DOCUMENT RESUME

ED 101 008 TM 004 135

AUTHOR Echternacht, Gary J.
TITLE Sample Design for "Other Nations, Other Peoples." Research Memorandum No. 74-4.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO RM-74-4
PUB DATE Apr 74
NOTE 17p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Age; Analysis of Variance; Educational Experience; *Elementary School Students; Elementary Secondary Education; *Ethnocentrism; Geographic Location; *Research Design; *Sampling; *Secondary School Students; Socioeconomic Status; Statistical Analysis; Urbanization
IDENTIFIERS *Other Nations Other Peoples

ABSTRACT

"Other Nations, Other Peoples", a research project of the U. S. Office of Education, studied the degree of ethnocentrism present in elementary and secondary public school students. The research design was three-way, calling for stratified samples according to county, school and students. The purpose of the study was to determine how ethnocentrism related to age and sex, knowledge of other peoples, educational experiences and geographical and cultural factors. Counties were classified by population size. Fifty counties were selected, of which 25 had populations over 500,000 persons. The smaller counties were stratified according to educational level, income level and region. Two schools were selected from each county and ten students were selected from each school. The resulting formulas for estimation are presented. (SM)

RM-74-4

# RESEARCH MEMORANDUM

## SAMPLE DESIGN FOR "OTHER NATIONS, OTHER PEOPLES"

Gary J. Echternacht





Educational Testing Service
Princeton, New Jersey

April 1974

# Sample Design for "Other Nations, Other Peoples"

Gary J. Echternacht

Educational Testing Service

## Introduction

Ethnocentrism, or ethnic self-love and out-group hostility, may not be a universal trait of man although past research has continually found evidence of it in many human populations. The U. S. Office of Education has sponsored a research project, titled "Other Nations, Other Peoples," whose purpose is to determine the degree of ethnocentrism present in various public school subpopulations and the correlations of facets of ethnocentrism with other variables.

Specifically, how ethnocentric are elementary and secondary pupils in United States public schools? How ethnocentric are their teachers? And how do measures of ethnocentrism (including stereotyping of "out-groups") relate to: (1) knowledge of other nations and peoples, (2) personal characteristics such as age and sex, (3) educational experiences, and (4) the geographical and cultural context.

It is reasonable to speculate that many American school children--and their teachers--are ethnocentric. They tend to apply stereotypes to "foreigners" and maintain considerable social distance from them. Theory, data, and informal observation support this contention.

Of particular interest is the relationship between ethnocentrism and age. The egocentrism of very young children is well documented in the research literature. The loss of egocentrism, in both the cognitive and personal-social sense, is viewed as a worthy objective of early education and a sign of increasing maturity. The literature contains little information on the relationship between the two constructs, egocentrism and

ethnocentrism. Yet, on the basis of knowledge about egocentrism in child development, elementary pupils could be expected to be more ethnocentric than older students. Two resulting hypotheses are: (1) attitudes toward other peoples becomes more differentiated with age, and (2) attitudes expressed by young children can be related more directly to school and home experiences than can those of older subjects.

In this research memorandum, the sampling design, its rationale, and the resulting formulas for estimation are presented. The results of this study should throw some light on the interrelationships among attitudes toward, interest in, and knowledge about other nations and other peoples. In addition, the study should result in information that will be helpful in revising school curricula and teacher training programs in the direction of fostering increased understanding of other peoples.

## Sample Design

Specifications for this study required that questionnaires be administered to a sample of 10 students from each of 100 randomly selected public schools for students in the 4th, 8th, and 12th grades. Thus, each grade level must have a sample of 1,000 students. Students were selected using a three-stage design. The first stage resulted in the selection of 50 counties from the counties and divisions making up the 50 states and the District of Columbia. In the second, two schools were selected within each selected county independently for each grade level. Ten students were selected within the appropriate grade level from each selected school for the final stage.

The 50 county first-stage sample was obtained using a stratified random sampling procedure. Counties in the United States and the District of Columbia were classified into 25 strata. One stratum consisted of the

District of Columbia and 25 counties containing a central city having more than 500,000 inhabitants, according to the 1970 Census of Population. The remaining counties were classified into eight groups defined by combinations of geographic region (four regions using the Census classification) and whether a county belonged to a Standard Metropolitan Statistical Area (SMSA).

For each county, statistics on the number of people between the ages of 5 and 17 years inclusive, median family income, and median years of schooling completed were obtained from the County and City Data Book, 1972. Within each of the eight defined groups, counties were ranked on the basis of median family income. Counties ranking in the lowest third of median family income were identified and grouped; operationally, they were termed low income counties. The remaining counties were ranked on the basis of median years of schooling completed, with the upper half of those counties grouped and termed high education counties. The remaining counties were simply termed group 3. The result of this subgrouping is illustrated in Figure 1 showing the bivariate distribution of median income and years of schooling completed. The result was a stratification scheme where all counties were classified into 25 strata defined by: (1) counties with central cities having more than 500,000 inhabitants and (2) all combinations of region, whether in an SMSA, and subgrouping (low income, high education, group 3). Construction of the strata is illustrated in Figure 2.

---

Insert Figures 1 and 2 about here

---

Two counties were selected within each stratum. Selections were made with replacement and with probability proportional to the size of

the school age population in the counties. The number of people aged 5 through 17 served as the measure of school age population.

After the 50 counties were selected, three lists of schools were compiled for each county, one for each grade level. All schools located geographically within the selected county containing the grades of interest were listed with the estimate of the number of students in that grade. The lists of schools, with their enrollment and grade spans, were obtained from state educational directories for 1972-73 or 1973-74, telephone calls to state departments of education, and county superintendents. Phone calls were made to state departments of education to verify the accuracy of the listings obtained from directories.

Two schools were selected from each of the three lists for selected counties. Selections were made with replacement and, again, with probability proportional to estimated school size. When exact figures of the number of students in a grade were available, those figures were used as a measure of size. If no figures were given, the number of students in a grade was estimated by dividing the enrollment for the school by the number of grades included in the school.

After schools were identified, a simple random sample of 10 students was selected from rosters provided by the schools. The total number of students in the school at the particular grade levels was obtained and retained for the analysis.

Design Rationale

The most important consideration in forming this sample design was the need to quickly produce a sampling design with a minimum expenditure since the project budget did not provide for an extensive or comprehensive design. Statistics were to be reported by geographic region, and no other

breakdowns were required. Thus, it was desirable to sample approximately the same number of students from each of the geographic regions.

The sample selection could have been performed using either a two- or three-stage design with existing on-site data files. If a two-stage design were used--selecting schools in the first stage and students in the second--the U. S. Office of Education's elementary and secondary school universe tape would have been used. This was undesirable since: (1) the data on that tape were collected for the 1968-69 school year and were considered out of date, (2) significant coverage errors had been found (see Hilton, et al.,1973), and (3) past experience had resulted in inconsistencies in grade span and enrollment statistics. Thus, a three-stage design, such as the resulting one, was believed the best approach.

The number of schools in the sample design was specified as 100, regardless of the sample design. Thus, the only variables in the design were the number of strata, counties within a stratum, and schools selected within a county. Since orthogonal designs were believed desirable both from the point of view of computation and analysis potential, the same number of schools were selected within each county, and the same number of counties were selected within each stratum. The analysis of variance table, showing the sources of variation and the degrees of freedom is given in Table 1 for an arbitrary grade level.

---

Insert Table 1 about here

---

As can be seen by examining the table, the product of the numbers s, r, and t must be 100. The different possible designs meeting this requirement are given in Table 2. Since it was desirable to retain the potential to make comparisons between schools within counties and between counties within strata or make estimates of the variance of county and stratum means, at least two counties and two schools were selected at each stage.

---

Insert Table 2 about here

---

Since many rural counties have very few schools--in some cases there are county high schools only--the number of schools selected within a county must be a minimum. For this reason, designs 3, 4, 7, 8, 9, 10, 11, and 12 were eliminated as possibilities. Designs 5 and 6 were eliminated from contention because they did not make sufficient use of any county stratification. The decision to use design 1 rather than design 2 was made because (1) design 1 offered a greater opportunity to stratify counties, and (2) if five counties were selected from a stratum with replacement, the selection of a county more than once was more likely, and that was not considered desirable.

Since one of the study purposes was to make regional comparisons, the number of strata created for each region was designated to be approximately equal. This resulted in each of the four geographic regions of the county being allocated six strata. The remaining stratum, not accounted for by region, was designated to contain counties with extremely large central

cities. It was believed that students in these counties would possibly have more contact with people from other cultures and thus have different attitudes than persons from other counties. An arbitrary population of 500,000 inhabitants in a city was set as the criterion for the large-city counties. This resulted in counties containing the 25 largest cities and the District of Columbia being placed in this stratum.

Two other factors were believed to have an effect on attitudes towards other nations and peoples. The degree of urbanization of a county was believed significantly related to attitudes. Children from predominantly rural counties were likely to have different attitudes than those from predominantly urban or suburban counties. Thus, within each region, strata were created for predominantly rural counties (counties <u>not</u> belonging to an SMSA) and predominantly urban counties (counties belonging to an SMSA).

Socioeconomic status was also believed related to attitudes towards other nations and peoples. Both income and education variables have been used to define socioeconomic status in the past, and it was desired to use both variables in this design. People from lower income families were believed to have less opportunity for travel and meeting new people and cultures. On the other hand, people who had much schooling were believed more likely to have traveled more extensively than those with less schooling. The most desirable strategy would have been to develop an adequate composite measure of socioeconomic status, but time and cost requirements did not permit such an effort. Thus, the resulting plan called for creating strata within region and SMSA designation by taking the lowest ranking

third of the counties within that grouping in terms of median income as one stratum and the upper half of the remaining counties in terms of median years of school completed as a second stratum and relegating the remaining counties to a third stratum.

Estimation

The estimation problem in this study was one of estimating stratum totals and their variances. These estimates can be combined over the strata to form population estimates by the methods given in sampling texts (see Cochran, 1963, p.88 ). Observations take the form of $y_{ijk}$ , where the subscript i indexes counties, j indexes schools within counties, and j indexes students within schools. The problem can be formulated more generally as one of estimating the stratum total, Y , given a design where r counties are selected with replacement, each with probability $z_r$ ; s schools are then selected from the counties with replacement and with probabilities $z_{rs}$ ; and finally, t students are selected from the schools with simple random sampling (without replacement) from the selected schools. Suppose, further, there are N students in the stratum and the students' responses are designated $y_i$ (that is, the $y_{ijk}$ are transformed to $y_{i'}$ , and for convenience the prime is dropped and responses are denoted as simply $y_i$).

If the notion of indicator variables, developed by Cornfield (1944), is used, the desired estimate takes the form of

$$\hat{Y} = \sum_{i=1}^{N} w_i a_i y_i \qquad (1)$$

where $y_i$ represents the response of the ith person, $w_i$ is a weight

given the ith person, and $a_i$ is a variable indicating the number of times the ith person appears in the sample, $a_i = 0, 1, \ldots, rs$ .

With repeated sampling, only $a_i$ is a variable, and its expectation is

$$E(a_i) = \sum_{x=0}^{rs} x \text{ prob } (a_i = x)$$

$$= \sum_{x=0}^{rs} x \binom{rs}{x} \phi_i^x \ (1-\phi_i)^{rs-x}$$

$$= rs\phi_i \tag{2}$$

where $\phi_i = z_r z_{rs} t/N_{rs}$ , the probability the ith unit appears in the sample, and $N_{rs}$ indicates the number of students in school (r,s).

Similarly, the variance and covariance of the $a_i$'s is given by,

$$V(a_i) = rs\phi_i \ (1-\phi_i) \tag{3}$$

and

$$Cov(a_i, a_j) = rs\phi_i\phi_j \ , \ i \neq j \ . \tag{4}$$

If the expected value of the estimate $\hat{Y}$ is taken, using (2), the result becomes

$$E(\hat{Y}) = E\left(\sum_{i=1}^{N} w_i a_i y_i\right)$$

$$= \sum_{i=1}^{N} w_i E(a_i) y_i$$

$$= \sum_{i=1}^{N} w_i (rs) \phi_i y_i$$

and $E(\hat{Y}) = \sum_{i=1}^{N} y_i = Y$ if $w_i = (rs\phi_i)^{-1}$ . Choosing the weight, $w_i$ , to be that value, the variance of the estimate becomes

$$V(\hat{Y}) = V\left(\sum_{i=1}^{N} \frac{a_i y_i}{rs\phi_i}\right)$$

$$= \sum_{i=1}^{N} \frac{V(a_i)y_i}{(rs)^2\phi_i} + \sum_{i\neq j}^{N} \frac{Cov(a_i a_j)}{(rs)^2\phi_i\phi_j} y_i y_j$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{rs\phi_i} - \frac{Y^2}{rs} \quad .$$

Since the variance and covariance of the indicator variables is known, the two expectations, $E(a_i^2)$ and $E(a_i a_j)$ can be obtained directly. Using the definitions of variance and covariance with (3) and (4), these expectations become

$$E(a_i^2) = rs\phi_i\,(1 + (rs-1)\,\phi_i)$$

and

$$E(a_i a_j) = rs\,(rs-1)\,\phi_i\phi_j \quad , \quad i\neq j$$

so that

$$E(\hat{Y}^2) = E\left(\left(\sum_{i=1}^{N} \frac{a_i y_i}{rs\phi_i}\right)^2\right)$$

$$= \sum_{i=1}^{N} \frac{E(a_i^2)y_i^2}{(rs)^2\phi_i^2} + \sum_{i\neq j}^{N} \frac{E(a_i a_j)}{(rs)^2\phi_i\phi_j} y_i y_j$$

$$= \sum_{i=1}^{N} \frac{(1 + (rs-1)\phi_i)}{rs\phi_i} y_i^2 + \sum_{i\neq j}^{N} \frac{rs-1}{rs} y_i y_j$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{rs\phi_i} + \frac{rs-1}{rs} Y^2 .$$

Now, one can see that $Y^2$ can be estimated using

$$E\left(\sum_{i=1}^{N} \frac{a_i}{rs\phi_i} y_i^2\right) = \sum_{i=1}^{N} y_i^2$$

and

$$E\left(\sum_{i\neq j}^{N} \frac{a_i a_j}{rs(rs-1)\phi_i\phi_j} y_i y_j\right) = \sum_{i\neq j}^{N} y_i y_j$$

so that

$$\tilde{Y}^2 = \sum_{i=1}^{N} \frac{a_i y_i^2}{rs\phi_i} + \sum_{i\neq j}^{N} \frac{a_i a_j}{rs(rs-1)} \frac{y_i y_j}{\phi_i\phi_j}$$

provides an unbiased estimate of $Y^2$ and

$$V(\hat{Y}) = \hat{Y}^2 - \frac{2rs-1}{rs} \tilde{Y}^2$$

provides an unbiased estimate of $V(\hat{Y})$ .

## References

Cochran, W. G. Sampling techniques (2nd ed.). New York: John Wiley & Sons, Inc., 1963.

Cornfield, J. On samples from finite populations. Journal of the American Statistical Association, 1944, 39, 236-239.

Hilton, T. L., Rhett, H., Broudy, I. L., Bower, C., Carter, M. M., Creech, F. R., & Echternacht, G. The base-year survey of the national longitudinal study of the high school class of 1972. Princeton, N. J.: Educational Testing Service, 1973. Final report for Contract OEC-0-72-0903, U. S. Office of Education, June 1973.

Table 1

Analysis of Variance

| Source | Degrees of Freedom |
|---|---|
| Strata | s-1 |
| Counties/Strata | s(r-1) |
| Schools/Counties, Strata | sr(t-1) |
| Students/Schools, Counties, Strata | 900 |
| Total | 999 |

Table 2

Possible Designs for the Numbers of Strata, Counties, and Schools

| Design Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Strata = s | 25 | 10 | 10 | 4 | 2 | 5 | 2 | 5 | 2 | 5 | 2 | 5 |
| Counties = r | 2 | 5 | 2 | 5 | 25 | 10 | 10 | 4 | 2 | 2 | 5 | 5 |
| Schools = t | 2 | 2 | 5 | 5 | 2 | 2 | 5 | 5 | 25 | 10 | 10 | 4 |

Figure 1

Subgrouping Within Groups Defined by Region and SMSA

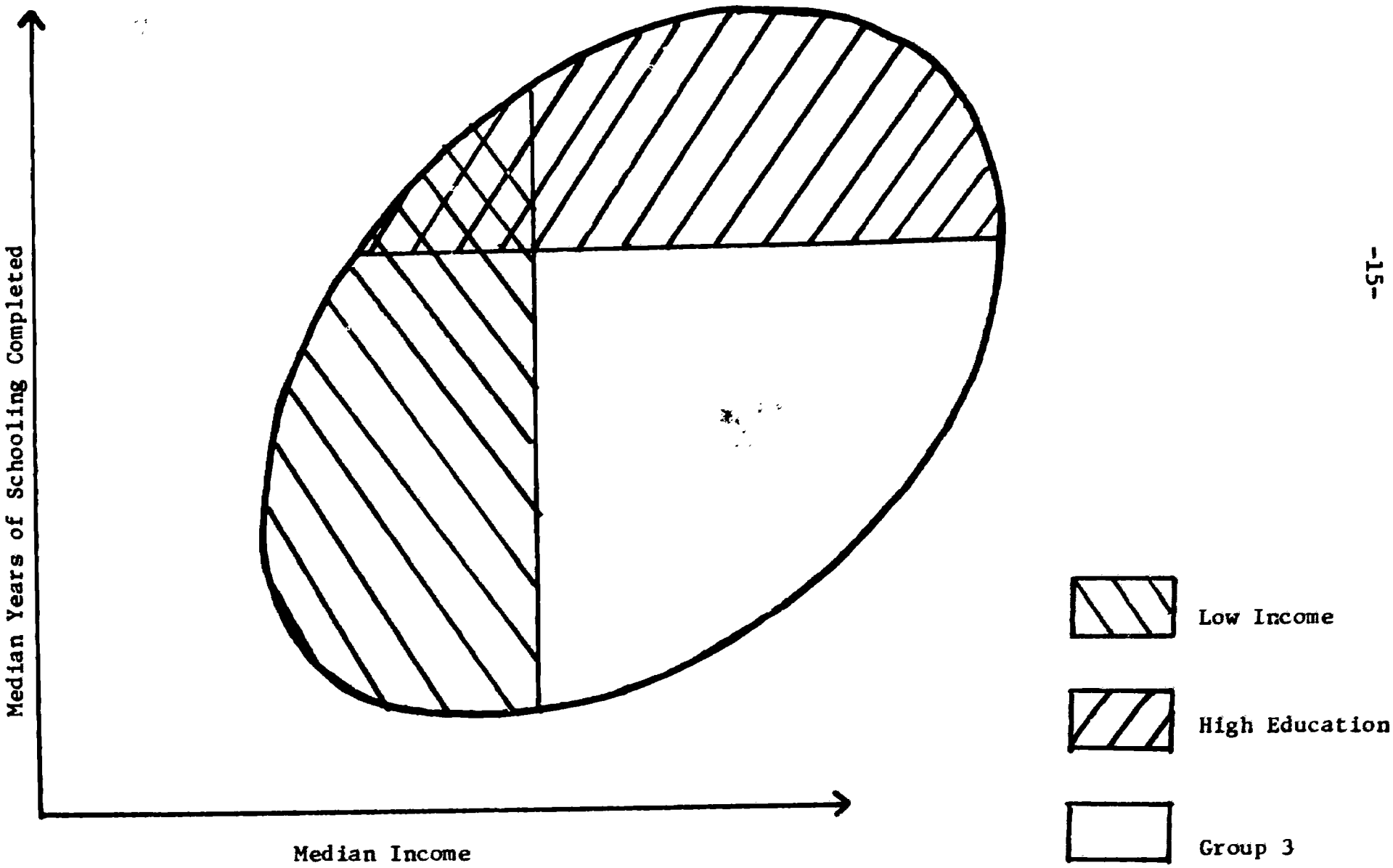
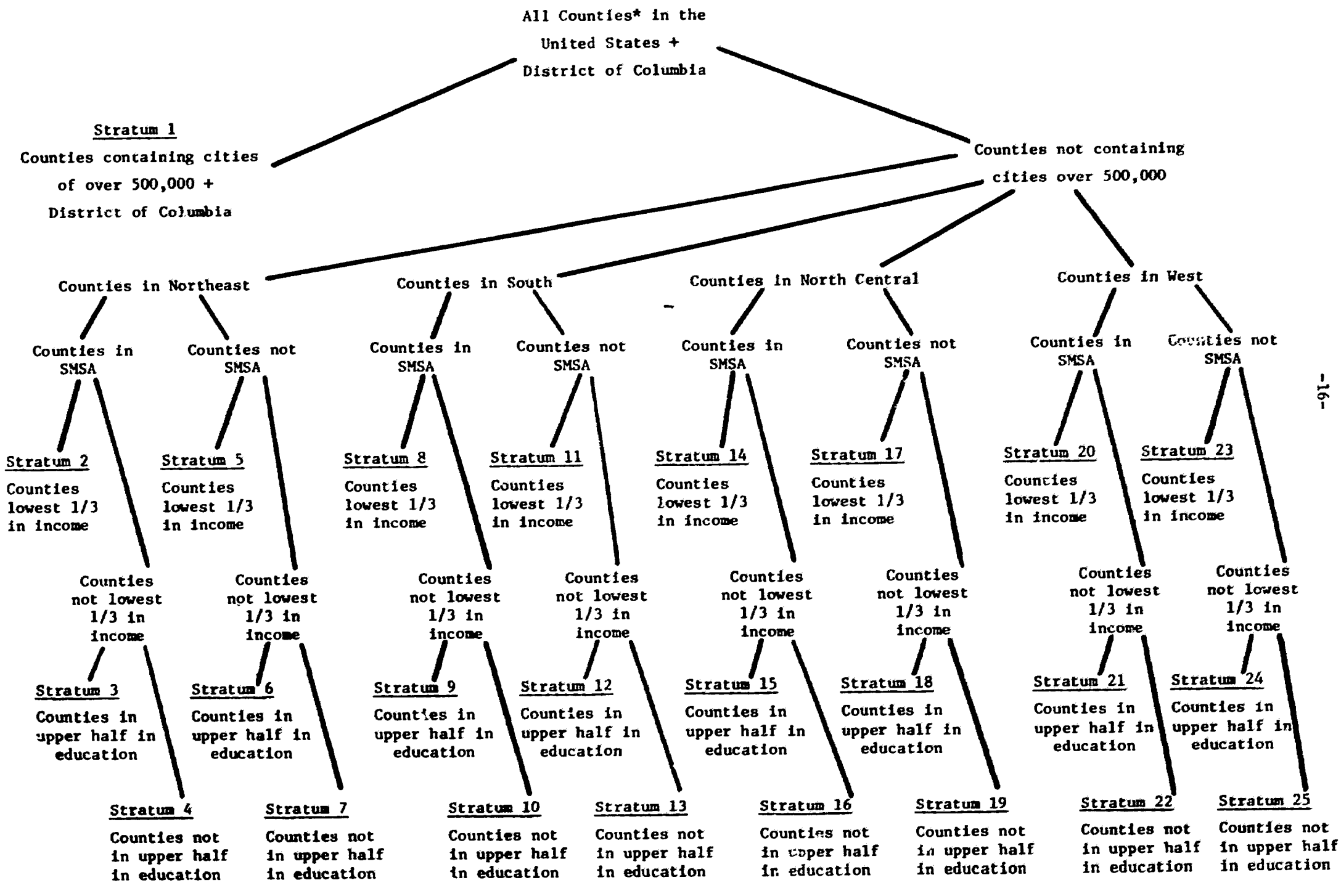


Figure 2

Formation of County Strata




*Includes Census Divisions in Alaska and Parishes in Louisiana.