This research was supported in part by an American Cancer Society Cancer Institutional Sub-Grant No. IN-15-M and a Public Health Service Research Career Development Award (No. GM 70004) from the National Institute of General Medical Sciences (G.G.K.)

A LINEAR MODELS APPROACH TO THE ANALYSIS OF SURVIVAL IN MULTI-DIMENSIONAL CONTINGENCY TABLES

by

Gary G. Koch, H. Dennis Tolley, and William D. Johnson

Department of Biostatistics
University of North Carolina

Institute of Statistics Mimeo Series No. 763

JULY 1971

A LINEAR MODELS APPROACH TO THE ANALYSIS OF SURVIVAL IN MULTI-DIMENSIONAL CONTINGENCY TABLES

by

Gary G. Koch, H. Dennis Tolley, and William D. Johnson
University of North Carolina

1. INTRODUCTION

One problem of interest in the study of terminal disease has been the measurement of the effectiveness of therapy. A number of authors have suggested that survival is an objective criterion for this purpose. The merit of survival in this context (for example, survival on a 5-year basis) has been aptly illustrated by Cutler and Ederer [1]. These authors have also presented procedures for determining survival rates in medical follow-up studies. Both the problem of patients being lost to follow-up before their termination dates and that of patients' terminating dates being dispersed outside the interval of study have been considered. The resulting estimates of survival rate use information from lost cases as well as from those cases that continue through the study to their termination dates. Standard error estimates are those derived by Greenwood [2].

In typical studies of human disease, individuals are generally cross-classified according to several variables, with survival rates differing across the resulting groups. Examples of such classification variables include age, social class, stage of disease at diagnosis, and treatment. Statistical techniques have been developed to adjust for the effect that such variables may have on overall survival rates, and thereby allow the analyst to compare survival rates for the different treatments with greater precision. One approach to this problem is formulated in terms of a distribution for survival rates (Meyers, Axtell, and Zelen [6]), while another involves procedures for pairing treated

patients with controls (Mantel and Haenszel [4]). Finally, in certain situations a simple arithmetic correlation of results is sufficient. The latter procedure is often applied for age adjustment.

This paper indicates how a general method for analyzing qualitative data can be used to estimate the survival rate for each of several groups and to test the statistical significance of the effects of the underlying variables. In particular, one can investigate whether certain categories (e.g., different treatments) have an overall effect on the survival rate after accounting for other categorical effects (e.g., stage of disease, age, etc.).

2. GENERALIZED APPROACH TO LINEAR CATEGORICAL MODELS

When data are of a categorical nature, they can be presented in a contingency table. Grizzle, Starmer and Koch [3] (subsequently abbreviated GSK) have indicated how general weighted least squares can be applied to estimated proportions in complex contingency tables to test appropriate hypotheses. If the categorical data follow the product of several multinomial distributions, both linear and non-linear functions of the cell proportions can be estimated together with a corresponding covariance matrix. These functions can then be analyzed in terms of a linear regression model by weighted least squares. The resulting chi-square (χ^2) statistics used to test various hypotheses belong to the class of minimum modified chi-square statistics due to Neyman [7] which is equivalent to the general quadratic criteria of Wald [9]. (See appendix for a more detailed explanation).

3. GENERAL LINEAR MODEL APPROACH APPLIED TO THE SURVIVAL RATE

The GSK approach assumes that the product of several multinomial distributions is an appropriate underlying model for the data. To show that it can be

applied to analyze survival rates, we first construct a contingency table for each group category as follows:

	Exposure Year	Live	Die	Withdrawn or Lost	Total
1st year	0-1	n_{11}	n_{12}	n_{13}	$n_{1\cdot}$
2nd year	1-2	n_{21}	n_{22}	n_{23}	$n_{2\cdot}$
3rd year	2-3	n_{31}	n_{32}	n_{33}	$n_{3\cdot}$
4th year	3-4	n_{41}	n_{42}	n_{43}	$n_{4\cdot}$
5th year	4-5	n_{51}	n_{52}	n_{53}	$n_{5\cdot}$

Each year of exposure is treated separately and the number surviving the year is entered into the "live" column for that year. Entries are made similarly for the number dying and the number withdrawing or lost to follow-up. In this sense, an individual is classified into one of these three cells for each year of inclusion in the study. Thus, treating each year independently, we have that the data follow the product of several multinomial distributions. Consequently the cell probabilities are estimated by $p_{kj} = n_{kj}/n_{k\cdot}$ for the j^{th} column of the k^{th} year.

We now construct the necessary vectors and matrices required by the GSK method to produce a function of cell proportions which can be used to generate the estimate of the probability of survival suggested by Cutler and Ederer [1]. Hence, it will be assumed that all patients withdrawn or lost in a given year were exposed to the risk of dying, on the average, for half of that year. With this in mind, we have for each group category i:

$$\underset{\sim}{p}_{(i)} = \begin{pmatrix} p_{11i} \\ p_{12i} \\ p_{13i} \\ p_{21i} \\ \vdots \\ p_{52i} \\ p_{53i} \end{pmatrix}$$

where the p_{kji} are determined from the contingency table corresponding to the i^{th} group. In addition, we choose

$$\underset{\sim}{A}_{(i)} = \begin{pmatrix} \underset{\sim}{A}^* & 0 & 0 & 0 & 0 \\ 0 & \underset{\sim}{A}^* & 0 & 0 & 0 \\ 0 & 0 & \underset{\sim}{A}^* & 0 & 0 \\ 0 & 0 & 0 & \underset{\sim}{A}^* & 0 \\ 0 & 0 & 0 & 0 & \underset{\sim}{A}^* \end{pmatrix}$$

where

$$\underset{\sim}{A}^* = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 1 & 1 & \frac{1}{2} \end{pmatrix}$$

and

$$\underset{\sim}{K}_{(i)} = (1,-1,1,-1,1,-1,1,-1,1,-1)$$

Combining the matrices and vectors for all of the r groups we obtain

$$\underset{\sim}{p}' = (\underset{\sim}{p}'_{(1)}, \underset{\sim}{p}'_{(2)}, \ldots, \underset{\sim}{p}'_{(r)})$$

$$\underset{\sim}{A} = \begin{pmatrix} \underset{\sim}{A}_{(1)} & 0 & \cdots & 0 \\ 0 & \underset{\sim}{A}_{(2)} & & \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \underset{\sim}{A}_{(r)} \end{pmatrix}$$

$$\underset{\sim}{K} = \begin{pmatrix} \underset{\sim}{K}_{(1)} & 0 & \cdots & 0 \\ 0 & \underset{\sim}{K}_{(2)} & & \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \underset{\sim}{K}_{(r)} \end{pmatrix}$$

Now if $\underset{\sim}{F}(\underset{\sim}{p}) = \underset{\sim}{K} \underset{\sim\sim}{\log} \underset{\sim\sim}{Ap}$ as in GSK, we have

$$\underset{\sim}{F}(\underset{\sim}{p}) = \begin{pmatrix} \log \alpha_1 \\ \log \alpha_2 \\ \vdots \\ \log \alpha_r \end{pmatrix}$$

with estimated covariance matrix

$$\underset{\sim}{V} = \begin{pmatrix} \text{var}(\log \alpha_1) & 0 & \cdots & 0 \\ 0 & \text{var}(\log \alpha_2) & & \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \text{var}(\log \alpha_r) \end{pmatrix}$$

Here α_i is the same estimate of survival rate for the i^{th} category given in Cutler and Ederer. However, $\underset{\sim}{F}(\underset{\sim}{p})$ as derived in this discussion pertains actually

to the logarithms of the survival rates. To obtain the survival rate vector we transform $\underset{\sim}{F}(\underset{\sim}{p})$ by the exponential function e^x to get $\underset{\sim}{G}(\underset{\sim}{p})$. Corresponding to this transformation, we have a diagonal covariance matrix $\underset{\sim}{W}$ for $\underset{\sim}{G}(\underset{\sim}{p})$ with elements of the type

$$\text{var } \alpha_i = \alpha_i^2 \text{ var}(\log \alpha_i)$$

These calculations can be performed by matrix operations.

The researcher can choose to use either $\underset{\sim}{F}(\underset{\sim}{p})$ or $\underset{\sim}{G}(\underset{\sim}{p})$. There are advantages in an analysis using $\underset{\sim}{F}(\underset{\sim}{p})$ if one wants a model on a relative (or multiplicative) scale. In either choice, the results are asymptotically equivalent to one another.

As an example, let us consider the data given in the Cutler and Ederer paper. These data involve only one group and, as a result, only one contingency table.

YR	LIVE	DIE	WITHDRAWN OR LOST	TOTAL
0-1	60	47	19	126
1-2	38	5	17	60
2-3	21	2	15	38
3-4	10	2	9	21
4-5	4	0	6	10

The corresponding probability vector is

$$\underset{\sim}{p} = \begin{bmatrix} .48 \\ .37 \\ \vdots \\ .40 \\ .00 \\ .60 \end{bmatrix}$$

Accordingly, the following results are obtained.

$$\underset{\sim}{K}\ \underset{\sim\sim}{\log}\ \underset{\sim\sim}{Ap} = -.82$$

$$\alpha = e^{-.82} = .44$$

$$\text{Var}\ \ \alpha = \alpha^2 \cdot \text{Var}(\log \alpha)$$

$$= .003590$$

$$\text{s.e. of } \alpha = .060$$

These results are consistent with those of Cutler and Ederer. The advantage of this approach to survival rate analysis, however, is that this estimation process can be combined with a general method for testing the statistical significance of the effects of the various groups.

For example, only one group has been considered and the vector $\underset{\sim}{F}(\underset{\sim}{p})$ is a single number. On the other hand, if we had directed this analysis at the raw data corresponding to regional and localized cancers of the kidney and breast, the following results given by Cutler and Ederer would have been obtained.

$$\underset{\sim}{F}(\underset{\sim}{p}) = \begin{bmatrix} -.82 \\ -1.43 \\ -.43 \\ -.94 \end{bmatrix} = \begin{bmatrix} \log .44 \\ \log .24 \\ \log .65 \\ \log .39 \end{bmatrix}$$

$$\underset{\sim}{V} = \begin{pmatrix} .018570 & 0 & 0 & 0 \\ 0 & .085069 & 0 & 0 \\ 0 & 0 & .001044 & 0 \\ 0 & 0 & 0 & .002630 \end{pmatrix}$$

where $\underset{\sim}{V}$ is the covariance matrix.

4. TESTS OF GROUP FACTOR EFFECTS

After the $\underset{\sim}{F}(\underset{\sim}{p})$ (or $\underset{\sim}{G}(\underset{\sim}{p})$) vector has been obtained, a linear regression model of the form $\underset{\sim\sim}{X}\underset{\sim}{\beta}$, where $\underset{\sim}{X}$ is an appropriate (r×n) coefficient matrix and $\underset{\sim}{\beta}$ is a vector of n unknown parameters, is fitted to the estimated survival rates for the r groups by weighted least squares. Under the hypothesis that $\underset{\sim}{F}(\underset{\sim}{p})$ is characterized by the linear model $\underset{\sim\sim}{X}\underset{\sim}{\beta}$, the residual sum of squares is approximately Chi-square distributed with (r-n) degrees of freedom.

In addition to testing the goodness of fit for the linear model, the researcher can test appropriate hypotheses pertaining to the parameters in the $\underset{\sim}{\beta}$ vector.

As an example of this application, we use the results pertaining to the four cancer groups discussed in the previous section. For the resulting $\underset{\sim}{F}(\underset{\sim}{p})$ vector (i.e., the logarithms of the respective survival rates), let us consider the linear model

$$\underset{\sim\sim}{X}\underset{\sim}{\beta} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

where β_1 represents an overall average, β_2 is the effect due to different locations (kidney vs breast), and β_3 is the effect due to the different kinds of

infection (localized vs regional).

The lack of fit statistic is $SS(\underset{\sim}{F}(\underset{\sim}{p}) = \underset{\sim\sim}{X}\beta) = .085$ with one degree of freedom. This result supports the hypothesis that the model provides an adequate fit and hence that there is no interaction between type of infection and location. Estimates of the parameters are given by

$$\underset{\sim}{\beta} = \begin{pmatrix} -.89 \\ .20 \\ -.26 \end{pmatrix}$$

For the hypothesis of no effect due to location of disease we have

$$H_0: \ \underset{\sim\sim}{C}\beta = 0 \quad \text{where} \quad \underset{\sim}{C} = (0,1,0)$$

$SS(\underset{\sim\sim}{C}\beta = 0) = 10.36$ with D.F. = 1. Hence, after adjusting for type of infection the location has a significant effect on survival rate.

For the hypothesis of no difference in effect due to type of infection we choose $\underset{\sim}{C} = (0,0,1)$ and obtain $SS(\underset{\sim\sim}{C}\beta = 0) = 74.50$ with D.F. = 1. Thus, after adjusting for the location effect, type of infection has a significant effect on survival.

5. SUMMARY

By use of the methods developed by Grizzle, Starmer and Koch [3], we have illustrated an approach to estimation of survival rates for individual groups of patients. First of all, survival rates are determined for each group. The variation of the resulting estimates are then investigated in terms of linear models using the method of weighted least squares. The use of linear models provides a straightforward and unified approach to testing various hypotheses of interest. Moreover, computer programs which permit the efficient calculation of the corresponding estimates and test statistics are not difficult to prepare.

One such computer program may be obtained by writing the Program Librarian, Department of Biostatistics, University of North Carolina.

For those situations involving incomplete data in the sense that certain individuals have not been classified according to all variables, a generalization of this approach can be applied (see Reinfurt [9]).

Another problem of interest in the evaluation of clinical data is the estimation of trends in survival rates through time. Although comparisons of survival rates from non-overlapping time intervals are easily computed, several non-overlapping intervals are not always available. More efficient use of the data can be accomplished by first reconstructing the $\underset{\sim}{A}$ matrix described here to form appropriate overlapping intervals. Correlation patterns among such rates can be accounted for by proper use of the covariance matrix determined in the analysis. Further aspects of this problem will be presented in a future report.

6. BIBLIOGRAPHY

[1] Cutler, S. J. and Ederer, F. (1958). Maximum utilization of the life table method in analyzing survival, J. CHRON. DIS. 6:8, pp. 699-712.

[2] Greenwood, M. (1926). Reports on Public Health and Medical Subjects, No. 33, Appendix 1, The "Errors of Sampling" of the Survivorship Tables, London, H. M. Stationary Office.

[3] Grizzle, J. E., Starmer, C. F., and Koch, G. G. (1969). Analysis of categorical data by linear models. Biometrics, 25, pp. 489-504.

[4] Mantel, N. and Haenszel, W. (1959). Statistical aspects of analysis of data from retrospective studies of disease, J. NAT. CANCER INST., 22, pp. 719-48.

[5] Merrell, M. and Shulman, L. E. (1955). Determination of prognosis in chronic disease, illustrated by systematic lupus erythematosus, J. CHRON. DIS. 1:12, pp. 12-32.

[6] Meyers, M. H., Axtell, L. M., and Zelen, M. (1966). The use of prognostic factors in predicting survival for breast cancer patients, J. CHRON. DIS. 19, pp. 923-33.

[7] Neyman, J. (1949). Contributions to the theory of the χ^2 test. pp. 239-73 in Proc. Berkeley Symp. Math. Stat. Prob. University of California Press, Berkeley and Los Angeles.

[8] Reinfurt, D. W. (1970). The analysis of categorical data with supplemented margins including applications to mixed models. Unpublished Ph.D dissertation, North Carolina State University (Mimeo Series No. 697).

[9] Wald, A. (1943). Tests of statistical hypothesis concerning several parameters when the number of observations is large, Trans. Amer. Math. Soc. 54, pp. 426-82.

7. APPENDIX

In this section we present the theory required to analyze categorical data with survival as a response variable. In this general framework, a wide variety of problems analogous to those discussed earlier become a series of special cases which are easily handled. First, consider the hypothetical data shown in Table 1A. One may observe that, within a particular group, an individual is identified with one of five sub- populations according to number of years of exposure so that a total of 5r populations are defined. Moreover, the groups may be determined by several factors which combine to characterize a particular individual. Hence, a population may be in reality a combination of levels of treatment together with the year of exposure, and the analysis may be viewed in terms of traditional factorial experiments.

Table 1A. Categories of Response

	Exposure Year	Pop.	Survive	Die	Withdrawn or Lost	Total
Group 1	1	1	n_{11}	n_{12}	n_{13}	$n_{1\cdot}$
	2	2				
	3	3	.	.	.	.
	4	4	.	.	.	.
	5	5	n_{51}	n_{52}	n_{53}	$n_{5\cdot}$
Group 2	1	6	.	.	.	.
	2	7	.	.	.	.
	3	8	.	.	.	.
	4	9	.	.	.	.
	5	10	.	.	.	.
Group r	1	.	.	.	.	.
	2	.	.	.	.	.
	3	.	.	.	.	.
	4	.	.	.	.	.
	5	5r	$n_{5r,1}$	$n_{5r,2}$	$n_{5r,3}$	$n_{5r,\cdot}$

With this general setting fixed, define Π_{kj} to be the probability that an individual in the k^{th} row (i.e., k^{th} population) gives the j^{th} response.

Also, let

$$\underset{\sim}{\Pi}_k' = (\Pi_{k1}, \Pi_{k2}, \Pi_{k3});$$

$$p_{kj} = \frac{n_{kj}}{n_{k\cdot}} = \text{sample estimate of } \Pi_{kj};$$

$$\underset{\sim}{p}_k' = (p_{k1}, p_{k2}, p_{k3}).$$

If (n_{k1}, n_{k2}, n_{k3}) has the multinomial distribution with parameters $n_{k\cdot}$ and $\underset{\sim}{\Pi}_k$, then

$$\underset{\sim}{V}(\underset{\sim}{\Pi}_k) = \underset{\sim\sim}{\text{Var}}(\underset{\sim}{p}_k) = \frac{1}{n_{k\cdot}} \begin{bmatrix} \Pi_{k1}(1-\Pi_{k1}) & -\Pi_{k1}\Pi_{k2} & -\Pi_{k1}\Pi_{k3} \\ -\Pi_{k1}\Pi_{k2} & \Pi_{k2}(1-\Pi_{k2}) & -\Pi_{k2}\Pi_{k3} \\ -\Pi_{k1}\Pi_{k3} & -\Pi_{k2}\Pi_{k3} & \Pi_{k3}(1-\Pi_{k3}) \end{bmatrix}$$

On combining these quantities from the respective groups, we have

$$\underset{\sim}{\Pi}' = (\underset{\sim}{\Pi}_1', \underset{\sim}{\Pi}_2', \ldots, \underset{\sim}{\Pi}_{5r}')$$

$$\underset{\sim}{p}' = (\underset{\sim}{p}_1', \underset{\sim}{p}_2', \ldots, \underset{\sim}{p}_{5r}')$$

Here, $\underset{\sim}{p}$ is an unbiased estimate of $\underset{\sim}{\Pi}'$ and $\underset{\sim}{V}(\underset{\sim}{p}_k)$ where $\underset{\sim}{\Pi}_k$ has been replaced by $\underset{\sim}{p}_k$, is a consistent estimate of $\underset{\sim}{V}(\underset{\sim}{\Pi}_k)$. The variance of $\underset{\sim}{p}$ is estimated by $\underset{\sim}{V}(\underset{\sim}{p})$, a block diagonal matrix of the form

$$\underset{\sim}{V}(\underset{\sim}{p}) = \begin{bmatrix} \underset{\sim}{V}(\underset{\sim}{p}_1) & \cdots & \cdots & \cdots \\ \cdots & \underset{\sim}{V}(\underset{\sim}{p}_2) & \cdots & \cdots \\ \cdots & \cdots & \ddots & \cdots \\ \cdots & \cdots & \cdots & \underset{\sim}{V}(\underset{\sim}{p}_{5r}) \end{bmatrix}$$

For any function of the cell probabilities that can be estimated by $\underset{\sim}{F}(\underset{\sim}{p}) = \underset{\sim}{K}\ \underset{\sim\sim}{\log_e}(\underset{\sim}{A}\underset{\sim}{p})$ where $\underset{\sim}{K}$ and $\underset{\sim}{A}$ are known matrices, a consistent estimate of the

covariance matrix is given by $\underset{\sim}{S} = \underset{\sim\sim}{K}\underset{\sim\sim}{D}^{-1}\underset{\sim\sim}{A}\underset{\sim\sim}{V}(\underset{\sim}{p})^{-1}\underset{\sim}{A}'\underset{\sim}{D}^{-1}\underset{\sim}{K}'$ where

$$\underset{\sim}{D} = \begin{pmatrix} \underset{\sim}{a}_1'\underset{\sim}{p} & \cdots & \cdots & \cdots \\ \cdots & \underset{\sim}{a}_2'\underset{\sim}{p} & \cdots & \cdots \\ \cdots & \cdots & \ddots & \cdots \\ \cdots & \cdots & \underset{\sim}{a}_i'\underset{\sim}{p} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

where $\underset{\sim}{a}_i$ is i^{th} row of $\underset{\sim}{A}$. One should note that the natural logarithm is used in the transformation. By the proper choice of $\underset{\sim}{A}$ and $\underset{\sim}{K}$ the analyst may generate many different functions $\underset{\sim}{F}(\underset{\sim}{p})$ depending upon the structure of the particular data in a given situation.

The variation in the elements of the resulting $\underset{\sim}{F}(\underset{\sim}{p})$ vector is analyzed by fitting the linear regression model $\underset{\sim\sim}{X}\underset{\sim}{\beta}$ where $\underset{\sim}{X}$ is a known matrix of coefficients, sometimes called the design matrix, and $\underset{\sim}{\beta}$ is a vector of unknown parameters. The estimate $\underset{\sim}{b}$ of $\underset{\sim}{\beta}$ is obtained by minimizing $(\underset{\sim}{F}(\underset{\sim}{p})-\underset{\sim\sim}{X}\underset{\sim}{b})'\underset{\sim}{S}^{-1}(\underset{\sim}{F}(\underset{\sim}{p})-\underset{\sim\sim}{X}\underset{\sim}{b})$, and belongs to the class of best asymptotic normal (BAN) estimators. The resulting expression for $\underset{\sim}{b}$ is $\underset{\sim}{b} = (\underset{\sim}{X}'\underset{\sim}{S}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{S}^{-1}\underset{\sim}{F}(\underset{\sim}{p})$. If the data are adequately described by this model, a test of the hypothesis H_0: $\underset{\sim\sim}{C}\underset{\sim}{\beta}=\underset{\sim}{0}$, where $\underset{\sim}{C}$ is a coefficient matrix, which determines a comparison of interest, is produced by the conventional methods of weighted multiple regression. Specifically, the test statistic is $SS(\underset{\sim\sim}{C}\underset{\sim}{\beta}=\underset{\sim}{0}) = \underset{\sim}{b}'\underset{\sim}{C}'[\underset{\sim}{C}(\underset{\sim}{X}'\underset{\sim}{S}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{C}']^{-1}\underset{\sim\sim}{C}\underset{\sim}{b}$ which has approximately a Chi-square distribution in large samples. The degrees of freedom for this test is determined by subtracting the rank of the matrix $\underset{\sim}{C}$ from the rank of the matrix $\underset{\sim}{X}$; i.e., D.F. = Rank $(\underset{\sim}{X})$ - Rank $(\underset{\sim}{C})$. Finally, a goodness of fit test to determine the validity of the model is

$$SS(\underset{\sim}{F}(\underset{\sim}{\Pi})=\underset{\sim\sim}{X}\underset{\sim}{\beta}) = \underset{\sim}{F}'\underset{\sim}{S}^{-1}\underset{\sim}{F} - \underset{\sim}{b}'(\underset{\sim}{X}'\underset{\sim}{S}^{-1}\underset{\sim}{X})\underset{\sim}{b},$$

which under the hypothesis that the model fits, has approximately a Chi-square distribution in large samples with D.F. = (number of elements in $\underset{\sim}{F}(\underset{\sim}{p})$) - Rank $(\underset{\sim}{X})$.

We note that in the main part of this paper $\underset{\sim}{F}(\underset{\sim}{p})$ was formulated to give estimates of survival for each group category of patients. However, this method gives results applicable to a general class of functions of the vector $\underset{\sim}{p}$. For example, the logit function $\log_e \{\frac{p_{kj}}{1-p_{kj}}\}$ can be represented in matrix form with similar resulting least squares analysis.