

Paper ST01
Analysis of Dosage-Response Data in Agricultural Research
K. Bondari

Experimental Statistics, Coastal Plain Station, University of Georgia, Tifton, GA 31793-0748

ABSTRACT

Agricultural researchers, especially research entomologists, are interested in determining how mortality rates change with increasing dose levels of a certain stimulus (e.g., pesticide, drug). A researcher may have a particular interest in determining the pesticide dose at which 50% (LD_{50}), 75% (LD_{75}), or 90% (LD_{90}) of insect populations respond. Several procedures have been suggested for estimating the best dosage choice for an insecticide. When the response variable is binary or measured ordinally rather than continuously, PROBIT or LOGIT analyses based on Maximum Likelihood procedure are appropriate. This report will explore the use of several dose-response models including LOGISTIC, GENMOD, PROBIT, and CATMOD procedures of SAS/STAT which can all be used for statistical modeling of the dose-response data. All these methods predict the probability of a positive response (insect death) as a function of the pesticide dosage applied. Similar approaches have been widely used in many fields of research including medicine, laboratory animals research, economics, sociology, and genetics.

INTRODUCTION

Entomologists are interested in determining how mortality rates change with respect to the dosage of a pesticide. For instance, an entomologist may have a particular interest in determining the pesticide dose at which 50% (LD_{50}), 75% (LD_{75}), or 90% (LD_{90}) of an insect population respond. The median lethal dosage (LD_{50}) is defined as the statistically derived exposure dosage of a pesticide expected to cause death in 50% of an insect population. Similar definitions apply to LD_{75} and LD_{90} . Several procedures have been suggested for estimating the best dosage choice for an insecticide. Response variables in entomological research may be binary or measured ordinally rather than continuously. For binary or ordinal response, PROBIT or LOGIT procedures based on Maximum Likelihood estimation method are appropriate. For details refer to SAS (1999), Allison (1999), and Bishop et al. (1975).

Binary responses are measured as “yes, no” for insecticide application or “0, 1” showing response=1 if insect damage is present and response=0 if not. The binary response may be coded as “1, 2” representing

“death, life” outcomes (the response is 1 if an insect is dead and 2 otherwise). Sometimes it may even be useful to recode a continuous variable as a binary variable and then analyze the data. Ordinal responses are measured as “high, medium, or low” insect damage or as “standard, mild, or severe” degree of severity of plant disease. “Dose” in this report is defined as the amount of a stimulus (drug, pesticide, etc.) administered, and “response” as the number of insects responding to the administered drug/pesticide dosage. In a dose-response study, increasing dosage levels are sometimes compared with the control dose (zero dose) to assess the effect.

The LOGISTIC, GENMOD, PROBIT, and CATMOD procedures of SAS/STAT can all be used for statistical modeling of the dose-response (categorical variables which can assume only a limited number of discrete values) data (SAS, 1999; Stokes et al., 1995). Proc-Log-Xact from Cytel fits exact Logistic and Poisson regression models to data sets where the usual maximum likelihood methods fail (for details visit www.cytel.com/SASchallenge).

Although the LOGISTIC procedure of SAS is specifically designed for Logistic Regression, many other options for performing Logistic Regression are also available in the SAS System. CATMOD, GENMOD, and PROBIT Procedures of SAS can all perform unconditional likelihood inference for LOGIT models (SAS, 1999). However, differences exist in the way that models are parameterized which might result in different parameter estimates when LOGISTIC Regression is performed using these different methods. Model differences include differences in coding of the data, differences in the sign of the parameter estimates, and the type of the maximum-likelihood algorithm used (for details consult SAS Version 8 online documents). Ordinary least squares regression methods are reported to be inadequate when the dependent variable is discrete (Collett, 1991; Agresti, 1990; SAS, 1999) but Probit analysis is more appropriate for the analysis of qualitative variables within the regression framework. The objectives of this report include: (1) determining how mortality rates change with respect to the dosage applied, (2) modeling Probability of the positive response as a function of the dosage applied, (3) determining the pesticide dosage at which 50% (LD_{50}), 75% (LD_{75}), or 90% (LD_{90}) of insect population

responds, and (4) estimating the best dosage choice for a pesticide/drug application.

METHODS

To determine the effect of increasing dosage levels of a chemical compound (e.g., pesticide), the applied chemical levels are often compared with the zero-dose control. Modeling the response distributions might become the focus of the dose-response study when the response variables are measured at different times and variable doses.

Application

Consider the hypothetical example in which an entomologist wishes to test the effectiveness of a pesticide at 10 dosage levels. To conduct this test 247 insects are assumed to have been randomly divided into 10 groups (assume that the researcher initially intended to use 250 insects, 25 insects/dose) and the number of dead insects (positive response to pesticide= #dead) is recorded for each dosage group (Table 1). Further, assume that the entomologist is interested in determining LD₅₀, LD₇₅, and LD₉₀ values for this pesticide.

TABLE 1. Dosage levels of a pesticide, number of dead insects per dose, and # of test insects per dose (247 insects total)

Dosage level										
0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	
No. dead per dose										
4	6	6	4	8	6	8	16	14	20	
Total # of insect per dose										
25	25	24	25	25	23	24	25	25	26	

The Model

Proc Probit of the SAS version 8 (SAS, 1999) will first be used to analyze the data summarized in Table 1. The ratio of the number dead/total number per dose and the logarithm (Log₁₀) of the dosage levels (Dose) will be included in the PROBIT procedure to model the data and to compare predicted probabilities from various dosage levels. For a given level of Dose, the probability of a binary response with values 0 and 1 (the probability of the lower response level = 0 in this example) is computed using the PROBIT equation:

$$p = \text{Probability (Response=0)} = C + (1 - C)F(X'\$)$$

where, C is the natural (threshold) response rate which can be either set as an initial value or specified from the observed control values, and F is the normal cumulative distribution function determined as:

$\int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2} dz$, $-4 < z < x$, and $z = (x - \mu)/\sigma$, $-4 < x < 4$, where σ is the constant (3.14...), μ is the mean, σ is the standard deviation, and e is the base of the natural logarithm. X in the PROBIT equation shown above is a vector of explanatory variables, and \$ is a vector of parameter estimates (SAS, 1999). The threshold dose may also be regarded as the minimum dose required to produce a detectable response or sometimes as the no observed effect dose in the test population. PROBIT equation computes the cumulative probabilities of the response categories rather than their individual probabilities. The equation could easily be expanded for ordinal response models (e.g., 1=standard, 2=mild, and 3=severe insect damage).

The data in Table 1 can be analyzed using the Logistic option of the Probit procedure (d=Logistic option in the Model statement using SAS) which fits linear logistic regression models, $\pi_i = \sigma(X'\beta)$, to binary or ordinal response data by the method of maximum likelihood (π represents intercept parameters, β is the vector of slope parameters, and X is a vector of explanatory variables). The distribution function may also take the form $[1/(1 + e^{-x})]$. In this report, both Probit and Logistic models will be fit to the data of Table 1 for demonstration.

The "right hand side" of the Logit and Probit are the same but the difference lies on the left-hand side of the equation. For the LOGIT model (a regression model tailored to fit categorical dependent variables), the left hand side is the log of the odds that an insect mortality occurs or does not occur. For example, if Y designates the insect mortality due to a pesticide application, the Logit model would express the effects of the dose of the pesticide (D) on the log of the odds of the occurrence of the insect death versus the absence of the death of an insect. For the Probit model, the left hand side of the equation can be assumed a Z score (the cumulative normal distribution) and the regression slope as a unit change in D corresponding to a unit change in the cumulative normal probability that an insect mortality will occur or not. PROC LOGISTIC of SAS/STAT is a tool for fitting Logistic models to analyze the log odds of an outcome and PROC PROBIT is used for fitting Probit models. As mentioned earlier, several options are available within each procedure to alternate the function of these

procedures. For instance, the Link function in LOGISTIC, has options of LOGIT for the standard Logistic regression model and the NORMIT for the Probit model. They differ from standard regression in substituting maximum likelihood estimation technique for regression's use of least squares estimation of the dependent variable.

Data and SAS Syntax

a) Probit Analysis:

```
Data Insect;
Input Dose Total No_dead @@;
0.0 25 4 0.5 25 6
      .
      .
4.0 25 14 4.5 26 20
;
Proc Probit data=Insect log10 optc;
Model No_dead/Total=Dose;
Output out=New p=p_hat;
Run;
```

b) Logistic/Probit Analysis:

The same data can be analyzed using the LOGISTIC option of Probit:

```
Proc Probit data=Insect log10 optc;
Model No_dead/Total=Dose/d=LOGISTIC;
```

Or using the NORMIT option of Logistic:

```
Proc Logistic data=Insect log10;
Model No_dead/Total=Dose/LINK=NORMIT;
```

Both approaches usually lead to the same conclusions for the same data. GENMOD Procedure can also be used to perform Logistic Regression when the response probability distribution function is binomial and the link function is LOGIT:

```
Proc GENMOD data=Insect log10;
Model No_dead/Total=Dose/DIST=
Binomial LINK=LOGIT;
```

Changing to Link=PROBIT will perform Probit analysis.

RESULTS

The objective of the analysis is to model the probability of insects' positive response (death) as a function of the pesticide dosage applied. The p-values for both intercept (-3.904; SE=1.448; P=0.007) and slope of the Log₁₀ of the dosage level (6.679; SE=2.444; P=0.006) parameters resulting from the PROBIT analysis are significant, yielding the following predicted equation:

$$Y = 0.213 + 0.787(F(-3.904 + 6.679X))$$

where, Y is the cumulative probability estimate of the proportional response, F is the normal cumulative distribution function, and X represents Log₁₀ (Dose) of the pesticide applied. The value 0.213 (SE=0.041) in the Probit equation is the proportion of individuals responding at zero dose (control group), also known as the natural response threshold, and 0.787 in the prediction equation is computed from $1 - 0.213 = 0.787$.

The distribution function F determines the normal or logistic specifications of the model. When probabilities are estimated using the normal distribution function, the LD₅₀ (median lethal dose) for Log₁₀(Dose) is 0.584 (95% confidence interval of 0.505 to 0.656), which corresponds to a probability of 0.50 of the standard normal distribution (Table 2). Corresponding values for LD₇₅ and LD₉₀ are 0.685 (95% confidence interval of 0.626 to 0.951) and 0.776 (0.688 to 1.263). These probabilities on the original scale correspond to 3.84 (LD₅₀), 4.85 (LD₇₅), and 5.97 (LD₉₀), respectively. The upper limit dose used in this experiment (4.5) corresponds to the probability level of 68% of the standard normal distribution (Table 2). For the discussion of normal and log-normal distributions consult Limpert, Stahel, and Abbt (2001).

Table 2. Results of Probit analysis of 10 doses of a pesticide based on standard normal distribution

Probability										
.10	.20	.30	.40	.50	.60	.70	.75	.80	.90	.95
Actual Dose										
2.5	2.9	3.2	3.5	3.8	4.2	4.6	4.8	5.1	6.0	6.8
Log10(Dose)										
.39	.46	.51	.55	.58	.62	.66	.69	.71	.78	.83

PROC LOGISTIC uses a different parameterization from that of PROC PROBIT, which might result in different parameter estimates (-6.426 for intercept and 11.006 for slope parameters). The standard errors of the intercept and slope estimates computed from Logistic procedure are 2.322 and 3.910,

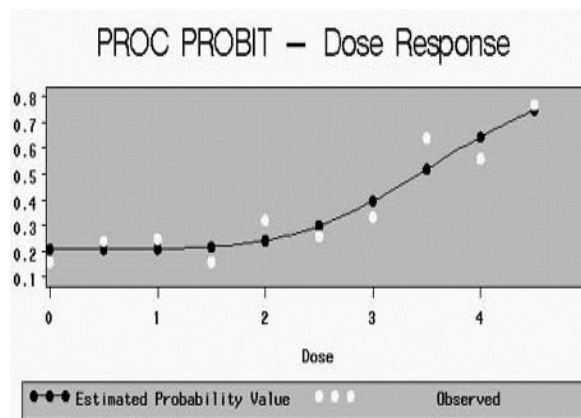
respectively. Parameter estimate differences exist between the two procedures because of the different algorithms used. However, a close correspondence exists in most cases between the Probit and Logistic probability estimates (Table 2 vs Table 3). The LD₅₀, LD₇₅, and LD₉₀ values computed from both Probit (Table 2) and Logistic (Table 3) procedures are very similar in values.

Table 3. Results of Logistic analysis of 10 doses of a pesticide

Probability										
.10	.20	.30	.40	.50	.60	.70	.75	.80	.90	.95
Actual Dose										
2.4	2.9	3.2	3.5	3.8	4.2	4.6	4.8	5.13	6.1	7.1
Log10(Dose)										
.38	.46	.51	.55	.58	.62	.66	.68	.71	.78	.85

Fig. 1 indicates a plot of the PROBIT results of the predicted probability of the dead insects as a function of the dosage levels (vertical axis indicates probabilities and horizontal axis shows dosage levels). The p-values for the Goodness-of-Fit tests (0.6669 for

Fig. 1. Modeling of the relationship between pesticide dose and insect mortality response



the Pearson chi-square and 0.6667 for the likelihood ratio chi-square) indicate an adequate fit for the model describing the relationship between dosage levels and observed and fitted probabilities. The graph of the relationship between dose and response (Fig. 1) shows

rapid increases in probability of insect death as the pesticide dosage increases beyond 2.0. Note that between minimum detectable and maximum, the response varies continuously with the dose. The shape of the dose-response curve is frequently sigmoidal/hyperbolic (not linear) and symmetric when cumulative % responding is plotted against the logarithm of the dose. Fig. 1 allows predictions about the proportion of the insect population responding to given doses of the insecticide applied.

DISCUSSION

Both LOGIT and PROBIT procedures are appropriate to analyze binary, ordered, and categorical data obtained from entomological research. These approaches have been reported frequently (e.g., Singh, 1989; Mehta and Patel, 1995; Derr, 2000) and are best summarized in SAS (1999). They are widely used in many fields of research including agriculture (Throne et. al, 1995a, 1995b; Cairns, 1979), medicine (Devidas et. al, 1993), laboratory animals (Zhang and Zelterman, 1999), economic (Heckman and Willis, 1977; Lo, 1986), sociology (Allison, 1982), and genetics (Lockhart et. al, 1992).

Logistic regression (LR) and ordinary least squares (OLS) regression differ in that LR applies maximum likelihood estimation (MLE) to the natural log of the odds of the dependent variable (LOGIT) and thus LR determines the changes in the log odds of the dependent variable and not changes in the dependent variable itself as does OLS (SAS, 1999). In contrast to the OLS estimation which is based on minimizing the sum of squared distances of the data points to the regression line, MLE seeks to maximize the log likelihood showing the odds that the observed values of the dependent variable may be predicted from the observed values of the independent variables (SAS, 1999). Researchers applying Logistic regressions need not to assume a linear relationship between the dependent and independent variables (a linear relationship between the Logit of the independent variables and the dependent variable is assumed), or assume that the dependent variables or error terms are normally distributed (error terms are assumed to be independent), as they do using the OLS. Since both OLS and Logistic regressions do not assume a linear relationship between the dependent and independent variables, interactions may also be included in the model especially when they are created as additional variables in the analysis as cross products (consult Bondari, 1999 for the discussion of interactions in entomological research).

The hypothetical example presented here demonstrates that the higher the dose the greater the death rate (Fig. 1). The statistical predictions (PROBIT results) of the mortality response varies continuously when plotted against the dose. The response plot is sigmoidal in this example using PROBIT and LOGIT analyses but Calabrese and Baldwin (1999) have presented evidences of U-shaped dose-response curves for toxicological and pharmacological data. The median lethal dose (LD_{50}) computed from the relationship shown in Fig. 1 represents the statistically derived single dose of a pesticide that can be expected to cause death in 50% of a given population of insects. This concept is very important to an entomologist but does not provide sufficient information about the shape of the dose-response curve on which it is based. One reason is that most insects are expected to cluster around the median lethal dose and very few require the extremes of a dosage to respond to pesticide. Model selection will play an important role in the analysis of dose-response data. Hardin (2000) provides analyses of several datasets in which statistical inference depends on the method of estimation.

The data for the dose-response study may be obtained from a variety of different designs with different objectives. In animal and human studies, the risk assessor may choose the study to determine safe exposure levels for carcinogens (the lowest dose that produces a detectable response) which differs from the objectives of a researcher wishing to examine all possible degrees of response between minimum detectable response and a maximum response for an insecticide. Some studies may divide the test population into groups and administer only one of a series of increasing doses of a drug/pesticide to each group (as assumed in the example given in this report). In other studies, increasing doses of a drug/pesticide may be administered to each individual member of a test population until a response is elicited. In either case, sample size and the statistical techniques applied to analyze the dose-response data should receive important considerations. According to Zhang and Zeltermann (1999), estimation of a safe exposure level to a known toxin is one of the most difficult problems that statisticians can face.

An entomologist involved with dose-response research may consult the "probit download" web site by Throne (2000) if not sure which models to use to analyze the dose-response data under consideration (<http://bru.usgmrl.ksu.edu/throne/probit/>). The web site examines data from a variety of study plans and computer programs are presented for statistical analysis of data from several bioassay studies: (1) a program for

the analysis of serial time-mortality data, (2) accessory programs for transforming probit-transformed data back to proportion of test organisms responding to a stimulus (the program will also allow graphing of observed and predicted data to assess goodness-of-fit of the probit line), (3) a program for calculating relative potency of two lethal doses; and (4) a program for determining whether slopes and intercepts from two probit equations are different, and all these programs are online (<http://bru.usgmrl.ksu.edu/throne/>).

Because the function used in log-linear is the logarithm of the dependent variable, in LOGIT is the natural log of the odds ratio, and in PROBIT is the inverse of the standard normal cumulative distribution function, and because the Maximum Likelihood Estimation method computes coefficients which maximize the odds that a dependent variable equals a given value, the interpretation of the results is sometimes difficult. One approach would be to use these techniques to: (1) look to see which variables are significant and to eliminate those which are not significant from the model (the significance of a LOGISTIC or PROBIT model is tested using "-2 log likelihood"), and (2) compare the relative importance of the independent variables. Logistic procedure has been enhanced in SAS Version 9 through the new STRATA and SCORE statements (Stokes et al., 2002), and the new SURVEYLOGISTIC procedure (An, 2002).

REFERENCES

- Agresti, A. 1990. *Categorical Data Analysis*. John Wiley & Sons, Inc. New York, NY.
- Allison, P. D. 1982. Discrete-Time Methods for the Analysis of Event Histories. In: Leinhardt S, editor. *Sociological Methods and Research*, 15: 61-98. San Francisco: Jossey-Bass.
- Allison, P. D. 1999. *Logistic Regression Using the SAS System: Theory and Application*. SAS Institute Inc., Cary, NC.
- An, A. B. 2002. Performing Logistic regression on survey data with the new SURVEYLOGISTIC procedure. Proceedings of the 27th Annual SAS Users Group International Conference, Paper 258-27.
- Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. 1975. *Discrete Multivariate Analysis: Theory and Practice*. The MIT Press, Cambridge, MA.
- Bondari, K. 1999. Interactions in entomology:

- Multiple comparisons and statistical interactions in entomological experimentation. *Journal of Entomological Science* 34: 57-71.
- Cairns, T. 1979. The ED₀₁ study: Introduction, objectives, and experimental design. *Journal of Environmental Pathology and toxicology* 3: 1-7.
- Calabrese, E. J. and Baldwin, L. A. 1999. Reevaluation of the fundamental dose-response relationship. *BioScience* 49: 725-732.
- Collett, D. 1991. *Modelling Binary Data*. Chapman & Hall, London.
- Devidas, M., George, E. O., and Zelterman, D. 1993. Generalized logistic models for low-dose response data. *Statistics in Medicine* 12: 881-892.
- Derr, R. E.. 2000. Performing exact logistic regression with the SAS system. *Proceedings of the 25th Annual SAS Users Group International Conference* 1323-1332.
- Hardin, J. W. 2000. LogXact 4.1 for windows. *The American Statistician*: 54: 320-321.
- Heckman, J. J. and Willis, R. J.. 1977. A beta-logistic model for the analysis of sequential labor force participation by married women. *Journal of Political Economy* 85: 27-58.
- Lockhart, A. M., Piegorsch, W. W., and Bishop, J. B. 1992. Assessing overdispersion and dose response in the male dominant lethal assay. *Mutation Research* 272:35-58.
- Limpert, E., Stahel, W. A., and Abbt, M. 2001. Log-normal distributions across the sciences: Keys and clues. *BioScience* 51: 341-352.
- Lo, A. W. 1986. Logit versus discriminant analysis: A specification test with applications to corporate bankruptcies. *Journal of Econometrics* 31:151-178.
- Mehta, C. R. and Patel, N. R. 1995. Exact Logistic regression: Theory and examples. *Statistics in Medicine* 14: 2143-2160.
- SAS Institute Inc. 1999. *SAS/STAT[®] User's Guide*, Version 8. SAS Institute Inc., Cary, NC.
- Singh, K. P. 1989. Statistical models for analysis of dose-response data. *Proceedings of the 1989 Kansas State University Conference on Applied Statistics in Agriculture* 47-58.
- Stokes, M. E., Davis, C. S., and Koch, G. G. 1995. *Categorical Data Analysis Using the SAS System*. SAS Institute Inc., Cary, NC, 499pp.
- Stokes, M. E., Rodriguez, R. N., and Tobias, R. D.. 2002. A preview of SAS/STAT[®] Version 9: Moving in new directions and building on old favorites. *Proceedings of the 27th Annual SAS Users Group International Conference*, Paper 257-27.
- Throne, J. E., Weaver, D. K., Chew, V., and Baker, J. E. 1995a. Probit analysis of correlated data: Multiple observations over time at one concentration. *Journal of Economical Entomology*. 88: 1510-1512.
- Throne, J. E., Weaver, D. K., and Baker, J. E. 1995b. Probit analysis: Assessing goodness-of-fit based on back transformation and residuals. *Journal of Economical Entomology* 88: 1513-1516.
- Zhang, H. and Zelterman, D. 1999. Binary regression for risks in excess of subject-specific thresholds. *Biometrics* 55: 1247-1251.

CONTACT INFORMATION

Contact the author at:

K. Bondari, Experimental Statistics, Coastal Plain Experiment Station, University of Georgia, Tifton, GA 31793-0748.

Phone: (229) 386-3385

Fax: (229) 391-2501

E-mail: Bondari@Tifton.UGA.EDU