# Query: An Affine Linear Model for the Relation between Two Sets of Frequency Counts

Joel E. Cohen; Peter D'eustachio

*Biometrics*, Vol. 34, No. 3. (Sep., 1978), pp. 514-521.

Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28197809%2934%3A3%3C514%3AQAALMF%3E2.0.CO%3B2-M

*Biometrics* is currently published by International Biometric Society.

# QUERY: *An Affine Linear Model for the Relation Between Two Sets of Frequency Counts*

In fetal development, different functional cell populations arise at different times and proliferate at different rates to give rise to the relative numbers and distributions of cell types ultimately observed in the adult. We recorded the numbers of nucleated cells ($X$-cells) and the numbers of cells ($Y$-cells) which formed rosettes with trinitrophenyl-derivatized sheep red blood cells, in the spleens of individual fetal mice 18 days after conception (Cohen, D'Eustachio and Edelman 1977). For example, in a litter of *BALB/c* × *CBA/J* fetuses, the numbers of $X$-cells and of $Y$-cells, presented as $(X, Y)$ pairs, in 5 individuals were (337, 52), (141, 6), (177, 14), (116, 5), (88, 5). In this case, 3 aliquots of the cell suspension were sampled for all individuals. In other cases, the number of aliquots sampled varied from one individual to another. The number of aliquots was always the same for the $X$-cells and $Y$-cells of a given individual.

When these data are plotted within individual litters, the points fall roughly along a straight line with a positive $X$-intercept. This positive $X$-intercept, or threshold, may indicate a delayed start in the expansion of the $Y$-cell compartment of the spleen relative to the $X$-cell compartment. We would like to evaluate the hypothesis of linearity in detail, in order to determine whether the observed deviations from an exact linear relationship could be accounted for by Poisson sampling fluctuation.

# RESPONSE:

JOEL E. COHEN and PETER D'EUSTACHIO[1]
The Rockefeller University, 1230 York Avenue, New York, New York, 10021, U.S.A.

Suppose that $\eta_i, \xi_i, i = 1, \ldots, k$ are the total numbers of $X$-cells and $Y$-cells in each of $k$ individuals. The experimental design determines the number $a_i$ of aliquots of cells sampled from each individual. We assume $a_i$ is chosen so that the observed numbers $X_i$ and $Y_i$ of each kind of cell are "large" and are subject to Poisson variation. We wish to fit the model

$$E(X_i) = \lambda_i, 4\ E(Y_i) = \mu_i = c(\lambda_i - a_i d), 4\ \lambda_i \geq a_i d \tag{1}$$

where only $a_i$ is known. Model (1) is a so-called "structural" regression for Poisson variates because the observed abscissae $X_i$ include sampling variation.

Other models for regression of Poisson variates or for analysis of the usual hypothesis of proportionality have been described by Fleiss (1973, p. 97), El-Sayyad (1973) and Simon (1974). An alternative to the present approach would be to stabilize variances by an appropriate square root transformation and then carry out a structural regression (Dolby 1976).

[1] Current address: Department of Biology, Yale University, New Haven, Connecticut 06520, U. S. A.

When $d = 0$, model (1) is the usual hypothesis of proportionality of rows and columns in a $2 \times k$ contingency table with $X_i$ in the first row and $Y_i$ in the second. When $d \neq 0$, model (1) may be interpreted as supposing that the average number $\mu_i$ of $Y$-cells in $a_i$ aliquots from an individual $i$ is proportional (with constant $c$) to the excess in the average number $\lambda_i$ of $X$-cells over some threshold $a_i d$ where $d$ is the threshold per aliquot.

If $X_i/a_i$ is plotted on the abscissa of a graph and $Y_i/a_i$ on the ordinate, then the points should approximate a straight line with slope $c$ according to the model (1). Assuming $c \neq 0$, the line should pass through the $X$-axis at $X = d$. Preparing such a graph is recommended as a preliminary test of the reasonableness of the model and a rough way of estimating $c$ and $d$.

Point estimates of the parameters, which can be obtained from the maximum likelihood ($ML$) equations, are solutions of

$$\hat{c} = \Sigma Y_i / \Sigma(\hat{\lambda}_i - a_i \hat{d}), \tag{2}$$

$$\hat{c}\Sigma a_i = \Sigma a_i Y_i / (\hat{\lambda}_i - a_i \hat{d}), \tag{3}$$

and

$$X_i/\hat{\lambda}_i + Y_i/(\hat{\lambda}_i - a_i \hat{d}) = \hat{c} + 1. \tag{4}$$

When $d = 0$, (2) and (4) give the conventional estimators for the $2 \times k$ table, $\hat{c} = \Sigma Y_i / \Sigma X_i$ and $\hat{\lambda}_i = (X_i + Y_i)\Sigma_j X_j / \Sigma_j (X_j + Y_j)$.

A suggested procedure for solving (2) to (4) numerically is to estimate an initial value for $d$ graphically or by ordinary least squares, using $X_i$ as an initial value for $\lambda_i$ and using (2) to obtain an initial $c$. Then (*i*) find a value of $d$ which satisfies (3) by the Newton-Raphson procedure, (*ii*) find an improved value of each $\lambda_i$ from (4), which is an explicitly soluble quadratic equation, (*iii*) improve $c$ via (2), and go to step (*i*). Stop when all parameter values quasi-converge. When a visual inspection of the data warrants using the model in the first place, the procedure gives reasonable results, e.g., Cohen, D'Eustachio and Edelman (1977, Tables IV and V).

The variance-covariance matrix of the parameter estimates is, asymptotically, i.e. for large counts $X_i$ and $Y_i$, the inverse of $(-1)$ times the expected value of the matrix of second partial derivatives of $\ln L$. We estimate this matrix by replacing the parameters by the $ML$ estimates. We enumerate the parameters as before in the order $c, d, \lambda_1, \ldots, \lambda_k$. Hence the first two rows of the estimated inverse variance-covariance matrix are

$$\begin{array}{ccccc} \Sigma(\hat{\lambda}_i - a_i\hat{d})/\hat{c} & -\Sigma a_i & +1 & \ldots & +1 \\ -\Sigma a_i & \Sigma a_i^2 \hat{c}/(\hat{\lambda}_i - a_i\hat{d}) & -a_1\hat{c}/(\hat{\lambda}_1 - a_1\hat{d}) & \ldots & -a_k\hat{c}/(\hat{\lambda}_k - a_k\hat{d}). \end{array} \tag{5}$$

The lower right $k \times k$ submatrix of the inverse variance-covariance matrix has

$$1/\hat{\lambda}_i + \hat{c}/(\hat{\lambda}_i - a_i\hat{d}), \tag{6}$$

as the *ith* diagonal element and off-diagonal elements zero.

When model (1) is fitted to the data with $a_i = 3$, we estimate $c = 0.17818$ and $d = 26.5853$. Goodness of fit to the expected values can be assessed using the conventional chi-squared, $\chi^2 = 5.13$, or $-2$ log likelihood ratio measure, $G^2 = 5.21$, with 3 degrees of freedom. The fit is quite acceptable. The estimated variance-covariance matrix (the inverse of the matrix from (5) and (6)) of the parameter estimates can be obtained (Table 1). The 99% confidence intervals for $c$ and $d$ are respectively from 0.0993 to 0.257 and from 16.3 to 36.9.

Since each aliquot of nucleated cells ($X$-cells) represented $10^{-4}$ of the total nucleated cells in a fetal spleen, we infer that there were $10^4 d = 265{,}853$ or approximately a quarter million

TABLE 1
Variance-Covariance Matrix of the Estimated Parameters

| | c | d | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
|---|---|---|---|---|---|---|---|
| c | 0.0009 | 0.0835 | −0.2152 | −0.0087 | −0.0595 | 0.0285 | 0.0767 |
| d | | 16.1467 | −14.1500 | 7.3373 | 1.2540 | 12.4308 | 19.7132 |
| $\lambda_1$ | | | 330.7938 | 6.5683 | 17.2873 | −0.9163 | −10.2331 |
| $\lambda_2$ | | | | 99.1315 | 6.6060 | 8.9243 | 11.2699 |
| $\lambda_3$ | | | | | 138.1067 | 5.5987 | 4.8013 |
| $\lambda_4$ | | | | | | 82.3421 | 16.9512 |
| $\lambda_5$ | | | | | | | 74.9106 |

nucleated cells in the fetal spleen before rosette-forming cells ($Y$-cells) began to appear in this litter. Since each aliquot of $Y$-cells represented 0.05 of the total $Y$-cells in a fetal spleen, each additional $X$-cell above the threshold of a quarter million was accompanied by $20c/10^4 = 3.56 \times 10^{-4}$ additional $Y$-cells. More biologically, for each additional million nucleated cells over the threshold, there were roughly 350 additional rosette-forming cells.

### *Acknowledgments*

Gerald M. Edelman provided biological and editorial guidance. We thank P. M. E. Altham, W. G. Cochran, S. E. Fienberg, N. Mantel, J. Wittes, and numerous referees and editors for critical comments. This work was supported in part by U. S. National Science Foundation grant BMS74-13276 and U. S. Public Health Service grants AI 11378 and AI 09273 from the National Institutes of Health.

### *References*

Cohen, J. E., D'Eustachio, P. and Edelman, G. M. (1977). The specific antigen-binding cell populations of individual fetal mouse spleens: repertoire composition, size and genetic control. *Journal of Experimental Medicine 146,* 394–411.

Dolby, G. R. (1976). A note on the linear structural relation when both residual variances are known. *Journal of the American Statistical Association 71,* 352–353.

El-Sayyad, G. M. (1973). Bayesian and classical analysis of Poisson regression. *Journal of the Royal Statistical Society, Series B 35,* 445–451.

Fleiss, J. L. (1973). *Statistical Methods for Rates and Proportions.* John Wiley and Sons, Inc., New York.

Simon, G. (1974). Alternative analyses for the single-ordered contingency table. *Journal of the American Statistical Association 69,* 971–976.

## RESPONSE:

WAYNE A. FULLER

Department of Statistics, Iowa State University, Ames, Iowa 50011, U.S.A.

The query on the relationship between the counts of two kinds of cell raises interesting problems because the measures of both kinds of cells for the *ith* individual are subject to sampling variation. Cohen and D'Eustachio present the maximum likelihood (*ML*) solution under the assumption of Poisson variation. We present analyses based upon the square roots of the original data. The square root transformation for the Poisson distribution has a considerable history. See Bartlett (1936), Cochran (1940), and Thöni (1967). The square root transformation typically has the advantage that it simplifies the computation of tests and estimators. In the present situation the square root transformation permits us to use existing computer software.

Let $\eta_i$ be the total number of cells of $X$-type and $\xi_i$ the total number of cells of $Y$-type in the $ith$ individual. Let $X_i$ be the observed number of cells of $X$-type in $a_i$ aliquots and $Y_i$ the observed number of cells of $Y$-type in $a_i$ aliquots selected from the $ith$ individual. The query postulated a Poisson distribution for the observed counts and a linear relation between the means of the two types of counts. To obtain such a model one might assume that the $\xi_i$ satisfy the regression equation

$$\xi_i = \alpha_0 + \alpha_1\eta_i + v_i,$$

where $E\{v_i \mid \eta_i\} = 0$.

If a small fraction of the fetal spleen is sampled it is reasonable to consider the sampling variation in the cell counts to be Poisson. Under the Poisson model the conditional mean (conditional on $\eta_i$) of the counts $X_i$ is $E\{X_i \mid \eta_i\} = f_{xi}\eta_i$, and the conditional variance is $Var\{X_i \mid \eta_i\} = f_{xi}\eta_i$, where $f_{xi}$ is the fraction of the spleen sampled for $X$-cells.

The conditional mean of $Y_i$ given $\xi_i$ is $E\{Y_i \mid \xi_i\} = f_{yi}\xi_i$, where $f_{yi}$ is the fraction of the spleen sampled for $Y$-cells. The conditional mean of $Y_i$ given $\eta_i$ is

$$E\{Y_i \mid \eta_i\} = x_{yi}(\alpha_0 + \alpha_1\eta_i)$$

and the conditional variance given $\eta_i$ is

$$Var\{Y_i \mid \eta_i\} = x_{ui}(\alpha_0 + \alpha_1\eta_i) + f_{ui}^2\ Var\{v_i\}.$$

The model of the query can be obtained by postulating $f_{yi}^2\ Var\{v_i\}$ to be small relative to $f_{yi}(\alpha_0 + \alpha_1\eta_i)$ so that $Y_i$, conditional on $\eta_i$, is approximately Poisson with mean $f_{yi}(\alpha_0 + \alpha_1\eta_i)$. The query specifies no distribution for $\eta_i$, so we treat $\eta_i$ as fixed unknown, parameters in our analysis. If the total number of aliquots of $X$-cells is $T_x$ and the total number of aliquots of $Y$-cells is $T_y$, the model for the observed counts becomes

$$E\{Y_i\} = T_y^{-1}a_i\alpha_0 + \alpha_1 T_y^{-1}T_x E\{X_i\}. \tag{1}$$

If we set $E\{X_i\} = \lambda_i$, $c_1 = T_y^{-1}T_x\alpha_1$ and $d_0 = -\alpha_0\alpha_1^{-1}T_x^{-1}$, we obtain the parameterization used by Cohen and D'Eustachio in their response, where we have added subscripts to $c$ and $d$ for later identification.

Let $Z_i$ denote the square root of the count of $X$-cells and $W_i$ the square root of the count of $Y$-cells for the $ith$ individual. By Taylor's theorem we have

$$Z_i = \lambda_i^{.5} + \tfrac{1}{2}\lambda_i^{-.5}(X_i - \lambda_i) - 8^{-1}\lambda_i^{-1.5}(X_i - \lambda_i)^2 + R_i,$$

where $R_i$ is the remainder. Thus, to a first order of approximation, $E\{Z_i\} = \lambda_i^{.5}$ and to a second order of approximation, $E\{Z_i\} = \lambda_i^{.5} - 8^{-1}\lambda_i^{-.5}$, where we have used the fact that the variance of a Poisson random variable is equal to the mean.

It follows that the first order approximation to the model linear in the original variables postulated in the query is

$$w_i = c_1^{.5}(z_i^2 - a_i d_0)^{.5}, \tag{2}$$

$$Z_i = z_i + u_i, \tag{3}$$

and

$$W_i = w_i + e_i, \tag{4}$$

where $z_t = E\{Z_t\}$, $w_t = E\{W_t\}$ and it is assumed that $(e_t, u_t)'$ are independently distributed with mean zero and diagonal covariance matrix, diag(0.25,0.25). The second order approximation to the model linear in the original observations is

$$w_t = c_1^{.5}(\lambda_t - a_t d_0)^{.5} - 8^{-1} c_1^{-.5}(\lambda_t - a_t d_0)^{-.5} \tag{5}$$

$$z_t = \lambda_t^{.5} - 8^{-1}\lambda_t^{-.5},\ Z_t = z_t + u_t, \quad \text{and} \quad W_t = w_t + e_t.$$

As alternative models for the relationship between $w_t$ and $z_t$ we consider

$$w_t = \beta_1 z_t, \tag{6}$$

$$w_t = \beta_0 a_t^{.5} + \beta_1 z_t, \tag{7}$$

and

$$w_t = a_t^{.5}\delta_0[\exp\{a_t^{-.5}\delta_1 z_t\} - 1]. \tag{8}$$

A method of constructing estimators suitable for small sample sizes can be obtained by considering the model in a nonlinear least squares framework. This approach is particularly suitable if the model is nonlinear, e.g. models (2) and (8). The data arrangement associated with the use of nonlinear least squares is given in Table 1. Note that the five $(Z, W)$ pairs obtained from the five $(X, Y)$ pairs of the query have been arranged in a column of ten observations. In terms of Table 1 the five models (2), (5), (6), (7), and (8) become

$$\psi_t = \sum\varphi_{it} z_i + c_1^{.5}[\sum\varphi_{5+i,t}\, z_i^2 - d_0 a_{i\varphi 0t}]^{.5} + \epsilon_t, \tag{9}$$

$$\psi_t = \sum \varphi_{it}(\lambda_i^{.5} - 8^{-1}\lambda_i^{-.5}) + c_1^{.5}[\sum \varphi_{5+i,t}\,\lambda_i - d_0 a_i\varphi_{0t}]^{.5}$$
$$- 8^{-1} c_1^{-.5}[\sum \varphi_{5+i,t}\,\lambda_i - d_0 a_i\varphi_{0t}]^{-.5} + \epsilon_t, \tag{10}$$

$$\psi_t = \sum\varphi_{it} z_i + \beta_1\sum\varphi_{5+i,5} z_i + \epsilon_t, \tag{11}$$

$$\psi_t = \sum\varphi_{it}\, z_i + \beta_0 a_i^{.5}\varphi_{0t} + \beta_1\sum\varphi_{5+i,t} z_i + \epsilon_t, \tag{12}$$

TABLE 1
Data Tableau for Nonlinear Least Squares Estimation

| Index<br>t | Original<br>obs. | $\psi$ | $\varphi_0$ | $\varphi_1$ | $\varphi_2$ | $\varphi_3$ | $\varphi_4$ | $\varphi_5$ | $\varphi_6$ | $\varphi_7$ | $\varphi_8$ | $\varphi_9$ | $\varphi_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $X_1^{1/2}$ | 18.358 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | $X_2^{1/2}$ | 11.874 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | $X_3^{1/2}$ | 13.304 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | $X_4^{1/2}$ | 10.770 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | $X_5^{1/2}$ | 9.381 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 6 | $Y_1^{1/2}$ | 7.211 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 7 | $Y_2^{1/2}$ | 2.449 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 8 | $Y_3^{1/2}$ | 3.742 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 9 | $Y_4^{1/2}$ | 2.236 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 10 | $Y_5^{1/2}$ | 2.236 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

TABLE 2
Nonlinear Least Squares Estimates

| Model | Parameter index<br>1 | <br>0 | Residual<br>mean<br>square |
|---|---|---|---|
| (9) | 0.178<br>(0.040) | 26.8<br>(5.2) | 0.44 |
| (10) | 0.178<br>(0.040) | 26.5<br>(5.2) | 0.44 |
| (11) | 0.300<br>(0.040) | | 1.27 |
| (12) | 0.608<br>(0.082) | −2.41<br>(0.62) | 0.23 |
| (13) | 0.528<br>(0.166) | 0.206<br>(0.029) | 0.10 |

and

$$\psi_t = \sum \varphi_{it} z_i + a_i^{.5}\delta_0[\exp\{\delta_1 a_i^{-.5} \sum \varphi_{5+i,t}\, z_i\} - 1] + \epsilon_t, \tag{13}$$

where each summation is from $i = 1$ to 5 and the $\epsilon_t$ are independent (0,0.25) random variables. The parameters of the nonlinear regression models (9)–(13) can be estimated using a nonlinear regression program, for example *NLIN* of *SAS* 76.

The observed $X$-values can be used as start values for $z_i^2$. The ordinary least squares regression of $Y$ on $X$ will provide start values for (9) and (10) and the ordinary least squares regression of $W$ on $Z$ will provide start values for (11) and (12). Graphical methods can be used to obtain start values for (13).

The nonlinear least squares estimates of the four models are given in Table 2. The sample standard errors in parentheses are those output by *NLIN* of *SAS* 76. They are computed using the regression residual mean square. To compute the standard errors under the Poisson model ($\sigma^2 = 0.25$), multiply the table standard errors by one half of the reciprocal of the square root of the regression residual mean square given in the last column of the table. The estimates obtained from the first and second order approximations to the linear model are nearly identical and very similar to the *ML* estimates presented by Cohen and D'Eustachio. However the estimated standard errors associated with nonlinear least squares are somewhat larger than the *ML* estimates.

Under the assumption that the original ($X$, $Y$) random variables are independent Poisson random variables, the model linear in the original values (9) and (10), the model linear in the square roots (12) and the exponential model (13) would all be judged acceptable by the usual $F$-test constructed as the residual mean square divided by 0.25. Model (11), the proportional model, would be rejected at the 1% level by the lack of fit $F_\infty^4$-statistic of $(1.27)/(0.25) = 5.08$. The model linear in the square roots (7) and the exponential model (8) give somewhat better fits than the model linear in the original variables. Both of these models have a concave shape, as does the plot of the original data. Clearly additional observations and observations from somewhat younger fetuses are required if one is to choose among the alternative models.

Because this is a nonlinear problem all distributional statements are approximations. The problem is further complicated by the fact that the number of parameters is approximately proportional to the number of observations. The conditions under which such approximations are adequate is difficult to establish. Wolter (1974), studying the nonlinear errors in variables problems, obtained a limiting distribution for $n^{.5}(\hat{d}_0 - d_0, \hat{c}_1 - c_0)$ of models such as (9) by considering a sequence of samples wherein the variance of $\psi_t$ decreased at the rate $n^{-.5}$. He also suggested a modification of the estimator that had a limiting distribution under slightly weaker conditions.

The model defined by the three equations (7), (3), (4) and the model defined by the three equations (6), (3), (4) are examples of the classical linear functional model. See, for example, Kendall and Stuart (1967, Ch. 29). The *MLE* of the parameter of model (7) for normally distributed $(e_i, u_i)$ is

$$\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix} = \left[\begin{pmatrix}\Sigma a_i & \Sigma a_i^{.5}Z_i\\ \Sigma a_i^{.5}Z_i & \Sigma Z_i^2\end{pmatrix} - \hat{\lambda}\begin{pmatrix}0 & 0\\ 0 & 0.25\end{pmatrix}\right]^{-1}\begin{bmatrix}\Sigma W_i\\ \Sigma Z_i W_i\end{bmatrix}, \qquad (14)$$

where $\hat{\lambda}$ is the smallest root of

$$\left|\begin{bmatrix}\Sigma W_i^2 & \Sigma a_i^{.5}W_i & \Sigma W_iZ_i\\ \Sigma a_i^{.5}W_i & \Sigma a_i & \Sigma a_i^{.5}Z_i\\ \Sigma W_iZ_i & \Sigma a_i^{.5}Z_i & \Sigma Z_i^2\end{bmatrix} - \lambda\begin{bmatrix}0.25 & 0.00 & 0.00\\ 0.00 & 0.00 & 0.00\\ 0.00 & 0.00 & 0.25\end{bmatrix}\right| = 0. \qquad (15)$$

The smallest root $\hat{\lambda}$ of (15) is equal to the regression residual sum of squares associated with model (12) divided by 0.25.

A program has been developed at Iowa State University to construct the estimator (14) and estimators of other errors in variables models (Hidiroglou, Fuller and Hickman 1978). The program is called *SUPER CARP* and is designed for the linear regression problem with multiple explanatory variables. The algorithm in *SUPER CARP* replaces $\hat{\lambda}$ in (14) with $\hat{\lambda}(n - p - 1)(n - p)^{-1}$, where $p$ is the number of parameters estimated. The modification produces an estimator with smaller mean square error. Also the program uses a method of computing variances that is applicable, in large samples, to observations $(e_i, u_i)$ selected from distributions possessing finite moments of order greater than four. The estimated variance matrix of the estimator is

$$\hat{V}\{(\hat{\beta}_0, \hat{\beta}_1)'\} = \hat{\mathbf{M}}^{-1}\mathbf{G}\hat{\mathbf{M}}^{-1},$$

where

$$\hat{\mathbf{M}} = \begin{pmatrix}\Sigma a_i & \Sigma a_i^{.5}Z_i\\ \Sigma a_i^{.5}Z_i & \Sigma Z_i^2\end{pmatrix} - \lambda\begin{pmatrix}0 & 0\\ 0 & 0.25\end{pmatrix}$$

$$\mathbf{G} = \begin{pmatrix}\Sigma a_i\hat{v}_i^2 & \Sigma a_i^{.5}Z_i\hat{v}_i^2\\ \Sigma a_i\cdot^{5}Z\hat{v}_i^2 & \Sigma Z_i^2\hat{v}_i^2\end{pmatrix}, \text{ and } \hat{v}_i = W_i - \hat{\beta}_0 - \hat{\beta}_1 Z_i\,.$$

It has been demonstrated under mild conditions that $n^{.5}(\hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1)'$ converges to a normal random variable as the sample size $n$ increases and that $n\hat{V}\{(\hat{\beta}_0, \hat{\beta}_1)\}$ is a consistent estimator of the variance of $n^{.5}(\hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1)'$. The limiting distribution can be obtained for the linear model by assuming that the error variances are becoming small relative to the mean of the random variables or by assuming that the number of observations is becoming large. The estimated model (7) obtained using the program *SUPER CARP* is

$$\begin{aligned} w_i = {} & -\underset{(0.63)}{2.38}a_i^{.5} + \underset{(0.071)}{0.605}z_i. \end{aligned}$$

The estimated parameters differ from those of Table 2 because $\hat{\lambda}$ is replaced by $(n-p)^{-1}(n-p-1)\lambda$ in the computation. The estimated standard errors differ because of the different methods of computation.

Because the error variance in $Z$ is small relative to the total variation, the estimates obtained by the errors in variables techniques are close to those obtained by ordinary least squares. However, the test for model fit requires the computation of $\hat{\lambda}$ or an equivalent statistic.

## *Acknowledgements*

We were privileged to read an earlier manuscript by Cohen and D'Eustachio. In that manuscript Cohen and D'Eustachio mentioned the possibility of analyzing the square roots of the original data. We thank W. G. Cochran, C. P. Cox, and D. F. Cox for useful comments.

## *References*

Bartlett, M. S. (1936). The square root transformation in analysis of variance. *Journal of the Royal Statistical Society Supplement 3,* 68–78.

Cochran, W. G. (1940). The analysis of variance when experimental errors follow the Poisson or binomial laws. *Annals of Mathematical Statistics 11,* 335–347.

Cohen, J. E., D'Eustachio, P. and Edelman, G. M. (1977). The specific antigen-finding cell populations of individual fetal mouse spleens: repertoire composition, size and genetic control. *Journal of Experimental Medicine 146,* 396–411.

Hidiroglou, M. A., Fuller, W. A. and Hickman, R. D. (1978). *SUPER CARP.* Statistical Laboratory, Iowa State University, Ames, Iowa.

Kendall, M. G. and Stuart, A. (1967). *The Advanced Theory of Statistics Vol. 2.* Hafner, New York.

Thöni, H. (1967). Transformations of variables used in the analysis of experimental and observational data; a review. Iowa State University Statistical Laboratory Technical Report No. 7, Ames, Iowa.

Wolter, K. M. (1974). Estimators for a nonlinear functional relationship. Unpublished Ph.D. thesis, Iowa State University Library, Ames, Iowa.