SOME MATHEMATICAL ASPECTS OF PROBIT ANALYSIS

by

PHIL DEAN GILLILAND

B. S., Kansas State University, 1959

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1964

Approved by:

Gary L. Krause
Major Professor

TABLE OF CONTENTS

INTRODUCTION

The area in which probit analysis is generally utilized is that of biological assay, where biological assay should be understood to mean the determination of potency of a stimulus, whether physical, chemical, biological, psychological or physiological, by means of the reactions which it produces in living matter. Biological assay is most commonly considered as referring to the assessment of the potency of vitamins, hormones, toxicants, and drugs of all types by means of the responses produced when doses of these preparations are given to suitable experimental animals.

One type of assay which has been found valuable in many different fields, but especially in toxicological studies, is dependent on a quantal response. Though quantitative measurement of response is always to be preferred when available, there are certain responses which permit no graduation and which can only be expressed as "occurring" or "not occurring". The most obvious example of this is death, although workers with insects have often found difficulty in deciding precisely when an insect is dead. In many investigations, the only practical interest lies in whether or not a test insect is dead or perhaps in whether or not it has reached a degree of inactivity such as is thought to be followed by early death.

One feature possessed by all biological assays is the variability in the reaction of the test subjects and the consequent impossibility of reproducing, at will, the same results in successive trials, however carefully the experimental conditions are controlled.

The statistical treatment of quantal-response data has been much aided by the development of probit analysis. This method has been widely adapted as the standard method of reducing the data to simple terms.

Unfortunately, the procedures involved in the maximum likelihood method of estimating the normal equations are not given in detail in any of the standard reference texts. Therefore, the purpose of this report is to outline these procedures and to show their derivations.

APPLICATION TO A GENERAL PROBLEM

The need for probit analysis arises from a general toxicology experiment where various concentrations of the chemical are prepared and a batch of insects is assigned at random to each concentration level. The chemical is applied and, for each batch, a count is made of the total number of insects (n) and the number killed (r). The ratio of the number killed to the total number gives the sample death proportion, $p = r/n$.

For any one subject, there is a level of intensity of the stimulus below which a response does not occur and above which it does occur. This level is referred to as the tolerance (or threshold) for that subject. Let this tolerance (or threshold) of a subject be represented by λ; then, in a population of subjects, the main concern is with the distribution of λ. This distribution of tolerances may be expressed by

$$d\pi = f(\lambda)\, d\lambda; \tag{1}$$

this equation states that a proportion, $d\pi$, of the whole population consists of individuals whose tolerances lie between λ and $\lambda + d\lambda$, where $d\lambda$ represents a small interval on the dose scale, and that $d\pi$ is the length of this interval multiplied by the appropriate value of the distribution function, $f(\lambda)$.

If a dose λ_o is given to the whole population, all individuals will respond whose tolerances are less than λ_o, and the proportion of these is

π, where

$$\pi_o = \int_o^{\lambda_o} f(\lambda)\, d\lambda; \tag{2}$$

the measure of dose is here assumed to be a quantity which can conceivably range from zero to $+\infty$, response being certain for very high doses, so that

$$\int_o^{\infty} f(\lambda)\, d\lambda = 1. \tag{3}$$

The distribution of tolerances, as measured on the natural scale, may be markedly skew, but it is often possible, by a simple transformation of the scale of measurement, to obtain a distribution which is approximately normal. Although the distribution of tolerance concentration of a toxic agent is usually far from symmetrical, on account of a few individuals with high tolerances providing an extended "tail" to the distribution, normalization can often be effected by expressing the tolerances in terms of the logarithms of the concentrations, instead of the absolute values; this transformation is now accepted as standard practice for expressing the results of such trials. The use of the log concentration for measuring the dosage in quantal trials requires no more justification than that it introduces a simplification into the analysis. (Finney 1952a)

It is convenient to take x as representing the intensity of the stimulus on the scale on which the tolerances are normally distributed, and λ as the untransformed value of concentration. Thus, for much quantal response work,

$$x = \log_{10} \lambda; \tag{4}$$

x will be referred to as the dosage and λ will be referred to as the dose.

In an investigation for which tolerance can be satisfactorily defined, so that, for any given dose, all individuals with equal or lower tolerance values will respond, a graph of the sample death proportion responding against the dose will give a steadily rising curve. The rate of increase in response per unit increase in dose is frequently very low with minimal and maximal dosage, but higher with intermediate values, so that the curve is sigmoidal. When the stimulus is measured in dosage units, the curve takes the characteristic normal sigmoid form. This curve does not attain the 0% or 100% response except at infinitely low or infinitely high dosage, a situation that does not truly arise (except that, when the measure of dosage intensity is logarithmic, an infinitely low value represents zero dose).

Assuming that tolerance measures in a population are normally distributed, on the dosage scale the relationships between the normal curve and the probit transformation can be seen from Figure 1.

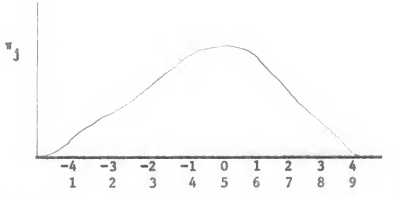

Figure 1. Relationship Between the N.E.D. and the Probit Transformation

The dosage deviations from the mean $(x_j - \bar{x})$ are replaced on the base line by $\frac{(x_j - \bar{x})}{\sigma}$ or what is called the normal equivalent deviate (N.E.D.)

This transformation to normal equivalent deviates was first made by Fechner (1860), but was not considered seriously until it was made again by Gaddum (1933). Bliss (1934) suggested adding a constant (five) to the normal equivalent deviate to remove negative numbers and also suggested the name of the transformation, probit. Therefore, the probit of the proportion π is defined as the abscissa which corresponds to a probability π in a normal distribution with mean = 5 and variance unity; in symbols, the probit of π is y where

$$\pi = \frac{1}{\sqrt{2c}} \int_{-\infty}^{y-5} e^{-\frac{1}{2}u^2} du, \qquad (5)$$

where c is 3.14159 .

If $d\pi = \frac{1}{\sqrt{2c}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} dx$ represents the element of probability from the distribution of tolerances on the x scale of dosages, the expected proportion of insects killed by a dosage x_o is

$$\pi_o = \frac{1}{\sqrt{2c}\,\sigma} \int_{o}^{x_o} e^{-\frac{1}{2}\left(\frac{x_o-\mu}{\sigma}\right)^2} dx_o \qquad (6)$$

Comparison of the two preceding formulas for π shows that the probit of the expected proportion killed is related to the dosage by the linear equation

$$y = 5 + \frac{1}{\sigma}(x - \mu). \qquad (7)$$

which says that the probit equals the N.E.D. + 5. A probit of 5 implies a mortality of .5 . (Finney, 1952a).

Since a straight line graph is expected when probits are plotted against dosage, the methods of linear regression are suggested. The measures that need to be extracted from the linear regression are those of potency and sensitivity of dosage. The potency generally regarded as best for making comparisons among drugs is that giving a 50% kill, and is referred to as the LD 50. (lethal dose) In experiments in which death is not the response, the ED 50 (median effective dose) is used.

The sensitivity is the range of dosage required for a given range of percentage kill. If a small change in concentration gives a wide range in the percentage kill, the sensitivity is high. The slope of the regression line, b, is associated with the sensitivity. The greater the slope, the narrower the range in dosage for a given range in the percentage kill.

THE METHOD OF MAXIMUM LIKELIHOOD

The Basic Assumptions

The model used is

$$y_j = \mu + \beta (x_j - \bar{x}) + \varepsilon_j , \quad j = 1, 2, \ldots, k \qquad (8)$$

= the number of doses.

η_j will be defined as

$$E\left[y_j\right] = \mu + \beta (x_j - \bar{x}). \qquad (9)$$

It is assumed that $E\left[\varepsilon_j\right] = 0$. η_j is the dosage which gives a π_j response or may also be defined as the point on the x axis which gives x_j.

π_j will be defined as

$$E\left[p_j\right] = \frac{1}{n_j} \quad E\left[r_j\right] \tag{10}$$

where p_j is the sample death proportion at dosage x_j, r_j is the number affected and n_j is the number in the trial.

Now let a tolerance distribution be represented by (1), so that the probability of response to a dose λ_o is π_o as defined by (6). If a batch of n subjects is exposed to the stimulus at a dose λ_o, and if the subjects react independently of one another, the probability of r responses is given by the binomial distribution as

$$\prod_{j=1}^{k} (r_j) = \binom{n_j}{r_j} \pi_j^{r_j} \left(1 - \pi_j\right)^{n_j - r_j} \quad . \tag{11}$$

Assume that a series of k doses is tested in an experiment; then the probability of a particular number killed in each group is proportional to e^L where

$$L = \sum_{j=1}^{k} r_j \log \pi_j + \sum_{j=1}^{k} (n_j - r_j) \log (1 - \pi_j) \quad . \tag{12}$$

The quantity e^L or more strictly, a quantity proportional to it but having a maximum value of 1, has been called the likelihood of the observations. Or if the likelihood function (L) is defined as

$$L = \prod_{j=1}^{k} \binom{n_j}{r_j} \pi_j^{r_j} \left(1 - \pi_j\right)^{n_j - r_j} \quad , \tag{13}$$

$$\log L = s = \sum_{j=1}^{k} \log \binom{n_j}{r_j} + \sum_{j=1}^{k} r_j \log \pi_j + \sum_{j=1}^{k} (n_j - r_j) \log (1 - \pi_j). \quad (14)$$

The Estimation of μ and β

Now π_j and $(1 - \pi_j)$ are functions of the dose which contain certain parameters, and the next problem is that of estimating the parameters from the experimental data. The likelihood is a maximum when s is a maximum; therefore if μ and β are parameters of the distribution of individual tolerances, the maximum likelihood estimates of μ and β must satisfy the equations

$$\frac{\partial s}{\partial \mu} = \frac{\partial s}{\partial \beta} = 0. \quad (15)$$

A composite derivative must be taken to estimate μ and β, since s is a function of π_j, π_j is a function of η_j and η_j is determined by these two parameters.

$$\frac{\partial s}{\partial \mu} = \sum_{j=1}^{k} \frac{\partial s}{\partial \pi_j} \cdot \frac{\partial \pi_j}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial \mu} \cdot$$

$$\frac{\partial s}{\partial \pi_j} = \frac{r_j}{\pi_j} - \frac{(n_j - r_j)}{(1-\pi_j)} = \frac{(1-\pi_j)(r_j - \pi_j)(n_j - r_j)}{\pi_j(1 - \pi_j)} = \frac{r_j - \pi_j r_j - \pi_j n_j + \pi_j r_j}{\pi_j(1-\pi_j)}$$

$$= \frac{r_j - \pi_j n_j}{\pi_j(1-\pi_j)} \cdot \quad (16)$$

Since $p_j = r_j/n_j$, $r_j = n_j p_j$; so that

$$\frac{\partial s}{\partial \pi_j} = \frac{n_j p_j - \pi_j n_j}{\pi_j(1-\pi_j)} = \frac{n_j(p_j-\pi_j)}{\pi_j(1-\pi_j)}$$

$$\frac{\partial \pi_j}{\partial \eta_j} = \frac{dF(\eta_j)}{d\eta_j}\Bigg|_{\eta_j} = \frac{1}{\sqrt{2c}\,\sigma} \; e^{-\frac{1}{2}\eta_j^2}$$

$$= Z_j, \text{ an ordinate of a } N(0,1)$$

$$\frac{\partial \eta_j}{\partial \mu} = \frac{\partial\left[\mu + \beta(x_j - \bar{x})\right]}{\partial \mu} = 1$$

Therefore

$$\frac{\partial s}{\partial \mu} = \sum_{j=1}^{k} \frac{n_j(p_j-\pi_j)}{\pi_j(1-\pi_j)} \cdot Z_j \tag{17}$$

To find the estimate of β, a similar derivative will be used.

$$\frac{\partial s}{\partial \beta} = \sum_{j=1}^{k} \frac{\partial s}{\partial \pi_j} \cdot \frac{\partial \pi_j}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial \beta} \cdot \tag{18}$$

Utilizing the results from the previous function, it can be seen that

$$\frac{\partial g}{\partial \pi_j} = \frac{n_j(p_j - \pi_j)}{\pi_j(1-\pi_j)}$$

and

$$\frac{\partial \pi_j}{\partial \eta_j} = Z_j \quad .$$

The solution for $\frac{\partial \eta_j}{\partial \beta}$ is

$$\frac{\partial \eta_j}{\partial \beta} = \frac{\partial \left[\mu + \beta(x_j - \bar{x})\right]}{\partial \beta} = (x_j - \bar{x})$$

Therefore $$\frac{\partial g}{\partial \beta} = \sum_{j=1}^{k} \frac{n_j Z_j (p_j - \pi_j)(x_j - \bar{x})}{\pi_j(1 - \pi_j)} \qquad (19)$$

The Weighting Coefficient

The reliability of the probit for an observed percentage kill depends not only on how many individuals were counted to determine this percentage but also upon the corresponding probit value of the regression line, or, in actual practice, upon that of the provisional regression line. It is customary to consider the reliability of a percentage as proportional to the number of individuals tested. Thus the justification for weighting by the number of individuals, rather than by the square root of the number of individuals, is that the reliability of a measure is inversely proportional

to the square of its standard error---the variance--- and not to the standard error itself. The variance, in turn, is a function not only of the number of cases, but also of several other important factors which will now be considered.

The standard error needed is not that for a proportion π_j but rather that for the corresponding inferred dosage or probit, x_j, which is equivalent to the percentile. The formula for the variance of a percentile is given by Kelley (cited in Bliss, 1935 page 150)

$$\frac{\sigma^2 \pi_j (1 - \pi_j)}{z_j^2 n_j} \tag{20}$$

where σ is the standard deviation, z_j is the ordinate of the normal curve and is given in tables of the probability integral, and the other terms have their previous significance. This will also be the variance for the probit of a single observed percentage mortality, but since the probit is already in terms of the standard deviation, σ^2 is always equal to 1 and the variance of a probit may be simplified to the form

$$\frac{\pi_j (1 - \pi_j)}{z_j^2 n_j} \quad . \tag{21}$$

In order, therefore, to give each observation a weight proportional to its true reliability, instead of multiplying it by n, we multiply by the reciprocal of the variance as the weight, w_j. Hence

$$w_j = n_j \frac{z_j^2}{\pi_j(1 - \pi_j)} \tag{22}$$

where n_j, z_j^2, π_j and $1 - \pi_j$ have their previous significance. The term

$$\frac{z_j^2}{\pi_j(1 - \pi_j)}$$

will be called the weighting coefficient. (Bliss, 1935)

Therefore $$\frac{\partial s}{\partial \mu} = \sum_{j=1}^{k} \frac{n_j w_j (p_j - \pi_j)}{z_j} \tag{23}$$

and $$\frac{\partial s}{\partial \beta} = \sum_{j=1}^{k} \frac{n_j w_j (p_j - \pi_j)(x_j - \bar{x})}{z_j} \tag{24}$$

The Re-definition of the Normal Equations

Equations (23, (24), cannot be solved as they are, so that a new definition will be used in an effort to solve them. Let

$$\frac{\partial s}{\partial \mu} = \phi_1(\mu, \beta)$$

and (25)

$$\frac{\partial s}{\partial \beta} = \phi_2(\mu, \beta)$$

The original condition

$$\frac{\partial s}{\partial \mu} = \frac{\partial s}{\partial \beta} = 0$$

still holds, so that the two new equations must equal zero or

$$\begin{bmatrix} \phi_1(\mu,\beta) \\ \phi_2(\mu,\beta) \end{bmatrix} = 0. \tag{26}$$

From a Taylor's expansion, any function can be written as

$$f(x,y) \triangleq f(x_o,y_o) + \frac{\partial f}{\partial x}\bigg|_{x_o y_o} \Delta x + \frac{\partial f}{\partial y}\bigg|_{x_o y_o} \Delta y + h(\Delta) = 0.$$

Therefore, (if we neglect $h(\Delta)$ which involves higher powers of Δ_x and Δ_y),

$$\begin{aligned} \phi_1(\mu,\beta) &\triangleq \phi_1(\mu_o,\beta_o) + \frac{\partial \phi_1}{\partial \mu_o}\bigg|_{\mu_o,\beta_o} \Delta\mu + \frac{\partial \phi_1}{\partial \beta}\bigg|_{\mu_o,\beta_o} \Delta\mu = 0 \\ \phi_2(\mu,\beta) &\triangleq \phi_2(\mu_o,\beta_o) + \frac{\partial \phi_2}{\partial \mu_o}\bigg|_{\mu_o,\beta_o} \Delta u + \frac{\partial \phi_2}{\partial \beta}\bigg|_{\mu_o,\beta_o} \Delta\beta = 0 \end{aligned} \tag{27}$$

Thus, in matrix notation,

$$\begin{bmatrix} \phi_1(\mu_o,\beta_o) \\ \phi_2(\mu_o,\beta_o) \end{bmatrix} + \begin{bmatrix} \frac{\partial \phi_1}{\partial \mu_o}\Big|_{\mu_o,\beta_o} & \frac{\partial \phi_1}{\partial \beta_o}\Big|_{\mu_o,\beta_o} \\ \frac{\partial \phi_2}{\partial \mu_o}\Big|_{\mu_o,\beta_o} & \frac{\partial \phi_2}{\partial \beta_o}\Big|_{\mu_o,\beta_o} \end{bmatrix} \begin{bmatrix} \Delta\mu \\ \Delta\beta \end{bmatrix} = 0 \tag{28}$$

or,

$$\begin{bmatrix} \Delta\mu \\ \\ \Delta\beta \end{bmatrix} = - \begin{bmatrix} \dfrac{\partial \phi_1}{\partial \mu_o}\Big|_{\mu_o,\beta_o} & \dfrac{\partial \phi_1}{\partial \beta_o}\Big|_{\mu_o,\beta_o} \\ \\ \dfrac{\partial \phi_2}{\partial \mu_o}\Big|_{\mu_o,\beta_o} & \dfrac{\partial \phi_2}{\partial \beta_o}\Big|_{\mu_o,\beta_o} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \phi_1(\mu_o,\beta_o) \\ \\ \phi_2(\mu_o,\beta_o) \end{bmatrix} = 0 \quad (29)$$

For convenience, the right hand side will be designated as

$$= - \left[V \right]^{-1} \phi \quad . \qquad (30)$$

But

$$\begin{bmatrix} \Delta\mu \\ \\ \Delta\beta \end{bmatrix} = \begin{bmatrix} \hat{\mu} - \mu_o \\ \\ \hat{\beta} - \beta_o \end{bmatrix} \qquad (31)$$

A good guess of $\hat{\mu}$ is μ_o and of $\hat{\beta}$ is β_o. So consider

$$\begin{aligned} \Delta\mu &= \mu_1 - \mu_o \\ \Delta\beta &= \beta_1 - \beta_o \quad . \end{aligned} \qquad (32)$$

Then

$$\begin{bmatrix} \mu_1 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \mu_o \\ \beta_o \end{bmatrix} + \begin{bmatrix} \Delta\mu \\ \Delta\beta \end{bmatrix} \quad \text{where} \quad \begin{bmatrix} \Delta\mu \\ \Delta\beta \end{bmatrix} = - \begin{bmatrix} V \end{bmatrix}^{-1} \phi \,. \tag{33}$$

The Evaluation of the V Matrix

The next problem is the evaluation of the V matrix.

Since $\frac{\partial s}{\partial \mu} = \phi_1(\mu,\beta)$

and $\frac{\partial s}{\partial \beta} = \phi_2(\mu,\beta)$,

$$V = \begin{bmatrix} \frac{\partial \phi_1}{\partial \mu} & \frac{\partial \phi_1}{\partial \beta} \\ \frac{\partial \phi_2}{\partial \mu} & \frac{\partial \phi_2}{\partial \beta} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 s}{\partial \mu^2} & \frac{\partial^2 s}{\partial \mu \partial \beta} \\ \frac{\partial^2 s}{\partial \mu \partial \beta} & \frac{\partial^2 s}{\partial \beta^2} \end{bmatrix} \tag{34}$$

Composite derivatives must be used to evaluate each of the elements in this matrix .

$$\frac{\partial s}{\partial \mu^2} = \sum_{j=1}^{k} \left[\frac{\partial \phi_1}{\partial \pi_j} \cdot \frac{\partial \pi_j}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial \mu} + \frac{\partial \phi_1}{\partial Z_j} \cdot \frac{\partial z_j}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial \mu} \right] \tag{35}$$

From the fact that $\frac{\partial s}{\partial \mu} = \phi_1(\mu,\beta)$

$$\phi_1 = \sum_{j=1}^{k} \frac{n_j w_j (p_j - \pi_j)}{Z_j} = \sum_{j=1}^{k} \frac{n_j Z_j (p_j - \pi_j)}{\pi_j (1-\pi_j)}$$

$$\frac{\partial \phi_i}{\partial \pi_j} = \frac{\partial}{\partial \pi_j}\left[\frac{n_i Z_i (p_i - \pi_i)}{\pi_j (1-\pi_j)}\right] = \frac{-\pi_i(1-\pi_i)\, n_i Z_i - n_i Z_i (p_i - \pi_i)(1-2\pi_i)}{\left[\pi_j(1-\pi_j)\right]^2}$$

$$= \frac{-\pi_i(1-\pi_i) n_i Z_i}{\left[\pi_j(1-\pi_j)\right]^2} - \frac{n_i Z_i (p_i - \pi_i)(1-2\pi_i)}{\left[\pi_j(1-\pi_j)\right]^2}$$

The right hand term involves $(p_j - \pi_j)$ and is assumed for our purposes, to be very small. Therefore

$$\frac{\partial \phi_i}{\partial \pi_j} \doteq - \frac{n_i Z_i}{\pi_j (1-\pi_j)}$$

From previous work,

$$\frac{\partial \pi_i}{\partial \eta_j} = Z_j$$

and

$$\frac{\partial \eta_i}{\partial \mu} = 1.$$

Therefore the left hand term of the composite derivative

$$\sum_{j=1}^{k} - \frac{n_i Z_i^2}{\pi_j (1-\pi_j)} \qquad (36)$$

$$\frac{\partial \phi_j}{\partial Z_j} = \frac{\partial \left[\frac{n_j Z_j (p_j - \pi_j)}{\pi_j (1-\pi_j)} \right]}{\partial Z_j} = \frac{n_j (p_j - \pi_j)}{\pi_j (1-\pi_j)}$$

$$\frac{\partial Z_j}{\partial n_j} = \frac{\partial \left[\frac{1}{\sqrt{2c}\,\sigma} \; e^{-\frac{1}{2} n_j^2} \right]}{\partial n_j} = \frac{1}{\sqrt{2c}\,\sigma} \; e^{-\frac{1}{2} n_j^2} \left(-\frac{1}{2} 2 n_j \right) = -Z_j n_j$$

$$\frac{\partial n_j}{\partial \mu} = 1 \quad \text{(from previous work)}$$

Therefore the right hand term becomes

$$- \sum_{j=1}^{k} \left[\frac{n_j (p_j - \pi_j)}{\pi_j (1-\pi_j)} \cdot n_j Z_j \right] \tag{37}$$

The entrie composite derivative is

$$\frac{\partial^2 s}{\partial \mu^2} = \sum_{j=1}^{k} \left[- \frac{n_j Z_j^2}{\pi_j (1-\pi_j)} - \frac{n_j n_j Z_j (p_j - \pi_j)}{\pi_j (1-\pi_j)} \right]$$

The right hand term involves $(p_j - \pi_j)$ and is again assumed to be very small.

Therefore,

$$\frac{\partial \phi_1}{\partial \mu} = - \sum_{j=1}^{k} \frac{Z_j^2 n_j}{\pi_j(1-\pi_j)} = - \sum_{j=1}^{k} n_j w_j \tag{38}$$

This is the (1, 1) element of the V matrix.

The next element to be found is $\frac{\partial \phi_1}{\partial \beta} = \frac{\partial \phi_2}{\partial \mu}$

$$\frac{\partial^2 \ell}{\partial \mu \partial \beta} = \left[\sum_{j=1}^{k} \frac{\partial \phi_1}{\partial \pi_j} \cdot \frac{\partial \pi_j}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial \beta} + \frac{\partial \phi_1}{\partial Z_j} \cdot \frac{\partial z_j}{\partial \eta_j} \cdot \frac{\partial \eta_j}{\partial \beta} \right] \tag{39}$$

Each of the terms of this composite derivative has been found in previous work.

Therefore

$$\frac{\partial \phi_1}{\partial \beta} = \sum_{j=1}^{k} \left[\left(- \frac{n_j Z_j}{\pi_j(1-\pi_j)} \right) \left(Z_j \right) \left(x_j - \bar{x} \right) + \left(\frac{n(p_j - \pi_j)}{\pi_j(1-\pi_j)} \right) \cdot \left(-Z_j n_j \right) \left((x_j - \bar{x}) \right) \right] \tag{40}$$

This function equals zero since $\sum (x_j - \bar{x}) = 0$, thereby reducing to 0 the off diagonal elements (1, 2, and 2, 1) of the V matrix.

The last element of the V matrix is given by

$$\frac{\partial \phi_2}{\partial \beta} = \sum_{j=1}^{k} \left[\frac{\partial \phi_2}{\partial \pi_j} \cdot \frac{\partial \pi_1}{\partial \eta_j} \cdot \frac{\partial \eta_1}{\partial \beta} + \frac{\partial \phi_2}{\partial Z_j} \cdot \frac{\partial z_1}{\partial \eta_j} \cdot \frac{\partial \eta_1}{\partial \beta} \right] \tag{41}$$

$$\frac{\partial \phi_2}{\partial \pi_j} = \frac{\partial \left[\frac{(n_1 Z_1 (x_1 - \bar{x})(p_1 - \pi_1)}{\pi_1 (1-\pi_1)} \right]}{\partial \pi_j} = \frac{-n_1 Z_1 (x_1 - \bar{x})}{\pi_j (1-\pi_j)} = \frac{-n_1 w_1 (x_1 - \bar{x})}{Z_j}$$

From previous work, applying the weights (22),

$$\frac{\partial \pi_1}{\partial \eta_j} = Z_j$$

and

$$\frac{\partial \eta_1}{\partial \beta} = (x_j - \bar{x})$$

$$\frac{\partial \phi_2}{\partial z_j} = \frac{\partial \left[\frac{n_1 w_1 (x_1 - \bar{x})(p_1 - \pi_1)}{Z_1} \right]}{\partial Z_j} = \frac{-n_1 w_1 (x_1 - \bar{x})(p_1 - \pi_1)}{Z_j^2}$$

Again from previous work.

$$\frac{\partial Z_1}{\partial \eta_j} = -\eta_j Z_j$$

$$\frac{\partial \eta_1}{\partial \beta} = (x_j - \bar{x})$$

$$\frac{\partial \phi_2}{\partial \beta} = \sum_{j=1}^{k} \left[\left(- \frac{n_j (x_j - \bar{x}) w_j}{Z_j} \right) \left(Z_j \right) \left(x_j - \bar{x} \right) + \left(\frac{n_j w_j (x_j - \bar{x}) (p_j - \pi_j)}{Z^2_j} \right) \left(- n_j Z_j \right) \left(x_j - \bar{x} \right) \right]$$

The right hand term involves $(p_j - \pi_j)$ and is assumed to be very small. Therefore,

$$\frac{\partial \phi_2}{\partial \beta} \doteq \sum_{j=1}^{k} - n_j w_j (x_j - \bar{x})^2 \tag{42}$$

This is the (2,2) term in the V matrix and completes the computations for the V matrix.

The V matrix now becomes

$$V = \begin{bmatrix} - \sum_{j=1}^{k} n_j w_j & 0 \\ 0 & - \sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2 \end{bmatrix}, \tag{43}$$

and

$$V^{-1} = - \begin{bmatrix} \frac{1}{\sum_{j=1}^{k} n_j w_j} & 0 \\ 0 & \frac{1}{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2} \end{bmatrix}. \tag{44}$$

Finding Values for μ_1 and β_1

Now substituting the proper values into (33) gives

$$\begin{bmatrix} \mu_1 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \mu_o \\ \beta_o \end{bmatrix} + \begin{bmatrix} \frac{1}{\sum_{j=1}^{k} n_j w_j} & 0 \\ 0 & \frac{1}{\sum_{j=1}^{k} n_j w_j (x-\bar{x})^2} \end{bmatrix} \cdot \begin{bmatrix} \sum_{j=1}^{k} n_j w_j \frac{(p_j - \pi_j)}{z_j} \\ \sum_{j=1}^{k} n_j w_j (x_j - \bar{x}) \frac{(p_j - \pi_j)}{z_j} \end{bmatrix} \tag{45}$$

From (45), values of μ_1 and β_1 are given by

$$\mu_1 = \mu_1 + \left(\frac{1}{\sum_{j=1}^{k} n_j w_j} \right) \left(\sum_{j=1}^{k} n_j w_j \frac{(p_j - \pi_j)}{z_j} \right)$$

and, (46)

$$\beta_1 = \beta_o + \left(\frac{1}{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2} \right) \left(\sum_{j=1}^{k} n_j w_j (x_j - \bar{x}) \frac{(p_j - \pi_j)}{z_j} \right) .$$

Estimates for the values of μ_o and β_o can be found from the sight line. These are

$$\hat{\mu}_o = \frac{\sum_{j=1}^{k} n_j w_j \eta_1^{\,o}}{\sum_{j=1}^{k} n_j w_j} \quad , \tag{47}$$

$$\hat{\beta}_o = \frac{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x})\ \eta_j^{\,o}}{\sum_{j=1}^{k} (x_j - \bar{x})^2 \sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2} \quad .$$

Then an estimate for μ_1 would be given by

$$\hat{\mu}_1 = \frac{\sum_{j=1}^{k} n_j w_j \eta_j^{\,o}}{\sum_{j=1}^{k} n_j w_j} + \left(\frac{1}{\sum_{j=1}^{k} n_j w_j}\right)\left(\sum_{j=1}^{k} n_j w_j \frac{(p_j - \pi_j)}{z_j}\right) \tag{48}$$

$$= \frac{\sum_{j=1}^{k} n_j w_j \left[\eta_j^{\,o} + \frac{(p_j - \pi_j)}{z_j}\right]}{\sum_{j=1}^{k} n_j w_j}$$

$$\hat{\mu}_1 = \frac{\sum_{j=1}^{k} n_j w_j\ y^{*o}}{\sum_{j=1}^{k} n_j w_j} \tag{49}$$

where $y_j^{*o} = \eta_j^o + \frac{(p_j - \pi_j)}{Z_j}$ is called the working probit for $\hat{\mu}_j$

Similarly an estimate for β_1 would be given by

$$\hat{\beta}_1 = \frac{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x}) \eta_j^o}{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2} + \left(\frac{1}{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2} \right) \left(\sum_{j=1}^{k} n_j w_j (x_j - \bar{x}) \frac{(p_j - \pi_j)}{Z_j} \right)$$

$$= \frac{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x}) \left[\eta_j^o + \frac{(p_j - \pi_j)}{Z_j} \right]}{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2} \tag{50}$$

$$\hat{\beta}_1 = \frac{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x}) y_j^{*o}}{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2} \tag{51}$$

where $y_j^{*o} = \eta_j^{*o} + \frac{(p_j - \pi_j)}{Z_j}$ is called the working probit for $\hat{\beta}_j$.

Maximum and Minimum Working Probits

The working probits for $\hat{\mu}_j$ and $\hat{\beta}_j$ are given above but $\eta_j^o - \pi_j/Z_j$ is tabulated as the minimum working probit and $\frac{1}{Z_j}$ as the "range". Therefore, the working probit is equal to the minimum working probit + (p_j) (range).

$$y^{*o}_j = w \cdot p = \eta^o_j - \frac{\pi_j}{Z_j} + \frac{p_j}{Z_j}$$

$$= \eta^o_j + \frac{p_j - \pi_j}{Z_j} \qquad (52)$$

Or the working probit is equal to the maximum working probits minus $(1 - p_j)$ (range), where the maximum working probit is defined as

$$\eta^o_j + \frac{1+\pi_j}{Z_j} \quad . \qquad (53)$$

$$y^{*o}_j = \text{working probit} = \eta^o_j + \frac{(1-\pi_j)}{Z_j} - \frac{(1-p_j)}{Z_j}$$

$$= \eta^o_j + \frac{1 - \pi_j - 1 + p_j}{Z_j} \qquad (52)$$

$$= \eta^o_j + \frac{p_j - \pi_j}{Z_j} \quad .$$

The preceding mathematical treatment of probit analysis would be used by starting with some initial values for μ_o and β_o and substituting them into equation (45) to obtain values for μ_1 and β_1. Then μ_1 and β_1 would be re-entered into (45) to obtain values for μ_2 and β_2. This re-iteration would continue until values for μ_n and β_n are within pre-set tolerance limits of μ_{n-1} and β_{n-1}. Each of the stationery values,

$$\sum_{j=1}^{k} nw_j \ , \ \sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2 \ , \ \sum_{j=1}^{k} n_j w_j \frac{(p_j - \pi_j)}{Z_j} \quad \text{and}$$

$\sum_{j=1}^{k} nw_j \frac{(x_j-\bar{x})(p_j-\pi_j)}{z_j}$ could be computed and would remain as constants in the calculations. Usually only one or two iterations are needed to bring the values very close together.

Heterogeneity

If the reactions of the individuals in a batch are not independent of one another, the weights nw, though still proportional to the true weights, will be too large, and the estimated variances will therefore be too small. This will be indicated by a large value of a statistic, χ^2, which will be seen to be a weighted sum of squares of the discrepancies between the expected and observed number killed. Since the expected value of χ^2 is its number of degrees of freedom, a significantly large factor indicates that all weights have been over estimated by a factor $\chi^2/k-2$, where k is the number of dosages tested. All variance should therefore be multiplied by this heterogeneity factor as compensation for the overweighting.

The χ^2 test for the heterogeneity of the discrepancies between observed and expected numbers is only valid when the expected numbers are not small, usually less than five. At the more extreme dosages tested either π or $1 - \pi$ is often nearly zero, so that, with the usual number of insects exposed to the stimulus, either the expected number killed, $n\pi$, or the expected number surviving, $n(1 - \pi)$, is too small for a χ^2 calculated in the usual manner.

Using the results from the maximum-likelihood method, the following equation (Finney, 1952b) may be used to compute χ^2:

$$\chi^2_{(k-2)} = \sum n_j w_j (y_j - \bar{y})^2 - \frac{\left\{ \sum n_j w_j (x_j - \bar{x})(y_j - \bar{y}) \right\}^2}{\sum n_j w_j (x_j - \bar{x})^2} \qquad (54)$$

The Variance of the Estimates

$$V(b) = \frac{1}{\sum_{j=1}^{k} n w_j (x_j - \bar{x})^2} \quad \text{(Finney, 1952a page 54)} \qquad (55)$$

$$V(a) = \frac{\sum_{j=1}^{k} n_j w_j x^2}{\left[\sum_{j=1}^{k} n_j w_j \right] \left[\sum_{j=1}^{k} n_j w_j (x - \bar{x})^2 \right]} \quad \text{(Irwin and Cheeseman, 1939 page 157)} \qquad (56)$$

$$V(\bar{y}) = \frac{1}{\sum_{j=1}^{k} n_j w_j} \quad \text{(Finney, 1952a page 54)} \qquad (57)$$

$$V(\hat{Y}) = \frac{1}{\sum_{j=1}^{k} n_j w_j} + \frac{(x_j - \bar{x})^2}{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2} \quad \text{(Finney, 1952a page 61)} \qquad (58)$$

$$V(m) = \frac{1}{b^2} \left[\frac{1}{\sum_{j=1}^{k} n_j w_j} + \frac{(m - \bar{x})^2}{\sum_{j=1}^{k} n_j w_j (x_j - \bar{x})^2} \right] \quad \text{(Finney, 1952a page 32)} \qquad (59)$$

Fiducial Limits

If no allowance has been made for heterogeneity, the variance of Y is given as in (58). But if the heterogeneity factor is significantly greater than 1, V(Y) must be multiplied by this factor. Therefore, fiducial limits to Y are

$$Y \pm s_Y \quad t \tag{60}$$

where s_Y is the square root of V(Y) and t is the normal deviate for the level of probability to be used. If there is significant heterogeneity, the t - value corresponding to this probability should be used.

If the fiducial limits of Y are plotted for each x, they will be found to lie on two curves which are convex to the regression line and which approach this line most closely at the dosage $\bar{x}$. The further x is removed from $\bar{x}$ in either direction the greater is the contribution to the variance of Y from the second term of (58), which represents the effect of the errors of estimation of the regression coefficient b, and consequently, Fiducial Limits premorewidely spread. These limits are the values of x for which the boundaries of the fiducial band attain the selected value of Y. (Finney 1952a)

Exact fiducial limits to x, the dosage giving a kill whose probit is Y, is found by solving an equation so as to obtain the value of x for which Y has a selected fiducial limit. These limits are

$$x + \frac{g}{1-g}(x - \bar{x}) \pm \frac{t}{b(1-g)} \sqrt{\frac{1-g}{\sum_{j=1}^{k} nw_j} + \frac{(x-\bar{x})^2}{\sum_{j=1}^{k} xx}}, \tag{60}$$

where $g = \frac{t^2}{b^2 \sum_{j=1}^{k} xx}$. Significant heterogeneity must be allowed for by increasing both g and the expression within the square root by the heterogeneity factor. (Finney, 1952a)

If g is small (less than about 0.1), the exact equations for confidence limits simplify to those obtained from the approximate variance. Specifically, g will be small under all conditions that reduce slope variance. For example, this will obtain if the (a) error variance is small; (b) slope itself is too steep; (c) dose range is large; (d) desired level of confidence is not too rigorous; and/or (e) number of observations is large so that t will be small. Inasmuch as the terms comprising g have to be found anyway for use in other parts of any data analysis, the actual computation of g will entail little additional work and should be performed routinely. If it is found to be smaller than 0.1, it can be dropped; otherwise, it is retained and the full equations for exact confidence limits must be used. If $g \geq 1$, then the slope will not differ significantly from zero and no confidence interval can be found. (Goldstein, 1964)

ACKNOWLEDGEMENT

The writer wishes to express his sincere appreciation to his major professor, Dr. Gary F. Krause, for suggesting this topic and for his advice and assistance during the preparation of this report.

REFERENCES

Biometrika tables for statisticians, Volume I. Cambridge: University Press, 1958.

Black, A. N.
Weighted probits and their uses. Biometrika 37: 158-167, 1950.

Bliss, C. I.
The method of probits. Science 79: 38-39, 1934.

_____.
The calculation of the dosage-mortality curve. Annals of Applied Biology. 22: 134-167, 1935.

Finney, D. J.
The estimation of the parameters of tolerance distributions. Biometrika 36: 239-256, 1949.

_____.
Probit analysis. Cambridge: University Press, 1952a.

_____.
Statistical method in biological assay. New York: Hafner, 1952b.

_____ and Stevens, W. L.
A table for the calculation of working probits and weights in probit analysis. Biometrika 35: 191-201, 1948.

Gaddum, J. H.
Reports on biological standards. III. Methods of biological assay depending on a quantal response. Special Report Series of the Medical Research Council, No. 183, 1933.

Garwood, F.
The application of maximum likelihood to dosage-mortality curves. Biometrika 32: 46-58, 1941.

Goldstein, Avram
Biostatistics: an introductory text. New York: Macmillan, 1964.

Goulden, C. H.
Methods of statistical analysis. New York: John Wiley and sons, 1952.

Irvin, J. O.
Statistical method applied to biological assays. Supplement to the Journal of the Royal Statistical Society. 4: 2-30, 1937.

_____and Cheeseman, E. A.
On the maximum-likelihood method of determining dosage-response curves and approximations to the median-effective dose, in cases of a quantal response. Supplement to the Journal of the Royal Statistical Society 6: 174-185, 1939.

APPENDIX

A Monte Carlo technique was used to generate the data to show the procedures involved in a probit analysis from a toxicological experiment.

It was theorized that a certain relationship existed. This was that

$$Y = 3 + 2 \log x,$$

where Y represented the probit of π_j, 3 was the intercept and 2 was the slope of the regression line. Log x was found by placing Y equal to the probit at 10%, 20%, . . . , 90%, and solving. p_j, the number affected, is found by using a table of random numbers. As an example, for the 10% case, 100 numbers were scanned and all numbers less than 10 were noted. These constituted the number killed by the lowest dosage. When this number is divided by 100, the result will be p_1. Then using a different set of 100 random numbers each time, p_2, . . . , p_9, may be found.

These values of p_j are then transformed into initial probits by using Table VI in the Biometrika Tables. This will be the column of the Y's.

The values of log x (X) and initial probits (Y) are plotted against each other to give a scatter diagram of the points. Log x is on the abscissa and the initial probits are plotted on the ordinate. A provisional regression line is then drawn by sight. The line should be close to all of the points but where they are widely divergent, get the best fit on the points between four and six probits.

The points on the provisional regression line intersecting with the log x values are the values for the conditional probit, η^o_j. These conditional probits are then used to find the minimum working probit, the range, and the weighting coefficient, from Table VI in the Biometrika Tables.

The sum of the minimum working probit and the product of p_j and the range gives the values of y_j^*. Then values for nw_j, nw_jx_j, $(nw_jx_j)x_j$, $nw_jx_jy_j^*$, $nw_jy_j^*$, and $(nw_jy_j^*)y_j^*$ are found for each log x.

The first iteration values for a_1 and b_1, which are the estimators for μ and β, are given by

$$a = \bar{y} - b\bar{x}$$

and

$$b = \frac{\sum xy}{\sum xx} \quad ;$$

where

$$\bar{y} = \frac{\sum_{j=1}^{k} nw_j y_j^*}{\sum_{j=1}^{k} nw_j}$$

$$\bar{x} = \frac{\sum_{j=1}^{k} nw_j x_j}{\sum_{j=1}^{k} nw_j} \quad ,$$

$$\sum xy = \sum_{j=1}^{k} nw_j x_j y_j^* - \frac{\left(\sum_{j=1}^{k} nw_j x_j \quad \sum_{j=1}^{k} nw_j y_j^*\right)}{\sum_{j=1}^{k} nw_j} \quad ,$$

and

$$\sum xx = \sum_{j=1}^{k} (nw_j x_j) x_j - \frac{\left(\sum_{j=1}^{k} nw_j x_j\right)^2}{\sum_{j=1}^{k} nw_j}$$

Then the values for a_1 and b_1 are substituted into $\eta^1_j = a+bx$ (or $a + b \log x$) to give the formula for the next iteration.

To start the next iteration, η^1_j is evaluated and produces that column in the table; log x and p_j stay the same. The same procedure as before is followed to give values for a_2 and b_2 in $\eta^2_j = a + bx$. η^2_j is evaluated and if the resultant values are within a pre-set tolerance limit, halt iterations and determine values for m and b. If the tolerance limit is not met, another iteration must be performed.

The values for the probits in the Biometrika Tables are given to two places so that the pre-set tolerance limit is met after the second iteration, where there is no differance between η^1_j and η^2_j.

For this example, m, the log of the LD 50 is 1 and the slope of the line, b, is 2.10742.

Table 1. Worksheet For Computation of Probit Analysis

log x (X)	probit (Y)	η^{o}_{j}	M.W.P.	Range	P_j	P_j • range
0.36000	3.72	3.66	3.1056	6.1518	0.10	0.61518
0.58000	4.08	4.13	3.4268	3.6597	0.18	0.65875
0.73750	4.48	4.48	3.6148	2.8695	0.30	0.86085
0.87375	4.64	4.72	3.7040	2.6068	0.36	0.93845
1.00000	5.00	5.00	3.7467	2.5066	0.50	1.25330
1.12625	5.25	5.29	3.6846	2.6143	0.60	1.56858
1.26250	5.58	5.55	3.4831	2.9159	0.72	2.09945
1.42000	5.84	5.89	2.8608	3.7247	0.80	2.97976
1.64000	6.41	6.35	0.6668	6.2351	0.92	5.73629

Table 1. (continued)

y^*_j	w_j	nw_j	nw_jx_j	$(nw_jx_j)x_j$	$(nw_jx_j)y^*_j$	$nw_jy^*_j$	$(nw_jy^*_j)y^*_j$
3.72078	0.3222	32.22	11.59920	4.17570	43.15807	119.88353	446.06024
4.08555	0.4810	48.10	27.89800	16.18084	113.97867	196.51496	802.87169
4.47565	0.5766	57.66	42.52425	31.36163	190.32366	258.06598	1155.01300
4.64245	0.6187	61.87	54.05891	47.23397	250.96579	287.22838	1333.44339
5.00000	0.6366	63.66	63.66000	63.66000	318.30000	318.30000	1591.50000
5.25318	0.6174	61.74	69.53468	78.31343	365.27819	324.33133	1703.77086
5.58245	0.5699	56.99	71.94988	90.83672	401.65661	318.14383	1776.02202
5.84056	0.4746	47.46	67.39320	95.69834	393.61403	277.19298	1618.96223
6.40309	0.3188	31.88	52.28320	85.74445	334.77404	204.13051	1307.06603
		461.58	460.90132	513.20508	2412.04906	2303.79150	11734.70946

$\sum_{j=1}^{k} xy = 111.64491$ $\quad \bar{x} = 0.99852$ $\quad \bar{y} = 4.99109$

$\sum_{j=1}^{k} xx = 52.98144$

$b = 2.10724$ $\quad a = 2.88697$

$\eta^1_j = 2.88697 + 2.10724 \cdot x$

Table 1. (continued)

log x (X)	probit (Y)	n'_j	M.W.P.	Range	p_j	p_j · range
0.36000		3.65	3.0981	6.2351	0.10	0.62351
0.58000		4.11	3.4145	3.7247	0.18	0.67045
0.73750		4.44	3.5963	2.9322	0.30	0.87966
0.87375		4.73	3.7068	2.5997	0.36	0.93589
1.00000		4.99	3.7466	2.5068	0.50	1.25340
1.12625		5.26	3.6977	2.5928	0.60	1.55568
1.26250		5.55	3.4831	2.9159	0.72	2.09945
1.42000		5.88	2.8875	3.6919	0.80	2.95352
1.64000		6.34	0.7426	6.1518	0.92	5.65966

Table 1. (continued)

y^*_j	w_j	nw_j	nw_jx_j	$(nw_jx_j)x_j$	$(nw_jx_j)y^*_j$	$nw_jy^*_j$	$(nw_jy^*_j)y^*_j$
3.72161	0.3188	31.88	11.47680	4.13165	42.71217	118.64493	441.55016
4.08495	0.4746	47.46	27.5268	15.96554	112.44560	193.87173	791.95632
4.47596	0.5675	56.75	41.85313	30.86668	187.33294	254.01073	1136.94187
4.64269	0.6199	61.99	54.16376	47.32559	251.46555	287.80035	1336.16781
5.00000	0.6368	63.68	63.68000	63.68000	318.40000	318.40000	1592.00000
5.25338	0.6211	62.11	69.95139	78.78275	367.48123	326.28743	1714.11186
5.58245	0.5699	56.99	71.94988	90.83672	401.65661	318.14383	1776.02202
5.84102	0.4778	47.78	67.84760	96.34359	396.29919	279.08394	1630.13488
6.40226	0.3222	32.22	52.84080	86.65891	338.30054	206.28082	1320.66344
		460.86	461.29016	514.59143	2416.09383	2302.52376	11739.54836

$\sum_{j=1}^{k} xy = 111.42092$ $\sum_{j=1}^{k} xx = 52.87070$ $b = 2.10742$

$\bar{x} = 1.00093$ $\bar{y} = 4.99614$ $a = 2.88676$

$n_j^2 = 2.88676 + 2.10742\ X$

Table 1. (continued)

log x (X)		η_j^2	
0.36000		3.65	
0.58000		4.11	
0.73750		4.44	
0.87375		4.73	
1.00000		4.99	
1.12625		5.26	
1.26250		5.55	
1.42000		5.88	
1.64000		6.34	

SOME MATHEMATICAL ASPECTS OF PROBIT ANALYSIS

by

PHIL DEAN GILLILAND

B. S., Kansas State University, 1959

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1964

The area in which probit analysis is generally utilized is that of biological assay. The type of assay found most valuable in many different fields, but especially in toxicological studies, is dependent upon a quantal response; that is, one which can be expressed as "occurring" or "not occurring". A feature possessed by all biological assays is the variability in the reactions of the test subjects and the consequent impossibility of reproducing the same results in successive trials.

The purpose of this report is to show, in greater detail, the procedures involved in the maximum likelihood method of estimating the normal equations in a probit analysis.

In a general toxicology experiment, various concentrations of the chemical are prepared and applied to a batch of insects which have been assigned at random to each concentration level. The total number (n) and the number killed (r) are counted and the ratio, $p = r/n$, gives the sample death proportion.

The tolerance for any one subject, the level below which a response does not occur and above which it does occur, is represented by λ. The main concern, for a population of subjects, is the distribution of λ, which may be expressed by $d\pi = f(\lambda)\, d\lambda$.

If a dose λ_o is given to the whole population, all individuals will respond whose tolerances are less than λ_o, and the proportion of these is π, where $\pi = \int_o^{\lambda_o} f(\lambda)\, d\lambda$.

When the distribution of tolerance concentrations is measured on a natural scale, the curve may be far from symmetrical, due to a few individuals with high tolerances. In such a case, normalization may be acheived by expressing the tolerances in terms of the logarithms of the concentrations. Letting

x represent the intensity of the stimulus and λ the concentration of the stimulus, then $x = \log_{10} \lambda$; x and λ will be referred to as the dosage and dose, respectively.

The relationship between the sample death proportion and the dose yields a monotonic function, and when the stimulus is measured in dosage units, the curve has the normal sigmoid form.

If the dosage deviations from the mean, in the normal curve, are replaced by the normal equivalent deviate + 5, the probit transformation is realized. The probit of the proportion π is defined as the abscissa which corresponds to a probability π in a $N(5,1)$.

The method of maximum likelihood was used to solve for the estimators for μ and β in the regression equation, $y_j = \mu + \beta (x_j - \bar{x}) + \varepsilon_j$. If a batch of n subjects is exposed to the stimulus at a dose λ_o, and if the subjects react independently of one another, the probability of r responses is given by the binomial distribution. The maximum likelihood estimates of μ and β, parameters of the distribution of individual tolerances, were found by a numerical solution of the normal equations.

The variance of the estimates and the fiducial limits are given for the maximum likelihood method.

A numerical example, solved by a Monte Carlo technique, is worked to show the procedures used in analyzing relevant data.