LOGISTIC REGRESSION ANALYSES IN DOSE RESPONSE STUDIES

FAQIR MUHAMMAD*
AJMAL KHAN*
SOHAIL AHMAD*

SUMMARY: In dosage response studies the two most commonly used techniques are i) Probit analysis, ii) Logit analysis. Logit analysis is easy to carry out and most practical situations are closer to logit than probit analysis. In logit analysis the function to fit is $\theta(t) = \frac{e^{\hat{a}} + \hat{a}t}{1 + e^{\hat{a}+\hat{a}t}}$

where it is the log dose and $\theta(t)$ *is the proportion killed. The parameters* α *and* β *in the function can be estimated by using following methods: Least squares, Method of moments, Weighted least squares, Maximum likelihood both the point and interval estimates for these methods are discussed by taking three practical examples.* LD_{50} *is estimated and the intervals constructed. It has been shown that method of weighted least squares gives quite good results.*

Key words: Probit analysis, logit analysis.

INTRODUCTION

In dosage response studies an attempt is made to describe relationship between θ_i the probability that a subject dies and t_i, dose of the drug usually taken in logarithm terms. The simplest model is to express θ as a linear function of t i.e.

$$\theta = \alpha + \beta t$$

In practice, this is generally not true because t, log concentration, can take any real value whereas θ must be between 0 and 1. To ovarcome this difficulty, there is need to express θ by some function g (θ) which can take any real value. This end is achieved by the following function.

$$g(\theta) = In\left[\frac{\grave{e}}{1-\grave{e}}\right] \quad \text{... (a)}$$

Where In. represents natural logarithm i.e. $\log_e$

The RHS of equa. (a) is In-odds i.e. logarithm of the odds ratio for death versus survival. We assume that In-odds is a linear function of the dose i.e.

$$In.\left[\frac{\grave{e}}{1-\grave{e}}\right] = \alpha + \beta t \quad \text{.. (1)}$$

or

$$\theta(t) = \frac{e^{\hat{a}+\hat{a}t}}{1+e^{\hat{a}+\hat{a}t}} \quad \text{... (2)}$$

This expression is called logistic regression because $\theta(t)$ is the cumulative distribution function of a continuous distribution known as logistic distribution. The logistic regression analysis is known as logit analysis.

At each concentration t_i, one can think of the experiment being carried out as a series of binomial trials with probability of kill θ_i. Given the dose for each experiment, it seems reasonable to propose a model linking the θ_i's to the dose t_i's. If this can be done successfully, we have a means of predicting the kill rate θ_i for any given dose t_i. One possibility in such a situation is the logistic regression model and the other is the probit analysis.

Probit analysis has been discussed in detail by Finney (7) and Busvine (4). Here we mainly emphasize on Logit analysis.

The probit and logit analysis provide bit similar results but logit is easier to use. The probit analysis uses the distribution function of normal distribution which can not be written in the simple form as we have for the logistic distribution.

* From Department of Mathematics and Statistics University of Agriculture, Faisalabad, Pakistan.

Some studies, regarding different aspects of logistic regression have been carried out by Berkson (2,3), Busvine (4), Finney (7), McCullagh and Nelder (8), Albert (1), Carroll and Spiegelman (6), Byron and Morgan (5), Wang (9), Wilson and Worcester (10) and Worcester and Wilson (11).

The equation (1) above contains two parameters and these can be estimated using different methods. Each method will give different values for these parameters which ultimately affect the LD_{50} and its interval estimates as $LD_{50} = -\alpha/\beta$. It looks sensible to see the point and interval estimation of α and β by using the following four methods taking examples with different number of concentrations and subjects. Here LD_{50} is log LD_{50} and we will carry on writing LD_{50} for log LD_{50}. Antilog of this LD_{50} will be the dose of the drug in original units. In probit analysis LD_{50} is calculated as $LD_{50} = (5 - \alpha)/\beta$.

METHODOLOGY

The methods of estimation of parameters in logistic regression are

1. Ordinary least-squares (OLS).
2. Method of moments (MOM).
3. Weighted least-squares (WLS).
4. Maximum likelihood (ML).

In the least squares, the estimates are calculated using relation (1) with t_i as fixed variable and

$$In\left[\frac{x_i/n_i}{1\text{-}x_i/n_i}\right]$$

as response variable. In this relation, θ_i, is estimated by x_i/n_i where x_i is number of insects killed out of total of n_i used for a dose t_i. It should be clear that the method of moments gives rise to the same normal equations as the least squares and ultimately the same values of α, β and LD_{50}.

To estimate α and β by using weighted least squares, the weights are calculated by relation.

$$w_i = n_i p_i (1\text{-}p_i)$$

where

$$p_i = In\left[\frac{x_i/n_i}{1\text{-}x_i/n_i}\right]$$

The estimates can be obtained by using weighted least squares manually or Minitab package can be used with the weights as above, the given values of t_i as explanatory variable and the

$$In\left[\frac{x_i/n_i}{1\text{-}x_i/n_i}\right]$$

as response variable.

To get the values of α and β using likelihood method, following procedure is adopted.

The likelihood using equation (2) and data (x_i, n_i, t_i); = 1, ….., k doses is

$$Lik\ (\theta, x) = \prod_{i=1}^{k} {}^{n_i}C_{x_i} (\theta_i)^{x_i}\ (1-\theta_i)^{n_i - x_i}$$

$$Lik = \prod_{i=1}^{k} {}^{n_i}C_{x_i}\left[\frac{\exp(\alpha + \beta t)}{1+\exp(\alpha + \beta t)}\right]^{x_i}$$

$$\left[\frac{1}{1+\exp(\alpha + \beta t)}\right]^{n_i x_i}$$

$$= \prod_{i=1}^{k} {}^{n_i}C_{x_i}\left[\prod_{i=1}^{k}\left[\frac{1}{1+\exp(\alpha + \beta t)}\right]^{n_i}\right.$$

$$\left.\left[\exp(\alpha\sum x_i + \beta\sum x_i t_i)\right]_1\right]$$

Let In. Lik (θ, x) = 1 (θ , x) or simply 1, then

$$1 = const. + \alpha\sum x_i + \beta\sum x_i t_i \text{ -}$$

$$\sum n_i In.\left[1+\exp(\alpha + \beta t_i\right]$$

To obtain the maximum likelihood estimators from the likelihood function 1, the partial derivatives w.r.t α and β are

$$\frac{\partial 1}{\partial \alpha} = \sum x_i - \sum n_i \frac{\exp(\alpha + \beta t_i)}{\left[1+\exp(\alpha + \beta t_i)\right]} = 0, \ldots (3)$$

$$\frac{\partial 1}{\partial \beta} = \sum x_i t_i - \sum n_i t_i \frac{\exp(\alpha + \beta t_i)}{\left[1+\exp(\alpha + \beta t_i)\right]} = 0 \quad (4)$$

As equations (3) and (4) can not be solved explicitly, so are solved numerically using modified Newton-Raphson method.

The approximate interval for LD_{50} using the above methods are as follows.

As LD_{50} is a non-linear function, its variance is approximated using Taylor series.

The approximate interval estimate of LD_{50} when estimates of α and β are by ordinary least squares is

$$-\frac{\hat{\alpha}}{\hat{\beta}} \pm 1.96\frac{1}{\hat{\beta}^2}\left[\frac{1}{n} + \frac{\{t + (\hat{\alpha}/\hat{\beta})\}^2}{\sum t_i^2 - \frac{(\sum t_i)^2}{n}}\right]$$

The approximate interval for LD_{50} using weighted least squares is

$$-\frac{\hat{\alpha}}{\hat{\beta}} \pm 1.96 \frac{1}{\hat{\beta}^2}\left[\frac{1}{\sum w_i}+\frac{\left\{t+(\hat{\alpha}/\hat{\beta})\right\}^2}{\sum w_i t_i^2-\frac{(\sum w_i t_i)^2}{\sum w_i}}\right]$$

where

$$t=\frac{\sum w_i t_i}{\sum w_i}, w_i = n_i p_i (1-p_i)$$

The interval estimate using MLE is

$$-\frac{\hat{\alpha}}{\hat{\beta}} \pm 1.96 \sqrt{J_1^T [K(x)]^{-1} J_1}$$

where

$$J_1 = \begin{bmatrix} \partial LD_{50} / \partial \hat{\alpha} \\ \partial LD_{50} / \partial \hat{\beta} \end{bmatrix} = \begin{bmatrix} -1/\hat{\beta} \\ \hat{\alpha}/\hat{\beta}^2 \end{bmatrix}$$

and $[K(x)]^{-1}$ is the information matrix

$$K(x) = \begin{bmatrix} \frac{\partial^2 1}{\partial \alpha^2} & \frac{\partial^2 1}{\partial \alpha \partial \beta} \\ \frac{\partial^2 1}{\partial \alpha \partial \beta} & \frac{\partial^2 1}{\partial \beta^2} \end{bmatrix}$$

Here, 1 is likelihood function

$$\frac{\partial^2 1}{\partial \alpha^2} = -\sum n_i \frac{e^{\acute{a}+\hat{a}ti}}{\left[1+e^{\acute{a}+\hat{a}t_i}\right]^2}$$

$$\frac{\partial^2 1}{\partial \alpha \partial \beta} = -\sum n_i t_i \frac{e^{\acute{a}+\hat{a}ti}}{\left[1+e^{\acute{a}+\hat{a}t_i}\right]^2}$$

$$\frac{\partial^2 1}{\partial \beta^2} = -\sum n_i t_i^2 \frac{e^{\acute{a}+\hat{a}ti}}{\left[1+e^{\acute{a}+\hat{a}t_i}\right]^2}$$

RESULTS AND DISCUSSION

Three practical examples with different situations are taken. The data given in appendix has been taken from Entomology Department of University of Agriculture Faisalabad. The first example has 4 concentrations and each concentration is used for 20 insects i.e. $n_i = 20$; the second example has 6 concentrations and 60 insects per concentration and the third example has 5 concentrations and 10 insects per concentration.

Table 1: Comparison of estimates by logit and probit analyses for example 1.

Method	Estimates of Parameters			
Logit	$\hat{\alpha}$	$\hat{\beta}$	LD_{50}	C.I. for LD_{50}
OLS	-1.2236	4.7573	0.2572	(0.1362, 0.3782)
WLS	-1.2307	4.9186	0.2501	(0.1303, 0.3700)
MLE	-1.2382	4.9480	0.2501	(0.1257, 0.3752)
Probit	4.2776	2.8285	0.2554	(0.1325, 0.3783)

Table 2: Comparison of estimates by logit and probit analyses for example 2.

Method	Estimates of Parameters			
Logit	$\hat{\alpha}$	$\hat{\beta}$	LD_{50}	C.I. for LD_{50}
OLS	0.3653	2.6459	-0.1381	(-0.2316, -0.0455)
WLS	0.4231	2.5438	-0.1663	(-0.2618, -0.0708)
MLE	0.4249	2.5824	-0.1645	(-0.2597, -0.0693)
Probit	5.2476	1.5572	-0.1527	(-0.2411, -0.0642)

The methods applied to estimate α and β, LD_{50} and approximate interval estimates. Results are given in Tables 1, 2 and 3.

It is clear from Table 1 that LD_{50} is same uptil four decimal places using MLE and WLS estimates of α and β. As a whole the values of LD_{50} agree by the probit and logit analysis. The intervals also overlap.

It is clear from the Table 2 that the LD_{50} is almost same by logit analysis using the estimates $\hat{\alpha}$ and $\hat{\beta}$ by WLS and MLE and these values are also close to LD_{50} estimated by probit analysis.

The confidence interval for LD_{50} in logit analysis overlap by using any of the estimates of α and β and also agree with the interval estimate of probit.

The LD_{50} using OLS is different from others and the interval estimate is also wider as compared with all the other interval estimates.

The Table 3 indicates that all the methods give almost same values of estimates of α and β in logit analysis. The intervals for LD_{50} are very close by using MLE or WLS in

Table 3: Comparison of estimates by logit and probit analyses for example 3.

Method	Estimates of Parameters			
Logit	$\hat{\alpha}$	$\hat{\beta}$	LD_{50}	C.I. for LD_{50}
OLS	33.510	12.9720	-2.5833	(-5.5225, 0.3560)
WLS	32.958	12.7507	-2.6632	(-2.7776,-2.5488)
MLE	33.048	12.7857	-2.5848	(-2.6413, -2.5283)
Probit	24.9044	7.6997	-2.5851	(-2.6348,-2.5354)

logit analysis and are also close to probit analysis. The interval estimate for LD_{50} using OLS is very wide and is quite different from the other interval estimates.

It can be concluded from these examples that the logistic regression analysis gives quite good results when using weighted least squares method. The intervals are usually shorter and sensible. On the basis of point and interval estimates it is suggested that weighted least squares method in logit analysis should be preferred in such bioassay situations.

REFERENCES

1. Albert A : On the existance of maximum likelihood estimates in logistic regression model. Biometrika 71(1): 1-10, 1084.

2. Berkson J : Application of the logistic function to bioassay. J Amer Statist Assoc, 39:357-365, 1944.

3. Berkson J : Why I prefer logit to probits. Biometrics 7: 327-339, 1951.

4. Busnive JR : The critical review of techniques for testing insecticides. Henry Ling Limited. The Dorset Press Dorchester, 1971.

5. Byron J, Morgan T : The cubic logistic model for quantal assay data. Appl Statist, 34(2):105-113, 1985.

6. Carroll RJ, Spiegelman CH : On errors in variables for binary regression model. Biometrika 73(1):19-25, 1984.

7. Finney DJ : Probit analysis. Cambridge University Press, Cambridge, 1971.

8. McCullagh P, Nelder JA : Generalized linear models. Monographs on Statistics and Probability, London, 1983.

9. Wang PC : A simple method of analyzing binary data from orthogonal arrays. J Quality Technology 20(4):230-232, 1988.

10. Wilson EB, Worcester J : The determination of LD_{50} and its sampling error to bio-assay. Proc Nat Acad Sci Wash 29: 79-85, 1943.

11. Worcester J, Wilson EB : A table determining LD_{50} or the 50% end point. Proc Nat Acad Sci Wash 29:207-212, 1943.

Correspondence:
Faqir Muhammad
Dept. of Maths. And Stats.
University of Agriculture,
Faisalabad, PAKISTAN.

APPENDIX: The data taken from the dept. Entomology University of agriculture Faisalabad (Pakistan).

EXAMPLE I:	Conc (%)	n_i	x_i
	0.01	20	4
	0.02	20	12
	0.04	20	17
	0.08	20	19
EXAMPLE II:	Conc	n_i	x_i
	0.12	60	5
	0.25	60	15
	0.50	60	27
	1.00	60	39
	2.00	60	45
	4.00	60	51
EXAMPLE III :	Conc	n_i	x_i
	0.0018	10	1
	0.0022	10	3
	0.0026	10	5
	0.0030	10	7
	0.0034	10	8