Analysis of Thematic Map Classification Error Matrices

George H. Rosenfield
U. S. Geological Survey, National Mapping Division, Reston, VA 22092

ABSTRACT: The classification error matrix expresses the counts of agreement and disagreement between the classified categories and their verification. Thematic mapping experiments compare variables such as multiple photointerpretation or scales of mapping, and produce one or more classification error matrices. Analysis of categorical data by linear models can analyze a set of one or more classification error matrices in a multivariate sense by utilizing all of the cell frequencies in the matrix, not just the diagonal elements. This paper presents a tutorial to implement a typical problem of a remotely sensed data experiment for solution by the linear model method. The results of the application analysis indicate that two independent photointerpreters are interpreting mapped categories in the same manner at the 95 percent probability level; and at the 59 percent probability level, the categories of oak and cottonwood cannot be separated by interpretation.

INTRODUCTION

THE CLASSIFICATION ERROR MATRIX expresses the counts of agreements and disagreements between the classified categories and their verification. An accuracy evaluation of a thematic map will usually have as its product a single classification error matrix. A thematic mapping experiment to compare such variables as multiple photointerpreters, scales of mapping, algorithms, or instruments for classification will produce several classification error matrices. In the past, analysis of the results of such experiments have been restricted to the data contained in the cells of the main diagonal of these matrices, representing the number of sample points in each category which have been correctly interpreted. Analysis of variance was recently applied by Rosenfield (1981), comparing the diagonal elements of the classification error matrices resulting from a test of three scales of land-use and land-cover mapping.

Such an analysis neglects the non-diagonal cell data which represent the errors by commission in the classification, and the errors of omission from the verification. A beginning has already been made to use the entire classification error matrix in studies of such experiments (Congalton, 1980) using methods of discrete multivariate analysis (Bishop *et al.*, 1975). Rosenfield (1981) suggested investigating other techniques of multivariate analysis of variance for analyzing the entire matrix. Chrisman (1982) also calls for this approach and discusses an example.

BACKGROUND

A general and unified approach to the analysis of categorical data by linear models was published by Grizzle *et al*. (1969). The procedure is tested by the linearized modified chi-square statistic, and ". . . has the advantages of giving the analyst more latitude in choosing models and testing hypotheses which are precisely tailored to specific data." (Grizzle *et al.*, 1969, p. 489). The procedure is based on a weighted least-squares adjustment. The 1969 paper contains an example which is extendable to making comparisons between more than one classification error matrix, as in thematic mapping experiments. A more recent version of the 1969 paper was published by Koch *et al.* (1977). This recent paper extends the mathematical and statistical basis of weighted least-squares model fitting and of hypothesis testing to additional analysis and functional relationships. "The resulting methodology represents a categorical data analogue to more well-known counterparts for quantitative data, like multivariate analysis of variance," (Kock *et al*., 1977, p. 135). The 1977 paper tests hypotheses on the basis of the first-order marginal probabilities, and provides an adjusted estimate of the relative effects associated with the respective categories and variables being studied.

The FUNCAT procedure, for functions of categorical responses as a linear model, of the Statistical Analysis System (Ray, 1982) performs the necessary calculations and hypothesis testing.

PURPOSE AND SCOPE

The purpose of this paper is to present a tutorial to implement a typical problem of a remotely sensed data experiment for solution by the linear model method. An example will use two classification error matrices of two photointerpreters, obtained from the paper by Congalton and Mead (1983). These data are used as the example because the publication is current, the data set is small, yet the experimental design is involved enough to illustrate all the

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 52, No. 5, May 1986, pp. 681–686.

0099-1112/86/5202-189$02.25/0

hypotheses desired. The data are illustrated in Table 1.

The linear model method of analysis will test the following three hypotheses: first, that there is no significant difference between the two photointerpreters; second, that there is no significant difference in classification among all categories; and third, that there is no significant difference in the interpretation between any pair of categories.

Starting with the data set of cell frequencies within the classification error matrix, expected cell probabilities are estimated by dividing each cell frequency value by its respective row sum. Within each row of the matrix, a general logit function is formed in two steps: (1) by the division of each cell probability value by the value in the last cell of the row (except for the last cell in the row), and (2) by the subtraction of the log of the denominator from the log of the numerator. The relation of the generalized logit functions to variables that were used to form the table is examined by means of the use of a design matrix. A vector of parameters expresses the relation of the generalized logits to the variables of interest. The vector of unknown parameters for the example is described, and the design matrix for the example is developed by matrix partitioning. Contrast matrices are developed for testing certain hypotheses about the problem. Development of these matrices is needed for entry into the FUNCAT computer program in order to solve the problem and test the hypotheses. Lastly, the results of the hypotheses tests for the example are given.

LOG LINEAR ANALYSIS

The data for the experimental design of the classification error matrices of the photointerpreters 1 and 2 used in this research are contained in Table 1. In Table 1, the rows represent the interpreted categories and the columns represent the verification (or ground truth). The column data are considered to be the responses in the linear model sense. The data may be considered as the form of a split-plot experimental design (Steel and Torrie, 1960, p. 236), as in Table 1.

THE VECTOR OF ESTIMATED CELL PROBABILITIES, p

Let the observed cell frequency of the jth response for the kth photointerpreter-ith interpreted category be

$$f_o = n_{kij}\cdot$$

The observed probability for each cell is then

$$p_{kij} = n_{kij}/n_{ki\cdot}$$

where the dot notation represents summation over all values of the index.

For the set of cell probabilities for both photointerpreters,

$$\underset{(32 \times 1)}{p'} = \underset{(16 \times 1)}{[p_1' \quad p_2']}\cdot$$

For the set of cell probabilities for all i interpreted categories for each photointerpreter,

$$\underset{(16 \times 1)}{p_k'} = \underset{(4 \times 1)}{[p_{k1}' \quad p_{k2}' \quad p_{k3}' \quad p_{k4}']}\cdot$$

For the set of cell probabilities for all j responses for each photointerpreter-interpreted category combination,

$$\underset{(4 \times 1)}{p_{ki}'} = \underset{(1 \times 1)}{[p_{ki1} \quad p_{ki2} \quad p_{ki3} \quad p_{ki4}]}\cdot$$

RATIONALE FOR LOG LINEAR MODELS

The classification error matrix may contain cells with large disparity in the numbers; that is, some cells contain very large frequencies and others very small frequencies. If a linear function is used for such a table, then the sums of the probabilities predicted from the model (summed by rows and colums) may exceed the probability limits of 0 and 1.

A dichotomous variable which has the probabilities p and q (q = 1-p) may be expressed in terms of

TABLE 1. TWO CLASSIFICATION ERROR MATRICES OF TWO PHOTOINTERPRETERS, FROM CONGALTON AND MEAD (1983), P. 72.

			Verification (responses)				
	PI	CA	1	2	3	4	Total	Category
Interpretation	1	1	35	14	11	1	61	pine
		2	4	11	3		18	cedar
		3	12	9	38	4	63	oak
		4	2	5	12	2	21	cottonwood
	2	1	32	15	5	3	55	pine
		2	7	8	5		20	cedar
		3	7	8	38	2	55	oak
		4	6	7	15	1	29	cottonwood

PI = photointerpreter
CA = interpreted category

its logit,

$$x = \ln(p/q),$$

and when $p = 0$, $x = -\infty$; and when $p = 1$, $x = \infty$. Because division by zero is not defined, the logit can be readily expressed in linearized form by its logs,

$$x = \ln(p) - \ln(q).$$

Thus, if the table frequencies represent a large disparity, the sums of the probability values can be greater than unity or less than zero. Then the linear function of the logs will accomodate values between the limits of plus and minus infinity.

The Logit Function of the Cell Probabilities

Define $\pi_{i1}, \pi_{i2}, \ldots, \pi_{i4}$ to be the expected probabilities of observing the $j = 1, 2, \ldots, r$ ($r = 4$) responses, respectively, for each interpreted category of each photointerpreter (Grizzle *et al.*, 1969, p. 499).

Write a generalized logit form of the log function of these probabilities for the *i*th interpreted category for the *k*th photointerpreter as

$$l_{ki1} = \ln(\pi_{i1}/\pi_{i4})_k \, , \ldots , \, l_{ki3} = \ln(\pi_{i3}/\pi_{i4})_k$$

where $i = 1, 2, \ldots, s$ ($s = 4$) for the four interpreted categories, and $k = 1, 2$ for the two photointerpreters.

If the logit functions $l_{ki1}, l_{ki2}, l_{ki3}$ can be considered as additive functions of the mean effect, the photointerpreter effect, and the interpreted category effect, there is no interaction. Then, given the additive model, tests can be made on the photointerpreter and interpreted category effects (Grizzle *et al.*, 1969, pp. 499–500).

TREATMENT FOR ZERO ELEMENTS OF THE MATRICES

The problem of handling studies with large numbers of cells with zero cell frequencies has been considered by several authors. Upton (1978, p. 65) reports the recommendation for the addition of 0.5 to every cell frequency before fitting the saturated model so that the problem resulting from the log 0 $= -\infty$ is obviated. Upton also recommends adding 0.5 when fitting the saturated model irrespective of whether there are zero cell frequencies and refers to Gart and Zweifel (1967) and to Plackett (1974, Chapter 1) for certain desirable features. Both Plackett (1974, p. 3) and Gart and Zweifel (1967, pp. 181–182) indicate that adding the term 0.5 reduces bias in the log or logit estimate to the second order, and in its variance to the third order.

According to Kock *et al.*, (1977, p. 157), for the case where the number of multivariate response profiles $r = L^d$ is large, L and d each greater than 3, d = number of measurement conditions (categories), L = number of response levels (columns), the properties of the rule to add 0.5 to the zero cell frequencies are largely unknown. They recommend to replace only these zero cell frequencies by 0.5 which are necessary to the construction of a non-singular covariance matrix.

The Matrix Form of the Linear Equations

We set the matrix form of the set of linear equations for the additive model as

$$\underset{(24 \times 1)}{\mathbf{F}(\pi)} = \mathbf{L} = \underset{(24 \times 15)}{\mathbf{X}} \; \underset{(15 \times 1)}{\boldsymbol{\beta}} \, ,$$

where $\boldsymbol{\beta}$ is the vector of unknown parameters, and $\mathbf{X}$ is the design matrix.

Then for the set of $j = 1, 2, 3$ logit functions,

$$\underset{(15 \times 1)}{\boldsymbol{\beta}'} = \underset{(5 \times 1)}{[\boldsymbol{\beta}_1' \;\; \boldsymbol{\beta}_2' \;\; \boldsymbol{\beta}_3']}$$

where, for the set of parameters for each logit function,

$$\underset{(5 \times 1)}{\boldsymbol{\beta}_j'} = [u_j \; \alpha_j \; \beta_{1j} \; \beta_{2j} \; \beta_{3j}]$$

in which

u_j = general mean for each logit function,
α_j = differential effect of photointerpreter 1,
β_{1j} = differential effect of interpreted category 1,
β_{2j} = differential effect of interpreted category 2, and
β_{3j} = differential effect of interpreted category 3.

The second photointerpreter effect $= -\alpha_1$; the fourth interpreted category effect $= -\Sigma\, \beta_{ij}$'s, $i = 1, 2, 3$; and similarly for all remaining logit functions. This process is a reparameterization of the photointerpreter differential effect, and of the interpreted category differential effect.

The Linear Operator Matrices $\mathbf{A}_1$ and $\mathbf{A}_2$

The linear operator matrices are used to summarize succinctly the functional relationship of the probabilities to the photointerpreter and interpreted category variables. The family of functions of the linear model for logarithmic relationships has the form (Grizzle and others, 1969, p. 492):

$$\underset{(24 \times 1)}{\mathbf{F}(\pi)} = \underset{(24 \times 32)}{\mathbf{A}_2} \; \ln \; \underset{(32 \times 32)}{\mathbf{A}_1} \; \underset{(32 \times 1)}{(\boldsymbol{\pi})}$$

(see also Koch *et al.* (1977), p. 139 and 155).

The vector $\boldsymbol{\pi}$ represents the vector of expected probabilities of all observations, and $\ln \mathbf{A}_1(\boldsymbol{\pi})$ is the vector of logarithms of the elements of $\boldsymbol{\pi}$ with $\mathbf{A}_1 = \mathbf{I}$ (Grizzle *et al.*, 1969, p. 500).

A generalized logit is defined as a linear combination of the logarithms as

$$\ln(\pi_{ij}/\pi_{ir}) = \ln(\pi_{ij}) - \ln(\pi_{ir})$$

The $\mathbf{A}_2$ matrix is needed to express this relationship.

The linear operator matrix, $\mathbf{A}_2$:

for each photointerpreter group, k: $k = 1, 2, \ldots, t$; $t = 2$

$$\underset{(24 \times 32)}{\mathbf{A}_2} = \begin{bmatrix} \underset{(12 \times 16)}{(\mathbf{A}_2)_1} & \\ & (\mathbf{A}_2)_2 \end{bmatrix}$$

for photointerpreter group k, for each interpreted category, i: $i = 1, 2, \ldots, s$; $s = 4$

$$\underset{(12 \times 16)}{(\mathbf{A}_2)_k} = \begin{bmatrix} \underset{(3 \times 4)}{(\mathbf{A}_2)_{k1}} & & & \\ & (\mathbf{A}_2)_{k2} & & \\ & & \ddots & \\ & & & (\mathbf{A}_2)_{k4} \end{bmatrix}$$

for photointerpreter-interpreted category combination, ki, for each log linear equation: $j = 1, 2, \ldots, r - 1$; $r = 4$

$$\underset{(3 \times 4)}{(\mathbf{A}_2)_{ki}} = \begin{bmatrix} (\mathbf{A}_2)_{ki1} \\ (\mathbf{A}_2)_{ki2} \\ \underset{(1 \times 4)}{(\mathbf{A}_2)_{ki3}} \end{bmatrix}$$

where

$$(\mathbf{A}_2)_{ki1} = [1\ 0\ 0\ -1]$$

$$(\mathbf{A}_2)_{ki2} = [0\ 1\ 0\ -1]$$

$$\underset{(1 \times 4)}{(\mathbf{A}_2)_{ki3}} = [0\ 0\ 1\ -1]$$

The Design Matrix, **X**

The design matrix, **X**, shown in Figure 1, represents the linear equation coefficient matrix in full parameterized form. The matrix is blocked by sets of three rows. The first set represents the three logit functions for the first subpopulation, being a set of photointerpreter-k, interpreted category-i effects. The following two sets represent through the photointerpreter-1, interpreted category-3 effect. The fourth set of three rows represent the photointerpreter-1, interpreted category-4 effect. This last effect is the negative sum of the other category effects, on the basis that the sum of the parameters for category equals 0. The entire set is repeated for the photointerpreter-2 effects. In this last case, the photointerpreter-2 effect is the negative of the photointerpreter-1 effect, on the basis that the sum of the parameters for photointerpreter equals zero.

The design matrix, **X**:

Category	logit			
	1	1 1 1 0 0		
1	2		1 1 1 0 0	
	3			1 1 1 0 0
	1	1 1 0 1 0		
2	2		1 1 0 1 0	
	3			1 1 0 1 0
	1	1 1 0 0 1		
3	2		1 1 0 0 1	
	3			1 1 0 0 1
	1	1 1 -1 -1 -1		
4	2		1 1 -1 -1 -1	
	3			1 1 -1 -1 -1
	1	1 -1 1 0 0		
1	2		1 -1 1 0 0	
	3			1 -1 1 0 0
	1	1 -1 0 1 0		
2	2		1 -1 0 1 0	
	3			1 -1 0 1 0
	1	1 -1 0 0 1		
3	2		1 -1 0 0 1	
	3			1 -1 0 0 1
	1	1 -1 -1 -1 -1		
4	2		1 -1 -1 -1 -1	
	3			1 -1 -1 -1 -1

Fig. 1. Design matrix for two photointerpreters ($t = 2$), four categories ($s = 4$), and four logit equations ($r = 4$).

for each photointerpreter group, k: $k = 1, 2, \ldots, t$; $t = 2$

$$\underset{(24 \times 15)}{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 \\ \underset{(12 \times 15)}{\mathbf{X}_2} \end{bmatrix}$$

for photointerpreter group k, for each interpreted category, i: $i = 1, 2, \ldots, s$; $s = 4$

$$\underset{(12 \times 15)}{\mathbf{X}_k} = \begin{bmatrix} \mathbf{X}_{k1} \\ \mathbf{X}_{k2} \\ \mathbf{X}_{k3} \\ \underset{(3 \times 15)}{\mathbf{X}_{k4}} \end{bmatrix}$$

for photointerpreter-interpreted category combination, ki, for each

logit function, j: $j = 1, 2, \ldots, r - 1$; $r = 4$

$$\underset{(3 \times 15)}{\mathbf{X}_{ki}} = \begin{bmatrix} \underset{(1 \times 5)}{\mathbf{X}_{kij}} \otimes \mathbf{I}_3 \end{bmatrix}$$

where for each photointerpreter-interpreted category-logit j combination:

$$\begin{aligned} \mathbf{X}_{11j} &= [1 \quad 1 \quad 1 \quad 0 \quad 0] \\ \mathbf{X}_{12j} &= [1 \quad 1 \quad 0 \quad 1 \quad 0] \\ \mathbf{X}_{13j} &= [1 \quad 1 \quad 0 \quad 0 \quad 1] \\ \mathbf{X}_{14j} &= [1 \quad 1 \quad -1 \quad -1 \quad -1] \\ \mathbf{X}_{21j} &= [1 \quad -1 \quad 1 \quad 0 \quad 0] \\ \mathbf{X}_{22j} &= [1 \quad -1 \quad 0 \quad 1 \quad 0] \\ \mathbf{X}_{23j} &= [1 \quad -1 \quad 0 \quad 0 \quad 1] \\ \mathbf{X}_{24j} &= [1 \quad -1 \quad -1 \quad -1 \quad -1] \end{aligned}$$

Tests of Hypotheses

Test statistics are obtained from the linear model by using weighted least-squares as a computational algorithm, where the solution vector **b** is a Best Asymptotic Normal estimator of $\boldsymbol{\beta}$ (Grizzle *et al.*, 1969, p. 491). An appropriate test statistic for the goodness of fit of the model is then

$$\mathbf{Q} = (\mathbf{F} - \mathbf{Xb})' \, \mathbf{V}_F^{-1} \, (\mathbf{F} - \mathbf{Xb}),$$

which corresponds to the sum of the squares in normal regression. The value **Q** is approximately chi-square distributed if the sample sizes are large enough for the Central Limit Theorem to hold (not necessarily so for the smaller sample sizes in this example).

If the model fit is adequate, tests of linear hypotheses pertaining to the parameter $\boldsymbol{\beta}$ can then be performed. For a general hypothesis in the form:

$$\text{Ho: } \mathbf{C}\,\boldsymbol{\beta} = \mathbf{0}$$

C is a known matrix called a "contrast" matrix, and **0** is a $c \times 1$ vector of zeros. A suitable test statistic is then

$$\begin{aligned} Q_c &= (\mathbf{Cb})'[\mathbf{C}(\mathbf{X}'\ \mathbf{V}_F^{-1}\ \mathbf{X})^{-1}\ \mathbf{C}']^{-1}\ \mathbf{Cb} \\ &= (\mathbf{Cb})'\ [\mathbf{CN}^{-1}\ \mathbf{C}']^{-1}\ \mathbf{Cb} \end{aligned}$$

which has a chi-square distribution with $df = c$ in large samples under the general hypothesis Ho (Koch *et al.*, 1977, pp. 153–154). The hypotheses to be tested are

- The hypotheses of no differences between the photointerpreters based on the marginal distributions (Koch *et al.*, 1977, p. 136).
- The hypothesis of no differences among the categories based on the marginal distributions (Koch and others, 1977, p. 136).
- The hypothesis of no difference in the responses for the pairs of categories (Landis and others, 1976, p. 38).

The Contrast Matrices, C

To test the effect of photointerpreter on the set of j logit functions

$$\underset{(3 \times 15)}{\mathbf{C}} = \begin{bmatrix} \underset{(1 \times 5)}{\mathbf{C}_1} & & \\ & \mathbf{C}_2 & \\ & & \mathbf{C}_3 \end{bmatrix}$$

for the set of parameters for each logit function:

$$\underset{(1 \times 5)}{\mathbf{C}_j} = [0\ 1\ 0\ 0\ 0]$$

which yields χ^2 with three degrees of freedom for the photointerpreter group effect.

To test for effect of interpreted category on the set of j logit functions:

$$\underset{(9 \times 15)}{\mathbf{C}} = \begin{bmatrix} \underset{(3 \times 5)}{\mathbf{C}_1} & & \\ & \mathbf{C}_2 & \\ & & \mathbf{C}_3 \end{bmatrix}_{\text{P}}$$

for the set of parameters for each logit function:

$$\underset{(3 \times 5)}{\mathbf{C}_j} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

which yields χ^2 with nine degrees of freedom for the interpreted category effect.

To test for the pairwise effect of interpreted category on the set of j logit functions simultaneously:

$$\underset{(3 \times 15)}{\mathbf{C}} = \begin{bmatrix} \underset{(1 \times 5)}{\mathbf{C}_1} & & \\ & \mathbf{C}_2 & \\ & & \mathbf{C}_3 \end{bmatrix}$$

for the set of parameters for each interpreted category pair:

category pair	$\mathbf{C}_j$
1–2	[0 0 1 −1 0]
1–3	[0 0 1 0 −1]
1–4	[0 0 2 1 1]
2–3	[0 0 0 1 −1]
2–4	[0 0 1 2 1]
3–4	[0 0 1 1 2]

which yields χ^2 with three degrees of freedom for the paired effect.

The simplest SAS program for the FUNCAT procedure for this example is

```
PROC FUNCAT;
WEIGHT COUNT;
MODEL C = A B/ADDCELL = .5;
CONTRAST '1–2' @1 B 1 −1 0, @2 B 1 −1 0, @3
  B 1 −1 0;
CONTRAST '1–3' @1 B 1 0 −1, @2 B 1 0 −1, @3
  B 1 0 −1;
CONTRAST '1–4' @1 B 2 1 1, @2 B 2 1 1, @3 B
  2 1 1;
CONTRAST '2–3' @1 B 0 1 −1, @2 B 0 1 −1, @3
  B 0 1 −1;
CONTRAST '2–4' @1 B 1 2 1, @2 B 1 2 1, @3 B
  1 2 1;
CONTRAST '3–4' @1 B 1 1 2, @2 B 1 1 2, @3 B
  1 1 2;
```

where

A = photointerpreter;
B = interpreted categories; and
CONTRAST '1–2', etc., represents the hypothesis test for the pair of categories.

RESULTS AND DISCUSSION

The data from Table 1 was analyzed using the FUNCAT procedure, and the hypothesis tests were made as indicated by the various contrast matrices. Each of the three hypotheses were tested, and the chi-square values, degrees of freedom, and probability values were computed. The results of these hypothesis tests are given in Table 2.

The hypothesis on photointerpreters is not significant at $\alpha = 0.05$, indicating that there is no evidence to reject the hypothesis of no difference between the two photointerpreters. The conclusion is that both photointerpreters are interpreting the categories in the same manner. This same conclu-

TABLE 2. RESULTS OF HYPOTHESIS TESTS FOR DIFFERENCE BETWEEN PHOTOINTERPRETER, DIFFERENCES AMONG ALL CATEGORIES, AND DIFFERENCES BETWEEN PAIRS OF CATEGORIES.

H_o	χ^2	*df*	*P*
Photointerpreter	0.1128	3	0.9903
categories	78.3803	9	0.0*
Category Pair	χ^2	*df*	*P*
pine—cedar	10.52	3	0.0146*
pine—oak	54.28	3	0.0001*
pine—cottonwood	26.15	3	0.0001*
cedar—oak	23.82	3	0.0001*
cedar—cottonwood	11.32	3	0.0101*
oak—cottonwood	2.87	3	0.4124

*Significant at $\alpha = 0.05$.

sion, as expected, was reached by Congalton *et al.* (1983) using comparison of the two values for Cohen's coefficient of agreement.

The hypothesis on categories is significant at $\alpha = 0.05$, indicating that the hypothesis of no differences among all the categories is rejected. The conclusion is that at least two of the categories are being interpreted differently.

This conclusion is what one would have logically expected. The purpose of interpretation is to differentiate between categories. But this conclusion tells us nothing about individual pairs of categories. It is hoped that all pairs of categories are being interpreted differently. It is the purpose of multiple comparison tests to determine which pair of categories are not being separately interpreted. That is the next step.

The purpose of the hypothesis for category pairs is to apply a multiple comparisons test to all pairs of interpreted categories to determine which pairs, if any, are not being interpreted differently. The null hypothesis is that the two categories of any pair are being interpreted the same. If two categories of a pair are being interpreted differently, then the null hypothesis will be rejected. We note in Table 2 that only the category pair oak-cottonwood is not significant, and that there is no evidence to reject the hypothesis that the categories are being interpreted the same. The conclusion is that oak and cottonwood are being misinterpreted in the same manner at the 59 percent probability level. This means that something is wrong in this experiment relative to separating oak and cottonwood. Possibly their reflectance is too similar in this environment to be differentiated. Or maybe the sensor does not have the sensitivity to differentiate oak and cottonwood in this environment. In any event, confining oak and cottonwood into one class would increase the accuracy of the classification.

REFERENCES

Bishop, Y. M. M., S. E. Feinberg, and P. W. Holland, 1975. *Discrete Multivariate Analysis: Theory and Practice*: Cambridge, Massachusettes, The MIT Press.

Chrisman, N. R., 1982. Beyond accuracy assessment: correction of misclassification: *Proceedings International Society for Photogrammetry and Remote Sensing Commission IV*, pp. 123–132.

Congalton, R. G., 1980. Statistical techniques for analysis of landsat classification accuracy: Paper presented at meeting of the American Society of Photogrammetry, St. Louis, Missouri; 11 March.

Congalton, R. G., and R. A. Mead, 1983. A quantitative method to test for consistency and correctness in photointerpretation: *Photogrammetric Engineering and Remote Sensing*, Vol. 49, No. 1, pp. 69–74.

Gart, J. J., and J. R. Zweifel, 1967. On the bias of various estimates of the logit and its variance: *Biometrika*, Vol. 54, No. 1 and 2, pp. 181–187.

Grizzle, J. E., C. F. Starmer, and G. G. Koch, 1969. Analysis of categorized data by linear models, *Biometrics*, Vol. 25, pp. 489–504.

Koch, G. G., J. R. Landis, J. L. Freeman, D. H. Freeman, Jr., and R. G. Lehnan, 1977. A general methodology for the analysis of experiments with repeated measurement of categorical data, *Biometrics*, Vol. 33, pp. 133–158.

Plackett, R. L., 1974. *The Analysis of Categorical Data*: London, Griffin.

Ray, Alice A. (Ed), and others, 1982. *SAS User Guide, Statistics*: SAS Institute Incorporated, Raleigh, North Carolina.

Rosenfield, G. H., 1981. Analysis of variance of thematic mapping experiment data: *Photogrammetric Engineering and Remote Sensing*, Vol. 47, No. 12, pp. 1685–1692.

Steel, R. G. D., and J. H. Torrie, 1960. *Principles and Procedures of Statistics*: New York, McGraw-Hill Book Company, Inc.

Upton, G. J. G., 1978. *The Analysis of Cross-Tabulated Data*: John Wiley & Sons, New York.

(Received 7 April 1984; accepted 18 October 1985; revised 11 February 1986)