Article

Mathematical methodology and *MATLAB* computer program to calculate the effective-dose of percent response using the probit analysis technique

M. Tlas, **B. A. Ghani**
Scientific Services Department, Atomic Energy Commission of Syria (AECS), P. O. Box 6091, Syria
E-mail: pscientific31@aec.org.sy

Received 31 October 2022; *Accepted 10 December 2022*; *Published online 20 December 2022*; *Published 1 March 2023*

Abstract
Using the probit analysis technique, mathematical algorithm and *MATLAB* computer program have been implemented in this paper to calculate both the Log-Dose (*LD*) and the Effective-Dose (*ED*) for any given percent. The probit analysis uses a successive weighted simple linear regression of experimental or row data. The kind of row data obtained from the bioassays should be generally in percent response (mortality or affected) at the corresponding doses (or concentrations). The response should be always in binomial form (e.g. death/no death) and the relationship between the response and the various doses or concentrations is always sigmoid or non-linear. The probit analysis here acts as a transformation from sigmoid or non-linear relationship to linear one and then uses a successive weighted simple linear regression on the linear relationship of the observed data. It is necessarily to note that, the probit analysis always assumes that the relationship between number responding (not proportion response) and dose (or concentration) should be normally distributed. Two simple examples are explained in this paper to prove the validity and the consistency of both the proposed mathematical methodology and the concerning computer program.

Keywords probit analysis; weighted simple linear regression; bioassay; transformation; Log-Dose; Effective-Dose; Log-Concentration; Effective-Concentration.

Computational Ecology and Software
ISSN 2220-721X
URL: http://www.iaees.org/publications/journals/ces/online-version.asp
RSS: http://www.iaees.org/publications/journals/ces/rss.xml
E-mail: ces@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

The probit analysis technique is considered as the most popular approach and methodology to modeling binomial response variables. It used extensively in the biological sciences, biomedical research, toxicology, and engineering. Indeed, even in the social sciences binomial response variables are found to be plentiful. The basic distribution for the response is either Bernoulli or binomial. The probit analysis transforms the sigmoid

dose-response curve to a straight line that can then be analyzed by regression either through least squares method or maximum likelihood function.

The idea of probit analysis was originally published in Science by Chester Ittner Bliss, 1934. He worked as an entomologist for the Connecticut agricultural experiment station and was primarily concerned with finding an effective pesticide to control insects that fed on grape leaves (Greenberg, 1980). By plotting the response of the insects to various concentrations of pesticides, he could visually see that each pesticide affected the insects at different concentrations, i.e. one was more effective than the other. However, he didn't have a statistically sound method to compare this difference. The most logical approach would be to fit a regression of the response versus the concentration, or dose and compare between the different pesticides. Yet, the relationship of response to dose was sigmoid or non-linear in nature and at the time regression was only used on linear data. Therefore, Bliss developed the idea of transforming the sigmoid dose-response curve to a straight line. In 1952, a professor of statistics at the University of Edinburgh by the name of David Finney took Bliss' idea and wrote books called Probit Analysis (Finney, 1952, 1964). Today, probit analysis is still the preferred statistical method in understanding dose-response relationships.

Probit analysis is also commonly used in toxicology to determine the relative toxicity of chemicals to living organisms. This is done by testing the response of an organism under various concentrations of each of the chemicals in question and then comparing the concentrations at which one encounters a response. The response is always binomial (e.g. death/no death) and the relationship between the response and the various concentrations is always sigmoid or non-linear. Probit analysis acts as a transformation from sigmoid to linear and then runs a regression on the relationship.

Once a regression is run, the researcher can use the output of the probit analysis to compare the amount of chemical required to create the same response in each of the various chemicals. There are many endpoints used to compare the differing toxicities of chemicals, but the Log-Concentration (*percent*) or in abbreviation *LC* (*percent*) or Log-Dose (*percent*) and in abbreviation *LD* (*percent*) are the most widely used outcomes of the dose-response experiments. The *LC* (*percent*)/*LD* (*percent*) represent the concentration (*LC* (*percent*)) or dose (*LD* (*percent*)) at which percent of the population responds.

After knowing the *LC* (*percent*)/*LD* (*percent*), it easily to be calculated the effective- Concentration (*percent*) and in abbreviation *EC* (*percent*) and the Effective-Dose (*percent*) and in abbreviation *ED* (*percent*).

For example, consider comparing the toxicity of two different pesticides to aphids, pesticide *A* and pesticide *B*. If the *LC* (0.50) of pesticide *A* is 50 μg/L and the *LC* (0.50) of pesticide *B* is 10 μg/L, pesticide *B* is more toxic than *A* because it only takes 10ug/L to kill 50% of the aphids, versus 50 μg/L of pesticide *B*. The same explanation for the *EC* (0.50) which is easily results from *LC* (0.50).

For more details, profound and expanded explanations about the probit analysis, readers are invited to consult the papers Throne et al. (1995), Moermans and Nelis (1994), Sakuma (1998), Kumar et al. (2020) and Vincent (2008).

2 Methodology and Algorithm

Input Data: Input the vectors: concentration, proportion, total number

Imput the numbers: total iterations, percent

where

k is the number of iterations

m is the number of groups

total number (*i*) is the total number in the group (*i*) (*i*=1,...,*m*)

percent is 50%, 60%, ...

ε is a small positive number close to zero

Replacement: Set $c(i)$=concentration(i) for i=1,...,m

Set $p(i)$=proportion(i) for i=1,...,m

Set $n(i)$=total number(i) for i=1,...,m

Initial iteration **(**$k=1$**)**

Calculate the vector: $x(i)=\log_2\left(\frac{c(i)}{c(1)}\right) \quad (i=1,...,m)$

Calculate the vector: $y(i)=\text{norminv}[p(i)] \, (i=1,...,m)$

Calculate the vector: $z(i)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(y(i))^2}{2}} \quad (i=1,...,m)$

Calculate the vector: $w(i)=\frac{n(i)\left(z(i)\right)^2}{p(i)\left(1-p(i)\right)} \quad (i=1,...,m)$

Solve the following set of linear equations:

$$\begin{bmatrix} \sum_{i=1}^{m} w(i) & \sum_{i=1}^{m} w(i)x(i) \\ \sum_{i=1}^{m} w(i)x(i) & \sum_{i=1}^{m} w(i)\left(x(i)\right)^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} w(i)y(i) \\ \sum_{i=1}^{m} w(i)x(i)y(i) \end{bmatrix}$$

To find the intercept $\alpha^1=\alpha$ and the slope $\beta^1=\beta$

Put $k=1$

Iteration **(**k**)**

While ($k \le totaliterations$) **do**

Calculate the vector: $Y^k(i)=\alpha^k+\beta^k x(i) \quad (i=1,...,m)$

Calculate the vector: $P^k(i)=normcdf\,[Y^k(i)] \quad (i=1,...,m)$

Calculate the vector: $Z^k(i)=\frac{1}{\sqrt{2\pi}}e^{-\frac{\left(Y^k(i)\right)^2}{2}} \quad (i=1,...,m)$

Calculate the vector: $W^k(i)=\frac{n(i)\left(Z^k(i)\right)^2}{P^k(i)\left(1-P^k(i)\right)} \quad (i=1,...,m)$

Solve the coming set of linear equations:

$$\begin{bmatrix} \sum_{i=1}^{m} W^k(i) & \sum_{i=1}^{m} W^k(i)x(i) \\ \sum_{i=1}^{m} W^k(i)x(i) & \sum_{i=1}^{m} W^k(i)\left(x(i)\right)^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} W^k(i)y(i) \\ \sum_{i=1}^{m} W^k(i)x(i)y(i) \end{bmatrix}$$

To find the intercept $\alpha^{k+1}=\alpha$ and the slope $\beta^{k+1}=\beta$

Stopping criteria

If ($k = totaliterations$) **then stop** or,

If ($\sqrt{\left(\alpha^{k+1}-\alpha^{k}\right)^2+\left(\beta^{k+1}-\beta^{k}\right)^2} < \varepsilon$) **then stop**

Else put $k = k+1$

End while

Set intercept= α^k and slop= β^k

Calculate the Log-Dose by: $LD(percent) = \dfrac{\text{norminv}(percent) - \text{intercept}}{\text{slope}}$

Calculate the effective-Dose by: $ED(percent) = c(1) \times 2^{LD(percent)}$

Calculate the variance of the Log-Dose by:

$$\text{Variance}[LD(percent)] = \frac{1}{(\beta^{k})^2}\left[\frac{1}{\sum_{i=1}^{m} W^{k}(i)} + \frac{\left[\bar{x} - LD(percent)\right]^2}{\sum_{i=1}^{m} W^{k}(i)\left[x(i)-\bar{x}\right]^2}\right]$$

Where: $\bar{x} = \dfrac{\sum_{i=1}^{m} W^{k}(i)x(i)}{\sum_{i=1}^{m} W^{k}(i)}$ and

$$\sum_{i=1}^{m} W^{k}(i)\left[x(i)-\bar{x}\right]^2 = \sum_{i=1}^{m} W^{k}(i)\left[x(i)\right]^2 - \frac{\left[\sum_{i=1}^{m} W^{k}(i)x(i)\right]^2}{\sum_{i=1}^{m} W^{k}(i)}$$

Calculate the standard deviation of $LD(percent)$ by:

$$\sigma[LD(percent)] = \sqrt{\text{Variance}[LD(percent)]}$$

Calculate the standard deviation of the concentration or dose by:

$$\sigma[ED(percent)] = ED(percent) \times \sigma[LD(percent)]$$

3 Notes and General Remarks

1. Impute the dose or the concentration vector in increasing order from small to large concentration(dose) including the control concentration(dose)
2. If there is a zero in the vectors: concentration or proportion, it should be replaced by ε where ε is a small positive number close to zero to respect the logarithm function

3. If there is a one (one hundred percent) in the vectors: concentration and proportion, it should be replaced by $(1-\varepsilon)$ where ε is a small positive number close to zero to respect the logarithm function
4. There is not any changes if we replace $\log_2$ by $\log_e$ or $\log_{10}$ in the methodology, we can get the same results
5. We can use another weighting factors as: $W(i)=\frac{1}{\sqrt{n(i)P(i)[1-P(i)]}}$ $(i=1,...,.m)$ or any other mathematical formulas in the weighting simple linear regression
6. The reserved function "normcdf" in *MATLAB* is the normal cumulative distribution function and it is defined by:

$$p(i)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y(i)} e^{-\frac{t^2}{2}}dt \equiv normcdf\ [y(i)] \quad (i=1,...,m)$$

7. The reserved function "norminv" in *MATLAB* is the inverse of the normal cumulative distribution function and it is defined by:

$$prob[-\infty < Z\ < y(i)] = p(i) \equiv y(i) = \text{norminv}[p(i)] \quad (i=1,...,m)$$

where Z the standard normal variable with an arithmetic mean is zero and a standard deviation is one

4 Testing Validity and Accuracy of Both Methodology and Computer Code

Two examples are treated in this work to prove the validity and accuracy of the methodology and the concerning computer code (see supplementary material).

Example 1

The data set in Table 1 below shows the effect different doses of nicotine on the common fruit fly. The purpose of the experiment was to use the probit analysis to arrive at an appropriate model relating probability of "kill" to concentration. The analyst sought the *LD* (*percent*), that is, the concentration of nicotine that results in a certain probability. Of particular interest is the *LD* (*percent*), the concentration that produces a percent probability of "insect kill".

Table 1 Data set of example 1.

Concentration(dose) level	Number of insects In group	Number of insects killed In group	Proportion of insects killed in group
0.10	47	8	0.170
0.15	53	14	0.264
0.20	55	24	0.436
0.30	52	32	0.615
0.50	46	38	0.826
0.70	54	50	0.926
0.95	52	50	0.962

Imput data of the program are:

concentration=[0.1 0.15 0.20 0.30 0.50 0.7 0.95] "concentration or dose vector"

proportion=[0.170 0.264 0.436 0.615 0.826 0.926 0.962]

total number=[47 53 55 52 46 54 52]"numbers in groups"

total iterations=100

percent=0.50 and percent=0.70

Results produced by the program are:

$LD(0.50)=1.2143$ with $\sigma[LD(0.50)]=0.0918$

$ED(0.50)=0.2320$ with $\sigma[ED(0.50)]=0.0213$ Versus the effective dose value 0.23176 reported by the *SPSS* statistical package

$LD(0.70)=1.8183$ with $\sigma[LD(0.70)]=0.0982$

$ED(0.70)=0.3527$ with $\sigma[ED(0.70)]=0.0346$ Versus the effective dose value 0.35273 reported by the *SPSS* statistical package

Example 2

A typical set of data with groups' number of 20 larvae tested for each dose or concentration is given in Table 2 below.

Table 2 Data set of example 2.

Concentration(dose) level	Number of larvae In group	Number of larvae killed In group	Proportion of larvae killed In group
0.00375	20	0	0
0.0075	20	1	0.05
0.015	20	8	0.40
0.03	20	11	0.55
0.06	20	16	0.80
0.12	20	18	0.90
0.24	20	20	1.00

Imput data of the program are:

concentration=[0.00375 0.0075 0.015 0.03 0.06 0.12 0.24] "concentration or dose vector"

proportion=[0.001 0.05 0.40 0.55 0.80 0.90 0.999]

total number=[20 20 20 20 20 20 20]"numbers in groups"

total iterations=100

percent=0.50 and percent=0.80

Results produced by the program are:

$LD(0.50)=2.8391$ with $\sigma[LD(0.50)]=0.1885$

$ED(0.50)=0.0268$ with $\sigma[ED(0.50)]=0.0051$ Versus the effective dose value 0.0265 reported by

the *SPSS* statistical package

$LD(0.80)$=3.9090 with $\sigma[LD(0.80)] = 0.2391$

$ED(0.80)$=0.0563 with $\sigma[ED(0.80)] = 0.0135$ Versus the effective dose value 0.05913 reported by the *SPSS* statistical package

5 Conclusions

An easy mathematical methodology and interactive computer software to calculate both the Log-Dose (*LD*) and the Effective-Dose (*ED*) for any given percent have been proposed in this paper. The proposed methodology and the code algorithm are mainly based on famous mathematical methodology well-known in applied statistics as probit analysis. The probit analysis acts as a transformation from sigmoid or non-linear relationship to linear one and then uses a successive weighted simple linear regression on the linear relationship of experimental or row data.

Two simple examples are shown to prove the validity and accuracy of both the proposed methodology and the concerned computer code. The code is implemented and programmed in *MATLAB* environment; it is easy to use and easy to be developed. The software is appended and it can be loaded directly when it is needed.

Acknowledgments

Author wishes to thank Prof. I. Othman, the DG of the AECS for his valuable support and encouragement throughout this work. The anonymous reviewers are cordially thanked for their critics, remarks and suggestions that considerably improved the final version of this paper.

References

Finney DJ. 1952. Probit Analysis. Cambridge University Press, Cambridge, UK

Finney DJ. 1964. Probit Analysis: A Statistical Treatment of The Sigmoid Response Curve. Cambridge University Press, London, UK

Finney DJ, Stevens WL. 1948. A table for the calculation of working probits and weights in probit analysis. Biometrika, 35(1-2): 191-201

Greenberg BG. 1980. Chester I. Bliss, 1899-1979. International Statistical Review / Revue Internationale de Statistique, 8(1): 135-136

Hahn ED, Soyer R. Probit and Logit Models: Differences in a Multivariate Realm. http://home.gwu.edu/~soyer/mv1h.pdf

Kumar V, Sheoran OP, Rani S, Malik K. 2020. Development of a web-based tool for probit analysis to compute LC50/LD50/GR50 for its use in toxicology studies. Journal of Applied Natural Science, 12(4): 641-646

Moermans R, Nelis S. 1994. Determination of the LD50: choice of the underlying model. Journal of Applied Entomology, 117: 210-214

Sakuma M.1998. Probit analysis of preference data. Applied Entomology and Zoology 33(3): 339-347

Throne JE, Weaver DK, Chew V, Baker JE. 1995. Probit analysis of correlated data: Multiple observations over time at one concentration. Journal of Economic Entomology, 88(5): 1510-1512

Throne JE, Weaver DK, Baker JE. 1995. Probit analysis: Assessing goodness of fit based on back transformation of fit and residuals. Journal of Economic Entomology, 88(5): 1513-1516

Vincent K. 2008. Probit Analysis. San Francisco State University, San Francisco, USA