UNITED STATES
DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY

# ANALYZING THEMATIC MAPS AND MAPPING FOR ACCURACY

By George H. Rosenfield

---

Open File Report 82-239

Reston, Virginia
1982

## PREFACE

This paper concerns certain aspects of applied statistics which can be used for analyzing thematic maps and mapping for accuracy. The concepts for this application have been developed over the past several years and have been documented in a series of papers—most of which have been recently published (Rosenfield, 1981; Rosenfield, 1982; Rosenfield and Melly, 1980; and Rosenfield and others, 1982). This paper brings together these separate ideas into a cohesive body. The thoughts of this paper have been presented during the Symposium on Machine Processing of Remotely Sensed Data at Purdue University/LARS, June 3-6, 1980, at the conference dealing with techniques for assessing Landsat classification accuracy at the EROS Data Center, November 12-14, 1980 (Mead and Szajgin, 1982), at the March 1982 meeting of Mid-Atlantic Division, Association of American Geographers, George Mason University, Fairfax, Va., and at the meeting of URISA, August 22-25, 1982, Minneapolis, Minn.

Contents

Page

Abstract .......................................................... 1
Introduction ...................................................... 2
1. Sample design .................................................. 3
1.1 Sample distribution ........................................... 3
1.2 Sample size ................................................... 4
1.3 Sampling procedure ............................................ 5
1.4 Minimum sample size for a category ............................ 6
1.5 Size of the sample unit ....................................... 7
2. Testing accuracy for a category ................................ 7
3. Accuracy of a thematic map ..................................... 8
3.1 Overall accuracy based on total sample points ................. 9
3.2 Overall accuracy based on stratified sampling ................. 10
3.3 Results of accuracy checking .................................. 11
4. Comparing factors in thematic mapping experiments .............. 11
4.1 Two-way analysis of variance without replication .............. 12
4.2 Multiple comparisons tests .................................... 13
References ........................................................ 14

# ANALYZING THEMATIC MAPS AND MAPPING FOR ACCURACY

By George H. Rosenfield

---


## ABSTRACT

Two problems which exist while attempting to test the accuracy of thematic maps and mapping are: (1) evaluating the accuracy of thematic content, and (2) evaluating the effects of the variables on thematic mapping. Statistical analysis techniques are applicable to both these problems and include techniques for sampling the data and determining their accuracy. In addition, techniques for hypothesis testing, or inferential statistics, are used when comparing the effects of variables.

A comprehensive and valid accuracy test of a classification project, such as thematic mapping from remotely sensed data, includes the following components of statistical analysis: (1) sample design, including the sample distribution, sample size, size of the sample unit, and sampling procedure; and (2) accuracy estimation, including estimation of the variance and confidence limits. Careful consideration must be given to the minimum sample size necessary to validate the accuracy of a given classification category.

The results of an accuracy test are presented in a contingency table sometimes called a classification error matrix. Usually the rows represent the interpretation, and the columns represent the verification. The diagonal elements represent the correct classifications. The remaining elements of the rows represent errors by commission, and the remaining elements of the columns represent the errors of omission.

For tests of hypothesis that compare variables, the general practice has been to use only the diagonal elements from several related classification error matrices. These data are arranged in the form of another contingency table. The columns of the table represent the different variables being compared, such as different scales of mapping. The rows represent the blocking characteristics, such as the various categories of classification. The values in the cells of the tables might be the

counts of correct classification or the binomial proportions of these counts divided by either the row totals or the column totals from the original classification error matrices.

In hypothesis testing, when the results of tests of multiple sample cases prove to be significant, some form of statistical test must be used to separate any results that differ significantly from the others.

In the past, many analyses of the data in this error matrix were made by comparing the relative magnitudes of the percentage of correct classifications, for either individual categories, the entire map or both. More rigorous analyses have used data transformations and (or) two-way classification analysis of variance. A more sophisticated step of data analysis techniques would be to use the entire classification error matrices using the methods of discrete multivariate analysis or of multiviariate analysis of variance.

## INTRODUCTION

Two statistical problems which occur in analyzing thematic maps and mapping are: (1) evaluating the accuracy of thematic content, and (2) evaluating the effects of the variables on thematic mapping. Statistical procedures which apply to both these problems include techniques for sampling the data and determining their accuracy. In addition, techniques for hypothesis testing, or inferential statistics, are used when comparing the effects of variables.

A comprehensive and valid accuracy test of a classification project, such as thematic mapping from remotely sensed data, includes the following components of statistical analyses: (1) sample design which covers sampling within the population, and includes consideration of the sample distribution, determination of the sample size, size of the sample unit, the sampling procedure, and estimation of the population means, totals, variances, and confidence limits from the sample information; and (2) accuracy estimation which includes estimation of the variance and confidence limits. Careful consideration must be given to the minimum sample size necessary to validate the accuracy of a given classification category.

# 1. SAMPLE DESIGN

## 1.1 Sample Distribution

If a given measurement (y) of a variable (such as yield of a crop) is made at each sample unit, and if repeated random samples of a certain size (n) are taken from any population, the frequency distribution of the sample mean (y) tends to become normal as n increases. The central limit theorem (Snedecor and Cochran, 1967, p. 51) allows the normal distribution to be used with sample means, regardless of the frequency distribution of the original population. If the area belonging to a particular category is measured at each sample unit, then the area measurement can be considered equivalent to the variate (y) selected by simple random sampling from a normal distribution.

If the category of the sample unit is considered as being in agreement or disagreement with some particular category, and if a value of 1 is assigned to signify agreement and a value of 0 to signify disagreement, then the population of 1's and 0's would be described by the binomial distribution. In some cases, the binomial distribution would apply even when the observations are nominal data of particular category rather than count data of 1 or 0 if the unit of interest is the proportion of given category to all units sampled (Cochran, 1977, p. 50).

When a number of categories are involved (other than two classes such as agreement or disagreement), the probability distribution of the proportions in each category is that of the multinomial distribution. Cochran (1977, p. 60) explains that the multinomial distribution is the appropriate extension of the binomial distribution and is a good approximation to the probability of drawing the observed sample, if the sample size (n) is small in relation to the total number of units (A) in the category. In addition, Cochran (1977, p. 55 and 60) points out that the binomial and multinomial distributions of the proportions (p) and ($p_j$) respectively for each category are special cases of the hypergeometric distribution, which does not require that the population be large in relation to the sample. The hypergeometric distribution is based on the actual sample counts for each category (<u>a</u>), and the sample size (n), rather than on the respective proportions ($p_j$) for each category.

Cochran (1977, p. 64) also points out that the probabilities and methods of the binomial and multinomial distributions are not valid if each sample unit is a cluster of elements and if an estimate of the proportion of elements that fall in each category is desired. Note that the formulations for cluster sampling are similar to those for simple random sampling from the normal distribution, with the proportions (p) substituted for the random variables (y).

Thus, the difference between using the probabilities of the binomial or multinomial distributions versus those of the normal distribution for the proportions (p), is a function of the size and use of the sample unit. For point samples, the sample proportions are taken as belonging to the binomial or multinomial distributions. For cluster samples, the distribution of sample proportions can usually be approximated by the normal distribution.

## 1.2 Sample Size

Statistical formulas exist for determining sample size for sampling populations within each of the above distributions: the normal, binomial, and multinomial distributions (for example: Mace, 1974). Snedecor and Cochran (1967, p. 59) summarize that the parameters to be considered for estimating sample size are an upper limit to the amount of error that can be tolerated in the estimate, the desired probability that the estimate will lie within this error, and an a priori estimate of the population standard deviation.

Several formulas are applicable for determining sample size when the random variables are considered as having been sampled from a population having the normal distribution:

(1) When it is desired to define the true mean within certain limits of error, the concept of confidence intervals for normal means is used.

(2) When it is desired to define the true variance within certain limits of error, the concept of confidence interval for normal variances is used.

(3) When it is desired to estimate the standard deviation as a percent of its true value, the concept of using the half length of the confidence interval about the true standard deviation as a percentage of the true standard deviation is used (Greenwood and Sandomire, 1950). This procedure was simplified into chart form by Natrella (1963).

When the sample variates belong to the binomial distribution, and it is desired to define the true mean within certain limits of error, then the concept of confidence interval for proportions is used to determine the sample size.

Tortora (1978) gives a method and example for estimating the sample size for multinomial proportions based on the approximate large sample equations for the simultaneous confidence limits. The end points for the simultaneous confidence intervals for the category proportions ($p_j$), with the joint confidence coefficient being approximately $1-\alpha$, are reported by Johnson and Kotz (1969, p. 289).

## 1.3 Sampling Procedure

Berry (1962) used the stratified systematic unaligned sampling procedure to select samples in similar type studies and recommends this procedure (Berry and Baker, 1968, p. 91-100) for use in accuracy testing of the land use and land cover maps produced by the U.S. Geological Survey. Cochran (1977, p. 227-228) discusses systematic sampling in two dimensions, that it has been found that the square grid had about the same precision as simple random sampling in two dimensions, and that the unaligned pattern within the square grid will often be superior to both a systematic pattern within the square grid and to stratified random sampling. He cites (p. 221) Matern (1947) as proposing this function as a model for the natural populations for forestry and land use surveys.

Systematic sampling distributes the sample units equitably over the entire region of interest, and may be treated as if it were random provided that systematic effects in the population are made ineffective by the sampling (Freund and Williams, 1972, p. 416).

The same sample selection can be applied directly to the entire mapped region. If the region is large, a selection is first made of quadrangles or blocks of areas within the larger region, and then within those quadrangles or blocks, smaller regions are selected. This technique is called subsampling, or two-stage sampling.

## 1.4 Minimum Sample Size for a Category

Van Genderen and others (1978) report that researchers faced with the problem of adequately representing important minor categories on thematic maps have tended to use some form of stratified sampling rather than strictly random sampling without fully describing their methods for selecting sample sizes. Experience, however, (Fitzpatrick-Lins, 1980) has shown that the stratified systematic unaligned sampling technique is clearly area weighted (proportional allocation). That is, most of the sample points are selected in those categories that cover most of the map area, and the fewest points are from categories that cover the least area. Some small polygons in sparse categories might not be sampled at all. Sampling a limited number of points for some categories would give a poor estimate of the overall accuracy of the map. A sample of the map, with an adequate sample size from each category, is needed to represent all categories and to adequately evaluate the overall accuracy of the thematic map.

The theoretical development of the method to determine the minimum sample size to validate the accuracy of any category is based on the cumulative binomial distribution. Let p be the probability that a certain category in a thematic map was interpreted correctly. (Such a probability (p) will likely vary from category to category.) For example, if "the minimum level of interpretation accuracy in the identification of land use and land cover categories from remote sensor data should be at least

85 percent" (Anderson and others, 1976, p. 5), then $p \geq 0.85$. Solution for n is effected for given values of $p_0$, $\alpha$, and E (Rosenfield and others, 1982).

Given the preliminary estimate ($p_0$) for the expected accuracy of each category, a computer program based on the solution for the cumulative binomial probability can be used to calculate the minimum sample size n for each category. For example, for those categories with the assumed $p_0 \geq 85$ percent, the minimum sample size n would be 19 with E = 10 percent, with 95 percent confidence (Rosenfield and others, 1982).

### 1.5 Size of the Sample Unit

Cochran (1977, p. 233) states the well known principal that in cluster sampling, the size of the sample unit must be selected to give the smallest variance for a given cost or the smallest cost for a given variance. Cochran (1977, p. 243-244) discusses variance functions for surveys such as soil sampling which utilize an area sample unit. The problem of finding the optimum size unit requires predicting the variance between units in the population as a function of the unit size by the analysis of variance.

## 2. TESTING ACCURACY FOR A CATEGORY

The theoretical development of the method to determine whether a given category meets the expected accuracy value for the number of points in the sample, with specified confidence, is based on the cumulative binomial distribution. The critical level is defined as one less than the minimum number of points which must be correctly interpreted from any given sample, in order to accept the hypothesis at a given significance level that the category is interpreted within the tolerance for the specified accuracy. When the number of correctly interpreted sample points for the category is larger than the critical level, for a given sample size, the category accuracy equals or exceeds the expected limit with some predetermined probability (Rosenfield and others, 1982).

The critical level determination can be made based upon tests of two different hypotheses. The first hypothesis to test is that the category accuracy equals or exceeds an expected value, e.g. 85 percent, at some predetermined probability level, e.g. $\alpha = 0.05$. If the number correct does not exceed the critical level, find the largest integer c (the critical level), and reject the hypothesis at the 5-percent level. For example, for the sample size of 45, reject the hypothesis of 85 percent accuracy at the 5-percent level when the number correct does not exceed 33 (Rosenfield and others, 1982).

The second hypothesis to test is that the category accuracy is less than an expected value, e.g. 85 percent, at some predetermined probability level, e.g. $\alpha = 0.05$. If the number correct exceeds the critical level, find the smallest integer c (the critical level), and reject the hypothesis at the 5-percent level. For example, for the sample size of 45, we will reject the hypothesis of accuracy less than 85 percent at the 5-percent level when the number correct exceeds 43. The critical level values for test two are similar to those determined by Ginevan (1979) in developing his tables of optimum sample size (Rosenfield and others, 1982).

## 3. ACCURACY OF A THEMATIC MAP

The accuracy of a thematic map has been a very complex issue both in definition and measurement. For a polygon to be correctly interpreted, both its boundary and its classification must be correct. The thematic map accuracy can be expressed as either: (1) the probability ($E_1$) that any randomly selected point on the map is classified correctly (expressed as a percentage of area) or (2) by the probability ($E_2$) that any polygon in the map is classified correctly (expressed as a percentage of the total number of polygons). The probability $E_1$ is $A_1/A$--the ratio of the area with correct interpretation to the total area of the map (using point data as a surrogate for area data). The probability $E_2$ is R/N—the ratio of the number of polygons with correct interpretation to the total number of polygons in the map (Rosenfield and others, 1982).

These two accuracy measures have their advantages and disadvantages, depending on the users' needs. For some particular maps, such as those having one predominant category with an intermixture of several small polygons in different categories, these two accuracy measures differ greatly. Therefore, a carefully planned sampling procedure must be adopted to incorporate the advantages of both, before conducting an independent evaluation.

To simplify quantifying the accuracy of area data, point data are used as a surrogate for area data. The total number of points selected would be representative of the total area of the map, and the number of points correct would be representative of the area of correct interpretation. The analysis is then based on point data rather than on area data. The resulting percentage of correctly classified points is an estimate for the accuracy of the map as a whole, but not for the accuracy of the classification of individual categories. Not every polygon need be sampled, but every category should have the necessary number of points sampled to provide a reliable estimate of accuracy. A composite of these two measures of accuracy can provide a reliable estimate of map classification accuracy.

## 3.1 Overall Accuracy Based on Total Sample Points

In the classical method of estimating the accuracy of a thematic map, the overall accuracy is the ratio of the number of correctly interpreted sample points to the total number of sample points, expressed as a percentage. This ratio can be derived from an area-weighted sampling technique such as some form of stratified systematic sampling in two dimensions.

## 3.2 Overall Accuracy Based on Stratified Sampling

The sample obtained by the area-weighted method, the stratified systematic unaligned sampling procedure, may be augmented with additional sample points in the sparse categories so that there is at least the desired minimum number of points in each category. This amounts to sampling in two frames (A and B) where frame A is the area of the map, and frame B is the list of interior points for each category. Because the selected sample now has additional points for the sparse categories, the sample is no longer area weighted. An accuracy value computed as the simple average would give undue weight to these sparse categories in an overall value for the map. It is therefore necessary to weight the individual category accuracies by the proportion of its area in order to again achieve an area-weighted overall accuracy value.

For an area-weighted accuracy estimate, the weight $W_h$ is the ratio of the area of the $h_{th}$ category to the total area of the map. Furthermore, this single accuracy value is not, in itself, sufficient. The accuracy and confidence limits for each category in the map should also be reported as a table in the marginal information on the map.

Since the sample size for the entire map is large, the estimated mean will be approximately normally distributed, and the 95-percent confidence limits about the accuracy value can be computed in the normal manner.

When evaluating the accuracy value for an individual category the confidence limits about that accuracy value should be considered, based upon the number of points in the respective sample. The confidence limits indicate the interval which contains the true percentage in the sampled population, with some pre-established confidence value. Hord and Brooner (1976) have a table showing the upper and lower 95-percent confidence limits for sample sizes of 50 to 400 (in steps of 50), and for accuracy values from 80 percent to 100 percent, although they neglect to apply the "correction for continuity" (Snedecor and Cochran, 1967, p. 209-213).

For sample sizes in excess of 30, confidence limits which include the correction for continuity may be computed from the normal approximation (Snedecor and Cochran, 1967, p. 31, and p. 210-211).

For sample sizes of 30 or less, exact confidence limits are computed from the binomial distribution. A table for the 95 percent and 99 percent confidence interval for the binomial distribution is given in Snedecor and Cochran (1967, p. 6 and 7).

### 3.3 Results of Accuracy Checking

The results of the accuracy test of the observations are given in the form of a row by column (RxC) contingency table sometimes called a classification error matrix. Usually the rows represent the interpretation and the columns represent the verification. The diagonal elements represent the correct classifications. The remaining elements of the rows represent the errors by commission (row-wise for the erroneous classification commited by the interpreter). The remaining elements of the columns represent the errors of omission (column-wise for the true classifications that were omitted by the interpreter).

## 4. COMPARING FACTORS IN THEMATIC MAPPING EXPERIMENTS

Thematic mapping experiments are conducted to evaluate different variables that operate simultaneously to affect classification by thematic categories. Such variables might include three scales of aerial photographs, two types of images, several algorithms or equipment of digital classification, different physiographic regions, etc. Data obtained from thematic mapping experiments can be evaluated by the analysis of variance method, which has been defined by Scheffé (1959, p. 3) as a "statistical technique for analyzing measurements depending on several kinds of effects operating simultaneously, to decide which kinds of effects are important and to estimate the effects."

For tests of hypothesis that compare variables, the general practice has been to use only the diagonal elements from several related classification error matrices. These data are arranged in the form of another contingency table. The columns of the table represent the different variables being compared, such as different scales of mapping. The rows represent the blocking characteristics, such as the various categories of classification. The values in the cells of the tables might be the counts of correct classification or the binomial proportions of these counts divided by either the row totals or the column totals from the original classification error matrices.

In the past, many analyses of the data in this error matrix were made by comparing the relative magnitudes of the percentage of correct classifications, for either individual categories, the entire map or both. More rigorous analyses have used data transformations and (or) two-way classification analysis of variance.

## 4.1 Two-way Analysis of Variance Without Replication

Two-way analysis of variance provides an efficient procedure for comparing two or more sets of data. For the case in which there are only two sets without replication the t-test for comparing sets is equivalent to a one-way analysis of variance. However, two-way analysis of variance provides a measure of the variance component among the rows of data. Also, for more than two sets, the analysis of variance is more efficient than the t-test, and has a variance reducing effect.

Many computer programs are available that can perform an analysis of variance. The two-way analysis of variance without replication is described by Sokal and Rohlf (1969, p. 299). The hypothesis to be tested is that the several samples came from the same population. If the alternate hypothesis is that the population means are not equal, then the two-tailed test is applied. If the hypothesis is accepted, it may be concluded that the populations under study for that variable are not significantly different.

The binomial proportion, or the percentage of items that agree, is the mean for that category. This value represents a single observation per cell for the analysis of variance. However, the mean, which is a proportion bounded between 0 and 1, does not satisfy the assumption of normality required for analysis of variance. Therefore, the arcsine transformation replaces the proportion $p_{ij}$. Special tables and techniques have been developed for small sample sizes. The analysis of variance and the tests of significance are performed on the transformed data.

The analysis of variance study uses only the diagonal elements of the classification error matrices. All thematic mapping experiments peformed in the past have used only the diagonal elements for analysis. A more sophisticated step of data analysis techniques would be to use the entire classification error matrices in such studies. An attempt is already being made in this direction as evidenced by the presentation of Congalton (1980). His approach uses discrete multivariate analysis (Bishop and others, 1975). Another approach using the entire classification error matrices would be the techniques of multivariate analysis of variance.

## 4.2 Multiple Comparisons Tests

Multiple comparisons tests are described by Sokal and Rohlf (1969, p. 226-227 and p. 235-246) as a posteriori tests performed after the analysis of variance to distinguish differences between means or groups of means. They are performed only if the analysis of variance is significant.

One such method is the Duncan multiple range test for variable response (Steel and Torrie, 1960, p. 107-109; and Duncan, D.B., 1955). Duncan's multiple range test is a multiple comparisons test to compare each treatment mean with every other treatment mean. The method takes into account the number of treatments in the experiment, whereas previous methods based on the least significant difference (LSD) do not. When an F-test determines that the differences between one or more of treatment means are significant, it does not specify which, if any, are not, but these multiple range tests indicate which differences are significant, and which are not.

## REFERENCES


Anderson, J.R., Hardy, E.E., Roach, J.T., and Witmer, R.E., 1976: A land use and land cover classification system for use with remote sensor data: U.S. Geological Survey Professional Paper 964, 28 p.

Berry, B.J.L., 1962, Sampling, coding, and storing flood plain data: U.S. Department of Agriculture Handbook 237.

Berry, B.J.L. and Baker, A.M., 1968, Geographic sampling, in Berry, B.J.L., and Marble, D.F., eds., Spatial analysis—a reader in statistical geography: Inglewood Cliffs, New Jersey, Prentice-Hall, Inc.

Bishop, Y.M.M., Feinberg, S.E., and Holland, P.W., 1975, Discrete multivariate analysis—theory and practice: Cambridge, Massachusetts, MIT Press.

Cochran, W.G., 1977, Sampling techniques: New York, John Wiley and Sons, Inc.

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

Duncan, D.B., 1955, Multiple range and multiple F tests: Biometrics, v.11, p. 1-42.

Fitzpatrick-Lins, K., 1981, Comparison of sampling procedures and data analysis for the land use and land cover maps of Tampa, Florida: Photogrammetric Engineering and Remote Sensing, v. 47, no. 3, p. 343-351.

Freund, J.E., and Williams, F.J., 1972, Elementary business statistics, 2d ed.: Inglewood Cliffs, New Jersey, Prentice-Hall, Inc.

Ginevan, M.E., 1979, Testing land-use map accuracy—another look: Photogrammetric Engineering and Remote Sensing, v. 45, no. 10, p. 1371-1377.

Greenwood, J.A., and Sandomire, M.M., 1950, Sample size required for estimating the standard deviation as a percent of its true value: Journal of American Statistical Association, v. 45, no. 250, p. 257-260.

Hord, R.M., and Brooner, W., 1976, Land use accuracy criteria: Photogrammetric Engineering and Remote Sensing, v. 42, no. 5, p. 671-677.

Johnson, N.L. and Kotz, Samuel, 1969, Discrete distributions: Boston, Massachusetts, Houghton Mifflin Company.

Mace, A.E., 1974, Sample-size determination: Huntington, New York, Robert E. Krieger Publishing Company.

Matern, B., 1947, Methods of estimating the accuracy of line and sample plot surveys: Medd. fr. Statens Skogforsknings Instut., 36, p. 1-38.

Mead, R.A. and Szajzin, J., 1982, Landsat classification accuracy assesment procedures: Photogrammetric Engineering and Remote Sensing, v. 48, no. 1, p. 139-141.

Natrella, M.G., 1963, Experimental statistics: National Bureau of Standards Handbook 91, U.S. Government Printing Office.

Rosenfield, G.H., 1981, Analysis of variance of thematic mapping experiment data: Photogrammetric Engineering and Remote Sensing, v. 47, no. 12, p. 1685-1692.

Rosenfield, G.H., 1982, Sample design for estimating change in land use and land cover: Photogrammetric Engineering and Remote Sensing, v. 48, no. 5, p. 793-801.

Rosenfield, G.H., and Melley, M.L., 1980, Applications of statistics to thematic mapping: Photogrammetric Engineering and Remote Sensing, v. 46, no. 10, p. 1287-1294.

Rosenfield, G.H., Fitzpatrick-Lins, K., and Ling, H.S., 1982, Sampling for thematic map accuracy testing: Photogrammetric Engineering and Remote Sensing, v. 48, no. 1, p. 131-137.

Scheffé, H., 1959, The analysis of variance: New York, John Wiley and Sons, Inc.

Snedecor, G.W., and Cochran, W.G., 1967, Statistical methods, 6th Edition: Ames, Iowa State University Press.

Sokal, R.R., and Rohlf, J., 1969, Biometry: San Francisco, California, W.H. Freeman and Company.

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

Totora, R.D., 1978, A note on sample size estimation for multinomial populations: The American Statistician, v. 32, no. 3, p. 100-102.

Van Genderen, J.L., Lock, B.F., and Vass, P.A., 1978, Remote sensing; statistical testing of thematic map accuracy: Remote Sensing of Environment, v. 7, p. 3-14.