https://ntrs.nasa.gov/search.jsp?R=19670008436 2018-07-24T21:41:01+00:00Z

# ESTIMATION FOR A SIMPLE EXPONENTIAL MODEL

Richard G. Cornell

Department of Statistics

Florida State University

Janace A. Speckman

Statistical Engineering Laboratory

National Bureau of Standards

GPO PRICE $ _____

CFSTI PRICE(S) $ _____

Hard copy (HC) 3.00

Microfiche (MF) .65

ff 653 July 65

February 24, 1966

Technical Report Number 3

NASA Grant No. NGR-10-004-029

FACILITY FORM 602

N67-17765 (ACCESSION NUMBER)

(THRU)

28 (PAGES)

1 (CODE)

CR 70495 (NASA CR OR TMX OR AD NUMBER)

04 (CATEGORY)

[redacted]

# ESTIMATION FOR A SIMPLE EXPONENTIAL MODEL

Richard G. Cornell, Ph.D.

and

Janace A. Speckman, M.S.

There are many situations of interest in public health where an estimate is required for the parameter $\rho$ in the model

$$Y = 1 - e^{-\rho t} . \qquad (1)$$

In this situation Y represents the expected value of a proportion y calculated from a count of the number of "reactions"among n independent observations and t is a variable such as time or dosage.

This model has been applied to histoplasm sensitivity conversion rates by Manos[17] where in his paper Y is the theoretical proportion of positive reactors to histoplasm at time t in years after the initial conversions take place. It is assumed that ny,the number of positive reactors out of n people tested at time t, has a binomial distribution with mean nY. Then $\rho$ is interpreted as the instantaneous annual conversion rate. Manos developed a weighted least squares method of estimating $\rho$ by calculating a straight line relationship between $(-\log_e y)$ and t. This estimation procedure was also studied earlier by Fisher[12] and has been discussed by several other authors.

Peto[19] has discussed another situation where the model is applicable. Here Y is the proportional of test animals which would not be expected to survive after administration of a dose of t micro-organisms. The actual number who fail to survive, ny, is assumed to have a binomial distribution. Peto developed an iterative maximum likelihood estimation procedure for this

[redacted]

model and presented tables to facilitate calculations. Finney[8,9] considered the same bioassay model as Peto and developed slightly different iterative calculation schemes for maximum likelihood estimation. Johnson and Brown[15] and Fisher[11] considered a similar situation except that they specified the dosages to be those which would be used in a serial dilution biological assay. Both the Johnson and Brown and the Fisher papers present estimation procedures for such assays.

Another example is an experiment in genetics performed by Edington, Epler and Regan[7] to study the frequency-dose relation of X-ray induced Y-suppressed lethals in drosophila. The model was fitted to data on orthodox lethals where y is the proportion of orthodox lethals found at a dose of t roentgens of X-rays. For a given dosage it is again reasonable to assume that ny has a binomial distribution, where n is the number of chromosomes tested. In radiation genetics and virology the model (1) is called the one-particle or one-hit curve. The derivation of the model in these contexts was discussed more fully by Cochran.[4]

These examples illustrate a variety of situations in which the simple exponential model is encountered. The importance of this model in epidemiology was discussed more fully by Muench[18], who also gave a nomogram which can be used in estimating $\rho$ . Muench called the model (1) the simple catalytic curve. He points out that this model will arise whenever the change in Y for a small change in t is proportional to Y. The proportionality constant or rate of change is $\rho$ .

The references discussed above also indicate that a variety of methods

are being used to estimate the parameter $\rho$ in (1). In this paper several of these methods will be presented, illustrated and compared in order to guide the public health or medical investigator who encounters this model in his research. In the examples cited it is reasonable to assume that ny, the observed count, has a binomial distribution for each value of t. All of the methods discussed in this paper may be applied under this assumption although some of them do not require this assumption and one involves a different assumption.

METHODS OF ESTIMATION

Graphical. Values of $u = [-\log_e(1-y)]$ are plotted against the corresponding t values in the graphical method of estimating $\rho$. A straight line is drawn through the origin by eye to describe the graphical relationship. The slope of this line is an estimate $r_g$ of $\rho$.

Maximum Likelihood. As mentioned earlier, Peto[19] has derived the maximum likelihood estimator of $\rho$ for the situation where ny has a binomial distribution for each value of t. Consider k samples of size $n_i$ taken at various values of $t_i$, $i = 1,2,\ldots k$. The binomial probability at $t_i$ is $Y = 1-\exp(-\rho t_i)$. Let $y_i$ be the proportion of the $n_i$ at time $t_i$ exhibiting the characteristic of interest. Using this notation, the equation for the maximum likelihood estimator $r_{m\ell}$ of $\rho$ becomes

$$f(r_{m\ell}) = -\sum_i n_i t_i + \frac{1}{r_{m\ell}} \sum_i \frac{n_i y_i x_i}{1-e^{-x_i}} = 0 \qquad (2)$$

where $x_i = t_i r_{m\ell}$. Tables A and B of Peto's article aid in an iterative solution of equation (2) utilizing Newton's method, where in his notation $n_i(1-y_i) = r_i$.

Finney[9,10] also presented tables to aid in calculating $r_{m\ell}$ using a different iterative calculation scheme which is patterned after that usually used in probit analysis for a similar but different bioassay model. In his book Finney[10] briefly reviewed several papers on maximum likelihood estimation for this "dilution series" model.

Least Squares. The least squares estimation procedure gives a method for calculating the slope of a straight line through the origin relating $u = [-\log_e(1-y)]$ and t in such a way that the sum of squares of deviations of the u's about the line is minimized. The slope of this line is the least squares estimate $r_{\ell s}$ of $\rho$. It involves the assumption of homogeneity of variance of the $[-\log_e(1-y)]$ values about the fitted line. This procedure is also a maximum likelihood procedure when the $\log_e(1-y)$ values have a normal distribution with homogeneous variance. The formula for the estimator is

$$r_{\ell s} = \frac{-\sum_i t_i \log_e(1-y_i)}{\sum_i t_i^2} \quad . \tag{3}$$

Weighted Least Squares. If the assumption of homogeneity of variance of the u values which is inherent is using least squares estimation is not tenable, then a possible alternative is the use of weighted least squares calculations as suggested by Manos[17]. If we define $r_{w\ell s}$ to be the weighted least squares estimator of $\rho$, then

$$r_{w\ell s} = \frac{-\sum_i w_i t_i \log_e(1-y_i)}{\sum_i w_i t_i^2} \quad . \tag{4}$$

Each weight $w_i$ is chosen to be inversely proportional to the estimated variance of the corresponding $u_i$ variable. This estimated variance is obtained through approximating $u_i$ by the first two terms of a Taylor series expansion and it involves the assumption that each ny variable for a particular value of t has a binomial distribution. It also involves approximating E(y) by y for each value of t used. The weights are given by $w_i = n_i(1-y_i)/y_i$. Substitution of these weights in (4) yields

$$r_{w\ell s} = \frac{-\sum_i [n_i(1-y_i)t_i \log_e(1-y_i)]/y_i}{\sum_i n_i(1-y_i)t_i^2/y_i} . \qquad (5)$$

If a given $y_i = 0$ or 1 it is common practice (for example, see Berkson[1]) to replace it by 1/2n or (1-1/2n), respectively, in order not to become involved with an indeterminate result in (5). Another procedure has been proposed by Fisher[12] for avoiding this difficulty by using the data only to obtain an initial estimate in an iterative procedure. This is discussed by Cornfield[6] who also suggested the possibility of applying weighted least squares calculations to the (1-y) proportions instead of to their logarithms. In some instances it is possible to replace indeterminate terms in (4) by the appropriate limit as $y_i$ approaches either 0 or 1.

<u>Partial Totals</u>. The method of partial totals, which has been discussed by Speckman and Cornell[20], consists of equating $\Sigma(1-y_i)$ to its expected value $\sum_i(1-Y_i) = \sum_i e^{-\rho t_i}$. If the t values are equally spaced, that is, $t_{i+1}-t_i = d$,

then, letting $v = \exp(-r_{pt}d)$, the estimation equation is

$$f(v) = v^{k+t_1/d} - v^{t_1/d} + (1-v) \sum_{i=1}^{k} (1-y_i) = 0. \tag{6}$$

If $t_1 = 0$, this simplifies to

$$f(v) = v^k + (1-v) \sum_{i=1}^{k} (1-y_i) - 1 = 0. \tag{7}$$

Either equation (6) or (7) may be solved for v by Newton's iterative procedure and then $r_{pt} = (-\log_e v)/d$. Alternatively, when $t_1 = 0$, tables given by Speckman and Cornell[20] may be entered for values of $\bar{S} = \sum_{i=1}^{k} (1-y_i)/k$ and k to obtain the value of $r_{pt}d$ such that $r_{pt}$ satisfies (7). If $t_1 = d$ estimates $r_{pt}d$ may also be obtained by entering the tables with $\bar{S} + 1$ and $k + 1$. Tables of estimates are presented for k = 5,6,8,10,15,20 and 25. Partial totals estimates for other values of k can sometimes be obtained by interpolation.

<u>Moments</u>. Muench[18] has presented the method of moments for computing an estimate $r_m$ of the parameter $\rho$ in (1). He computed the area under a histogram drawn from the data as $\Sigma A = \sum_{i=1}^{k} y_k d_i$ where each $d_i$ is the width of the interval on the t scale for which the corresponding $y_i$ proportion is calculated. Muench assumed that the range of t values, for which the histogram with area $\Sigma A$ is drawn, if from 0 to $\tau$. He approximated $\Sigma A$ by the integral

$$\int_0^\tau Y\, dt = \int_0^\tau (1-e^{-\rho t})\, dt = \tau + (e^{-\rho t} - 1)/\rho \ .$$

He equated $\Sigma A$ to this integral in order to solve for his estimate $r_m$ of $\rho$. The estimation equation, after dividing both $\Sigma A$ and the integral by $\tau$, is

$$\Sigma A/\tau = 1 + (e^{-r_m \tau} - 1)/r_m \tau \ . \tag{8}$$

Muench's book[18] contains a nomogram which gives values of

$$r_m' = r_m \ \tau/100 \tag{9}$$

for various values of

$$\Sigma' A = (100) \ \Sigma A/\tau \ . \tag{10}$$

<u>Finite Differences</u>. The model (1) can be generated by the difference equation

$$Y_{i+1} - (1-e^{-\rho d}) \ -e^{-\rho d} \ Y_i = 0 \tag{11}$$

where it is assumed as for the partial totals method that the $t_i$ values are equally spaced at intervals of width d. Lipton and McGilchrist[16] have given two methods of estimating $\rho$ in (11). One is to choose $\exp(-\rho d)$ to minimize $\sum_{i=1}^{k-1} D_i^2$ where $D_i$ is obtained from the left side of (11) by replacing the Y expectations by the corresponding y proportions calculated from the data. This leads to an estimator $r_{fd1}$ of $\rho$ which can be calculated from the formula

$$r_{fd1} = -\frac{1}{d} \log_e \left[ \frac{\sum_{i=1}^{k-1} (1-y_i)(1-y_{i+1})}{\sum_{i=1}^{k-1} (1-y_i)^2} \right] \ . \tag{12}$$

The other finite differences method consists of minimizing $\sum_{j=1}^{k-1} (\sum_{i=1}^{j} D_i)^2$ which yields the estimate $r_{fd2}$ of $\rho$ given by

$$r_{fd2} = -\frac{1}{d} \log_e \left[ 1 - \frac{\sum_{j=2}^{k} S_j(y_j - y_1)}{\sum_{j=2}^{k} S_j^2} \right] \tag{13}$$

where

$$S_j = \sum_{i=1}^{j-1} (1-y_i) \ . \tag{14}$$

The estimator given by equation (13) is similar to that given by Hartley[14] under the heading of internal least squares.

<u>Fisher</u>. Fisher[11] presented another procedure for the model $E(y_i) = 1-\exp(-\rho a^{-(i-1)})$, $i = 1,2,\ldots,k$, for constant a. Letting $a = e^c$, we can regard his procedure as an estimation procedure for (1) with the t values spaced exponentially, that is, with $t = e^x$ for equally spaced x's, $-\infty < x < \infty$ . It is assumed that E(y) nearly ranges from 0 to 1. Fisher set

$$\sum_{i=1}^{k} y_i = \sum_{i=1}^{k} [1-\exp(-\rho a^{-(i-1)})] \tag{15}$$

and solved for an estimate $r_f$ of $\rho$ by a numerical procedure. Fisher and Yates'[13] book gives tables for a = 2, 4, and 10 which list values of a quantity K for various pairs of values of $\sum_{i=1}^{k} y_i$ and k. In their notation $\sum_{i=1}^{k} y_i = x$ and $k = s$.

Then the equation

$$\log_{10}(r_f) = (\sum_{i=1}^{k} y_i) \log_{10} a - K \tag{16}$$

may be solved to obtain $r_f$ .

Fisher and Yates' tables may be used even if the t values used are not equal to $a^{-(i-1)}$ for $i = 1,2,\ldots,k$ and $a = 2$, 4, or 10 provided that $t_i/t_{i+1} = a$. The procedure is exactly the same as that given above except that equation (16) yields $\log_{10}(r_f t_1)$ .

Spearman. Like the Fisher procedure, the Spearman method presented by Johnson and Brown[15] may be used for the model (1) when the t values used are such that $t = e^x$ for equally spaced values of x with intervals of width c. Cornell[5] has shown that the Spearman estimator may be derived by replacing the right side of equation (15) by a integral approximation under the assumption that k is large and c is correspondingly small. The resultant Spearman estimator $r_s$ of $\rho$ is given by

$$r_s = \exp(c \sum_{i=1}^{k} y_i - x_k - c/2 - \gamma) \tag{17}$$

where the range from $x_1$ to $x_k$ is assumed to be great enough so that $y_1 \doteq 0$ and $y_k \doteq 1$. In this equation, $\gamma = 0.57722$ is Euler's constant.

ILLUSTRATIONS

Two sets of data are used to illustrate the various methods discussed above. The first set of data has equally spaced t's, the second has

exponentially spaced t's. Since the graphical, maximum likelihood, least squares, weighted least squares and moment estimation procedures do not require any specific spacing of the t's, they will be illustrated in both examples. Partial totals and finite differences, which require equal spacing, will be illustrated only in Example 1. The Fisher and Spearman estimators, requiring exponential spacing, will be illustrated only in Example 2.

Example 1. Equally spaced t's.

The data for this example come from a genetics experiment discussed by Edington, Epler and Regan[7]. The purpose of the experiment was to establish a dose response relation of X-ray induced Y-suppressed lethals in drosophila. Five doses of X-rays, shown in column (1) of Table 1, were administered to n chromosomes as shown in column (2) and the numbers of orthodox and Y-suppressed lethals were counted. The number of orthodox lethals observed are shown in column (3) with the corresponding proportions, y, given in column (4). With model (1) and the assumption that ny is binomially distributed, we would expect that $y = 0$ when $t = 0$. Since this is not the case, due to the natural occurance of lethals of about 0.02 percent, the data are corrected for this discrepancy. Thus the y values in column (4) are computed by taking the ratios of the corresponding entries in column (3) to those in column (2) and subtracting 0.0002.

The eight estimates calculated for this example are listed in Table 4.

<u>Graphical Method</u>. Columns (5) and (1) of Table 2 enable $u = -\log_e(1-y)$ to be plotted against t as in Figure 1. The slope of this line, which was fitted by eye to relate u and t, is $r_g = 0.0000305$.

Maximum Likelihood. Peto's iterative calculations involve the computing of nyA and nyB for each t value during each iteration. He gave tables of A and B corresponding to various trial values of $x_i = t_i r_{m\ell}$. By following his instructions and using columns (3) and (4) of Table 2 as well as $r_g$ as an initial estimate, one cycle of calculations yields $r_{m\ell} = 0.00003048$.

Table 1 - Data for Example 1

| Dose t (1) | Chromosomes Tested, n (2) | Orthodox Lethals (3) | y (4) |
|---|---|---|---|
| 0 | 4358 | 1 | 0.0000 |
| 1082 | 3852 | 111 | 0.0286 |
| 2164 | 3605 | 232 | 0.0642 |
| 3246 | 2813 | 259 | 0.0919 |
| 4328 | 2206 | 292 | 0.1322 |

Figure 1 - Graph of $u = -\log_e(1-y)$ against t with the line fitted by eye for Example 1

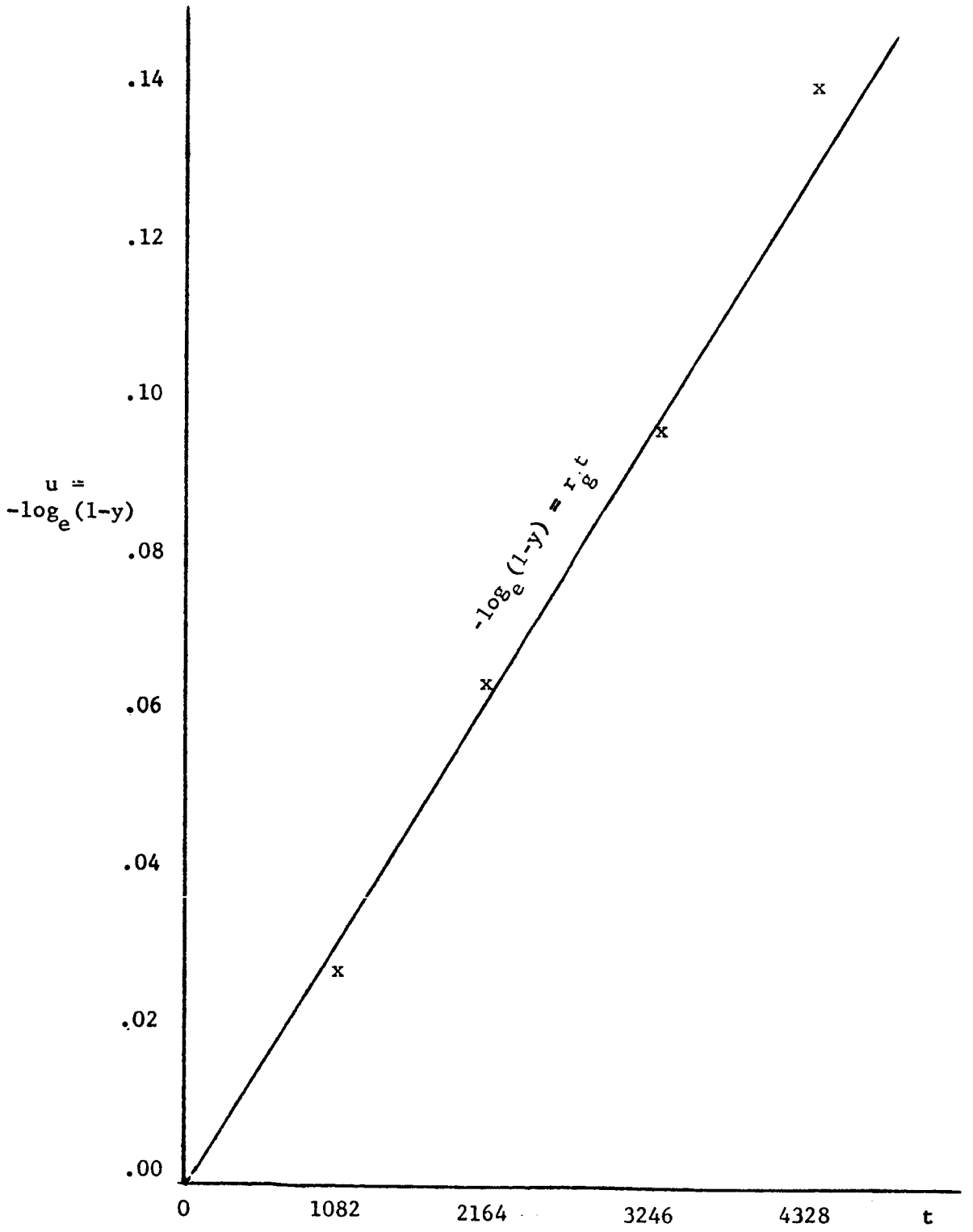

Table 2 - Calculations for Example 1 Leading to $r_g$, $r_{\ell s}$, $r_{w\ell s}$, and $r_{m\ell}$

| t | 1-y | nt | ny | u = $-\log_e(1-y)$ | $t^2$ | tu | w = $n(1-y)/y$ | $wt^2 x10^{-6}$ | wtu |
|---|---|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| 0 | 1.0000 | 0 | 0 | 0 | 0 | 0 | | 0* | 0* |
| 1082 | 0.9714 | 4,167,864 | 110 | 0.0290 | 1,170,724 | 31.3780 | 130,839 | 153,177 | 4,105,472 |
| 2164 | 0.9358 | 7,801,220 | 231 | 0.0663 | 4,682,896 | 143.4732 | 52,555 | 246,110 | 7,540,233 |
| 3246 | 0.9081 | 9,130,998 | 259 | 0.0964 | 10,536,516 | 312.9144 | 27,791 | 292,822 | 8,696,244 |
| 4328 | 0.8678 | 9,547,568 | 292 | 0.1418 | 18,731,584 | 613.7104 | 14,478 | 271,197 | 8,885,330 |
| TOTAL | 4.6831 | 30,647,650 | | | 35,121,584 | 1101.4760 | | 963,306 | 29,227,279 |

*L'Hospital's rule shows that the correct value for these terms is zero in the limit as y approaches zero.

Least Squares. The totals for columns (6) and (7) in Table 2, when substituted in equation (3), give $r_{\ell s} = 0.00003136$.

Weighted Least Squares. Column (8) in Table 2 gives the weights needed for these calculations. Reference to formula (4) shows that $r_{w\ell s}$ is the ratio of the totals of columns (10) and (9), that is, $r_{w\ell s} = 0.00003034$.

Partial Totals. The range of i from 1 to 5 in Table 3 shows that k = 5 and the total for column (4) gives $\sum_{i=1}^{5} (1-y_i) = 4.6831$ so that $\bar{S} = (4.6831)/5 = 0.9366$. Speckman and Cornell's[20] tables of $r_{pt}d$ values have entries for k = 5 and $\bar{S} = 0.93$ and 0.94. Linear interpolation yields $r_{pt}d = 0.0333$ and $r_{pt} = 0.00003078$ since d = 1082. Alternatively, $r_{pt}$ could have been determined by solving equation (7) iteratively.

Method of Moments. To calculate $r_m$, we first observe that the difference between two consecutive dosages is constant and equal to 1082. Taking each t value as the midpoint of an interval for the histogram, we have $d_i = 1082$ for all i, and $\tau = t_k + d/2 = 4328 + 1082/2 = 4869$. Thus,

$$\Sigma A = \sum_{i=1}^{5} y_i d_i = (1082) \sum_{i=1}^{5} y_i = 342.8858,$$

using the total for column (3) of Table 3. From equation (10),

$$\Sigma' A = (100)\Sigma A/\tau = (100)(342.8858)/(4869) = 7.0422.$$

Entering Chart I in Appendix A of Muench's book[18] with this value of $\Sigma' A$, we obtain $r'_m = 0.0015$. This leads to

$$r_m = (100)\ r'_m/\tau = (100)(0.0015)/(4869) = 0.0000308$$

from equation (9).

Finite Differences Method 1. Substituting the totals of columns (5) and (6) of Table 3 along with d = 1082 into equation (12) yields

$$r_{fd1} = -\frac{1}{(1082)} \log_e\left(\frac{3.5182}{3.6439}\right) = 0.00003245 .$$

Finite Differences Method 2. Substitution of the totals of columns (8) and (10) into equation (13) yields

$$r_{fd2} = -\frac{1}{(1082)} \log_e\left(1 - \frac{0.9268}{27.8947}\right) = 0.00003123 .$$

Table 3 - Calculations for Example 1 Leading to $r_{pt}, r_m, r_{fd1}$, and $r_{fd2}$

| $i$ | $t_i$ | $y_i$ | $1-y_i$ | $(1-y_i)^2$ | $(1-y_i)(1-y_{i+1})$ | $S_i = \sum_{h=1}^{i-1} (1-y_i)$ | $S_i^2$ | $y_i - y_1$ | $S_i(y_i - y_1)$ |
|---|---|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| 1 | 0 | 0.0000 | 1.0000 | 1.0000 | 0.9714 | | | | |
| 2 | 1082 | 0.0286 | 0.9714 | 0.9436 | 0.9090 | 1.0000 | 1.0000 | 0.0286 | 0.0286 |
| 3 | 2164 | 0.0642 | 0.9358 | 0.8757 | 0.8498 | 1.9714 | 3.8864 | 0.0642 | 0.1266 |
| 4 | 3246 | 0.0919 | 0.9081 | 0.8246 | 0.7880 | 2.9072 | 8.4518 | 0.0919 | 0.2672 |
| 5 | 4328 | 0.1322 | 0.8678 | | | 3.8153 | 14.5565 | 0.1322 | 0.5044 |
| TOTAL | | 0.3169 | 4.6831 | 3.6439 | 3.5182 | | 27.8947 | | 0.9268 |

Table 4 - Estimates of $\rho$ Computed by Several Different Methods for Example 1

| Method | Graphical | Maximum Likelihood | Least Squares | Weighted Least Squares | Partial Totals | Moments | Finite Differences Method 1 | Finite Differences Meth. 2 |
|---|---|---|---|---|---|---|---|---|
| Estimate $\times 10^7$ | 305 | 304.8 | 313.6 | 303.4 | 307.8 | 308 | 324.5 | 312.3 |

Example 2. Exponentially spaced t's.

For this example we use data obtained by Cîrstea and Suhaciu[3] in studies concerned with the passive sensitization of guinea-pigs. Table 5 gives their data on the mortality rates of guinea-pigs sensitized passively with intraperitoneal rabbit anti-ovalbumin serum and challenged 24 hours later with intracardial

antigen. Each rate is based on ten animals.

The values for the seven estimates calculated for this example are displayed in Table 6. The calculations for $r_g$, $r_{m\ell}$, $r_{\ell s}$ and $r_{w\ell s}$ are not shown since they were done in the same manner as for Example 1 except that $y_5 = 0$ was replaced by $1/2n = 0.05$ in computing the denominator of $r_{w\ell s}$ as given by equation (5) since this denominator approaches infinity as any $y$ approaches zero. The term in the numerator for $i = 5$ was replaced by its limit as $y$ approaches zero, which is $nt = 0.125$. This is essentially the same as the procedure followed in calculating $r_{w\ell s}$ for Example 1 except in that instance the smallest t value was zero.

Method of Moments. The calculation of this estimate is the same as in Example 1 except for the division of the range of t into intervals. In this calculation, equal intervals on the log t scale were used. That is,

$d_i = \text{antilog}\ [\log t_i + (\log 2)/2] - \text{antilog}\ [\log t_i - (\log 2)/2]$ and

$\tau = \text{antilog}\ [\log 0.2 + (\log 2)/2] = 0.283$ .

Fisher. To apply this procedure, we note that $t_i/t_{i+1} = 2$, $i=1,2,3,4$. However, $t_1 = 0.2$ instead of 1. Therefore, this method with $a = 2$ gives a solution for $t_1 r_f = (0.2) r_f$ . In Fisher's notation $s = k = 5$ and $x = \Sigma y = 2.5$. Entering Fisher and Yates'[13] Table VIII 2 with $a = 2$ (Two-fold), we find $K = 0.358$. Substitution in equation (16) yields 0.395 which for this example equals $\log_{10} (0.2) r_f$ . Solving for $r_f$ gives $r_f = 12.4$ as listed in Table 6.

Spearman. From Table 5 we see that the largest t value is 0.2 which is $t_k$ . The corresponding $x_k = \log_e t_k = -1.60944$. Also, $c = \log_e (t_i/t_{i+1}) = \log_e 2 = 0.69315$ and Euler's constant $= \gamma = 0.57722$. Substitution of these quantities in equation (17) yields the value of 11.2 which is given for $r_s$ in Table 6.

Table 5 - Data for Example 2

| Dose of Antiserum (mℓ) t (1) | Mortality Rate y (2) | Survival Rate 1 - y (3) |
|---|---|---|
| 0.2 | 0.9 | 0.1 |
| 0.1 | 0.8 | 0.2 |
| 0.05 | 0.7 | 0.3 |
| 0.025 | 0.1 | 0.9 |
| 0.0125 | 0 | 1.0 |
| | 2.5 | |

Table 6 - Estimates of ρ Computed by Several Different Methods for Example 2

| Method | Graphical | Maximum Likelihood | Least Squares | Weighted Least Squares | Moments | Fisher | Spearman |
|---|---|---|---|---|---|---|---|
| Estimate | 13.5 | 12.9 | 12.8 | 9.2 | 7.1 | 12.4 | 11.2 |

## DISCUSSIONS AND COMPARISONS

Ease of Computation. The examples investigated in this last section help compare some of the characteristics of the estimation procedures described in this paper. Also, a summary of the procedures considered in this paper with regard to the computation of estimates is given in Table 7. As is indicated in Table 7, the graphical procedure is simple and easy to carry out. It requires no assumption concerning the spacing of the t values provided at least two different t values are used or concerning the distribution of the deviations of the data about the model of expectations given by (1). However, the graphical method depends on the judgment of the person drawing the line to describe the data. It usually leads to a lack of reproducibility and the variability in the estimates determined by this method is impossible to assess. The graphical method ordinarily is used only to obtain a preliminary parameter estimate which is subsequently refined, perhaps using an iterative computational procedure. The unweighted and weighted least squares methods as well as the moment, Spearman and finite differences procedures are relatively simple computationally since they do not involve iterative calculations. The partial totals and Fisher methods require extensive calculation if the estimation formulas are evaluated entirely with computations, but, as mentioned earlier, tables are available which make such calculations unnecessary in many instances and make these estimation procedures the easiest to apply in these instances. For the maximum likelihood method, the tables given by Peto and Finney only assist in, but do not eliminate, iterative calculations and

therefore this method remains one of the more difficult computationally.

Distribution and Sampling Specifications, Large Sample Properties. As presented in this paper, only the maximum likelihood and weighted least squares methods take into account variability in the $n_i$ values. The others would not be preferred if the $n_i$ were very different from each other. However, some of them performed well on Example 1 where $n_1$ is nearly twice as large as $n_5$. The maximum likelihood and weighted least squares procedures are also the only ones which make explicit use of the assumption of a binomial distribution of the ny variable for each t value. The other estimation methods only use the functional form of the expected value of y as given by the model in equation (1). The maximum likelihood, least squares, weighted least squares and moment procedures share with the graphical method the advantage of being applicable regardless of the spacing of the t values while the partial totals, Fisher and Spearman methods require particular spacings of the t values. Like the partial totals method, the two finite differences methods require equally spaced t values. These two methods will not be discussed further because no attempt has been made to evaluate them for the model under consideration.

The evaluation of the least squares and weighted least squares procedures is difficult because of the necessity of replacing y by an arbitrary approximation when y = 0 or 1, making the properties usually associated with least squares procedures in doubt. Moreover, in order for the unweighted least squares method to yield an estimator with minimum variance, the variables $u_i = [- \log_e(1-y_i)]$ would have to have the same variance for all i, an assumption that is not fulfilled, for instance, under the binomial model proposed

Table 7 - Summary of Ten Procedures with regard to the Computation of Estimates

| Method | Restrictions on Spacing of t Values | Restrictions on Approximate Range of y | Distribution Assumptions for Derivation of Estimation Methods | Ease of Computation | Computational Aids |
|---|---|---|---|---|---|
| Graphical (g) | | | | Very easy | |
| Maximum Likelihood (mℓ) | | | Binomial | Difficult, iterative | Tables used in iterations |
| Least Squares (ℓs) | | | Constant variance of log y | Easy, direct | |
| Weighted Least Squares (wℓs) | | | Binomial | Easy, direct | |
| Partial Totals (pt) | Equal | | | Easy with tables | Estimates tabled, $t_1 = 0,d$ |
| Moments (m) | Equal Preferred | | | Easy with nomogram | Nomogram for estimates |
| Finite Differences Method 1 (fd1) | Equal | | | Easy, direct | |
| Finite Differences Method 2 (fd2) | Equal | | | Easy, direct | |
| Fisher (f) | Exponential | 0 to 1 | | Easy with tables | K tabled for a = 2,4,10 |
| Spearman (s) | Exponential | 0 to 1 | | Easy, direct | |

earlier for maximum likelihood estimation. Even the weighted least squares method does not necessarily lead to a minimum variance estimator under this model because the appropriate weights are not necessarily known exactly but are estimated from the data and because of the problems when $y = 0$ or 1. However, when the sample size $n_i$ for each i is large enough so that no $y_i = 0$ or 1, under our model the weighted least squares estimator is approximately normally distributed. In fact, the weighted least squares estimator is consistent and asymptotically efficient with a variance approximated by $1/\sum_{i=1}^{k} w_i t_i^2$ when each $n_i$ is sufficiently large.

Large sample results are also known for several of the other methods. The maximum likelihood estimator is consistent and asymptotically efficient and normally distributed provided that for given $t_i$ at least two of the $n_i$ become large as the total sample size becomes large for given $t_i$. Peto[19] gives a procedure for estimating the asymptotic variance of the maximum likelihood estimator. The situation where Muench's method of moments is most likely to be used, especially for large samples, is when $t_i$ values are equally spaced and E(y) comes close to ranging from 0 to 1. In this instance, Cornell[5] has shown that Muench's method of moments and the partial totals methods are equivalent when observations are taken for a large number of t values, that is, for k large. Moreover, Speckman and Cornell[20] showed that the partial totals estimator, $r_{pt}$, is approximately normally distributed for large sample sizes whether or not k is large. They give an expression for the asymptotic variance of $r_{pt}$ and they show that $r_{pt}$ is consistent but not asymptotically efficient.

It has also been shown by Cornell[5] that Fisher and Spearman estimators are equivalent for k large. Fisher[11], and Johnson and Brown[15], gave a formula for the asymptotic variance of these estimators which are consistent and asymptotically normally distributed. These procedures have an asymptotic relative efficiency of 88 per cent. In deriving these large sample properties independent sampling at different t values and finite variances of the y variables were assumed. The large sample results derived for the Spearman and Fisher procedures also require addition of the assumption that for any given t value, the corresponding y variable is binomially distributed.

Small Sample Properties. The large sample properties just reviewed are useful but can be very misleading for small samples. This is particularly true for the least squares and weighted least squares procedures since the large sample conclusions require a large sample for each t value for these methods. Therefore, to better evaluate these methods, it is necessary to examine the results of Monte Carlo studies. In their study, Speckman and Cornell[20] compared the least squares, weighted least squares, maximum likelihood and partial totals methods for the model discussed in this paper with t values a constant interval of width d apart for the following parameter combinations:

$$\rho d = 0.30;\ k = 5;\ n = 5,10,15,25,$$

$$k = 10;\ n = 5,10$$

$$k = 15;\ n = 5,25,$$

$$\rho d = 0.15;\ k = 10;\ n = 10,$$

$$\rho d = 0.10;\ k = 15;\ n = 5,25.$$

They included the assumption that ny has a binomial distribution for each value

of t in their model. They found that the partial totals estimator is slightly less biased than the maximum likelihood estimator, particularly for small n and k. The least squares and especially the weighted least squares estimators are considerably more biased. In their study for y = 0 and 1 in the two least squares procedures, an adjustment of 1/2n was made only when the indeterminate form of concern had no limit as y approached the appropriate end-point. If such a limit existed, it was used. In comparing the methods on relative efficiency, which equals the variance plus the square of the bias, for $\rho d = 0.30$ the partial totals estimator was superior for k = 5 and either the least squares or partial totals method was superior for k = 10 and 15. For $\rho d = 0.15$ and 0.10, for which only k values of 10 and 15 were studied, the maximum likelihood estimator was more efficient. The weighted least squares estimator had low efficiency relative to the other methods in most cases and never had the highest efficiency. In a comparison of the maximum likelihood and partial totals estimators, it was found that they performed very similarly, although the partial totals estimator had slightly higher efficiency for the small sample sizes while the opposite was true for the larger sample sizes studied empirically. A recommendation was also made that asymptotic variances be multiplied by 1.5 when used to estimate variances for small samples. Further work has shown that for some situations this factor may be too small.

Comparable evidence on the properties of the other estimators for small samples is not available. However, Johnson and Brown[15] presented a bias approximation for the Spearman estimator which can be used to obtain a less biased estimator. Also, in studies summarized by Brown[2], Spearman estimation has been

found to have high efficiency relative to maximum likelihood for small samples for models different but similar to the one considered here. However, since the Spearman method involves approximating a sum used by Fisher under the assumption that k is large, the Fisher method would seem preferable for small samples. Similarly, the method of partial totals for equally spaced t values would be preferred to Muench's method of mements since Cornell[5] has shown that they are related in the same way as the Fisher and Spearman methods in this instance. When the t values are not equally spaced the integral used by Muench is the same as that when they are equally spaced, so his method of moments would appear inappropriate then. In particular, if the t values are exponentially spaced, the Spearman method, which is similar to the Muench method except that it is explicitly devised for exponential spacing instead of equal spacing, would be more appropriate.

The examples given earlier illustrate several properties discussed. In Table 4, the estimates computed by several methods for Example 1 are given. The estimate computed by the first finite differences method is much larger than the others while those computed by least squares and the second finite differences method are close together and relatively large. The other estimates are close together with the weighted least squares estimate being the smallest. This supports the evidence given above that least squares estimators for the model being studied, whether unweighted or weighted, are biased more than maximum likelihood and partial totals estimators. The partial totals and moment methods, which are similar for equally spaced t values as used in this example, give essentially the same answer. The estimates given for the second example, which has exponential

spacing of t values, in Table 6 are similar except the estimates computed by the weighted least squares and moments methods are small relative to the others. Unsatisfactory performance of weighted least squares for small samples is illustrated here as is the inappropriateness of Muench's method of moments when the t values are not equally spaced. These two examples, of course, give no indication of the sampling variability of the various estimators.

## SUMMARY AND CONCLUSIONS

Several methods of estimating the parameter in a simple exponential model, which often arises in epidemiological studies and biological assay, are presented, illustrated and compared. The method of maximum likelihood, which can be used for any spacing of doses, is known to have very desirable large sample properties and behaved quite well in the limited Monte Carlo study. The simple method of partial totals is suggested as a possible alternative to maximum likelihood for small samples for equally spaced doses. The Fisher and Spearman methods, which are also computationally easy, are suggested as alternative methods regardless of the sample size for exponentially spaced dosages, that is, for dosages whose logarithms are equally spaced.

## REFERENCES

1. Berkson, J. Application of the Logistic Function to Bio-assay. J. of the Amer. Stat. Ass. 39:357-365, 1944.

2. Brown, B. W., Jr. Some Properties of the Spearman Estimator in Bioassay. Biometrika 48: 293-302, 1961.

3. Cîrstea, M. and Suhaciu, G. Studies on Passive Sensitization in Guinea-pigs. Int. Archives of Allergy and Appl. Immunology 23: 216-218, 1963.

4. Cochran, W. G. Estimation of Bacterial Densities by Means of the "Most Probable Number". Biometrics 6: 105-116, 1950.

5. Cornell, R. G. Spearman Estimation for a Simple Exponential Model. Biometrics 21, 1965.

6. Cornfield, J. Measurement and Comparison of Toxicities: The Quantal Response. Statistics and Mathematics in Biology (ed. by Kempthorne, Bancroft, Gowen and Lush). Ames: The Iowa State College Press, 1954, Chapter 25.

7. Edington, C. W., Epler, J. L. and Regan, J. D. The Frequency-dose Relation of X-ray Induced Y-suppressed Lethals in Drosophila. Genetics 47: 397-406, 1962.

8. Finney, D. J. The Principles of Biological Assay. J. of the Roy. Stat. Soc., Supplement 9: 46-91, 1947.

9. Finney, D. J. The Estimation of Bacterial Densities from Dilution Series. J. of Hygiene 49: 26-35, 1951.

10. Finney, D. J. Statistical Methods in Biological Assay (2nd ed.) New York: Hafner Publishing Company, 1964, Section 21.5.

11. Fisher, R. A. On the Mathematical Foundations of Theoretical Statistics. Phil. Trans. of the Roy. Soc. A 222: 309-368, 1921.

12. Fisher, R. A. The Case of Zero Survivors. Appendix in C. I. Bliss. The Calculation of the Dosage-mortality Curve. Ann. Appl. Biol. 22: 134-167, 1935.

13. Fisher, R. A. and Yates, F. Statistical Tables for Biological, Agricultural and Medical Research (6th ed.), Edinburgh: Oliver and Boyd, 1963.

14. Hartley, H. O. The Estimation of Non-linear Parameters by "Internal Least Squares." Biometrika 35: 32-45, 1948.

15. Johnson, E. A. and Brown, B. W., Jr. The Spearman Estimator for Serial Dilution Assays. Biometrics 17: 79-88, 1961.

16. Lipton, S. and McGilchrist, C. A. The Derivation of Methods for Fitting Exponential Regression Curves. Biometrika 51: 504-508, 1964.

17. Manos, N. E. Histoplasm Sensitivity Conversion Rates. Am. J. of Hygiene 58: 119-131, 1953.

18. Muench, H. Catalytic Models in Epidemiology. Cambridge: Harvard University Press, 1959.

19. Peto, S. A Dose-response Equation for the Invasion of Micro-organisms. Biometrics 9: 320-335, 1953.

20. Speckman, J. A. and Cornell, R. G. Estimation for a One-parameter Exponential Model. J. of the Amer. Stat. Ass. 60: 560-572, 1965.

Dr. Cornell is Associate Professor of Statistics and Director of the Biometry Training Program at Florida State University. During a portion of the work on this paper he received support from training grant number 5TI GM-913 from the National Institute of General Medical Sciences and from NASA grant number NGR-10-004-029.

Miss Speckman is a statistician with the Statistical Engineering Laboratory of the National Bureau of Standards.