2540

Survey Research Center
University of Michigan

DRAFT (11/60)
Sampling Section
I. Hess, K. K. Pillai

MULTIPLE CLASSIFICATION ANALYSIS

This paper summarizes some of the theoretical aspects of multiple classification analysis and its interpretation, with particular attention to the SRC's IBM 704 program.* We regard this draft as our first contribution to a paper which will deal more comprehensively with the subject and will be a joint product of the Economic Behavior Program, the Data Processing Section, and the Sampling Section.

1. INTRODUCTION

It is a common experience in the analysis of data from socio-economic surveys that the researcher will be facing the problem of predicting an individual's position on a given numerical dependent variable when he has knowledge about each individual's classification with respect to a number of independent characteristics believed to influence the dependent variable. Frequently the researcher develops a predictive model which utilizes the method of least squares to minimize the sum of squares of the errors of prediction. The multiple classification model is of this type, and we discuss the underlying assumptions in Section 2. Section 3 deals with the interpretation of the machine analysis with attention to questions that have been directed to us. Section 4 extends the discussion to analyses not immediately available from the IBM 704 program.

---

* The IBM 704 program, written by John Sonquist for the Economic Behavior Program, is an improved version of an IBM 650 program developed for Vernon Lippit at General Electric Company [6].

2. THE MODEL AND THE INITIAL ASSUMPTIONS

The model is

$$y_{ijk\ldots\alpha} = \bar{y} + a_i + b_j + c_k + d_l + \ldots + e_{ijk\ldots\alpha},$$

where $\bar{y}$ is the over-all mean

$$\sum_i \sum_j \sum_k \ldots \sum_\alpha w_{ijk\ldots\alpha}\ y_{ijk\ldots\alpha} \Big/ \sum_i \sum_j \sum_k \ldots \sum_\alpha w_{ijk\ldots\alpha}\ .$$

(See [7, pp. 91-96]). Here $w_{ijk\ldots\alpha}$ is the weight assigned to the αth element in the $ijk\ldots$ cell. If the sample is self-weighting, $w_{ijk\ldots\alpha} = 1$ and the sum is the number of sample elements in the analysis. The quantities a_i, b_j, c_k, etc., are effects of the various classes to which an individual belongs. The $e_{ijk\ldots}$ adjust for the error in $y'_{ijk\ldots}$ as an estimate of $y_{ijk\ldots}$,

where $y'_{ijk\ldots} = \bar{y} + a_i + b_j + c_k + d_l + \ldots$.

With the model we assume no interactions[1/] among the a_i's, b_j's, c_k's, etc; the $e_{ijk\ldots}$'s have an expectation of zero, a common variance σ^2, and are uncorrelated; the effects of the a_i's, b_j's, etc., are additive. There is no restriction of equality or proportionality on the class frequencies and no assumption of orthogonality.

Each individual is classified in p ways; he is assigned to one of the i classifications of characteristic a, one of the j classifications of characteristic b, etc., throughout p characteristics.

To illustrate, y_{ijkl} might be the income of a family with head in age class i, occupation j, education k, living in geographic region l.

If we knew absolutely nothing about a family, we would guess the income to be $\bar{y}$, but there might be substantial error in the prediction. However, knowing the specific cell to which the individual belongs, we can predict the family income with

---

1/ A variation of the model regards interaction as another characteristic. For example, to provide for interaction between characteristics a and b we may introduce the characteristic $(ab)_{ij}$, where i and j may apply in turn to all or to only some of the classes of a and b. The model is then

$$y_{ijk\ldots\alpha} = \bar{y} + a_i + b_j + (ab)_{ij} + \ldots + e_{ijk\ldots\alpha},$$

and we assume the interaction effects, together with the main effects, are additive.

error $e_{ijk\ldots}$, and the values of a_i, b_j, c_k, etc., will have been computed in such a manner that the error sum of squares is a minimum.

The analysis of multiple classification gives results consistent with those that would be obtained from a multivariate regression analysis in which each class is regarded as a separate characteristic with the property that it has the value 1 if an individual possesses this characteristic and 0 otherwise [5], [10]. The multiple classification analysis has some practical, operational virtues but no theoretical advantages over the multivariate regression analysis with "dummy" variables[2/]. It follows that the theoretical assumptions and interpretations must be identical for the two analyses.

3: INTERPRETING THE OUTPUT FROM THE MACHINE PROGRAM

The machine computes estimates of the parameters by an iterative process (see Appendix). These estimates we designate $\hat{a}_i$, $\hat{b}_j$, etc. Algebraically, the methods of matrix inversion and iteration are equivalent and differences between exact and approximate values are assumed negligible[3/].

Output from the program leads to the analysis of variance in Table 1.

2/ When using the multivariate regression analysis with dummy variables a common practice is to set equal to zero the regression coefficient for one class of each characteristic. As a consequence, computed coefficients are deviations from the omitted class of a characteristic, rather than deviations from the over-all mean as is the case with the multiple classification analysis. The two sets of coefficients are related by a linear transformation.

3/ The iterative solution is independent of the ordering of the characteristics and classes. That is, characteristics may be designated a, b, c, etc., in any convenient order. Conceivably one arrangement could require more iterations than another, but this consideration is unimportant.

Table 1--Analysis of Variance

Variation due to: (1)	Degree of freedom (2)	Sum of squares 4/ (3)	Mean square (col. 3/col. 2) (4)	Expected mean square (5)	F ratio (6)
$\hat{a}_i, \hat{b}_j, \hat{c}_k \ldots$	$r + s + t + \ldots - p$	$\Sigma y_{ai}\hat{a}_i + \Sigma y_{bj}\hat{b}_j + \ldots$	V	--	V/E
Remainder	$n - r - s - t \ldots + p - 1$	By subtraction	E	σ^2	--
Total	$n-1$	$\sum_i \sum_j \sum_k \ldots \sum_\alpha w_{ijk\ldots} y^2_{ijk\ldots\alpha}$ $- \frac{(\sum_i \sum_j \sum_k \ldots \sum_\alpha w_{ijk\ldots\alpha} y_{ijk\ldots\alpha})^2}{w}$	--	--	--

In the table p = number of characteristics

r, s, t, etc. = number of non-zero classes in characteristics a, b, c etc., respectively

n = number of elements in the sample

$y_{ai} = \sum_j \sum_k \sum_l \ldots w_{ijkl\ldots} y_{ijkl\ldots}$ = weighted total for the i-th subclass of a

$y_{bj} = \sum_i \sum_k \sum_l \ldots w_{ijkl\ldots} y_{ijkl\ldots}$ = weighted total for the j-th subclass of b

y_{ck} etc. are similar subtotals.

In order that the F test in column 6 be valid, it is necessary to add the assumption that the error terms, $e_{ijk\ldots}$, are normally distributed. Furthermore, if we wish to make valid

---

4/ The total sum of squares may be expressed as

$$\Sigma\Sigma\Sigma\ldots\Sigma\, w_{ijk\ldots}(y_{ijk\ldots} - \bar{y})^2 = \Sigma\Sigma\Sigma\ldots\Sigma\, w_{ijk\ldots}[(y_{ijk\ldots} - y'_{ijk\ldots}) + (y'_{ijk\ldots} - \bar{y})]^2.$$

An equivalent form of the right-hand member is

$\Sigma\Sigma\Sigma\ldots\Sigma\, w_{ijk\ldots} e^2_{ijk\ldots} + \Sigma\Sigma\Sigma\ldots\Sigma\, w_{ijk\ldots}(\hat{a}_i + \hat{b}_j + \hat{c}_k + \ldots)^2$, the cross-product term having vanished over the summation.

$\Sigma\Sigma\Sigma\ldots\Sigma\, w_{ijk\ldots} e^2_{ijk\ldots}$, the unexplained or error sum of squares, is not a convenient computational form and the numerical value is generally obtained by subtraction.

The explained sum of squares is easily calculated from the expression $\Sigma\, y_{ai}\hat{a}_i + \Sigma\, y_{bj}\hat{b}_j + \ldots$; since this is equivalent to a sum of squares, it is obvious that the quantity is always positive although individual terms or groups of terms in this computational form may be negative even for the summation over all classes of a characteristic.

inferences about a population represented by the data under analysis, it is necessary to assume that we have a simple random sample. To the extent that these conditions are not fulfilled, and to the extent that the iterative solution differ from the exact solution, the analysis of variance and the F test are approximate.

The researcher may view the computational form of the explained sum of squares (line 1, column 3 of Table 1) and be tempted to regard the summations by characteristic as the sum of squares attributable to the respective characteristics. THIS CANNOT BE. The machine program was written to print out these values to suggest lines of research rather than to serve immediate analytical purposes. The fact that some of the quantities may be negative precludes our interpreting the computational form as a partitioning of the explained sum of squares. In section 4 we discuss the calculation necessary to obtain the sum of squares attributable to one characteristic or to a subset of characteristics. Here it is important to recognize that such quantities are not products of this program.

Available from the program is the comparison of the explained sum of squares (in line 1 and column 3 of Table 1) with the total sum of squares (line 3, column 3 of the table),

R^2 = explained sum of squares/total sum of squares,

where R is the multiple correlation coefficient.[5/] Note that R^2 is the same as the "coefficient of joint determination" described in Ezekial [4, pp. 159, 217] and by Croxton and Cowden [3, pp 663, 774]. Even though this measure may be split into components--coefficients of separate determination--by taking ratios like $\Sigma \hat{a}_i y_i$/Total SS, $\Sigma \hat{b}_j y_j$/Total SS, etc.; each includes part of the joint determination of the other independent variables and some of these coefficients may be negative. This is due to nonorthogonality.

A significance test for R is identical with the single-tailed F test, in column 6 of Table 1, for the significance of the regression; appropriate degrees of freedom are (r + s + t +...- p) and (n - r - s - t...+ p-1). Note that the test may

---

5/ Although theoretically biased [8, p384], for all practical purposes, the R and R^2 can be taken as estimates of the corresponding population values.

not be exact because the F test assumes normality and simple random sampling, conditions which the data may not fulfill. Inasmuch as the effect on the F test is unknown, we suggest that the researcher require the value for significance to be somewhat larger than the tabular one.

An alternative is to obtain an estimate of the sampling variability[6] of R^2 by comparing the estimates from two half-samples (or from the whole sample and one half-sample). Although the technique is crude, estimates of R^2 and its sampling error would enable the researcher to judge the explanatory value of the set of characteristics and to make comparisons with other samples from the same universe.

The half-sample analyses (see preceding paragraph and Appendix) yield approximate sampling errors for the parameters ($\hat{a}_i$, $\hat{b}_j$, etc.) as well as for R^2. "Small" coefficients of variation (standard error of the estimate/estimate) and some stability from sample to sample will increase confidence in their predictive value. [The following is contributed by Dr. James N. Morgan of the Economic Behavior Program]

The "explained sum of squares" partitioned among the predictors is affected both by the dimensions of the dependent variable and by correlations among the predictors themselves. What we need is a measure of the relative importance of each predictor independent of the units of measurement of predictor or dependent variable, i.e., an analogue to the beta coefficient in multiple correlation. It is easy to compute beta coefficients for each class of each characteristic, but what we want is a measure of the importance of the whole set of classes of a characteristic taken as group, e.g., of age as a factor, not of belonging to one particular age group.

---

[6] For a simple random sample the variance of R^2 is approximated by

$$\text{variance } (R^2) \doteq \frac{4 R^2(1 - R^2)^2}{n} ,$$

where n is the sample size. However, the estimate breaks down for R near zero [8, Vol I, p 385].

Such a measure of importance is easy to compute by treating the adjusted coefficients as measured explanatory variables, and computing a multiple regression with these newly developed variables "age", etc. Each variable will have a regression coefficient b_i of 1.00.

The formula for beta is $\beta_i = \frac{b_i \sigma_{x_i}}{\sigma_y}$ where y is the dependent variable and x_i the predictor. We know the standard deviation of y from the 704 output, b_i is equal to 1, and the standard deviation of our artificial variable is $\sqrt{\frac{\sum w_i (x_i - \bar{x})^2}{\sum w_i}} = \sqrt{\frac{\sum w_i x_i^2}{\sum w_i}}$ since the weighted sum of the coefficients over any characteristic is zero.

The only difference between this and the standard beta coefficient is that it is always positive because all b_i's are positive (i.e., equal to 1).

* * * * * * * * * * * * * *

Although, at present, we cannot fully agree with the rationale of the above discussion leading to a β-measure, we feel that the expression $\left(\sqrt{\frac{\sum w_i x_i^2}{\sum w_i}} / \sigma_y \right)$ can serve as a very good measure of the "index of importance". Of course, we are looking at this as an "average effect" relative to the variability of the dependent variable; greater the index, the greater the importance of the characteristic. We would like to pursue the matter further.

4. Other Analyses

a. Testing effects of the characteristics

It is to be noted that the differences among the marginal (subclass) averages do not represent the "true" estimates of the parameter differences in the case of multiple classification with disproportionate subclass numbers. These reflect not only the effects of the particular classification but also include the residual effects of the other classifications. Thus the differences between income means represent not only all the effects of income differences but also include the fluctuations caused by the differences in occupations, age groups, etc., as a part of these differences; in other words, the main effects are nonorthogonal.

This nonorthogonality of the main effects implies that a "partition" of the total sum of squares into those exclusively attributable to the various individual characteristics and the "error" is impossible [7, pp 110-113]; that is to say that the "true" sum of squares of the characteristics will not add up to the total for such nonorthogonal data. Because of this result and the nonorthogonality of the main effects, the usual sum of squares attributed to a particular characteristic, say $\sum \hat{a}_i y_{a_i}$ for example, will not represent the true sum of squares attributable to that characteristic, a in the example. And the set of terms $\sum \hat{a}_i y_{a_i}$, $\sum \hat{b}_j y_{b_j}$, etc., are no more orthogonal. Note that one or more of these terms can even be negative! (Also note that this rules out any measure of homogeneity defined in terms of the components of variance.)

If the researcher is interested in finding out whether the various subclasses of a characteristic produce any differential effect on the independent variable, it is necessary to rerun the machine program deleting the characteristic in which the experimenter is interested.-/ If the new solutions are denoted by $\hat{a}_{io}$, $\hat{b}_{jo}$, etc., then the appropriate sum of squares to test our hypothesis will be given by the difference between the first and second "explained" sum of squares (see Table 2) namely $(\sum_i \hat{a}_i y_{a_i} + \sum_j \hat{b}_j y_{b_j} + \ldots) - (\sum \hat{a}_{io} y_{a_i} + \sum \hat{b}_{jo} y_{b_j} + \ldots)$. Note that the second set will not

-/ In this event convergence might be achieved in fewer iterations if the previous solutions are used as first approximations. However, the technique is impractical if its use involves changes in the machine program.

contain a term corresponding to the characteristic for which we are testing. This difference will have r-1 degrees of freedom if the characteristic has r subclasses. The appropriate test will be the single-tailed F test, the F-ratio being defined in column 5 of Table 2. (Remember that the sum of squares are approximate since we are using the approximate solutions obtained by iteration.)

Table 2--Analysis of Variance

Variation due to	Degrees of freedom	Sum of Squares	Mean square	F-ratio
(1)	(2)	(3)	(4)	(5)
b, c, etc., (ignoring a)	s + t +... -(p-1)	$\sum \hat{b}_{jo} y_{b_j} + \sum \hat{c}_{ko} y_{c_k} +..$	--	--
a	r-1	(By subtraction)	V_a	V_a/E
a, b, c, etc.	r + s + t +.. - p	$\sum \hat{a}_j y_{a_i} + \sum \hat{b}_j y_{b_j} +..$	--	--
Error	n - (r + s + t..-p)-1	(By subtraction)	E	--
Total	n-1	$\sum w_{ijk} y^2_{ijk} - \frac{(\sum w_{ijk} y_{ijk})^2}{\sum w_{ijk}}$	--	--

If the sum of squares attributable to characteristics a and b is desired, the machine program is again repeated, this time excluding the two characteristics. The analysis of variance table again has five lines:

Variation due to
c, etc. (ignoring a and b)
a and b
a, b, c, etc.
Error
Total

It is not valid to subtract the sum of squares due to a (from Table 2) from the sum of squares due to a and b, and then attribute the remainder to characteristic b. The latter is obtained only by repeating the analysis with characteristic b excluded.

At this point it is worth considering some alternative methods even though they are approximate. If the various subclasses have roughly equal numbers of observa-

tions (or at least reasonable to believe this would be true in the population from which the sample is drawn), a simple approximate method would be that of the method of unweighted means--the method of calculation being to ignore the subclass numbers and apply the ordinary analysis of variance to the means of the subclasses, treating them as single observations [9, pp 287-288, 294]. If the data conform, more or less, to the assumption stated above, it is advisable to perform this simple and easy analysis of variance to get an approximate idea of what the results would look like. (The error mean square is calculated using the whole data, subtracting between cell sum of squares from the total sum of squares and finally dividing it by the harmonic mean of the cell frequencies.)

An alternate, approximate procedure is that of expected subclass numbers; the technique to be used if the subclass numbers are quite different and this seems to be a characteristic of the population [9, p 295]. Here the observed numbers and totals of the cells are replaced by the "expected" numbers and "expected" totals based on these numbers and the observed means. Then analyse the data which now have proportionate subclass numbers. This method has been found useful and gives close approximation to the "exact" method of fitting constants. [9, pp 291-292, 299-300]

The method of weighted squares of means is yet another approximate method suited in cases where the marginal means are assumed representative of some populations. The reciprocals of the numbers in each cell are computed and the weights for subclass means (and their squares) are the reciprocals of the sums; i.e., the weight for each subclass means will be the harmonic mean of the numbers in that class, in other classifications. This method leads to a test of significance of the differences among the subclass effects.

If one is interested in comparison between subclasses of a characteristic the problem is more complicated. However, approximate procedures may be adopted to achieve these comparisons. A simple but crude method would be to apply an approximate t-test on the estimates obtained from the iterative process; the variance being approximately calculated by using the subclass numbers. (These will be in general underestimates and

so may result in rejection more often than is indicated by the significance level of the t-test.) The other approximate analysis of variance methods suggested above provide also convenient comparison methods corresponding to them, the methods being in line with the standard analysis of variance and individual comparisons.

b. Check on additivity assumption

It has been tacitly assumed in the above procedures that the additive model holds good for the data at hand. In many situations confronted in socio-economic surveys, this may be far from reality. So it would be advisable to first of all have a rough check of the validity of the additive model by a careful scrutiny of the data combined with accumulated experience in the behavior of characteristics under study. One need not be much alarmed if he finds that the assumption of additivity does not suit his data. In such cases it would be possible to find some suitable transformation so that the transformed variable satisfies the additivity assumption. Such transformations have the additional effect of making the distribution more nearly normal than before and thereby validating tests of significance as well as estimation. The most useful type of transformations that occur commonly are the logarithmic and the square root transformations [2, pp 39-52].

Analysis of the residuals offers a belated check on the additivity assumption. Marked departures from normality may result from a break-down of this assumption [5, p 372 ff]. It is then necessary to search for an appropriate transformation.

Appendix

These notes were prepared as an aid to writing the machine program. Although the notation is in terms of three characteristics, the program provides for a maximum combination of characteristics and classes. The normal equations and several steps in the iterative solution are included.

It is necessary to specify a convergence test to determine when a solution has been achieved. The researcher may choose either of two tests (where K and K + 1 denote the next to the last and the last iterations, respectively):

1. $\left| \frac{a_i^{K+1} - a_i^K}{a_i^{K+1}} \right| \leq Q$; $\left| \frac{b_j^{K+1} - b_j^K}{b_j^{K+1}} \right| \leq Q$ etc.

2. $\left| a_i^{K+1} - a_i^K \right| \leq Q \bar{Y}$; $\left| b_j^{K+1} - b_j^K \right| \leq Q \bar{Y}$

He must also specify the value of Q and the maximum number of iterations to be completed if convergence is not achieved in K + 1 or fewer iterations. (For some studies convergence has been achieved in around 10 iterations using test 2 and Q = .001.) Results of the last two iterations are available routinely and all may be had when specified in advance.

Necessary conditions for a solution are $\sum_i w_i a_i^{K+1} = \sum_j w_j b_j^{K+1} = \sum_k w_k c_k^{K+1} = 0$. If these quantities do not appear as part of the program output, the researcher should verify that the conditions have been met. Possibly this verfication may be added to the program in the near future.

The researcher may choose to run the analysis for either or both of two properly selected half-samples as well as for the entire sample.✓

The output format is identical for the whole sample and the half-samples. The number of non-zero classes, not necessarily known in advance, can be determined from the output.

✓ Possibly the program may be adjusted, at a later date, to repeat the calculations for several half-samples.

Notation

Model: $Y_{ijk\alpha} = \bar{Y} + a_i + b_j + c_k + e_{ijk\alpha}$ where

$Y_{ijk\alpha}$ = observed value of variable y for individual α in the ijk-th cell

$\bar{Y}$ = the over-all weighted mean = $\dfrac{\sum_i \sum_j \sum_k \sum_\alpha W_{ijk\alpha} Y_{ijk\alpha}}{\sum_i \sum_j \sum_k \sum_\alpha W_{ijk\alpha}} = \dfrac{Y}{W}$

$W_{ijk\alpha}$ = weight for observation $Y_{ijk\alpha}$ (= 1 if data are unweighted)

a_i = "effect" due to i-th subclass of characteristic a

b_j = "effect" due to j-th subclass of characteristic b

c_k = "effect" due to k-th subclass of characteristic c

$e_{ijk\alpha}$ = "error" (random fluctuation of observed value from predicted)

$$Y_i = \sum_j \sum_k \sum_\alpha W_{ijk\alpha} Y_{ijk\alpha} \quad ; \quad Y_j = \sum_i \sum_k \sum_\alpha W_{ijk\alpha} Y_{ijk\alpha} \quad ; \quad Y_k = \sum_i \sum_j \sum_\alpha W_{ijk\alpha} Y_{ijk\alpha}$$

$$W_{ij} = \sum_k \sum_\alpha W_{ijk\alpha} \quad ; \quad W_{jk} = \sum_i \sum_\alpha W_{ijk\alpha} \quad ; \quad W_{ik} = \sum_j \sum_\alpha W_{ijk\alpha}$$

$$W_i = \sum_j W_{ij} = \sum_k W_{ik} \quad ; \quad W_j = \sum_i W_{ij} = \sum_k W_{jk} \quad ; \quad W_k = \sum_i W_{ik} = \sum_j W_{jk}$$

$A_i = \dfrac{\sum_j \sum_k \sum_\alpha W_{ijk\alpha} Y_{ijk\alpha}}{\sum_j \sum_k \sum_\alpha W_{ijk\alpha}} = \dfrac{Y_i}{W_i} = \bar{Y}_{Ai}$, the mean of i-th subclass of a.

Similarly, $B_j = \dfrac{Y_j}{W_j} = \bar{Y}_{Bj}$ and $C_k = \dfrac{Y_k}{W_k} = \bar{Y}_{Ck}$ represent means of subclasses of b and c respectively.

$$Y = \sum_i \sum_j \sum_k \sum_\alpha W_{ijk\alpha} Y_{ijk\alpha} = \sum_i Y_i = \sum_j Y_j = \sum_k Y_k$$

$$W = \sum_i \sum_j \sum_k \sum_\alpha W_{ijk\alpha} = \sum_i W_i = \sum_j W_j = \sum_k W_k$$

Note $\bar{Y} = \dfrac{Y}{W} = \dfrac{\sum_i W_i A_i}{\sum_i W_i} = \dfrac{\sum_j W_j B_j}{\sum_j W_j} = \dfrac{\sum_k W_k C_k}{\sum_k W_k}$

$\sum_i \sum_j \sum_k \sum_\alpha W_{ijk\alpha} Y^2_{ijk\alpha} - W\bar{Y}^2$ = Total sum of squares.

[In all the expressions above, α will run through the values 1 to n_{ijk}, n_{ijk} being the number of observations in the ijk-th cell; i through 1 to r, the number of subclasses of a; j through 1 to s, the number of b subclasses; and k through 1 to t, the number of c subclasses.]

Normal equations to be solved:

$$a_i = A_i - \bar{Y} - \frac{1}{W_i} \sum_j W_{ij} b_j - \frac{1}{W_i} \sum_k W_{ik} c_k$$

$$b_j = B_j - \bar{Y} - \frac{1}{W_j} \sum_i W_{ij} a_i - \frac{1}{W_j} \sum_k W_{jk} c_k$$

$$c_k = C_k - \bar{Y} - \frac{1}{W_k} \sum_i W_{ik} a_i - \frac{1}{W_k} \sum_j W_{jk} b_j$$

Solution

Step 1. $a'_i = A_i - \bar{Y}$; $b'_j = B_j - \bar{Y}$; $c'_k = C_k - \bar{Y}$

Step 2.

$$a''_i = a'_i - \frac{1}{W_i} \sum_j W_{ij} b'_j - \frac{1}{W_i} \sum_k W_{ik} c'_k$$

$$b''_j = b'_j - \frac{1}{W_j} \sum_i W_{ij} a''_i - \frac{1}{W_j} \sum_k W_{jk} c'_k$$

$$c''_k = c'_k - \frac{1}{W_k} \sum_i W_{ik} a''_i - \frac{1}{W_k} \sum_j W_{jk} b''_j$$

Step K.

$$a_i^K = a'_i - \frac{1}{W_i} \sum_j W_{ij} b_j^{K-1} - \frac{1}{W_i} \sum_k W_{ik} c_k^{K-1}$$

$$b_j^K = b'_j - \frac{1}{W_j} \sum_i W_{ij} a_i^K - \frac{1}{W_j} \sum_k W_{jk} c_k^{K-1}$$

$$c_k^K = c'_k - \frac{1}{W_k} \sum_i W_{ik} a_i^K - \frac{1}{W_k} \sum_j W_{jk} b_j^K$$

Note: Superscripts denote primes NOT powers.

Step K + 1.

$$a_i^{K+1} = a'_i - \frac{1}{W_i} \sum_j W_{ij} b_j^K - \frac{1}{W_i} \sum_k W_{ik} c_k^K$$

$$b_j^{K+1} = b'_j - \frac{1}{W_j} \sum_i W_{ij} a_i^{K+1} - \frac{1}{W_j} \sum_k W_{jk} c_k^K$$

$$c_k^{K+1} = c'_k - \frac{1}{W_k} \sum_i W_{ik} a_i^{K+1} - \frac{1}{W_k} \sum_j W_{jk} b_j^{K+1}$$

Basic Output of IBM 704 Machine Program

1. $\bar{Y}$

2. $N = \sum_i \sum_j \sum_k \sum_\alpha N_{ijk\alpha}$

3. $W = \sum_i \sum_j \sum_k \sum_\alpha W_{ijk\alpha}$

4. $a_i' = A_i - \bar{Y}$, $b_j' = B_j - \bar{Y}$, $c_k' = C_k - \bar{Y}$

5. a_i^{K+1} , b_j^{K+1} , c_k^{K+1} (solutions to normal equations)

6. $K + 1$ = number of iterations or steps

7. $\sum_i \sum_j \sum_k \sum_\alpha W_{ijk\alpha} Y^2_{ijk\alpha} - W \bar{Y}^2$ = total sum of squares = TSS

8. $SSA = \sum_i Y_i a_i^{K+1}$, $SSB = \sum_j Y_j b_j^{K+1}$, $SSC = \sum_k Y_k c_k$

9. ESS = SSA + SSB + SSC = explained sum of squares

10. W_i , W_j , W_k

[Note: The machine prints out many other results best described in detail by the Data Processing Section.]

REFERENCES

1. Anderson, R. L. and Bancroft, T. A., "Statistical Theory in Research" McGraw-Hill Book Company, Inc., New York, 1952.

2. Biometrics, Vol. 3, No. 1, March 1947.

3. Croxton, Frederick E. and Cowden, Dudley J., "Applied General Statistics," Prentice-Hall, Inc., New York, 1940.

4. Ezekiel, Mordecai, "Methods of Correlation Analysis," Second Edition, John Wiley and Sons, Inc., New York, 1947.

5. Hill, T. P., "An Analysis of the Distribution of Wages and Salaries in Great Britain," Econometrica, Vol. 27, 3(July 1959).

6. Lippitt, Vernon, "Determinants of Consumer Demand for House Furnishings and Equipment," Harvard University Press, 1959.

7. Kempthorne, Oscar, "The Design and Analysis of Experiments," John Wiley and Sons, Inc., New York, 1952.

8. Kendall, Maurice G., "The Advanced Theory of Statistics," Vols. I and II, 5th edition, Hafner Publishing Company, New York, 1952.

9. Snedecor, G. W., "Statistical Methods," 4th edition, The Iowa State College Press, Ames, Iowa, 1948.

10. Suits, D. B., "Use of Dummy Variables in Regression Equations," Journal of the American Statistical Association, Dec. 1957.