George H. Rosenfield
Mary Lou Melley*
U.S. Geological Survey
Reston, VA 22092

Applications of Statistics to Thematic Mapping

In properly conducted tests, all of the methods of data analysis will lead to equivalent results.

Introduction

This paper concerns certain aspects of applied statistics which can be used for analyzing all types of thematic mapping, such as: forest-type mapping, soil-survey mapping, geological mapping, vegetation mapping, land-use mapping, etc.

Two problems of a statistical nature which occur in the effort to analyze thematic maps and mapping are (1) determining the accuracy of the thematic content, and (2) comparing factors studied in thematic mapping. Statistical procedures applicable to both these problems include techniques for sampling and determining accuracy. In addition, statistical techniques for hypothesis testing are used when comparing factors. This paper includes an introductory discussion of these three techniques together with demonstrations for their use in solution of the two problems. Since the authors are associated with the National Land Use and Land Cover Mapping Program of the U.S. Geological Survey, examples have been taken from data of that program.

Many factors affect the classification of a particular piece of land. The images of the land may be of different scales or may have been formed using different instrumentation. More than one interpreter or classification system may have been involved. Tests of thematic mapping are conducted to compare factors which might affect the classification and to arrive at specifications and techniques for producing the maps.

The thematic map is divided into regions, sometimes called polygons, according to the categories of the theme as determined from imagery. A number of test points which have been selected in the polygons of each category is called a sample. The interpreted category at the test points is compared with what is known from investigations on the ground. The data thus acquired result in a set of agreements and disagreements between the categories determined by ground investigation and the categories determined by image interpretation classification. This set is analyzed to determine accuracy of classification or to determine the relative effectiveness of the factors involved.

Abstract: *Two statistical problems occurring in the effort to analyze thematic maps and mapping are determining the accuracy of thematic content and comparing factors studied in thematic mapping. Statistical procedures applicable to thematic mapping involve sampling, determining accuracy, and comparing factors. A sampling procedure using an unaligned pattern within a square grid network is applicable for use with thematic maps. Sample size may be determined using the binomial distribution based upon the confidence interval to define the true mean of the population within certain limits. The confidence interval may also be used to define the upper and lower limits of the accuracy of the thematic map. Several tests for comparing factors using differences of paired samples are the t test, the signed rank test, and the sign test. When more than two factors are being studied, analysis of variance is the most efficient procedure to use.*

* Now with U.S. Environmental Protection Agency, Washington, D.C.

Photogrammetric Engineering and Remote Sensing,
Vol. 46, No. 10, October 1980, pp. 1287-1294.

Sampling

The manner of sampling does not have to be random. Sampling in a systematic manner 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). Cochran (1977, p. 227-228), in discussing systematic sampling in two dimensions, states that it has been found that a square grid pattern had about the same precision as simple random sampling in two dimensions, and that an unaligned pattern within the square grid will often be superior to both a systematic pattern and to stratified random sampling.

SAMPLE SELECTION

A square grid can be overlaid on each area of interest in the following manner. Assume that the area is located with an X, Y coordinate system. Let $X_{max}, X_{min}, Y_{max}, Y_{min}$ be respectively, the maximal and minimal X, Y coordinates of the area. The area is divided into squares each with side of dimension D, where D is calculated by

$$D = [(X_{max} - X_{min})\,(Y_{max} - Y_{min})/n]^{1/2}$$

where n is the initial desired sample size. The number of squares falling within the boundaries of the area are counted. If there are not enough to meet the desired sample size, the value of n is increased accordingly, until the number of squares within the area exceed the minimum desired sample size.

The origin for the unaligned pattern is selected by using a pair of random numbers to fix the coordinates of the upper left unit. An additional pair of random numbers determine the horizontal coordinates of the remaining units in the first column of grids, and the vertical coordinates of the remaining units in the first row of grids. Constant intervals (equal to the sides of the square grid) then fix the locations of all units.

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

If the sample selection does not provide enough data to check all the categories, then a larger sample size is required. One method to achieve this is to use a larger value of n in the equation for the grid dimension. If a larger sample size is desired within specific categories, it can be obtained by selection in a random manner within a list of all regions, or polygons, of those categories. These lists of regions, or polygons, of each category are easily obtained if the information is contained in a computer data base.

SAMPLE SIZE

The investigator is faced with the task of selecting an appropriate sample size. An adequate number of sample items must be selected from each category of classification.

A method of using the binomial distribution for determining sample size, based upon the confidence interval for the mean $\bar{x}$, is given by Hord and Brooner (1976), although they neglect to apply the "correction for continuity" (Snedecor and Cochran, 1967, p. 209-213). This correction may be significant for small values of n (the number of items in the sample) and low values of $\bar{x}$ but is insignificant for the ranges of "n" and $\bar{x}$ considered. The correction accommodates the difference between the discrete, binomial distribution of the proportion of correct interpretations in the data and the continuous, normal distribution used later in determining confidence limits. The correction amounts to widening the confidence interval, and allows for the skewness of the curve of the binomial distribution.

The concept of the confidence interval for proportions is used to determine the sample size when it is desired to define the true mean within certain limits of error. The standard meaning of confidence interval is the interval such that the true mean in the sampled population lies in the interval from the lower to the upper limits with specified statistical confidence (Snedecor and Cochran, 1967, p. 5).

Mace (1973, p. 61-62) gives another algorithm for determining sample size when the sample variables are considered as belonging to the binomial distribution, and it is desired to define the true mean within a specified confidence interval. It is based on the normal curve approximation and incorporates the arcsine transformation of the anticipated proportion of correct interpretations.

Van Genderen and Lock (1977) approach the problem of determining sample size from a different point of view. They choose the minimum sample size according to the probability of making incorrect interpretations at specifically prescribed accuracy levels. Their table is used to select the minimum sample size to meet a specified interpretation accuracy at a given level of probability.

Determining Accuracy

The 95 percent confidence interval may be used to determine the upper and lower limits of the accuracy of the thematic map. Hord and Brooner (1976) suggest using the lower limit of the confidence interval as an estimate of the accuracy of the map. If the lower confidence limit alone is to be used to express map accuracy, then a different statistical approach, a one-tailed test rather than a two-tailed test, is required.

DEVELOPMENT OF CONFIDENCE INTERVAL

To obtain the mean and variance of the set of correctly interpreted points of a thematic map, attach the number 1 to every success in the population and the number 0 to every failure. The population then becomes a large collection of 1's and 0's, and the population distribution of the variable x then takes only two values: 1 with relative frequency p and 0 with relative frequency q. The variate x thus has population mean p and population variance pq (Snedecor and Cochran, 1967, p. 208). Since this population has a binomial distribution, the mean of the set of successes, μ_x, the mean of the set of failures, q, and the variance of the population, σ_x^2, are, respectively,

$$\mu_x = p,\ q = 1 - \mu_x,\ \sigma_x^2 = \mu_x(1 - \mu_x). \quad (1)$$

If r is the number of items (in the sample) having the desired attribute, then according to the Central Limit Theorem, "the mean $\bar{x}$ of a random sample from any population with finite variance tends to normality. Hence, as n increases, the binomial distribution of r/n or of r approaches the normal distribution" (Snedecor and Cochran, 1967, p. 209). For large sizes of sample, the normal distribution can be substituted for the binomial distribution of the population. It is now possible to calculate a confidence interval about the estimated value of the mean, μ, of the set of successes. Given a sample of size n with $\bar{x}$ as the mean, the probability is 0.95 that $z \equiv \bar{x} - \mu_x/\sigma_x/\sqrt{n}$, will be less than 1.96, i.e.,

$$P_r\left(\frac{|\bar{x} - \mu_x|}{\sigma/\sqrt{n}} < 1.96\right) = 0.95. \quad (2)$$

The probability is 0.95 that the population mean, μ, will lie in the interval bounded by

$$\bar{x} \pm (1.96)\sigma/\sqrt{n}. \quad (3)$$

Since the binomial distribution of r/n is discrete, the correction for continuity may be applied to approximate the normal distribution. In accordance with Snedecor and Cochran (1967, p. 210), the correction for continuity applied to the inequality of Equation 2 is

$$\frac{|\bar{x} - \mu_x| - 1/2n}{\sigma/\sqrt{n}} < 1.96. \quad (4)$$

At this point, a distinction must be made between theoretical probabilities and observed proportions. In sampling the thematic map, the digit 1 is assigned to a correct interpretation and the digit 0 to an incorrect one. If "a random sample of size n contains r successes, the Σx, taken over the sample, is r, so that $\bar{x} = \Sigma x/n$ is r/n, the sample proportion of successes. But we know that the mean of a random sample from any distribution is an unbiased estimate of the population mean, and has variance σ^2/n" (Snedecor and Cochran, 1967, p. 208). Therefore, the mean of the proportion of success, the mean of the proportion of failures, and the variance of the proportion is determined theoretically and expressed as, respectively,

$$\mu = p = r/n,\ q = 1 - p,\ \sigma^2 = pq/n. \quad (5)$$

In sampling, as n increases, the estimate of the proportion $\hat{p} = r/n = \bar{x}$ in the sample is approximately normally distributed about the proportion, p, in the population. According to Snedecor and Cochran (1967, p. 210-211) the equation of z for the proportion in the sample, with correction for continuity, and for p not near 1/2, the probability is 0.95 that

$$\frac{|\hat{p} - p| - 1/2n}{\sqrt{(pq/n)}} < 1.96. \quad (6)$$

The inequality is solved by substitution, p, q, and n being known.

Substituting Equations 1 (without subscripts) and $\bar{x} = \hat{p}$ into Equation 6 with μ *not* determined and squaring the inequality in Equation 6 gives

$$\frac{n[|\bar{x} - \mu| - 1/2n]^2}{\mu(1 - \mu)} < (1.96)^2. \quad (7)$$

Change the inequality to an equality and solve the quadratic equation for μ; the roots μ^1, μ^2 are the limits of the confidence interval of the accuracy values for the sample of size n. The coefficients of the quadratic equation

$$A_1\mu^2 + A_2\mu + A_3 = 0, \quad (8)$$

are:

$$\begin{aligned} A_1 &= n + (1.96)^2, \\ A_2 &= 1 - 2\bar{x}n - (1.96)^2, \\ A_3 &= \bar{x}^2 n - \bar{x} + 1/4n. \end{aligned} \quad (9)$$

APPLICATION OF CONFIDENCE INTERVAL TABLES

The table of Hord and Brooner (1976) allows either the sample's accuracy or its size to be determined. If the sample's accuracy is known, the upper and lower confidence limits of the map's accuracy can be determined according to the sample's size. As the number of items in the sample increases, the size of the confidence interval decreases. A measure of the relative increase in accuracy for an increase in sample size is given by the decrease in a quantity that might be called the percent-of-true-value. This quantity is taken as one-half of the actual length of the confidence interval divided by the sample's accuracy. The percent-of-true-value is to be compared with the number of items in the sample for a given accuracy of the sample.

For example: if the expected accuracy of the sample is 0.98, then for a sample size of 50 items, there is 95 percent confidence that the accuracy of the map will lie between 0.89 and 0.99. One-half of this confidence interval will be 5.2 percent-of-true-value based on the sample's accuracy of 0.98. If the sample's size is increased to 150 items, the confidence limits of the map's accuracy will range between 0.94 and 0.99, and half of this interval

will be 2.5 percent-of-true-value. As the sample's size increases for a given accuracy, the percent-of-true-value decreases. The analyst can then weigh the increase in size against the decrease in percent-of-true-value in order to establish the desired size.

Once the data have been recorded and the sample's accuracy calculated, the confidence limits of the map's accuracy may be determined for a sample of size n. According to Hord and Brooner (1976) the lower limit would be the acceptable accuracy for the map. For example: if the sample's accuracy is determined by experimentation to be 0.91, and the sample size, n, is 350, then with 0.95 confidence the minimum probable accuracy of the map is 0.87, the lower limit of the confidence interval. However, the maximum probable accuracy does not exceed 0.93, the upper limit of the confidence interval.

Comparing Factors

The comparison of categories determined by interpretation of images with those determined by inspection in the field has been used in land-use and land-cover classification. Tests have been performed on land-use and land-cover maps prepared from different types of imagery, such as Landsat multispectral imagery and high-altitude aerial photographs. Similar tests have also been performed on the same kind of imagery at different scales. When the test includes only two sets of data, several types of techniques using differences between pairs are applicable: (1) parametric methods such as the t test, and (2) nonparametric methods such as Wilcoxon's signed rank test (Sokal and Rohlf, 1969, pp. 399-401) and the sign test (Snedecor and Cochran, 1967, p. 125). If more than two types of data are to be compared, tests which work on only one pair at a time are not suitable. Analysis of variance is a more powerful parametric statistical procedure for this purpose than the others, for use when the data either fulfills the necessary assumptions, or can be so transformed.

T TEST FOR COMPARISON OF PAIRS

The t test is the traditional method of determining the significance of the difference between two means. It is a parametric test based on the normal distribution. Given a number of pairs of data x_{1i} and x_{2i}, the difference D_i of the i^{th} pair is

$$D_i = (x_2 - x_1)_i. \tag{10}$$

The quantity

$$t \equiv (\bar{D} - \mu_D)/S_D, \tag{11}$$

follows Student's t-distribution with $n - 1$ degrees of freedom, where n is the number of pairs. The t-distribution may be used to test the hypothesis that $\mu_D = 0$, or to compute a confidence interval for μ_D (Snedecor and Cochran, 1967, p. 93). The alternate hypothesis is that $\mu_D \neq 0$.

The standard deviation of the sample may be expressed as

$$S_D = [(\sum D_i^2 - (\sum D_i)^2/n)/(n - 1)]^{1/2} \tag{12}$$

and;

$$S_{\bar{D}} = S_D/\sqrt{n}. \tag{13}$$

Refer to a table of critical values of Student's t-distribution for a two-tailed test to determine the significance. The computed value of t must be greater than the tabular value of t in order to claim significance for that probability. If the computed value is equal to or less than the tabular value, the hypothesis is accepted and it may be concluded that the difference between the means of the two populations is not significant.

In addition, the confidence limits at the 95 percent level for $n - 1$ degrees of freedom may be calculated as:

$$\bar{D} \pm t_{0.05(n-1)}S_{\bar{D}}. \tag{14}$$

In a recent investigation by Fitzpatrick-Lins (1978), two sets of imagery were compared for interpretation and classification of categories for land-use and land-cover mapping. These sets were high-altitude aerial photographs (*HA*), and Landsat multispectral imagery (*LI*), each at a scale of 1:250,000. The number of correct classifications in agreement with what was found by investigation on the ground (field checking) was determined for each set of images. In Table 1, the values for each category in columns HA and LI represent the number of items from high-altitude aerial photographs and from Landsat imagery, respectively, which were correctly classified. The values in column D represent the differences between the number of items in agreement between these two sets for each category. The hypothesis to be tested is that the two kinds of images produce an equal number of correct classifications, so that $\bar{D} = 0$.

Evaluation of the equations for the data in Table 1 lead to

$$\begin{aligned} \bar{D} &= 8.8333 \\ S_{\bar{D}} &= 4.2694 \\ t &= 2.0690 \end{aligned}$$

The tabular value of t, for 5 degrees of freedom at the 5-percent significance level (two-tailed test), is

$$t_{0.05(5)} = 2.571.$$

The computed value of t, 2.0690, is less than the tabular value, so the hypothesis that the two sets of imagery produce an equal number of correct classifications is accepted. It may then be concluded at the 5 percent significance level that there is no significant difference between the two sets of imagery in providing correct classifications.

TABLE 1. DATA FOR *t* TEST

[*HA:* High-altitude aerial photographs, *LI:* Landsat Imagery. Test data from Fitzpatrick-Lins (1978)]

Level I Category	Total Number of Points	Number of Correctly Classified Points		$D = HA - LI$	D^2
		HA	*LI*		
1	71	49	24	25	625
2	217	162	145	17	289
4	342	270	261	9	81
5	93	74	76	−2	4
6	36	26	22	4	16
7	1	0	0	0	0
Total				53	1,015

WILCOXON SIGNED RANK TEST FOR PAIRED SAMPLES

The Wilcoxon signed rank test is a nonparametric test used as a substitute for the *t* test in paired samples, and may be used with data from either normal or non-normal distributions. It is not as efficient a test as the *t* test, and requires at least five pairs of data. In performing the signed rank test, the differences between the paired items are obtained. Then ranks are assigned to the absolute values of these differences, the smallest difference being assigned rank 1. The signs are then restored to the rankings. The sums of all ranks having positive signs, and of all ranks having negative signs, are obtained. The test criterion is the smaller value of these two sums. The test hypothesis is that the frequency distribution of the original measurements is the same for the different members of a pair. The alternate hypothesis is that they are not the same (Snedecor and Cochran, 1967, p. 128 seq.).

Refer to a table of critical values of the Wilcoxon rank sum (Rohlf and Sokal, 1969, p. 246) for a two-tailed test to determine the significance. The computed value *T* must be equal to or less than the tabular value for the given number of pairs in order to claim significance for that probability. If the computed value is greater than the tabular value, the hypothesis is accepted, and it may be concluded that the two populations are not significantly different.

The following example again uses data from Fitzpatrick-Lins (1978). In Table 2, the values for each category in Column *HA* and *LI* represent the number of items from high-altitude aerial photographs and from Landsat imagery, respectively, which were correct. The values in column *D* represent the differences between the number of items in agreement between these two sets for each category. The differences were then ranked in ascending order, and assigned to rank 1 if positive or rank 2 if negative. The hypothesis to be tested is that the two sets of images come from populations having the same frequency distribution.

"Since the exact probability level desired cannot be obtained with integral critical values of *T*, two such values and their attendant probabilities bracketing the derived significance level are furnished" (Rohlf and Sokol, 1969, p. 245). Thus, the two tabular values of *T* (and their attendant probabilities), at the 5 percent significance level (two-tailed test) for six pairs of data, are

$$T_{0.05(6)} = \begin{matrix} 0\ (0.0156) \\ 1\ (0.0312) \end{matrix}$$

TABLE 2. DATA FOR WILCOXON SIGNED RANK TEST

[*HA:* High-altitude aerial photographs, *LI:* Landsat imagery. Test data from Fitzpatrick-Lins (1978)].

Level I Category	Number of Correctly Classified Points		$D = HA - LI$	Rank	Rank	
	HA	*LI*			R_1 (+)	R_2 (−)
1	49	24	25	6	6	
2	162	145	17	5	5	
4	270	261	9	4	4	
5	74	76	−2	2		2
6	26	22	4	3	3	
7	0	0	0	1	1	
TOTAL					19	2

The computed value of T, 2, is greater than the selected tabular value, so the hypothesis that the two populations have the same frequency distribution is accepted. It may then be concluded that at the 5 percent significance level there is no significant difference between the two kinds of images insofar as correct classifications are concerned.

SIGN TEST FOR PAIRED SAMPLES

The sign test is a non-parametric test which uses "the chi-square test in place of the t test for paired samples. It is known as the sign test, because the differences between the members of a pair are replaced by their signs (+ or −), the size of the differences being ignored" (Snedecor and Cochran, 1967, p. 127). Like the Wilcoxon signed rank test, it is not as efficient as the corresponding t-test.

To perform the sign test, the data must be tabulated as to success or failure (true or false when compared to a standard). To illustrate the use of this test, consider the classification of n items on each of the two sets of imagery, *HA* and *LI*. If an item is correctly classified, it is considered true (T), if incorrectly classified, it is considered false (F).

Point	*HA*	*LI*
1	T	T
2	T	F
3	F	T
4	F	F
.	.	.
n	.	.

Count the number of pairs of points which were both true, both false, *HA* false and *LI* true, and *HA* true and *LI* false.

The hypothesis is that the proportion of trues is the same for the two sets. The occurrences of both true and both false are ignored since they give no indication in favor of *HA* or *LI*. The explicit hypothesis is that the population contains as many *HA* true − *LI* false pairs (TF pairs) as *HA* false − *LI* true pairs (FT pairs). The alternate hypothesis is that the proportion of trues is not the same for the two sets.

Let a represent the number of TF pairs, and b represent the number of FT pairs.

The theoretical ratio of each, TF and FT, to the total of the two is $p = 1/2$, i.e.,

$$\frac{\text{number of TF pairs}}{\text{number of TF plus FT pairs}} = 1/2$$

The value of chi-square when $p = 1/2$ and with the correction for continuity is

$$\chi^2 = (|a - b| - 1)^2/n. \qquad (15)$$

with 1 degree of freedom (Snedecor and Cochran, 1967, p. 127 and 214).

Refer to a table of critical values of the chi-square distribution to determine the significance. The computed value of chi-square must be greater than the tabular value in order to claim significance at that probability. If the computed value is equal to or less than the tabular value, the hypothesis is accepted and it may be concluded that the two populations are not significantly different.

Since the original data from the investigation by Fitzpatrick-Lins (1978) are not now available, there is no example in this paper of application to land-use and land-cover mapping. In addition, when there are as many as several hundred items in the sample, as there were in this case, the sign test is impractical.

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, described below, is similar to that presented by Sokal and Rohlf (1969, p. 299). For rows, the degrees of freedom are $n - 1$; for columns, they are $m - 1$; for error they are $(m - 1)(n - 1)$. The hypothesis to be tested is that the two 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.

The quantity F, defined as the factor (A or B) mean square divided by the error mean square is used to test the hypothesis. Refer to a table of critical values of the F-distribution for a two-tailed test to determine the significance. The computed value of F for each factor must be greater than the tabular value in order to claim significance for that probability, and with $(n - 1)$ and $(m - 1)(n - 1)$ degrees of freedom. If the computed value is equal to or less than the tabular value, the hypothesis is accepted, and it may be concluded that the populations under study for that factor are not significantly different.

The following example uses data from Fitzpatrick-Lins (1978). For each category, the number of correctly interpreted items was divided by the total number of items in that category. This 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 pro-

portion 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}, or

$$\theta = \arcsin \sqrt{p_{ij}}, \qquad (16)$$

where θ is the arcsine transformed value. "The arcsine transformation stretches out both tails of a distribution of percentages or proportions and compresses the middle," (Sokal and Rohlf, 1969, p. 386) so that the curve of the distribution conforms more closely to that of the normal distribution. For example, the proportion value 1 becomes the value 90.00, thus making room for the tails of the normal distribution. 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.

In this experiment, factor B consists of the two types of images, and factor A consists of categories of land-use and land-cover classification. The two types of images are the two sets to be compared: high-altitude aerial photographs and Landsat imagery.

Table 3 gives the number and percent of items correctly classified from high-altitude aerial photographs and Landsat imagery.

Table 4 gives these data after the arcsine transformation for input into a program created for the Wang 2200 computer (Wang Laboratories, 1975).

Table 5 gives the analysis of variance table obtained from these data.

Analysis of the sets of high-altitude aerial photographs and Landsat imagery for land-use or land-cover classification is of prime importance. These two sets are included in factor B in the analysis of variance. For a two-tailed test, the tabular value of F, for 1 and 5 degrees of freedom at the 5 percent significance level, is (Rohlf and Sokal, 1969, p. 170)

$$F_{0.05(1,5)} = 10.$$

TABLE 3. NUMBER AND PERCENT OF CORRECTLY CLASSIFIED POINTS

[*HA:* High-altitude aerial photographs, *LI:* Landsat Imagery. Test data from Fitzpatrick-Lins 1978)].

Level I Category	Total Number Of Points	Correctly Classified Points			
		Number		Percent	
		HA	*LI*	*HA*	*LI*
1	71	49	24	69.0	33.8
2	217	162	145	74.7	66.8
4	342	270	261	78.9	76.3
5	93	74	76	79.6	81.7
6	36	26	22	72.2	61.1
7	1	0	0	0.0	0.0

TABLE 4. ARCSINE TRANSFORMED DATA

Factor A	Factor B	
	High-altitude Aerial Photographs	Landsat Imagery
1	56.17	35.55
2	59.80	54.82
3	62.65	60.87
4	63.15	62.67
5	58.18	51.41
6	0.0	0.0

The computed value of F, 2.739, is less than the selected tabular value, so that the hypothesis that the samples come from the same population is accepted. It may then be concluded, at the 5 percent significance level, that there is no significant difference between the two sets of imagery for providing correct classifications.

Analysis of the categories under factor A in the analysis of variance is also a two-tailed test. The tabular value of F for 5 and 5 degrees of freedom at the 5 percent significance level is 7.15. The computed value of F, 35.581, is greater than the selected tabular value, so that the hypothesis that the samples came from the same population is rejected. It may then be concluded at the 5 percent significance level that the success of classification into each of the various categories using these types of images varies with each category.

SUMMARY AND DISCUSSION OF STATISTICAL TESTS

The particular statistical tests described have been selected from a larger number that are available. Two-way analysis of variance is the most efficient and most powerful test. The t test is the traditional statistical method of solving the problem of determining significance of the differences between two means. It is mathematically equivalent to a two-sample, single-classification analysis of variance. The assumptions of the t test are that the computed mean of the observations is nor-

TABLE 5. TWO-WAY ANALYSIS OF VARIANCE WITHOUT REPLICATION (ONE OBSERVATION PER CELL): OUTPUT.

Summary of analysis of variance table

Level of Factor A = 6
Level of Factor B = 2

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F
A	5	5761.435	1152.287	32.581*
B	1	88.726	88.726	2.739
Error	5	161.921	32.384	
Total	11	6012.084		

* significant

mally distributed and that the variances of the two samples are equal. Fortunately, the *t* test is also effective for data from populations which are moderately non-normal.

The nonparametric tests (sometimes called distribution-free) are needed when the data are obtained from populations that are greatly non-normal; however, they can be used with data from normal distributions. Their popularity is partly due to the fact that these tests are rapid and simple to execute. The Wilcoxon signed rank test is already in accepted use in geographic science (McCullagh, 1974). For any continuous distribution, the significance levels of the rank tests remain about the same as the *t* test. In large normal samples, the rank tests have an efficiency of about 95 percent relative to the *t* test. In small normal samples, the efficiency of the signed rank test is slightly higher. For non-normal data from a continuous distribution, and in large samples, the rank tests have an efficiency of never below 86 percent relative to the *t* test. The sign test, when used in sampling from normal distributions, has an efficiency of about 65 percent relative to the *t* test (Snedecor and Cochran, 1967, p. 120 seq.).

Conclusions

This paper has given a number of statistical concepts and procedures which are applicable to thematic mapping. It has been shown that sampling can and should be placed on a firm statistical foundation. Selection of the sample size should be based upon predetermined levels of probability and confidence. If the selected sample size is too large for economic feasibility, then the effect of a reduction of the sample size in terms of confidence in the accuracy of the mapping must be understood. In the field of land-use and land-cover mapping, tests have been presented for solving problems such as use of different scales, different types of imagery, different algorithms or equipment, different approaches to classification, and different geographical regions. Parametric tests should be used if the necessary assumptions for their use to be valid can be met. Transformations such as the arcsine transformation help satisfy these assumptions if the original data are not normally distributed. The analysis of variance techniques are more powerful and efficient than the *t* test and the nonparametric tests, especially for comparisons involving more than two factors. For comparisons of pairs, the *t* test is preferred to the nonparametric tests if the assumptions of normality are reasonably met. This paper has shown that, in properly conducted tests, all of the methods of data analysis will lead to equivalent results.

References

Cochran, W. G., 1977. *Sampling Techniques:* John Wiley & Sons, Inc., New York.

Fitzpatrick-Lins, Katherine, 1978. Accuracy and consistency comparisons of land-use maps made from high altitude photography and Landsat imagery, *Journal of Research, U.S. Geological Survey*, v. 6, no. 1, pp. 23-40.

Freund, J. E., and F. J. Williams, 1972. *Elementary Business Statistics*, second edition: Prentice-Hall, Inc., Inglewood Cliffs, N. J.

Hord, R. M., and William Brooner, 1976. Land-use map accuracy criteria, *Photogrammetric Engineering and Remote Sensing*, v. 42, no. 5, pp. 671-677.

Mace, A. E., 1973. *Sample Size Determination:* Robert Krieger Publishing Co., Huntington, New York.

McCullagh, Patrick, 1974. Data use and interpretation, in *Science in Geography:* Brian Fitzgerald (General Editor), Oxford University Press.

Rohlf, F. J., and R. R. Sokal, 1969. *Statistical Tables*, W. H. Freeman and Co., San Francisco.

Snedecor, G. W., and W. G. Cochran, 1967. *Statistical Methods*, The Iowa State University Press, Ames, Iowa.

Sokal, R. R., and F. J. Rohlf, 1969. *Biometry*, W. H. Freeman and Co., San Francisco.

Van Genderen, J. L., and B. F. Lock, 1977. Testing land-use map accuracy, *Photogrammetric Engineering and Remote Sensing*, v. 43, no. 9, pp. 1135-1137.

Wang Laboratories, 1975. *Analysis of variance operator's manual and programs*, Tewksbury, Mass.

(Received 28 June 1979; revised and accepted 24 March 1980)