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

Sample Design for Estimating Change in Land Use and Land Cover

Principles of sample design considered are the sample distribution, the sample size, the sampling procedure, estimation of the population means, variances and confidence limits, estimation of change in the population values with variances and confidence limits, and extension of the parameters from subpopulations to the overall population.

Introduction

Determination of change in land use and land cover involves comparing present land use and land cover against a previous record of land use and land cover and calculating the areal changes between the various categories. The previous record may be in the form of land-use and land-cover maps. The present land use and land cover can be determined from a direct onsite inspection, by interpretation or classification of remote sensing records, or from already prepared current land-use and land-cover maps in either graphical or digital format. The land-use and land-cover information for the region of interest can be determined from a complete census (canvass), or by sampling and then estimating the regional values. Sample design is that field of applied statistics which covers sampling within the population, considers the sample distribution, and includes determination of the sample size, the sample selection procedure, and estimation of the population means, totals, variances, and confidence limits from the sample information.

Abstract: *Sample design is that field of applied statistics which covers sampling within a specified population, considers the sample distribution, and includes determination of the sample size, the sample selection procedure, and estimation of the population means, totals, variances, and confidence limits from the sample information. The methodology of sample design which is applied to estimating change in land use and land cover is general and extendable to determination of change in any type of thematic mapping that is time variant. Land-use maps of the State of Pennsylvania at a scale of 1:250,000 were compiled circa 1958 with land use classified into six categories. The more detailed land-use and land-cover mapping of the State of Pennsylvania at a scale of 1:250,000 was completed by the U.S. Geological Survey circa 1977. With some rearrangement of these categories, the recent maps are very nearly compatible with a combination of five categories of the earlier maps. An opportunity is presented to determine change in land use and land cover for an entire State over a 20-year period. A preliminary experiment of a sample selected in both the previous and current maps to determine change in land use and land cover is used as an example of the sample design considerations.*

Background

The U.S. Geological Survey has published land-use and land-cover maps at 1:250,000 scale for the entire State of Pennsylvania as part of the National Land Use and Land Cover Mapping Program. The land use and land cover indicated on these maps was interpreted from aerial photo-

0099-1112/82/4805-0793$02.25/0

graphs and other source materials representative of the 1974-76 time period and classified into Level II categories (Anderson *et al.*, 1976). Digitization of these land-use and land-cover maps in the Geographic Information Retrieval and Analysis System (GIRAS) (Mitchell *et al.*, 1977) is completed.

Land-use and land-cover maps for the State of Pennsylvania were compiled previously using aerial photographs having the approximate date of 1954 (Klimm, 1958). These maps were also compiled at 1:250,000 scale, with land use and land cover classified into the six categories of urban and industrial, mineral, cropland, non-crop and mixed, forest (not empty), and empty (mostly forested). With some schematic rearrangement of the Level II categories on the recent maps, they are very nearly compatible at Level I with a combination of five categories of the earlier maps. This situation is the first available instance of two sequential sets of similar land-use and land-cover classification and mapping for as large an area as an entire state. It presents an opportunity to determine change in land use and land cover for an entire state over a 20-year period.

The U.S. Geological Survey has entered into a cooperative research project with the Remote Sensing and Interpretation Laboratory at Florida Atlantic University to develop the "methodology for measuring and analyzing the changes perceived from comparing archival and recent categorizations of surface areas. . . . and to relate these significant readjustments to trends in resource utilization and population change . . ." (Latham, 1979).

Latham (1979) has reported on the results of a preliminary experiment of sampling both the 1958 and 1977 land-use and land-cover maps of Pennsylvania. Systematic sampling of point samples was used on both sets of maps to estimate the proportion of each category of land use and land cover. Sampling was used for the 1977 maps because the tabulation of land-use areas was not available in time for the preliminary experiment, although the maps had already been completed. Sampling was used for the 1958 maps because direct measurement of the area of each category is expensive and time consuming. More than 3,000 samples were selected on a grid spacing of 3 minutes of latitude equally distributed over the State. In addition to statewide comparisons, three State sectors of variable character were separately aggregated to provide preliminary insight into regional and areal variations. These sections were south-western, north-central, and southeastern Pennsylvania. Latham (1979) does not explicitly state the total number of samples, nor does he state the number of samples in each of the three sectors. Several bar-charts are included which give the separately aggregated percent of surface of land-use categories for both the 1958 and the 1977 land-use and land-cover maps for the three sectors and for the entire State. The six categories of the 1958 map were collapsed for the comparison to the five categories of urban and built-up, barren/mineral land, agricultural, forest and brush, and water. Numerical values have been interpolated from these bar-charts and are given in Table 1.

PURPOSE AND SCOPE

The purpose of this study is to compile and discuss the methodology of sample design which is applicable to determination of change in area of land-use and land-cover category in an experiment of this type. Although Latham (1979) did not use the methodology developed herein, the experiment and data developed by Latham are used as an example, where applicable, since no others are extant.

This report discusses first the probability distribution, sample size, and sampling procedure to estimate the proportion of category area using point samples to meet certain designated probability requirements. The report then discusses methodology for estimating the category area from the proportion. Estimating change in area for the category is then discussed from two points of view. First, when point sampling at the time periods of both populations as performed in the preliminary experiment; and second, when the category area is known at the recent time period and point sam-

TABLE 1. VALUES OF PERCENTAGES FOR EACH CATEGORY OF LAND USE AND LAND COVER, FOR THREE SECTIONS AND THE STATE OF PENNSYLVANIA, INTERPOLATED FROM THE BAR-CHARTS OF LATHAM (1979), FOR THE MAPS OF 1958 AND 1977

	South-western		North-central		South-eastern		State	
Category	1958	1977	1958	1977	1958	1977	1958	1977
1 Urban and Built-up	5.2%	9.2%	1%	1%	7%	15%	3.3%	6.8%
2 Barren/Mineral Land	3.3	3.9	0.5	1.2	2.2	1.8	2.5	2.0
3 Agricultural	42.0	33.0	14.0	17.0	56.0	50.0	37.5	34.0
4 Forest and Brush	49.0	53.0	84.0	78.0	33.0	31.0	56.0	55.0
5 Water	0.6	1.3	0.0	1.0	2.5	2.2	1.5	2.0

pling is performed at the earlier time period. Lastly, the extension of the estimates for category area and change from subpopulations to a super population is discussed. Again in this last section, point sampling at both time periods and known values at the recent time period are covered. In all cases under consideration, not only are the total values estimated, but estimates of their variances and confidence limits are also given. The results and data from the preliminary experiment are used as illustrations and discussed.

The end points of simultaneous confidence intervals for the p_j, with the joint confidence coefficient being approximately $1 - \alpha$, are reported by Johnson and Kotz (1969, p. 289). They further report that Goodman (1965) improved on the expression of the end points (for $k > 2$), replacing the upper $\alpha \times 100$-th percentage point of the chi-square distribution with $k - 1$ degrees of freedom by the upper $(\alpha/k) \times 100$-th percentile of the chi-square distribution with one degree of freedom. The resulting expression has the form

$$\frac{\chi^2_{1,1-\alpha/k} + 2a_j \pm \left[\chi^2_{1,1-\alpha/k}\left\{\chi^2_{1,1-\alpha/k} + 4a_j\, n^{-1}\,(n - a_j)\right\}\right]^{1/2}}{2(n + \chi^2_{1,1-\alpha/k})},\ (j = 1, 2, \ldots, k).$$

Values for $\chi_{1,1-\alpha/k}$ can be found from tables of percentage points for the unit normal distribution. "..., the χ^2 distribution, with 1 d.f., is the distribution of the square of a normal deviate: the 5% significance level of χ^2, 3.84, is simply the square of 1.96" (Snedecor and Cochran, 1967, p. 212).

The individual categories of the multinomial sample are considered as if they were sampled from a binomial distribution. In this case, the sample variate y_i has the value 1 if the corresponding sample point is in the category of interest and 0 if not, and the sample mean $\bar{y}$ has the form according to Cochran (1977, p. 51, 3.3) as

$$\bar{y} = \sum_{i=1}^{n} y_i/n = a/n = p,$$

where the proportion p is determined from the ratio of the number of units (a) in the category to the total number tested (n). The sample proportion p is an unbiased estimate of the population proportion P.

An unbiased estimate of the variance of p is derived from the sample and is given in the form of Cochran (1977, p. 52, 3.11) as

$$v(p) = s_p^2 = [p(1 - p)/(n - 1)](1 - f),$$

where $f = n/N$ is called the sampling fraction and $(1 - f) = (N - n)/N$ is called the finite population correction (fpc), and can be ignored when n/N does not exceed 5 or 10 percent (Cochran, 1977, p. 25) and where n is the number of sample units and N is the total number of units in the population. If N is very large relative to n, so that the fpc is negligible, the factor $1 - f$ is ignored.

Sample Design

THE MULTINOMIAL DISTRIBUTION

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_j in each category, j. This probability is given in the form of Cochran (1977, p. 60, 3.20) as

$$\Pr(a_j) = \frac{n!}{a_1!\, a_2! \ldots a_k!}\, p_1^{a_1} p_2^{a_2} \ldots p_k^{a_k},$$

where $j = 1, 2, \ldots, k$ (k = the number of categories),
a_j = the number of units of the jth category in the sample,
$n = \sum a_j$,
$P_j = A_j/N$,
A_j = the number of units in the jth category in the population, and
$N = \sum A_j$.

The binomial distribution and its extension, the multinomial distribution, are developed from the procedure of simple random sampling (Cochran, 1977, p. 50). For systematic sampling in two dimensions (advocated in this report), the binomial and multinomial distributions are only approximations. Cochran (1977, p. 229) states that no trustworthy method for estimating the variance of the mean of a systematic sample from the sample data is known.

Johnson and Kotz (1969, p. 288-289) give expressions for estimating the parameters p and n of the multinomial distribution. For the situation in which n and k are known, the maximum likelihood estimate of the population proportion $P_j = A_j/N$ is

$$p_j = a_j/n\ (j = 1, 2, \ldots, k).$$

SAMPLE SIZE

Statistical algorithms exist for determining sample size for sampling populations within the multinomial distribution. Snedecor and Cochran (1967, p. 59) summarize that the parameters to be considered for estimating sample size are an upper limit L to the amount of error that can be tolerated in the estimate, the desired probability that the estimate will lie within this limit of error, and an a priori estimate of the population standard deviation σ.

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 equation has the form

$$n = \chi^2_{1,1-\alpha/k} p_j(1 - p_j)/\delta_j^2,$$

for a large population in which the finite population correction can be ignored. The parameter δ_j is the half width of the desired confidence interval. When $\delta_j = \delta$, only one calculation is required for the p_j which is closest to 0.5. When each p_j has a different δ_j, then a separate calculation is made for each pair $(\delta_j, p_j), j = 1, 2, \ldots, k$, and the largest n is selected as the desired sample size.

If the sample variables are considered as belonging to the multinomial distribution, i.e., classification into more than two categories, and a priori estimates of the category proportions are available, then the necessary sample size can be estimated in order to establish the true proportions for the categories of the population within certain simultaneous limits of error. If a priori estimates of proportions do not exist, a conservative sample size can be estimated using $p = 0.50$.

The following calculations for sample size are based on the multinomial distribution and use the proportion values of Table 1 for each sector, or State, which is closest to 0.5. The remaining parameters of the calculation are: $\delta = \pm 10\%$, and $P = 0.90$; $k = 5$, the number of categories:

$$\alpha = 1 - 0.90 = 0.10, \alpha/k = 0.10/5 = 0.02$$

then $\chi^2_{1,0.98} = 5.4149$.

From Table 1, for the 1958 data, the proportions p_j closest to 0.5, and the computed sample sizes are are

	p_j	n
Southwestern	0.49	135
North-central	0.84	73
Southeastern	0.56	133
State of Penn.	0.56	133

Note that the value for the sample size based on the section nearest 0.50 is 135.

For the Pennsylvania project, the population of interest has been selected to be the county; and information on land-use and land-cover change is desired for each county. The county populations can then be aggregated for the entire State. Since there are 67 counties in the State of Pennsylvania, the total sample size needed for the entire State is $67 \times 135 = 9{,}045$ based on the multinomial distribution. Since the preliminary sample points are locatable, the proportions for each category can be computed on a county basis, as they were for the three sections. A sample size estimate can be determined for each county, then the total aggregate for the State may be a smaller number.

Sample size to establish a common standard deviation. It might be that it is desired to sample to achieve a particular variance or standard deviation. A preliminary experiment would then be conducted using a preliminary sample size, and the estimated standard deviation for the desired population parameter determined. It may be noted in the variance equation, that the standard deviation is functionally related to the inverse of the square root of the degrees of freedom: i.e.,

$$s_p = f(1/\sqrt{\text{d.f.}}).$$

If it is desired to decrease the standard deviation by a particular proportion, then the sample size must be increased by the square of that proportion. Thus, to halve the standard deviation, it is required to take four times the sample size.

It may also be noted in the sample size equation of Tortora that the sample size n is functionally related to the inverse square of the half width of the desired confidence interval for the particular a priori parameters selected: i.e.,

$$n = f(1/\delta^2).$$

In the Pennsylvania project, the sample units were taken as point samples, and the land-use and land-cover category determined for that point, thus the multinomial distribution. It is possible to estimate an a priori standard deviation for each proportion for the population. Therefore, it is possible to vary the sample size for each county (or class of counties) based on standard deviation, in order to obtain a common standard deviation for the population estimates which may be desired for further analysis (such as economic studies). The value for the standard deviation of the proportion is the measure to be standardized, and is computed in the form:

$$s_p = \sqrt{[p(1 - p)/(n - 1)]}.$$

The maximum value of the minimum sample size would be for a proportion of $p = 0.50$. Using this value and an estimated sample size of 135, the estimated variance is

$$s^2 \cong (0.50)\,(0.50)/134 = 0.0018656$$

and the standard deviation is $s \cong 0.043$.

To decrease the standard deviation to some desired value, say 0.03, or a factor of 1.43, would require a sample size increased by a factor of 2.05, or 277 samples. A check on the evaluation is provided by

$$s^2 \cong (0.50)\,(0.50)/277 = 9.0 \times 10^{-4}$$
$$s \cong 0.03.$$

SAMPLING PROCEDURE

Berry (1962) used the stratified systematic unaligned sampling procedure to select samples in similar type studies and has recommended this procedure (Berry and Baker, 1968, pp. 91-100) for use in accuracy testing of the land-use and land-cover maps produced by the U.S. Geological Survey. Cochran (1977, pp. 227-228) in discussing systematic sampling in two dimensions states 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. Cochran (1977, p. 221) cites 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).

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

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

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 squares to meet the desired sample size, the value of n is increased accordingly until the number of squares within the area exceed the desired minimum sample size.

The origin for the unaligned sample is selected by using a pair of random numbers to fix the coordinates of say the upper left unit. An additional pair of random numbers determine the horizontal coordinates of the remaining units in the first column of strata, and the vertical coordinates of the remaining units in the first row of strata. The constant intervals k_i (equal to the sides of the squares) then fixes the locations of all points.

Latham (1979) reported that in the preliminary experiment a systematic sampling system was applied, using a square grid with sides of the square equal to a distance equivalent to 3 minutes of latitude. The land-use and land-cover map was divided into 30-minute sections and the sampling grid was applied to each of these sections. The sample points were taken at the grid intersections, and the land-use and land-cover category recorded at each point and identified for later study and aggregation into one of the State sectors or for the entire State.

If in an experiment the land-use and land-cover category at a sample point are to be directly compared with the digitized land-use and land-cover data of the GIRAS information system, then the sample point locations must be recorded in the UTM coordinate system in a manner simulating the grid cells of the GIRAS data. If the sampling unit is taken as a point, the coordinates are recorded to the nearest 200 metres, which would correspond to the southeast corner of the 200-metre-square grid cell of the GIRAS data for the State of Pennsylvania. The recorded category could then be compared with the category of the grid cell.

If in such an operational experiment the land-use and land-cover category at the sampling unit are not to be directly compared with the GIRAS data, then the sample data do not need to be UTM located. For a point sample, the category at the point only need be recorded.

Since the 1958 and the 1977 land-use and land-cover maps are each on a map base of different map projections, and are each of different accuracies, it would be highly unlikely that corresponding points or areas on the two maps would truly represent the same ground location. However, land-use change determination in the future might use comparable maps or digitized data.

ESTIMATING LAND USE AND LAND COVER

The proportions of the land-use and land-cover category in the population are estimated from the sample. For sampling in the multinomial distribution, the sample proportion $p = a/n$ is an unbiased estimate of the population proportion. Cochran (1977, p. 60) indicates that when the sample units are classified into more than two classes, such as categories of land use and land cover, the sample mean for each category in the multinomial distribution is

$$p_j = a_j/n.$$

When the sample units are taken as points, an unbiased estimate of the variance of p_j for the multinomial population as $N \to \infty$, is given in the form of Cochran (1977, p. 52, 3.11) as

$$v(p_j) = s^2(p_j) = p_j(1 - p_j)/(n - 1),$$

Approximate simultaneous confidence limits are given in the form of Tortora (1978, 2.1 and 2.2) as

$$(p_L)_j = p_j \pm \sqrt{[\chi^2_{1,1-\alpha/k} v(p_j)]}.$$

A more exact formula for the end points of simultaneous confidence intervals, incorporating the correction for continuity, is that reported by Johnson and Kotz (1969, p. 289), and has been given above.

Confidence limits about the unbiased estimate for the population variance can be computed in an approximate manner by the method of Snedecor and Cochran (1967, pp. 74-76) for the variance in the normal distribution in the form

$$\frac{(n-1)v(p_j)}{\chi^2_{1,\alpha/k}} \leq v(p_j) \leq \frac{(n-1)v(p_j)}{\chi^2_{1,1-\alpha/k}} .$$

An unbiased estimate of the population total Y_j of the area for the jth category is computed in the manner

$$\hat{Y}_j = Np_j,$$

where N is the total area of the region taken as the population.

An unbiased estimate of the variance of the population total Y_j for the jth category is computed by error propagation as

$$v(\hat{Y}_j) = N^2\, v(p_j).$$

Confidence limits of the unbiased estimate of the population total Y_j may be computed in the normal form (Cochran, 1977, p. 95, 5.15, or p. 156, 6.15) as

$$(\hat{Y}_L)_j = \hat{Y}_j \pm Z\, \sqrt{[v(\hat{Y}_j)]},$$

assuming that the sample size is large enough that p_j is normally distributed and that $v(p_j)$ is well determined and "where Z is the normal deviate corresponding to the chosen confidence probability." If the sample size is less than 50, the percentage points of Student's distribution with $n - 1$ degrees of freedom are used.

Confidence limits about the unbiased estimate of the variance of the population total Y_j can be computed by the method of Snedecor and Cochran (1967, pp. 74-76) in the form

$$\frac{(n-1)v(y_j)}{\chi^2_{\alpha}} \leq v(\hat{Y}_j) \leq \frac{(n-1)v(y_j)}{\chi^2_{1-\alpha}} .$$

For values of χ^2 when degrees of freedom f exceed 100, Greenwood and Sandomire (1950) use the formula

$$\chi^2_f = f\left[1 - 2/9f + Z_{\alpha/2}\sqrt{(2/9f)}\right]^3.$$

This above formularization is based on simple random sampling in the normal distribution. The remarks of Cochran (1977, pp. 227-228) given earlier for the stratified systematic unaligned sampling technique indicate the applicability of these equations.

Latham (1979) reported on the preliminary experiment of the Pennsylvania project that point samples were taken to obtain proportions of each category within each of three particular sections of the State and for the entire State. Latham (1979) does not report on the number of samples for each category, either in the sections or in the State, but only that more than 3,000 samples were used. Since the preliminary experiment used only point samples for proportions, the estimates for land use and land cover on a proportion basis only was computed. If these samples are to be utilized in a follow-on experiment, then the use of point samples must be continued. The number of samples in each category in each county must be recorded. Unbiased estimates can then be obtained for the population proportions, totals, and variances, and confidence limits about these estimates computed.

ESTIMATING CHANGE IN LAND USE AND LAND COVER

Sampling at the time periods of populations X and Y. In the context of estimating change in land use and land cover, the designations x and y represent the categories of land-use and land-cover classification at the earlier time period (population X) and the later time period (population Y). The sample estimate of the population ratio R is given by Cochran (1977, p. 31, 2.38) as

$$\hat{R} = \bar{y}/\bar{x}.$$

Cochran (1977, p. 151) states that "if x_i is the value of y_i at some previous time, the ratio method uses the sample to estimate the relative change Y/X that has occurred since that time." Thus, the population ratio represents the change in land use and land cover.

According to Cochran (1977, p. 153) the ratio estimate is consistent and of negligible bias in sample sizes exceeding 30, and if the coefficient of variation of $\bar{x}$ and $\bar{y}$ are both less than 10 percent.

When point sampling for the proportion of land-use and land-cover categories in the multinomial distribution, the sample proportion p represents the sample mean $\bar{y}$. The sample estimate of the population ratio R_j for the jth category then is

$$\hat{R}_j = (p_y/p_x)_j,$$

where p_y, p_x are computed similarly to p_j given above, but for the Y and X populations, respectively.

An estimate of the variance of the population ratio R_j is computed by error propagation as

$$v(\hat{R}_j) = [1/(p_x^2)_j][v(p_y)_j + \hat{R}_j^2 v(p_x)_j],$$

where $v(p_y)_j$ and $v(p_x)_j$ are computed for the Y and X populations, respectively. Confidence limits are computed in the normal form.

An unbiased estimate of the change in the population total ΔY_j is computed in the manner

$$\Delta\hat{Y}_j = \hat{Y}_j - \hat{X}_j,$$

where $\hat{X}_j$ is computed similarly to $\hat{Y}_j$ given above, but for the X population.

An unbiased estimate of the variance of the change in population total ΔY_j is computed by error propagation as

$$v(\Delta\hat{Y}_j) = v(\hat{Y}_j) + v(\hat{X}_j)$$

where $v(\hat{X}_j)$ is computed for the X population. Confidence limits are computed in the normal form.

Population Y known and sampling at time period of population X. If the population values for the area of the land-use and land-cover categories at the present time are known, as the statistics obtained from the digitized values of the current land-use and land-cover maps produced by the U.S. Geological Survey, sampling might then be performed in the land-use and land-cover maps from the previous time.

An unbiased estimate of the change in the population total ΔY_j is computed in the manner

$$\Delta\hat{Y}_j = Y_j - \hat{X}_j,$$

where Y_j is known, and $\hat{X}_j$ is computed similarly to $\hat{Y}_j$ given above, but for the X population.

An unbiased estimate of the variance of the change in the population total ΔY_j is computed by error propagation as

$$v(\Delta\hat{Y}_j) = v(\hat{X}_j),$$

where $v(\hat{X}_j)$ is computed for the X population. Confidence limits are computed in the normal form.

If desired, the estimate of the population ratio R_j can be computed on the basis of proportions, as

$$\hat{R}_j = (P_y/p_x)_j,$$

where $(P_y)_j$ is the known population proportion, and $(p_x)_j$ is the sampled estimate of the population proportion for population X.

An estimate of the variance of the population ratio R_j is computed by error propagation as

$$v(\hat{R}_j) = [1/(p_x^2)_j][\hat{R}_j^2 v(p_x)_j],$$

where $v(p_x)_j$ is computed for the X population. Confidence limits are computed in the normal form.

When the change in land use and land cover is desired from one category to another, the change values must be determined from the sample units. This requires that the units be identified on both the previous and the current maps, and that the specific change information be recorded. The sample information would then no longer be one-dimensional by category alone, but two-dimensional on the basis of a from-to combination.

Latham (1979) reports on the change in land use and land cover for the various categories in the three sections and the State of Pennsylvania. In some instances the estimated p_j value for the categories is given as a percent of total area for both the 1958 and 1977 time periods. In other instances the percent of change is given. In no case is the number of samples for any of the categories, in any of the sections or in the State, given. Thus, it is not possible to determine estimates for variances or confidence limits from the data made available. Values of $\hat{R}$, the estimate of the ratio in land use and land cover, can be computed from the data in Table 1. As an example, for the Urban and Built-up category for the entire State,

$$\hat{R} = 6.8\%/3.3\% = 2.1.$$

Conversion to percentage change requires subtraction of the base of 1 from R. Thus, the example change is 110 percent.

EXTENSION OF ESTIMATES FROM SUBPOPULATIONS

Up to this point, this report has been concerned with the population. Cochran (1977, pp. 89-111) covers the situation where the population of N units is composed of nonoverlapping subpopulations, whose sum of units comprise the entire population in the form

$$N = N_1 + N_2 + \ldots + N_L.$$

These subpopulations are called strata. Samples are drawn independently within these subpopulations with the sample size $n_1, n_2, \ldots, n_L$, respectively.

In the equations to follow for sampling within subpopulations, the subscript h denotes the subpopulation and the subscript st is for stratified, or that the population has been divided into subpopulations and therefore the estimates of the population parameters are stratified estimates.

Stratified estimates for proportions. In accordance with Cochran (1977, p. 107), the proportions of points of each category in the sample from the hth stratum is

$$p_h = a_h/n_h,$$

and an unbiased estimate of the proportion in the whole population, based on stratified random sampling is in the form of Cochran (1977, p. 107, 5.52),

$$p_{st} = \sum_{h=1}^{L} [W_h p_h].$$

where W_h, the stratum weight, is computed in the form of Cochran (1977, p. 90) as

$$W_h = N_h/N,$$

and is the ratio of the area N_h of the subpopulation to the area N of the whole population.

Cochran (1977, p. 108, 5.56) indicates that the sample estimate of the variance of the population proportion p_{st} is

$$v(p_{st}) = \sum_{h=1}^{L} \frac{W_h^2\, p_h\, q_h}{n_h - 1},$$

where $q_h = 100 - p_h$. Confidence limits are computed in the normal form.

An unbiased estimate for the population total $(Y_{st})_j$ is computed by

$$(\hat{Y}_{st})_j = N(p_{st})_j$$

where N is the total area of the whole population.

An unbiased estimate of the variance of the population total $(Y_{st})_j$ is computed by

$$v(\hat{Y}_{st})_j = N^2 v(p_{st})_j.$$

Confidence limits are computed in the normal form.

'Combined' ratio estimate. According to Cochran (1977, p. 166) the so-called 'combined' ratio estimator is much less subject to risk of bias than another so-called 'separate' ratio estimator which is obtained by a summation of totals for each subpopulation. The estimate of the 'combined' ratio R_c is reported by Cochran (1977, p. 166, 6.48) to be computed from the stratified estimates of the population means, and when modified for proportions as

$$(\hat{R}_c)_j = [(p_y)_{st}/(p_x)_{st}]_j,$$

where $(p_y)_{st}$ and $(p_x)_{st}$ are computed for the Y and X populations, respectively.

An estimate of the variance of the combined ratio R_c is computed by error propagation in the form of Cochran (1977, p. 155, 6.13) as

$$v(R_c)_j = [1/(p_x^2)_{st}]_j[v(p_y)_{st} + \hat{R}_c^2 v(p_x)_{st}]_j.$$

An unbiased estimate of the change in the population total $(\Delta Y_{st})_j$ is computed in the manner

$$(\Delta\hat{Y}_{st})_j = (\hat{Y}_{st})_j - (\hat{X}_{st})_j,$$

where $(\hat{X}_{st})_j$ is computed for the X population.

An estimate of the variance and the confidence limits are computed as before.

If the population values for the present time are known, and sampling is performed in the maps of the previous time, then an unbiased estimate of the change in the population total $(\Delta Y_{st})_j$ is computed in the manner

$$(\Delta\hat{Y}_{st})_j = Y_j - (\hat{X}_{st})_j$$

where $(\hat{X}_{st})_j$ is computed for the X population.

An estimate of the variance and the confidence limits are computed as before.

If desired, the estimate of the stratified population ratio $(R_{st})_j$ can be computed on the basis of proportions as

$$(\hat{R}_{st})_j = [P_y/(p_x)_{st}]_j$$

where $(P_y)_j$ is the known population proportion.

An estimate of the variance and the confidence limits are computed as before.

In the Pennsylvania project, information is desired on a county basis first, and then cumulated for the entire State. In this case the county is the subpopulation and the State is the population. All of the preceding work for the sample design pertains to the individual counties, and finally the information is to be combined for the State totals. Since the report by Latham (1979) for the preliminary experiment contains only the percentage information for several representative sections of the State, an example cannot be developed. The information given by Latham (1979) for the whole State is on the basis of a sample size of 3,000 points within the State as the population.

Conclusion

This methodology is given as a guide to sample design of experiments leading to determination of change in area of land-use and land-cover category. The preliminary experiment of Latham (1979) has been used when applicable as an example of this methodology.

For the analyst embarking on an experiment of this type, the principles of sample design to be considered are the sample distribution, the sample size, the sampling procedure, estimation of the population means, variances and confidence limits, estimation of change in the population values with variances and confidence limits, and extension of the parameters from subpopulations to the overall population.

References

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer, 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, Agriculture Handbook no. 237.

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

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

Freund, J. E., and F. J. Williams, 1972. *Elementary business statistics,* 2nd Edition, Ingelwood Cliffs, New Jersey, Prentice-Hall.

Goodman, L. A., 1965. On simultaneous confidence intervals for multinomial proportions, *Technometrics,* v. 7, no. 2, pp. 247-254.

Greenwood, J. A., and M. M. Sandomire, 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, pp. 257-260.

Johnson, Norman and Samuel Kotz, 1969. *Discrete distributions,* Boston, Houghton Mifflin Company.

Klimm, L. E., 1958. *Description of a land-use map of Pennsylvania, scale 1:250,000,* University of Pennsylviana, Warton School of Finance and Commerce, Geography and Industry Department, Technical Report no. 2, October 1, 7 p.

Latham, J. P., 1979. Land use change analysis from archival and currently remotely sensed patterns in Pennsylvania, paper presented at annual meeting of

American Society of Photogrammetry, Washington, D.C., March 23.

Matern, B., 1947. Methods of estimating the accuracy of line and sample plots surveys, *Medd. fr. Statens Skogsforsknings Instut.*, 36, pp. 1-38

Mitchell W. B., S. C. Guptill, K. E. Anderson, R. G. Fegeas, and C. A. Hallam, 1977. *GIRAS—A geographic information retrieval and analysis system for handling land use and land cover data,* U.S. Geological Survey Professional Paper 1059, 16 p.

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

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

(Received 31 July 1981; accepted 25 December 1981; revised 14 January 1982)

Summer School
Remote Sensing for Land-Use Inventories

Joint Research Centre, Ispra, Italy
13 September-1 October 1982

Organized in association with the International Institute for Aerial Surveying and Earth Sciences, Enschede, The Netherlands, the program will include

- Review of existing methods and techniques of crop and land-used inventory and mapping used in Europe and in the Mediterranean Region.
- Air and space photography and imagery for land-use mapping. Digital image processing for land-use mapping.
- Place and value of field information and data collection methods and techniques.
- Case studies illustrating practical applications of remote sensing in the Mediterranean Region.

The program will be practical, lectures being provided only insofar as they are a necessary introduction to practical exercises.

For further information please contact

ISPRA-Courses Secretariat
Joint Research Centre
21020 Ispra (VA) Italy

Forthcoming Articles

Stan Aronoff, Classification Accuracy: A User Approach.
Stan Aronoff, The Map Accuracy Report: A User's View.
H. H. Elfick and *M. J. Fletcher*, The Qasco SD-4.
Lawrence W. Fritz, Testing Procedures for Analytical Plotters.
Clifford W. Greve, APPS-IV, Improving the Basic Instrument.
P. D. Grumstrup, M. P. Meyer, R. J. Gustafson, and *E. R. Hndrickson*, Aerial Photographic Assessment of Transmission Line Structure Impact on Agricultural Crop Production.
Alex Hasler, The Analytical Stereoplotter Wild Aviolyt BC 1.
M. L. Imhoff, G. W. Petersen, S. G. Sykes, and *J. R. Irons,* Digital Overlay of Cartographic Information on Landsat MSS Data for Soil Surveys.
J. A. Kirchner, S. Youkhana, and *J. A. Smith,* Influence of Sky Radiance Distribution on the Ratio Technique for Estimating Bidirectional Reflectance.
Donald G. Leckie, An Error Analysis of Thermal Infrared Line-Scan Data for Quantitative Studies.
Timothy P. Roberts and *Hans J. Wehrli*, A Unique Concept in Analytical Plotters.
Richard H. Seymour, US-2 Analytical Stereoplotter.
Brian J. Turner and *Donald N. Thompson*, Barrier Island Vegetation Mapping Using Digitized Aerial Photography.
Howard Turner, The Canadian Marconi ANAPLOT II System.
J. E. Unruh and *E. M. Mikhail*, Mensuration Tests Using Digital Images.
Kenneth C. Winterberger and *Karl M. Hegg*, A Low-Cost Photo-Scale Protractor for Small Scale Photography.