the parameters. The estimated value of ED_{100p} with the 95% confidence band is presented in Figure 2.

Table 2. Estimated value of ED_{100p} and 95% confidence intervals for ED_{100p}

p	Estimated value of ED_{100p}	Lower bound of ED_{100p}	Upper bound of ED_{100p}
0.10	0.50529	0.21229	0.69486
0.20	0.68710	0.44371	0.84458
0.30	0.80795	0.59752	0.94409
0.40	0.90701	0.72360	1.02567
0.50	0.99792	0.83931	1.10053
0.60	1.08882	0.95502	1.17539
0.70	1.18788	1.08110	1.30437
0.80	1.30873	1.21786	1.47607
0.90	1.49054	1.35813	1.73439

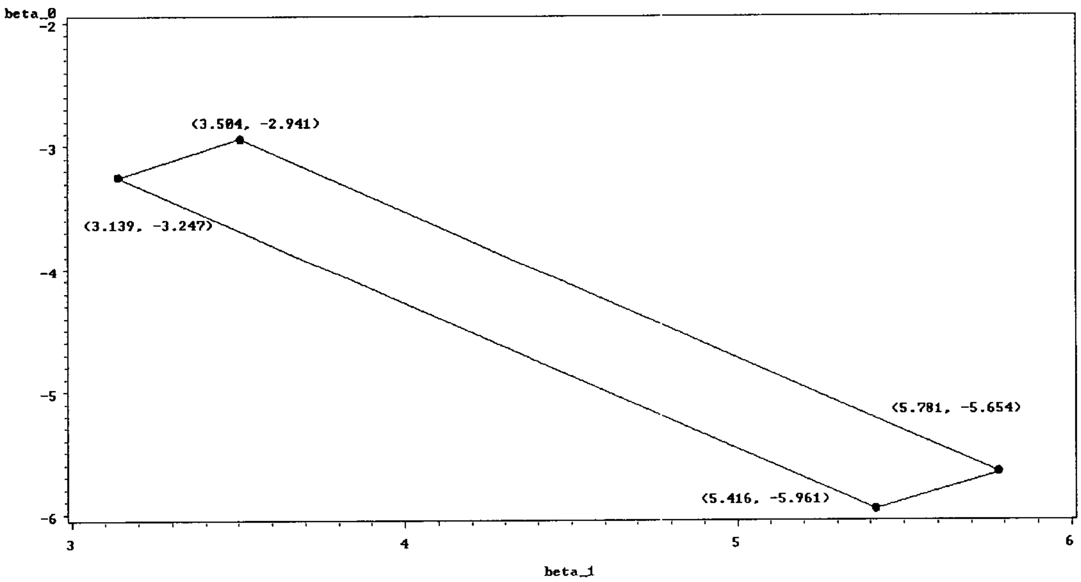

Figure 1. 95% confidence region for simple logistic parameters (β_0, β_1).

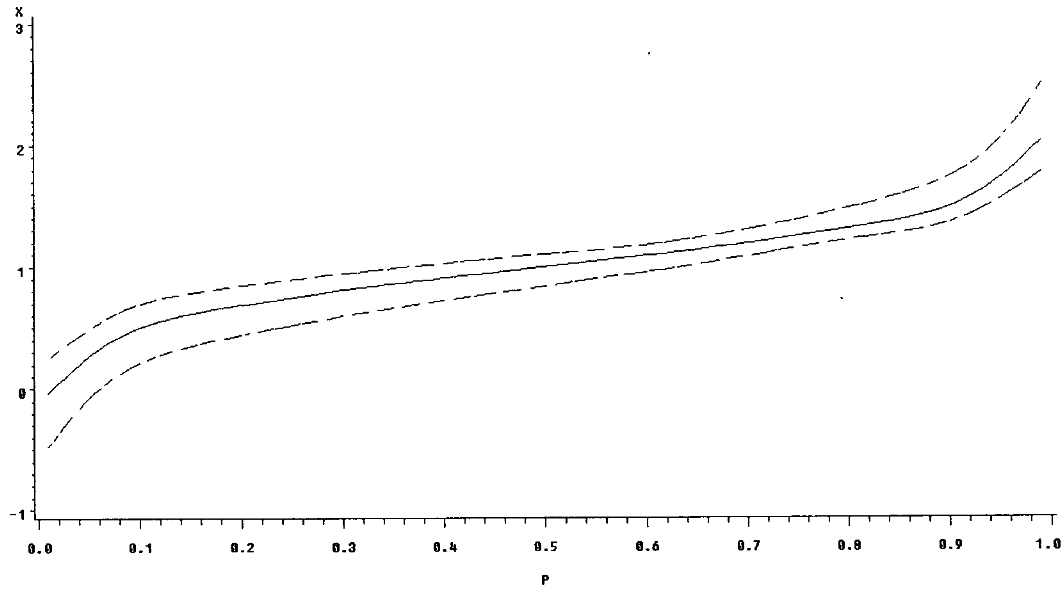

Figure 2. 95% confidence bands(---) of ED_{100p} and the estimated value(—) of ED_{100p}

A simulation study is carried out in order to illustrate the confidence bands. Suppose β_0 =-4 and β_1=4 and we have same dose levels as in the previous example. Using random numbers generated from binomial distribution with sample size 50, we get the 95% confidence intervals of ED_{100p}. Then 100 separate runs are made for each p. Finally, we have the means of 100 intervals for each p. We repeat same procedure for the sample size 100 and 200. The results of a simulation are given in Table 3.

Table 3. Simulation of 95% confidence intervals for ED_{100p}

		sample size=50		sample size=100		sample size=200	
p	True ED_{100p}	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
0.1	0.4506939	0.1071880	0.6482272	0.2197831	0.5929444	0.3172747	0.5699752
0.2	0.6534264	0.3685419	0.8160451	0.4623557	0.7711253	0.5423583	0.7514853
0.3	0.7881755	0.5422546	0.9275877	0.6235851	0.8895558	0.6919634	0.8721286
0.4	0.8986337	0.6846525	1.0190228	0.7557500	0.9866371	0.8145995	0.9710238
0.5	1.0000000	0.8153294	1.1029318	0.8770364	1.0757276	0.9271413	1.0617789
0.6	1.1013663	0.9460064	1.1911393	0.9983227	1.1648350	1.0396831	1.1525339
0.7	1.2118245	1.0884043	1.3451591	1.1304876	1.2944796	1.1623192	1.2691001
0.8	1.3465736	1.2333243	1.5423562	1.2616953	1.4692121	1.2852907	1.4267142
0.9	1.5493061	1.3892623	1.8390429	1.4303149	1.7321004	1.4596581	1.6638476

References

[1] Bradley, R. A. and Gart J. J. (1962). The Asymptotic Properties of ML Estimators When Sampling From Associated Population, *Biometrika*, 49, 205-214.

[2] Brand R. J., Pinnock D. E. and Jackson K. L.(1973). Large Sample Confidence Bands for the Logistic Response Curve and its Inverse, *The American Statistician*, 27, No. 4, 157-160

[3] Carter, W. H. Jr, Chinchilli, V. M., Willson J. D.. Campbell E. E., Kessler F. K., and Carchman R. A.(1986). An Asymptotic Confidence Region for the ED_{100p} From the Logistic Response Surface for a Combination of Agents, *The American Statistician*, 40, No. 2, 124-128.

[4] Copenhaver, T. W. and Mielke, P. W. (1977). Quantit analysis: a quantal assay refinement. *Biometrics*, 33, 175-186.

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

[6] Hauck W. H. (1983). A Note on Confidence Bands for the Logistic Response Curve, *The American Statistician*. 37. No. 2. 158-160.