Probit Analysis
By: Kim Vincent

Quick Overview

- Probit analysis is a type of regression used to analyze binomial response variables.
- It transforms the sigmoid dose-response curve to a straight line that can then be analyzed by regression either through least squares or maximum likelihood.
- Probit analysis can be conducted by one of three techniques:
    - Using tables to estimate the probits and fitting the relationship by eye,
    - Hand calculating the probits, regression coefficient, and confidence intervals, or
    - Having a stastitical package such as SPSS do it all for you.

Background

The idea of probit analysis was originally published in *Science* by Chester Ittner Bliss in 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 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 a book called *Probit Analysis* (Finney 1952). Today, probit analysis is still the preferred statistical method in understanding dose-response relationships.

The Basics

Probit Analysis is a specialized regression model of binomial response variables.

Remember that regression is a method of fitting a line to your data to compare the relationship of the response variable or dependent variable (Y) to the independent variable (X).

Y = a + b X + e

Where
- a = y-intercept
- b = the slope of the line
- e = error term

Also remember that a binomial response variable refers to a response variable with only two outcomes.

For example:

- Flipping a coin: Heads or tails
- Testing beauty products: Rash/no rash
- The effectiveness or toxicity of pesticides: Death/no death

Applications

Probit analysis is used to analyze many kinds of dose-response or binomial response experiments in a variety of fields. However, because my background knowledge of probit analysis stems only from toxicology, the examples from this webpage will only be of toxicology.

Probit Analysis is 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. As discussed above, the response is always binomial (e.g. death/no death) and the relationship between the response and the various concentrations is always sigmoid. 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 LC50 (liquids) or LD50 (solids) are the most widely used outcomes of the modern dose-response experiments. The LC50/LD50 represent the concentration (LC50) or dose (LD50) at which 50% of the population responds.

For example, consider comparing the toxicity of two different pesticides to aphids, pesticide A and pesticide B. If the LC50 of pesticide A is 50ug/L and the LC50 of pesticide B is 10ug/L, pesticide B is more toxic than A because it only takes 10ug/L to kill 50% of the aphids, versus 50ug/L of pesticide B.

How does probit analysis work? How to get from dose-response curve to an LC50?

Below you will find a step by step guide to using probit analysis with various methods. The easiest by far is to use a statistical package such as SPSS, SAS, R, or S, but it is good to see the history of the methodology to get a thorough understanding of the material.

Step 1: Convert % mortality to probits (short for probability unit)

Method A: Determine probits by looking up those corresponding to the % responded in Finney's table (Finney 1952):

Table 3.2 Transformation of percentages to probits

%	0	1	2	3	4	5	6	7	8	9
0	—	2.67	2.95	3.12	3.25	3.36	3.45	3.52	3.59	3.66
10	3.72	3.77	3.82	3.87	3.92	3.96	4.01	4.05	4.08	4.12
20	4.16	4.19	4.23	4.26	4.29	4.33	4.36	4.39	4.42	4.45
30	4.48	4.50	4.53	4.56	4.59	4.61	4.64	4.67	4.69	4.72
40	4.75	4.77	4.80	4.82	4.85	4.87	4.90	4.92	4.95	4.97
50	5.00	5.03	5.05	5.08	5.10	5.13	5.15	5.18	5.20	5.23
60	5.25	5.28	5.31	5.33	5.36	5.39	5.41	5.44	5.47	5.50
70	5.52	5.55	5.58	5.61	5.64	5.67	5.71	5.74	5.77	5.81
80	5.84	5.88	5.92	5.95	5.99	6.04	6.08	6.13	6.18	6.23
90	6.28	6.34	6.41	6.48	6.55	6.64	6.75	6.88	7.05	7.33
—	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
99	7.33	7.37	7.41	7.46	7.51	7.58	7.65	7.75	7.88	8.09

For example, for a 17% response, the corresponding probit would be 4.05. Additionally, for a 50% response (LC50), the corresponding probit would be 5.00.

Method B: Hand calculations (Finney and Stevens 1948):

The probit Y, of the proportion P is defined by:

$$P = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{Y-5} e^{-\frac{1}{2}u^2}\, du.$$

The standard method of analysis makes use of the maximum and minimum working probits:

$$Y_{max.} = Y + \frac{Q}{Z}$$

$$Y_{min.} = Y - \frac{P}{Z}:$$

And the range 1/Z where

$$Z = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(Y-5)^2}.$$

Method C: Computer software such as SPSS, SAS, R, or S convert the percent responded to probits automatically.

Step 2: Take the log of the concentrations.

This can either be done by hand if doing hand calculations, or specify this action in the computer program of choice.

For example, after clicking Analyze, Regression, Probit, choose the log of your choice to transform:
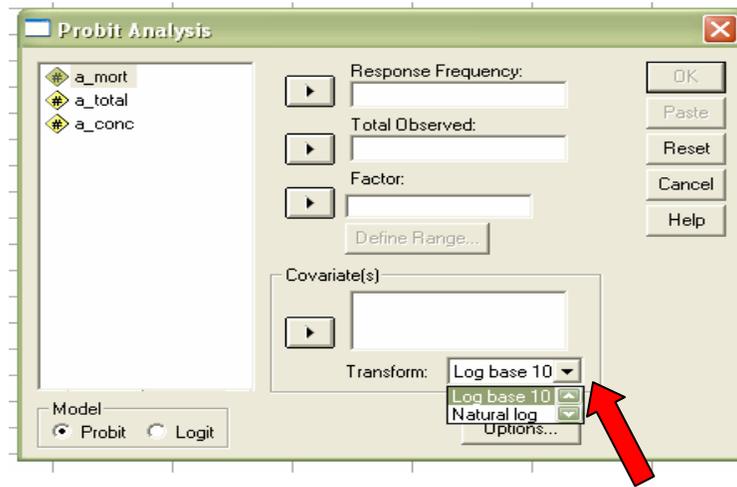
Step 3: Graph the probits versus the log of the concentrations and fit a line of regression.

Note: Both least squares and maximum likelihood are acceptable techniques to fitting the regression, but maximum likelihood is preferred because it gives more precise estimation of necessary parameters for correct evaluation of the results (Finney 1952).

Method A: Hand fit the line by eye that minimizes the space between the line and the data (i.e. least squares). Although this method can be surprisingly accurate, calculating a regression by hand or using computer program is obviously more precise. In addition, hand calculations and computer programs can provide confidence intervals.

Method B: Hand calculate the linear regression $Y = 5 + (x - \mu)/\sigma,$ by using the following method (Finney and Stevens 1948):

First set the proportion responding to be equal to p = r/n and the complement equal to q = 1-p.

The probits of a set value of p should be approximately linearly related to x, the measure of the stimulus, and a line fitted by eye may be used to give a corresponding set of *expected probits*, Y.

The working probit corresponding to each proportion is next calculated from either of the following equations:

$$y = Y + Q/Z - q/Z,$$
$$y = Y - P/Z + p/Z,$$

Next a set of expected probits is then derived from the weighted linear regression equation of working probits on x, each y being assigned a weight, nw, where the weighting coefficient, w, is defined as:

$$w = Z^2/PQ.$$

The process is repeated with the new set of Y values. The iteration converges to give you a linear regression.

Method C: Use a computer program. SPSS uses maximum likelihood to estimate the linear regression.

To run the probit anaylsis in SPSS, follow the following simple steps:

Simply input a minimum of three columns into the Data Editor
- Number of individuals per container that responded
- Total of individuals per container
- Concentrations

For example in the following screen, a_mort is the number of individuals that responded per container, a_total is the total number of individuals per container, and a_conc are the concentrations. Row 7 in the following example is data from the control where 0 out of 10 responded at a concentration of 0.

Screen 1:
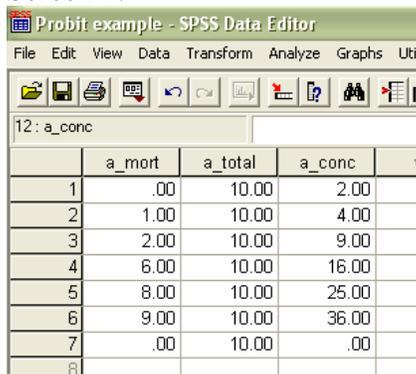
	a_mort	a_total	a_conc
1	.00	10.00	2.00
2	1.00	10.00	4.00
3	2.00	10.00	9.00
4	6.00	10.00	16.00
5	8.00	10.00	25.00
6	9.00	10.00	36.00
7	.00	10.00	.00
8			

After you columns are set, simply go to analyze, regression, probit:
Screen 2:
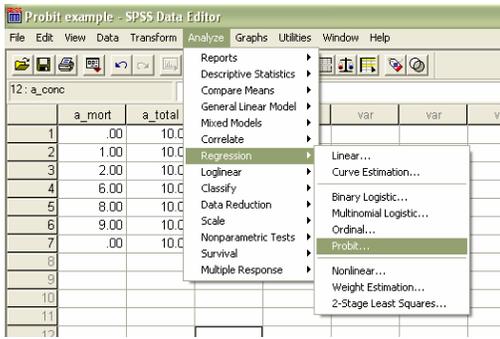
Then set your number responded column as the "Response Frequency", the total number per container as the "Total Observed", and the concentrations as the "Covariates". Don't forget to select the log base 10 to transform your concentrations.

Screen 3:
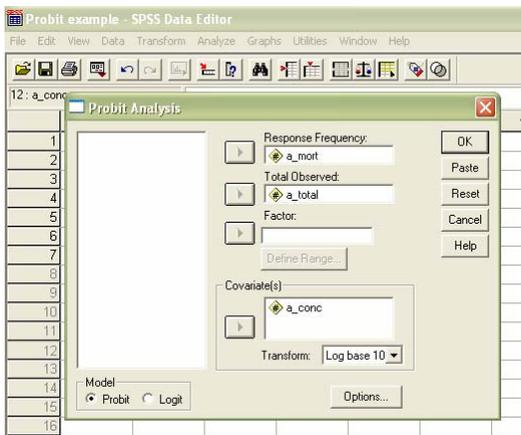
If you run the above example, you will see that SPSS determines an optimal solution after 18 iterations.

Step 4: Find the LC50

Method A: Using your hand drawn graph, either created by eye or by calculating the regression by hand, find the probit of 5 in the y-axis, then move down to the x-axis and find the log of the concentration associated with it. Then take the inverse of the log and voila! You have the LC50.
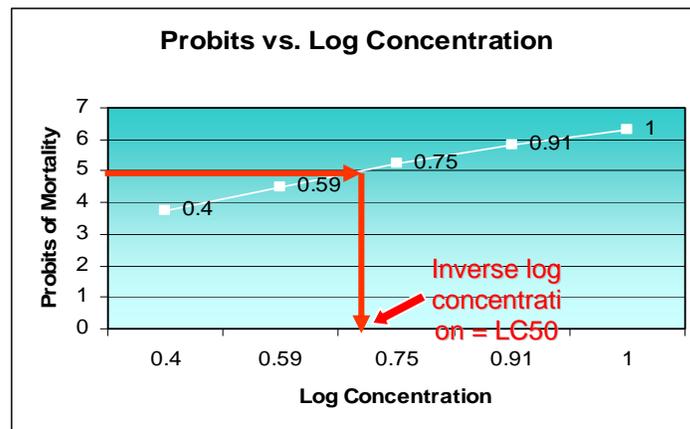
Method B: The LC50 is determined by searching the probit list for a probit of 5.00 and then taking the inverse log of the concentration it is associated with.

Step 5: Determine the 95% confidence intervals:

Method A: Hand calculate using the following equation:

The standard error is approximately: $\pm 1/b \sqrt{(Snw)}$

- b = estimate of the slope of the line
- Snw = summation of nw
- w = weighted coefficient from table III = Z^2/PQ Finney, 1952

Method B: SPSS and other computer programs calculate this automatically

Notes of Interest for Probit Analysis

- Probit analysis assumes that the relationship between number responding (not percent response) and concentration is normally distributed.
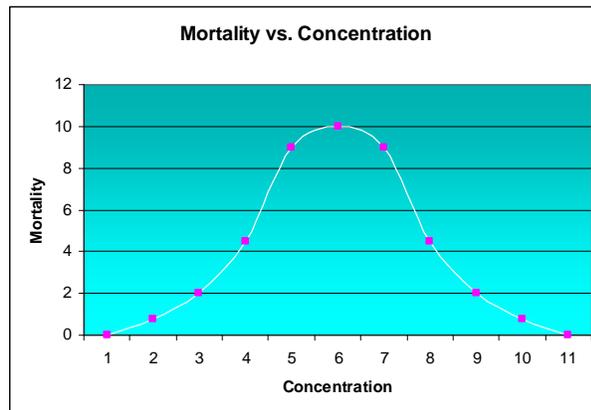
If data are not normally distributed, logit (more on this below) is preferred.

- Must correct data if there is more than 10% mortality in the control

One method is to use the Schneider-Orelli's (1947) formula:

$$\text{Corrected} = \frac{\% \text{ Responded} - \% \text{ Responded in Control}}{100 - \text{Responded } \% \text{ in Control}} \times 100$$

For example:
Let's say you have 20% mortality in the control and you are correcting the mortality rate for the concentration where 60% occurred.   Plug in the mortality rates into the equation above and you come up with a mortality of 50% instead of the original 60%.

$$\frac{60\% - 20\%}{100\% - 20\%} \qquad 40/80 = 50\%$$

Logit vs. Probit:

Logit is another form of transforming binomial data into linearity and is very similar to probit. Logit functions by taking the log of the odds: $logit(P) = \log P/ (1-P)$. Yet, the relationship between logit and probit is almost indistinguishable: $Logit \approx (\pi/\sqrt{3})$ x probit. In general, if response vs. dose data are not normally distributed, Finney suggests using the logit over the probit transformation (Finney, 1952). Although the multivariate usage of probit analysis is beyond the content of this webpage, it is worth noting that the similarity between probit and logit doesn't hold in a multivariate realm (Hahn and Soyer date unknown). Hahn and Soyer suggest that logit provides a better fit in the presence of extreme independent variable levels and conversely that probit better fit random effects models with moderate data sets (Hahn and Soyer date unknown).

Summary

- Probit Analysis is a type of regression used with binomial response variables. It is very similar to logit, but is preferred when data are normally distributed.

- Most common outcome of a dose-response experiment in which probit analysis is used is the LC50/LD50.

- Probit analysis can be done by eye, through hand calculations, or by using a statistical program.

Citations:

Finney, D. J., Ed. (1952). Probit Analysis. Cambridge, England, Cambridge University Press.

Finney, D. J. and W. L. Stevens (1948). "A table for the calculation of working probits and weights in probit analysis." Biometrika **35**(1-2): 191-201.

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

Hahn, E. D. and R. Soyer. (date unknown). "Probit and Logit Models: Differences in a Multivariate Realm."   Retrieved May 28, 2008, from http://home.gwu.edu/~soyer/mv1h.pdf.