ESTIMATING INTRACLUSTER HOMOGENEITY IN MULTISTAGE SAMPLES

by

Anne Fakler Clemmer

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1482T

February 1985

Estimating Intracluster Homogeneity In Multistage Samples

by

Anne Fakler Clemmer

A Dissertation submitted to the faculty of the
University of North Carolina at Chapel Hill in
partial fulfillment of the requirements for
the degree of Doctor of Public Health in the
Department of Biostatistics

Chapel Hill

1985

Approved by:

Advisor

Reader

Reader

ABSTRACT

CLEMMER, ANNE FAKLER, Estimating Intracluster Homogeneity In Multistage Samples. (Under the direction of WILLIAM D. KALSBEEK.)

Multistage cluster samples are often used in social science because clusters of households are cost effective. However, sampling elements within the clusters are homogeneous, thus negating independence assumptions and increasing the variance compared to simple random sampling. This research quantifies when each of two formulas is appropriate to estimate rate of homogeneity (roh). Roh is a measure of how much sampling elements are alike within clusters. It is speculated that one formula performs better when several different domain categories are present within each sampling cluster (cross-class domains) and the other better when only one domain category is present per cluster (segregated domains). As a secondary issue, this research investigates how the relationship between the domain and analysis variables may determine the appropriate formula for roh.

To investigate the primary issue, two-stage cluster samples were selected from the National Medical Care Expenditure Survey (NMCES) household file; sample values for roh were calculated using each formula and compared with the population value from the entire file. The secondary issue was explored in two steps. First, the average domain ratio (ADR) or ratio of average rate of homogeneity across all categories in a particular domain to roh for the population total was calculated for each sample. The ADR was calculated using each of the two estimation

procedures and compared with the true population value. Secondly, the domain degree of segregation or cross-classness in each sample was modeled and the fit evaluated separately for each estimation procedure.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

ACKNOWLEDGMENTS

I am especially indebted to my advisor, Dr. William D. Kalsbeek, for his patient guidance, strong encouragement, and helpful suggestions at each step of this research effort. The valuable comments and support of the other members of my committee, Dr. Shirley A. A. Beresford, Dr. James R. Chromy, Dr. Dennis B. Gillings, Dr. Barbara S. Hulka, and Dr. Gary G. Koch are also gratefully acknowledged.

Many thanks are additionally due to the Veterans Administration Medical Center in Durham, North Carolina for providing financial support for this research project.

I also am grateful to Ruth Drum for her skillful and conscientious typing of this manuscript.

Finally, very special thanks are due to my husband, Tom, my children, Hew and Nikki Dalhouse, and my parents, Mr. and Mrs. Paul C. Fakler, whose unwavering support, patience, and understanding, I have constantly relied on.

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

FIGURE | PAGE

CHAPTER 1

Statement Of The Problem

1.1 What Is Intracluster Homogeneity And Why Is It Important

Data collection costs must be considered in sample survey design. When costs are considered, invariably a cluster sample will be chosen over a simple random sample. Cluster sampling takes advantage of the fact that units of the population are often found in close geographic proximity. Sampling clusters rather than elements in the population reduces data collection costs.

If costs were the only consideration, the survey would be conducted in a few immense clusters. However, the important item to consider is the number of independent observations provided by each cluster. The number of independent observations per cluster is affected by the homogeneity of the cluster (Sudman, 1976; Kish, Groves, and Krotki, 1976). Less information is collected in a homogeneous cluster than in a nonhomogeneous one. The homogeneity of the cluster depends on the variable of interest in the survey. Sample clusters may be racially homogeneous but heterogeneous with respect to date of birth. Thus, one may not refer to cluster homogeneity for a whole design unless it is possible to combine the measures of cluster homogeneity for all variables of interest into a single index number expressive of average cluster homogeneity across all variables (Kish, Groves, and Krotki, 1976).

The importance of intracluster homogeneity has been observed by several authors. Kish and Frankel (1974) note that clustering induces large and positive correlations between element values with a resultant increase in variance. They report increases in variance for both linear statistics, such as simple means, proportions, and totals, and nonlinear statistics, like ratio means, regression coefficients, and correlations. Holt, Smith, and Winter (1980) and Nathan and Holt (1980) have investigated the effect of intracluster homogeneity on variance of regression coefficients in complex surveys. Ordinary least squares estimates of variance were found in both papers to be underestimates. It was speculated that this situation occurred because the particulars of the design such as intracluster homogeneity and stratification were not taken into account. The first authors observed that probability-weighted methods provided much better results for equal-probability designs with widely different selection probabilities. The best results were obtained using maximum likelihood theory where the estimator required that values of a design variable be known for all units in the population. This finding again points out the importance of including clustering and stratification effects in the estimation procedure. Hansen, Madow, and Tepping (1983) in their comparison of inference from model-dependent and probability-sampling surveys stress the relevance of design in sample surveys. They suggest that intracluster homogeneity is particularly important and that failure to recognize such effects may lead to underestimates of variance and understatement of confidence intervals. Lastly, the importance of including clustering effects in tests of hypotheses from complex survey data has been researched by Fellegi (1980) and Rao and Scott (1981). Both papers show that under a complex sample design, the usual chi-squared test statistics for goodness of fit and independence are asymptotically distributed as the weighted sum of independent chi-squared variables where the weights are functions of the design effects. The design effect is a function of cluster homogeneity as shown in equation 1.5.

The design effect (Kish, 1965) is defined as the variance of the sample under consideration, usually a cluster design, to the variance of a simple random sample of the same total size. Generally the design effect is greater than one implying the cluster sample provided a larger estimate of variance than the simple random sample. The larger estimated variance is normally caused by positive correlation among elements within the clusters. Sometimes the design effect is one or less than one implying the estimated variance from the cluster sample is the same as or less than the simple random sample. A design effect of one or less may imply an intracluster measure of homogeneity of zero or less.

1.2 Usefulness Of Intracluster Homogeneity

1.2.1 Population Measures For Cluster Homogeneity: Clusters Of Equal Sizes

Fisher (14th Edition, 1970) defines intraclass correlation as the extent to which elements from the same group (cluster) deviate from each other. For a specific population, Fisher considered pairs of brothers (elements) from the same family (cluster). In general, for a particular population composed of A clusters, B elements per cluster, and observation Y_{ij} on element-j from cluster -i, the intraclass correlation coefficient is defined as:

$$\text{Rho}\ (\bar{\bar{Y}}) = \frac{\Sigma_i^A \ \Sigma_{j\neq k}^B \ (Y_{ij} - \bar{\bar{Y}})\ (Y_{ik} - \bar{\bar{Y}})}{A\ B\ (B-1)\ \dfrac{\Sigma_i^A \ \Sigma_j^B \ (Y_{ij} - \bar{\bar{Y}})^2}{AB}} \tag{1.1}$$

$$= \frac{\Sigma_i^A \ \Sigma_{j\neq k}^B \ (Y_{ij} - \bar{\bar{Y}})\ (Y_{ik} - \bar{\bar{Y}})}{A\ B(B-1)\ \sigma^2}$$

$$\text{where } \bar{\bar{Y}} = \frac{\Sigma_i^A \ \Sigma_j^B \ Y_{ij}}{A\ B} \text{ and } \sigma^2 = \frac{\Sigma_i^A \ \Sigma_j^B \ (Y_{ij} - \bar{\bar{Y}})^2}{A\ B}.$$

It is interesting that this definition existed prior to the analysis of variance (Kotz and Johnson, 1984) and that Fisher introduced it in his book preceding the analysis of variance.

In sampling literature, equation 1.1 is defined for the special case of equal numbers of elements (B) per cluster (A). Suppose a particular cluster contains no elements. Does one count the cluster as part of A? Yates (1981, page 289) takes care to point out that the number of clusters (A) must be taken to be the number of first-stage units for which the number of elements (B) is not zero. This detail becomes important when domains are considered. We adopted this convention for computing Roh.

Domain variables, consisting of domain categories, are subgroups of the population for which separate estimates are planned in the sample design (Kish, 1965). For example, suppose the domain variable is region and the particular category is the Northeastern region of the United States; an estimate is desired of the proportion of people who are over age 65 in the Northeast. Then A is the number of clusters in the Northeast containing people who are over age 65. More complicated examples are provided later.

Harris (1913) provided an important simplification by observing that

$$B \Sigma_i^A \ (\overline{Y}_i - \bar{\bar{Y}})^2 = A\sigma^2 \ (1 + (B - 1) \text{ Rho } (\bar{\bar{Y}})) \tag{1.2}$$

where $\overline{Y}_i = \dfrac{\Sigma_j^B \ Y_{ij}}{B}$ is the mean for cluster -i.

Thus, an easier computing formula for Rho was obtained as

$$\text{Rho } (\bar{\bar{Y}}) = \frac{B\sigma_A^2 - \sigma^2}{(B - 1)\sigma^2} \tag{1.3}$$

where $\sigma_A^2 = \dfrac{\Sigma_i^A \ (\overline{Y}_i - \bar{\bar{Y}})^2}{A}$ was the between-cluster variance.

1.2.2 <u>Sample Estimate For Cluster Homogeneity: Clusters Of Equal Sizes</u>

We noted earlier that the formula for Rho was defined for the special case of equal-sized clusters. Suppose we desire a sample estimate of Rho and have selected a two-stage sample of a clusters and b elements per cluster using simple random sampling. Then rho is calculated by substituting expected sample estimates for population estimates in equation 1.3 (Kish, 1965, pages 170 to 172) as

$$\text{rho}(\bar{\bar{y}}) = \frac{s_a^2 - \dfrac{s_b^2}{b}}{s^2}$$

where $s_a^2 = \dfrac{\Sigma_i^a (\bar{y}_i - \bar{\bar{y}})^2}{(a - 1)}$ with $\bar{y}_i = \dfrac{\Sigma_j^b y_{ij}}{b}$ and $\bar{\bar{y}} = \dfrac{\Sigma_i^a \Sigma_j^b y_{ij}}{a\ b}$;

$$s_b^2 = \frac{\Sigma_i^a \ \Sigma_j^b (y_{ij} - \bar{y}_i)^2}{a\ (b - 1)}; \text{ and } s^2 = s_a^2 + \frac{b - 1}{b} \ s_b^2 .$$

The formula for rho may also be stated in terms of analysis of variance where a treatments (clusters) are selected and b replicates (elements) per treatment are observed. The design is balanced since there are an equal number of observations (b) per treatment. We have

$$\text{rho}(\bar{\bar{y}}) = \frac{\text{MSB - MSW}}{\text{MSB + (b-1) MSW}}$$

where MSB and MSW are defined in Table 1.1

A comparable table may also be prepared for the population value Rho (Sukhatme and Sukhatme, 1970).

1.2.3 Population Measures For Cluster Homogeneity: Clusters Of Unequal Sizes

In the introductory section, the design effect for a sample was defined (section 1.1). Similarly, the design effect for the population or Deff is the ratio of the true variance for the cluster sample to the true variance of an entirely random sample of the same size. Kish (1965) tells us that the design effect, for either the population or sample, may be applied to samples of equal or unequal clusters, samples

Table 1.1 Analysis of variance of the mean into the components of the total variance

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	Expected Mean Squares
Between clusters	a-1	$\frac{\Sigma_i^a (\Sigma_j^b \; y_{ij})^2}{b} - \frac{(\Sigma_i^a \; \Sigma_j^b \; y_{ij})^2}{ab}$	MSB = bs_a^2	$\left[\frac{B - b}{B}\right] S_b^2 + bS_a^2$
Within Clusters	a(b-1)	$\Sigma_i^a \; \Sigma_j^b \; y_{ij}^2 - \frac{\Sigma_i^a (\Sigma_j^b \; y_{ij})^2}{b}$	MSW = s_b^2	S_b^2
Total	ab - 1	$\Sigma_i^a \; \Sigma_j^b \; y_{ij}^2 - \frac{(\Sigma_i^a \; \Sigma_j^b \; y_{ij})^2}{ab}$	s^2	

where $S_b^2 = \frac{\Sigma_i^A \; \Sigma_j^B (Y_{ij} - \bar{Y}_i)^2}{A(B - 1)}$ and $S_a^2 = \frac{\Sigma_i^A (\bar{Y}_i - \bar{\bar{Y}})^2}{(A - 1)}$.

with more than two stages of selection, with diverse selection methods at each sampling stage, and with stratification for each stage. Similarly, a more general measure of homogeneity or Rate Of Homogeneity (Roh) was defined by Kish (1965) for the more widely applicable case of multistage stratified cluster samples. From the Harris simplification (1.2), we have in terms of average cluster size ($\overline{B}$) and Kish's Roh that the design effect is

$$\mathrm{Deff}(\bar{\bar{Y}}) = \frac{\mathrm{Variance}\ (\bar{\bar{Y}})}{\mathrm{Variance}\ (\bar{\bar{Y}}_{SRS})} = \frac{\overline{B}\,\sigma_A^2}{\sigma^2} = 1 + (\overline{B} - 1)\ \mathrm{Roh}\ (\bar{\bar{Y}}) \qquad (1.4)$$

$$\text{or } \mathrm{Roh}\ (\bar{\bar{Y}}) = \frac{\mathrm{Deff}(\bar{\bar{Y}}) - 1}{\overline{B} - 1}\,. \qquad (1.5)$$

For unequal clusters, the total variance (σ^2) and the between-cluster variance (σ_A^2) are defined similarly to the equal-cluster case of equations 1.1 and 1.3 as (Sukhatme and Sukhatme, 1970)

$$\sigma^2 = \frac{\Sigma_i^A\ \Sigma_j^{B_i}\ (Y_{ij} - \bar{\bar{Y}})^2}{A\,\overline{B}} \qquad (1.6)$$

$$\text{where } \overline{B} = \frac{\Sigma_i^A\ B_i}{A} \quad \text{and} \quad \bar{\bar{Y}} = \frac{\Sigma_i^A\ \Sigma_j^{B_i}\ Y_{ij}}{A\,\overline{B}}\ ;$$

$$\sigma_A^2 = \frac{\Sigma_i^A\ B_i\ (\overline{Y}_i - \bar{\bar{Y}})^2}{A} \qquad (1.7)$$

where $\bar{Y}_i = \dfrac{\Sigma_j^{B_i} \; Y_{ij}}{B_i}$ is the cluster-i mean.

The within-cluster variance is defined as

$$\sigma_B^2 = \frac{\Sigma_i^A \; \Sigma_j^{B_i} \; (Y_{ij} - \bar{Y}_i)^2}{A \; \bar{B}} . \tag{1.8}$$

Using equation 1.4 and the fact that $^{2} = {}_{A}^{2} + {}_{B}^{2}$, we obtain an easier computing formula for Roh, similar to equation 1.3 but for unequal clusters, as

$$\text{Roh } (\bar{Y}) = \frac{\sigma_A^2 - \dfrac{\sigma^2}{\bar{B}}}{\dfrac{\bar{B} - 1}{\bar{B}} \; \sigma^2} . \tag{1.9}$$

We can show algebraically that equation 1.9 reduces to the formula for computing Roh in Appendix A as follows.

From equations 1.6 to 1.8, we can re-express σ_A^2 and σ^2 in equation 1.9 as

$$\sigma_A^2 = \frac{\Sigma_i^A \; B_i \; (\bar{Y}_i - \bar{Y})^2}{A} = \frac{A - 1}{A} \; \frac{\Sigma_i^A \; B_i \; (\bar{Y}_i - \bar{Y})^2}{A - 1} = \frac{A - 1}{A} \; \beta^2 ,$$

because $\beta^2 = \dfrac{\Sigma_i^A \; B_i \; (\bar{Y}_i - \bar{Y})^2}{A - 1}$;

and $$\sigma^2 = \sigma_A^2 + \sigma_B^2 = \frac{A-1}{A}\beta^2 + \frac{\bar{B}-1}{\bar{B}}W^2 = V^2 ,$$

because

$$\sigma_B^2 = \frac{\Sigma_i^A \Sigma_j^{B_i} (Y_{ij} - \bar{Y}_i)^2}{A\bar{B}} = \frac{\bar{B}-1}{\bar{B}} \frac{\Sigma_i^A \Sigma_j^{B_i} (Y_{ij} - \bar{Y}_i)^2}{A(\bar{B}-1)}$$

$$= \frac{\bar{B}-1}{\bar{B}} W^2 \text{ where } W^2 = \frac{\Sigma_i^A \Sigma_j^{B_i} (Y_{ij} - \bar{Y}_i)^2}{A(\bar{B}-1)} .$$

Substituting these expressions for σ_A^2 and σ^2 into equation 1.9, we have

$$\mathrm{Roh}(\bar{Y}) = \frac{\sigma_A^2 - \frac{\sigma^2}{\bar{B}}}{\frac{\bar{B}-1}{\bar{B}}\sigma^2}$$

$$= \frac{\frac{A-1}{A}\beta^2 - \frac{V^2}{\bar{B}}}{\frac{\bar{B}-1}{\bar{B}}V^2} ,$$

but this expression is the computing formula for Roh from Appendix A.

From equation 1.5, we see that the range of Roh is from $-\frac{1}{\bar{B}-1}$, when Deff is zero, to positive one (1.0), when Deff is $\bar{B}$. When Deff is one (1.0), then Roh is zero. A high positive value for Roh implies that all responses from one cluster will be the same. Thus, one cluster provides only one independent observation (Sudman, 1976). When Roh is zero, the responses are arranged at random within the clusters, and one cluster may provide several independent observations, depending on the randomization. A large negative Roh implies that responses are arranged in clusters so that each cluster provides the true population value. Thus, only one cluster needs to be selected, and it will provide the maximum number of independent observations.

1.2.4 Sample Estimate For Cluster Homogeneity: Clusters Of Unequal Sizes

Substituting sample estimates of the design effect (deff) and average cluster size ($\bar{b}$), in equation 1.5, we obtain a sample estimate for Roh as

$$\text{roh}\ (\bar{\bar{y}}) = \frac{\text{deff}(\bar{\bar{y}}) - 1}{\bar{b} - 1}\ . \tag{1.10}$$

This estimator is called the design effect method in this dissertation. Kish (1965, p. 162) tells us that in clusters of unequal size, the average cluster size generally yields a good approximation for roh.

In this dissertation, the average sample cluster size was estimated for a cross-class domain as the number of sample elements in the domain category divided by the total number of clusters in the sample ($a = 20$). For a segregated domain, the average sample cluster size was estimated as the ratio of the number of sample elements in the domain category to the number of sample clusters containing nonzero elements for that domain category. Segregated domains are those in which only one domain category is present in a cluster; examples include region and urban or rural status. Cross-class domains are those where several domain categories are present within each cluster; for instance, age, sex, social status, income, and attitudinal variables. This definition for $\bar{b}$ or the average sample cluster size was adopted because; in cross-class domains, we can think of the proportion of the population in the domain as being constant for each cluster; while for segregated domains, we view the domain population proportion as that portion of clusters consisting entirely of domain members while the remaining clusters have no domain members (Kalsbeek, 1984, Chapter 8, page 7). These definitions for average sample cluster size in cross-class and segregated domains are also consistent with that provided by Yates (1981, pages 265 to 268) for the number of sample clusters as the number of first-step units for which the number of elements is not zero.

1.2.5 Current Uses For Intracluster Homogeneity

A number of methods for measuring or estimating intracluster homogeneity have been reviewed in this dissertation. The question now is how is intracluster homogeneity used. Roh is used in survey design to determine either the precision of estimates, the total sample size, or the optimum cluster size. An estimate for roh is often obtained from a previous survey or available from a pretest. Kish, Groves, and Krotki (1976) observed the ratio of domain to total roh to be relatively stable for many variables in different surveys. They advocate using roh to impute design effects, standard errors, and other rohs for additional variables in the same survey and for similar variables in other surveys. Using roh for imputation and for determination of the total sample size are discussed in the first section which follows. The two subsequent sections report using rohs to determine the optimum cluster size in multistage samples. The second section details the optimization formulas, and the third section talks about specific surveys where these formulas have been applied.

1.2.5.1 Impute Rohs, Design Effects, and Standard Errors For Other Variables

Kish, Groves, and Krotki (1976) have observed that roh is relatively stable for the same variable classified by different domains within the same sample and for similar variables from different samples. They termed this phenomenon portability, a property of an estimator that "facilitates its use far from its source." (Kish, Groves, and Krotki, 1976, p. 11). Other words similar to portability are generality, invariance, and robustness. They suggest using roh to impute rohs, design effects, and standard errors for other variables in the same survey and for similar variables in other surveys. Kish, Groves, and Krotki (1976) propose the indirect method of imputation as illustrated in Figure 1.1 for a mean estimator. In the diagram, deft is the square root of the design effect and ste is the square root of the variance. Values of ste are converted to deft, and then to roh; roh is then imputed for the new statistic; from this we transform to deft, and finally to the desired ste. For example,

$$
\begin{tikzcd}
\text{computed ste}(\bar{\bar{x}}) \arrow[d] & \text{imputed ste}(\bar{\bar{y}}) \\
\text{deft}(\bar{\bar{x}}) \arrow[d] & \text{deft}(\bar{\bar{y}}) \arrow[u] \\
\text{roh}(\bar{\bar{x}}) \arrow[r] & \text{roh}(\bar{\bar{y}}) \arrow[u]
\end{tikzcd}
$$

Figure 1.1 Indirect Imputation of Deft and Standard Error Via Roh.

if we desire to estimate the standard error for another similar estimate ($\bar{\bar{y}}$) in the same survey ($\bar{\bar{x}}$ is the original estimate), we have the following three step imputation:

1. Assume roh $(\bar{\bar{y}})$ = roh $(\bar{\bar{x}})$;

2. Compute deft $(\bar{\bar{y}}) = \sqrt{1 + \text{roh}\ (\bar{\bar{y}})\ (\bar{b}_{\bar{\bar{y}}} - 1)}$ where $\bar{b}_{\bar{\bar{y}}}$ is the estimated average cluster size for the new estimate;

3. Obtain ste$(\bar{\bar{y}})$ = deft $(\bar{\bar{y}})$ ste$(\bar{\bar{y}}_{SRS})$ where ste$(\bar{\bar{y}}_{SRS})$ is the standard error for the new estimate based on a simple random sample.

More often we know the rate of homogeneity for some total population estimate, but we actually want to know the rate of homogeneity for a domain estimate. In this case, Kish et al (1976) have suggested that we assume roh $(\bar{\bar{y}}) = \lambda_{\bar{\bar{y}}}$ roh $(\bar{\bar{x}})$ for step 1 and then proceed as before with steps 2 and 3. They assumed that the ratio of domain to total roh is a constant, $\lambda_{\bar{\bar{y}}}$. Here we also assume the domain estimate is $\bar{\bar{y}}$ and the total estimate is $\bar{\bar{x}}$. The standard error so imputed might be used to estimate the required sample size for the domain. For instance, suppose we want to determine the sample size required (n_d) to estimate the crude death rate among people aged 45 to 49(p_d). We know

$$\text{variance } (P_d) = \text{deff } (P_d) \frac{p_d (1 - p_d)}{n_d}$$

$$= [\ 1 + \text{roh } (p_d)\ (\bar{n}_d - 1)] \frac{p_d (1 - p_d)}{n_d}$$

where $\bar{n}_d$ is the average number of 45 to 49 year-olds per selected cluster. Using the imputed variance (p_d) and roh(p_d) for the domain 45- to 49-years of age, the above equation may be solved for n_d. Kalton (1983) stresses including the design effect in the sample size estimate for a complex design. The formula above as well as the imputation procedure illustrate this principle.

1.2.5.2 Determine Optimum Cluster Sizes For Multistage Samples

A second general use for roh is found in determining the optimum cluster size for a multistage sample. Suppose a two-stage sample is planned; then one must determine the number of primary sampling units (PSU's) or clusters to be selected (a) and the average cluster size ($\bar{b}$). The rate of homogeneity plays an important part in this decision.

Hansen, Hurwitz, and Madow (1953) discuss the case using the simple cost function

$$c = ac_1 + a\bar{b}\, c_2$$

where c = total cost of the survey;

c_1 = average cost per PSU;

c_2 = average cost per element.

Then the optimum cluster size is

$$\bar{b}_{opt} = \left[\frac{c_1}{c_2} \; \frac{1 - roh(y)}{roh(y)} \right]^{1/2},$$

and the number of PSU's (a) is

$$a = \frac{c}{c_1 + \bar{b}_{opt}\, c_2}.$$

Thus, roh determines the optimum cluster size and in turn the number of PSU's.

Hansen, Hurwitz, and Madow (1953) discuss more complicated cost functions as well as cost functions for more than two stages of sampling. A more complicated cost function is often justified because between PSU travel costs may be sizeable.

Cost and variance formulas may be considered for more than two stages of sampling. However, several authors suggest that estimates which view sampling to have been done in two stages are acceptable even though the sampling may actually have occurred in more than two stages.

Hansen, Hurwitz, and Madow (1953, pages 257-258) define an ultimate cluster as the aggregate of units included in the sample from a primary unit. They remark that, "The idea of the ultimate cluster makes it possible to consider two-or more stage sampling theory as a special case of the sampling of whole clusters." They further note that "the ultimate cluster approach yields a consistent estimate of the variance as long as there are two or more first-stage units in the sample, no matter how the subsampling is done, provided the subsample is a probability sample (for proof, see Volume II, Chapter 6, Section 2)."

Kish, Groves, and Krotki (1976) point out, "To compute all the components of the design would be difficult for individual statistics. Moreover, it seems impossible to present and utilize such detail for many... statistics."

In these approximations, either the first-stage is left intact and higher stages are collapsed, or the last stage remains intact and earlier stages are collapsed. Kalsbeek (1983), investigating rates of fertility in developing countries, used the first approach; whereas Lorah, Singh, and Tegels (1981) incorporated the second approach when they computed the optimum cluster size of the Current Population Survey for the Census Bureau.

Although the approximation may not be totally correct, this dissertation has followed the ultimate cluster approach (Hansen, Hurwitz, and Madow, 1953) and left the first stage intact and collapsed higher stages. The specific application for this dissertation is the four-stage National Medical Care Expenditure Survey (NMCES). The sampling design has been documented by Cohen and Kalsbeek (1981). A two-stage probability sample was selected for this dissertation from this four-stage sample. The first-stage units or primary sampling units in NMCES were retained as ultimate clusters of households or second-stage units. The sampling procedure for both the original NMCES sample and the two-stage sample for this dissertation are provided in section 2.1.

1.2.5.3 Determination Of Optimum Cluster Size For Several Large Surveys

Illustrating a third general use of roh, the optimum cluster sizes have been determined for several continuing national surveys (Krewski, Platek, and Rao, 1981) among them the Current Population Survey, the National Crime Survey, the Health Interview Survey, and the National Assessment of Educational Progress.

Current Population Survey. Design efficiency studies determine the optimum cluster size subject to minimum variance for a fixed cost. Previous design efficiency studies for the Current Population Survey (Lorah, Singh, and Tegels, 1981) have indicated that variance for the major labor force items would be minimized for a fixed cost by using sampling units at the final selection stage of size 4 (i.e., a cluster of adjacent housing units of expected size 4). It appeared that more reliable estimates for minorities might be obtained by reducing cluster sizes to 3 or 2. Some reduction in variance was observed when this reduced cluster size was simulated at the Census Bureau. The simulation was conducted assuming a two-stage design even though the actual sample was four-stage. It has been suggested that the cluster size be reduced in areas of high concentration of minorities.

National Crime Survey. The National Crime Survey was originally designed using the Current Population Survey as a model (Alexander and Kobilarcik, 1981; Krewski, Platek, and Rao, 1981). Optimum cluster size and stratification variables were determined from labor-related variables of the Current Population Survey. Uniform Crime Reports data are now available; stratification variables and optimum cluster size have been compared for variables from the Current Population Survey and Uniform Crime Reports. Few differences have been observed in either stratification variables or optimum cluster size. This anomaly might lend some credence to the idea of the stability of roh.

Health Interview Survey. Optimum cluster size for the Health Interview Survey in the final selection stage has been previously determined based on compact segments. A study was undertaken to see if variance of estimators were significantly reduced for noncompact segments (Waksberg, Hanson, and Jacobs, 1973). It was evident from small losses in variance that noncompact segments were not justified over compact segments. Also, small estimates for intraclass correlation (generally less than 0.05 for cluster size of 18) indicate that size of cluster may be increased while only minimally increasing variance due to clustering.

National Assessment of Educational Progress. The National Assessment of Educational Progress, a project of the Education Commission of The States, has conducted design efficiency studies to determine the optimum cluster size subject to minimum variance for a fixed cost. In these studies, costs and variance estimates were developed separately for each of the three stages in the sample design. A report was prepared prior to a major redesign effort in the eleventh year of the study and concluded that the design parameters used in the seventh through tenth years of the survey could be maintained without appreciably increasing the variance of estimates (Williams, Budescu, and Chromy, 1981).

1.2.6 Range Of Values For Intracluster Homogeneity

Empirical studies investigating intracluster homogeneity are summarized in the first portion of the subsequent sections. First, values for intracluster homogeneity from other health studies are reviewed. Then the fact that socioeconomic variables tend to produce higher rates of homogeneity than demographic variables is noted by examining roh from variables of both types. Next, the average to total domain ratio is examined for fertility measures in developing countries; conflicting results are noted. Lastly, a plan for empirically modeling the domain degree of segregation or cross-classness is discussed. This plan may aid in resolving the conflicting domain ratios noted in developing countries.

In the second portion, another formula to estimate Roh is first derived. Then, it is hypothesized that this formula may perform better than the design effect method or equation 1.10 when roh is large. It is proposed to test this hypothesis.

1.2.6.1 Empirical Results

All else constant, Roh decreases as size of cluster ($\overline{B}$) increases. The rate of decrease for Roh is generally less than the rate of increase for $\overline{B}$. Empirical studies (Cochran, 1977) have shown that with the model

$$Roh(y) = q\,\overline{B}^{t} ,$$

$0.1 < q < 0.5$ and $-0.5 < t < -0.2$.

In most settings, roh lies between 0.0 and 0.5 and in the majority of cases, roh is less than 0.1 (Kish, 1965). Sudman (1976) observed that values of roh for health statistics are small, averaging 0.05 or less. Waksberg, Hanson, and Jacobs (1973) also noted small rates of homogeneity for health variables; the maximum reported roh was 0.17 associated with an average cluster size of 6. Rates of homogeneity for crude birth and crude death rates in six developing countries from Africa and Asia were found to range between -0.001 and 0.013 (Marks, Seltzer, and Krotki, 1974).

For demographic variables such as age, Kish, Groves, and Krotki (1976) noted a smaller rate of homogeneity, ranging from 0.001 to 0.005, than with socioeconomic and attitudinal variables where the range was from 0.1 to 0.2. They also observed that the rate of homogeneity for domain variables was larger than for the total population. Kish et al (1976) calculated the average domain ratio (ADR) defined as the average value of roh across all categories in a domain divided by the corresponding population total estimate of roh. The definition of average domain ratio is attributable to Kalsbeek (1983). As a rule of thumb, Kish et al suggested that the domain average may be 1.2 times as large as the total estimate for demographic variables. Kalsbeek (1983) computed the average domain ratio for demographic variables and found a range from 2.2 to 3.6. It appeared from Kalsbeek's work that a domain average to total ratio of 2.5 might be more suitable for demographic variables. The major difference appeared to be in the computer software criteria used to suppress calculation of roh for certain small domains. Calculation of roh for clusters of size 6 or smaller were censored from the Kish et al data and included in the Kalsbeek data. Many of the rates of homogeneity computed for small domains were very large because a moderately-sized design effect was divided by a very small average cluster size. The fact that roh was computed for smaller cluster sizes with the Kalsbeek data might help to explain the larger ADRs observed by him. The computer algorithm SESUDAAN (Shah, 1981) was used to calculate roh for Kalsbeek's data; this procedure suppresses roh only for clusters smaller than 2. SESUDAAN, an acronym for Standard Errors for SUrvey DAta ANalysis, is a Statistical Analysis System (SAS) procedure. Kish et al rates of homogeneity were produced by the computer algorithm CLUSTERS (Verma, 1978) which suppresses roh for clusters of size 6 or smaller. Kalsbeek also noted that other factors may have contributed to the different ratio such as the fact that (1) the Kish et al data came from the World Fertility Survey (WFS), whereas the Kalsbeek data came from the POPulation LABoratory Survey (POPLAB); (2) the ADRs were defined for totally different reporting variables in the two surveys; and (3) POPLAB used mainly age groups to form domains, and WFS used both demographic and socioeconomic variables to form domains.

Verma, Scott, and O'Muircheartaigh (1980), using WFS data, found roh for fertility variables, such as nuptiality, fertility, and contraceptive knowledge and preference, ranged from 0.004 to 0.170 for several Central and South American countries. The authors theorized that the relationship between the domain and analysis variable may have a bearing on the magnitude of roh. An analysis variable is a reporting variable in a survey. They attempted to model the proportion of the total sample contained in a particular domain category as a function of the ratio of the design effect for the domain category to the design effect for the total population. The specific model they assumed was

$$\left[\frac{\bar{b}_d}{\bar{b}_t}\right]^{\alpha} = (M_d)^{\alpha} = \frac{\mathrm{deff}(y_d)-1}{\mathrm{deff}(y_t)-1}$$

where $\bar{b}_d$ and $\bar{b}_t$ denoted the average cluster size for the domain and total; M_d was the proportion of the total sample contained in the particular domain; $\mathrm{deff}(y_d)$ and $\mathrm{deff}(y_t)$ were the design effects for the domain and total; and α was an empirically determined parameter expected to be in the range 0 to 1 with values near 0 corresponding to segregated domains and those near 1 corresponding to cross-class domains. Segregated and cross-class domains were defined in section 1.2.4.

1.2.6.2 Derivation Of A New Estimator For Roh

In this section, we develop a new estimator for Rate of Homogeneity based on the definition of Roh in Hansen, Hurwitz, and Madow (1953). Since this formulation is based on the definition of Roh, we think it may do better than the design effect method or equation 1.10 which is based on the definition of the design effect.

If we assume the number of clusters in the population, A, is large, the definition for population Roh from a two-stage, unequal cluster (B_i) population is (Hansen, Hurwitz, and Madow, Volume I, 1953, p. 163)

$$\rho^* = \frac{S_a^2 - \bar{B}S_b^2}{S_a^2 + \bar{B}\,(\bar{B} - 1)\; S_b^2}$$

where

$$S_a^2 = \frac{\Sigma_i^A Z_i^2}{A - 1} \quad \text{and} \quad S_b^2 = \frac{\Sigma_i^A \frac{B_i}{B_i - 1} \Sigma_j^{B_i} (Z_{ij} - \bar{Z}_i)^2}{A \bar{B}} ;$$

$$Z_i = Y_i - RX_i; \qquad Z_{ij} = Y_{ij} - RX_{ij};$$

$$R = \frac{\Sigma_i^A \Sigma_j^{B_i} Y_{ij}}{\Sigma_i^A \Sigma_j^{B_i} X_{ij}} ; \qquad \bar{\bar{Z}}_i = \bar{\bar{Y}}_i - R\bar{\bar{X}}_i;$$

$$B = \frac{\Sigma_i^A B_i}{A} ; \qquad \bar{\bar{Y}}_i = \frac{\Sigma_j^{B_i} Y_{ij}}{B_i} \quad \text{and} \quad \bar{\bar{X}}_i = \frac{\Sigma_j^{B_i} X_{ij}}{B_i} .$$

considering ρ_h^* to be rate of homogeneity at the stratum -level, we have

$$\rho_h^* = \frac{S_{ha}^2 - \bar{B}_h S_{hb}^2}{S_{ha}^2 + \bar{B}_h(\bar{B}_h - 1) S_{hb}^2}$$

where $\bar{B}_h = \dfrac{N_h}{A_h}$;

$$S_{ha}^2 = \frac{\Sigma_i^{A_h} Z_{hi}^2}{A_h - 1} ;$$

$$S^2_{hb} = \frac{\Sigma_i^{A_h} \frac{B_{hi}}{B_{hi} - 1} \Sigma_j^{B_{hi}} (Z_{hij} - \bar{Z}_{hi})^2}{N_h} ;$$

$$Z_{hij} = Y_{hij} - R_h X_{hij} \text{ ; and } R_h = \frac{\Sigma_i^{A_h} \Sigma_j^{B_{hi}} Y_{hij}}{\Sigma_i^{A_h} \Sigma_j^{B_{hi}} X_{hij}} ;$$

$$Z_{hi} = \Sigma_j^{B_{hi}} Z_{hij} = Y_{hi} - R_h X_{hi} ;$$

$$\bar{Z}_{hi} = \frac{\Sigma_j^{B_{hi}} Z_{hij}}{B_{hi}} = \bar{Y}_{hi} - R_h \bar{X}_{hi} .$$

When estimating a mean or proportion $X_{hij} = 1$. Thus,

$$\bar{X}_{hi} = \frac{\Sigma_j^{B_{hi}} X_{hij}}{B_{hi}} = \frac{B_{hi}}{B_{hi}} = 1 ;$$

$$X_{hi} = \Sigma_j^{B_{hi}} X_{hij} = B_{hi} ;$$

$$Z_{hij} = Y_{hij} - R_h ;$$

$$Z_{hi} = Y_{hi} - R_h B_{hi} ;$$

$$\bar{\bar{Z}}_{hi} = \bar{\bar{Y}}_{hi} - R_h \text{ ; and}$$

$$S_{ha}^2 = \frac{\Sigma_i^{A_h} (Y_{hi} - R_h B_{hi})^2}{A_h - 1} ;$$

$$S_{hb}^2 = \frac{\Sigma_i^{A_h} \frac{B_{hi}}{B_{hi} - 1} \Sigma_j^{B_{hi}} \left\{ (Y_{hij} - R_h) - (\bar{Y}_{hi} - R_h) \right\}^2}{N_h}$$

$$= \frac{\Sigma_i^{A_h} \frac{B_{hi}}{B_{hi} - 1} \Sigma_j^{B_{hi}} (Y_{hij} - \bar{Y}_{hi})^2}{N_h} .$$

So, when estimating a mean or a proportion, we can express

$$\rho_h^* = \frac{S_{ha}^2 - \bar{B}_h S_{hb}^2}{S_{ha}^2 + \bar{B}_h (\bar{B}_h - 1) S_{hb}^2} \tag{1.11}$$

$$= \frac{\left[\frac{\Sigma_i^{A_h} (Y_{hi} - R_h B_{hi})^2}{A_h - 1}\right] - \bar{B}_h \left[\frac{\Sigma_i^{A_h} B_{hi} - 1 \Sigma_j^{B_{hi}} (Y_{hij} - \bar{Y}_{hi})^2}{N_h}\right]}{\left[\frac{\Sigma_i^{A_h} (Y_{hi} - R_h B_{hi})^2}{A_h - 1}\right] + \bar{B}_h (\bar{B}_h - 1) \left[\frac{\Sigma_i^{A_h} \frac{B_{hi}}{B_{hi} - 1} \Sigma_j^{B_{hi}} (Y_{hi} - \bar{Y}_{hi})^2}{N_h}\right]}$$

We want to find a reasonable estimator for ρ^*_h. Suppose we let

$$z_{hij} = y_{hij} - r_h x_{hij} ;$$

$$r_h = \frac{\Sigma_i^{A_h} \Sigma_j^{b_{hi}} y_{hij}}{\Sigma_i^{a_h} \Sigma_j^{b_{hi}} x_{hij}} = \frac{y_h}{n_h} ;$$

$$Y_{hi} = \Sigma_j^{b_{hi}} y_{hij} ;$$

b_{hi} = number of elements in cluster -i of stratum -h;

$$\bar{\bar{y}}_{hi} = \frac{\Sigma_j^{b_{hi}} y_{hij}}{b_{hi}} = \frac{y_{hi}}{b_{hi}} ; \; Б_h = \frac{n_h}{a_h} .$$

Substituting these sample estimates into equation 1.11, we obtain expression 1.12,

$$\overset{v}{\rho}{}^*_h = \frac{\left[\dfrac{\Sigma_i^{a_h} (Y_{hi} - r_h b_{hi})^2}{a_h - 1}\right] - Б_h \left[\dfrac{\Sigma_i^{a_h} \dfrac{b_{hi}}{b_{hi} - 1} \Sigma_j^{b_{hi}} (y_{hij} - \bar{\bar{y}}_{hi})^2}{n_h}\right]}{\left[\dfrac{\Sigma_i^{a_h} (y_{hi} - r_h b_{hi})^2}{a_h - 1}\right] + Б_h (Б_h - 1) \left[\dfrac{\Sigma_i^{a_h} \dfrac{b_{hi}}{b_{hi} - 1} \Sigma_j^{b_{hi}} (y_{hij} - \bar{\bar{y}}_{hi})^2}{n_h}\right]}$$

$$= \frac{s_{ha}^2 - \bar{b}_h\, s_{hb}^2}{s_{ha}^2 + \bar{b}_h\, (\bar{b}_h - 1)\, s_{hb}^2} \,. \tag{1.12}$$

We call this new estimation method the proportion variation method because it is based on the definition of roh, and roh is by definition, approximately the proportion of the total variation not accounted for within clusters. As noted earlier, since the proportion variation method or equation 1.12 is based on the definition for Roh, we think it will generally perform better than the design effect method or equation 1.10 which is based on the definition of the design effect.

Another form of the proportion variation estimate but specifically for proportion and assuming approximately equal clusters is

$$\text{roh } (p) = 1 - \frac{\Sigma_i^a \; b_i \; \check{p}_i(1 - \check{p}_i)}{a\, \bar{b}\, \check{p}(1 - \check{p})} \tag{1.13}$$

where $\check{p}_i$ is the sample attribute proportion for cluster -i; b_i is the number of sample elements in sample cluster -i; $\bar{b} = \frac{\Sigma_i^a \; b_i}{a}$ is the average sample cluster size; a is the number of selected clusters; and $\check{p}$ is the sample estimate of the attribute proportion. This formula is derived in Appendix D.

When roh is large, we think that equation 1.13 will perform better than the design effect method. For example, suppose we calculate the proportion urban. In this situation, roh should be large because PSUs will be either all urban or all rural. In urban PSUs, p_i will be near 1 making $(1 - P_i)$ zero; the second term in equation 1.13 will then be zero and roh approximately 1. Conversely, in rural PSUs p_i will be close to 0 but the second term again will be zero making roh about 1. Using the design effect method, it was found that roh for proportion urban ranged from 0.2 to 0.4 (Kalsbeek, 1983) which is not particularly close to one and implying that equation 1.13 may perform better. If this supposition turns out to be true, it would imply that the proportion variation method should be used for segregated domains where roh is large and that the design effect method should possibly be used for cross-class domains where roh is moderate or small. Cross-class and segregated domains were defined in section 1.2.4.

1.3 Recapitulation And Specification Of Objectives Of This Research

In preceding sections, it has been shown that intracluster homogeneity is both important and useful. Its importance has been noted by many authors who observed that variances are underestimated in complex surveys unless stratification and cluster effects are included in the estimation procedure (Hansen, Madow, and Tepping, 1983; Kish and Frankel, 1974; Holt, Smith, and Winter, 1980; Nathan and Holt, 1980). Fellegi (1980) and Rao and Scott (1981) have shown that intracluster homogeneity is important in specifying the asymptotic distribution of the chi-squared statistic when tests of hypotheses are performed on complex survey data. The rate of homogeneity is used either to determine the total sample size or the optimum cluster size. We have specified how these tasks are performed and have cited examples where they were utilized in large scale surveys. Considering the usefulness and importance of intracluster homogeneity, it appears that a simple formula to estimate roh would be helpful.

The design effect method or equation 1.10 is generally used to estimate intracluster homogeneity. There are cases where the design effect method does not perform well. When roh is large, theory suggests that equation 1.12 or the proportion variation method may perform better. Further support for this hypothesis was gained when roh for proportion urban was calculated by the design effect method and found to be in the range 0.2 to 0.4 which is not close to 1, the anticipated value. Quantifying when the design effect and proportion variation estimation methods are appropriate to estimate roh is the primary purpose of this dissertation and is discussed in more detail in section 1.3.1.

Values for the average domain ratio of 1.2 and 2.5 have been suggested by Kish et al (1976) and Kalsbeek (1983). Verma et al (1980) have hypothesized that the relationship between the domain and analysis variables may determine the magnitude of roh. These topics are explored as secondary purposes which are described in section 1.3.2.

1.3.1 Primary Objective

The purpose of this research is to determine under what circumstances equations 1.10 (design effect) and 1.12 (proportion variation) are appropriate to estimate intracluster homogeneity. To do this, we propose to select 10 independent samples from the study or limited population, to calculate roh by the design effect and proportion variation methods, and to compare each sample estimate with the true Roh calculated from the limited population. The results of the comparison will be examined separately for cross-class and segregated domains. It is speculated that the design effect method will perform better for cross-class domains and the proportion variation method better for segregated domains.

1.3.2 Secondary Objective

As a second issue, the relationship between the analysis variable and the nature of the domain will be investigated in section 2.4. It is first proposed to compute the average domain ratio by each estimation method and compare with the true value from the study population. Then following the approach of Verma et al (1980), the proportion of the domain contained in the particular category will be modeled, and the fit examined separately for each estimation method.

1.3.3 Orientation Of Analysis Toward Veterans

Analyses of use to the Veterans Administration (VA) will be undertaken in this dissertation. They include:

1. Using the complete NMCES file to calculate intracluster homogeneity for certain prespecified domain and analysis variables.
2. Comparing certain analysis variables between specified domains using the total NMCES file.

Both analyses have relevance for the Veterans Administration. Suppose an analysis variable in a particular domain has a high rate of homogeneity. Then the first analysis would tell us that research concerning this analysis variable and domain category has low generalizability. A recommendation could be made that such research be conducted by combining data from several locations. Conversely, confidence as to the generalizability of research findings would be gained from an analysis variable having a low rate of homogeneity within a different domain category.

The second analysis or comparison between domains will complement other studies of factors influencing use of medical care by veterans. Specific domain and analysis variables of interest to veterans are enumerated in section 2.

2. PLAN OF RESEARCH

To investigate whether the design effect or proportion variation method was superior for estimating the rate of homogeneity, ten independently replicated samples were selected from the NMCES household file which thus served as the population to which inference would be made. Independent replication enabled the variance to be calculated for sample estimates and thus to test hypotheses at known probability levels. These independently replicated samples were also used to test hypotheses determining whether the relationship between the analysis variable and nature of the domain affected the appropriate formula for roh.

The NMCES household file contained data from a large national probability sample consisting of about 40,320 individuals. A record existed for each individual, and each record contained about 250 identification or response variables. It was evident that computer costs and work load could be kept to a manageable level only by defining a subset of the total number of individuals and total variables and studying this subset intensively.

The creation of the random subset of observations from the total file and the selection of the ten independently replicated samples from this subset are detailed in section 2.1. The identification of the subset of domain and analysis variables from the total variable file is documented next in section 2.2. The methodology for investigating the primary and secondary purposes is then discussed in sections 2.3 and

2.4, respectively. Lastly, specific analyses of interest to the Veterans Administration are described in section 2.5.

2.1 Sample Selection

The primary and secondary purposes of this dissertation were investigated using the National Medical Care Expenditure Survey (NMCES) household data. NMCES was sponsored by the National Center for Health Services Research (NCHSR) with support from the National Center for Health Statistics (NCHS). Its objectives were to analyze how Americans use health care services and to determine the patterns and characteristics of health expenditures and health insurance. The NMCES household sample may be described as a four-stage stratified area probability design consisting of about 14,000 households, 40,320 individuals, and 142 primary sampling units (PSUs). The NMCES sample has been previously described (Cohen and Kalsbeek, 1981). National probability samples from Research Triangle Institute (RTI) and National Opinion Research Center (NORC) of the University of Chicago were utilized for the NMCES household sample. The survey population for NMCES was the civilian noninstitutionalized population of the 50 states and the District of Columbia. The reference period for household data collection was calendar year 1977. First-stage units (PSUs) were counties or groups of contiguous counties which were stratified by census region and division as well as level of urbanization. Secondary sampling units consisting of enumeration districts (EDs) and/or block groups (BGs) were stratified by residential area. Tertiary sampling units were clearly defined land segments within the selected EDs or BGs; and fourth-stage units were housing units in selected segments. All members of selected households were included in the sample.

In order to conserve computing costs in this dissertation, a probability sample was chosen from the NMCES file. The subsampled population was called the limited population. No inference will be made beyond this limited population. The limited population thus serves as the study population for this dissertation. Inference to veteran users of medical services will be made using the total NMCES file. From the limited population, ten independent samples of 500 individuals each were selected to determine the conditions under which either the design effect or the proportion variation method was appropriate to estimate roh, and how the relationship between the analysis variable and nature of the domain might determine the appropriate formula. The selection of the limited population and of the ten independent samples are described in the two sections which follow.

2.1.1 Subsample From Total To Limited Population

A random subset of the total NMCES file was created and called the limited population. The objectives in formulating the limited population were:

1. Limiting the total computer budget associated with this dissertation;
2. Appealing to the basic equal-cluster definition of intraclass correlation;
3. Controlling the average cluster size in small domains.

With these objectives in mind, a two-stage subset of the total NMCES file was determined consisting of all 142 primary sampling units from the total population and 50 individuals per PSU. Thus, we can think of the limited population as consisting of A = 142 clusters with B = 50 individuals in each cluster.

The limited population of 7,100 individuals (142 X 50) was approximately a 1/6 subset of the total population (7,100/40,000). The amount of computer time required to read the limited population was reduced by about 1/4 resulting in a considerable savings in computer costs. By creating the limited population consisting of equal clusters of 50 individuals, a small domain containing perhaps 5 percent of the population would still average 2 to 3 individuals per cluster.

The limited population was determined as follows. First, a count of the total number of individuals on the NMCES file was obtained by primary sampling unit number. Seven urban PSUs containing fewer than 100 individuals were identified. Urban PSUs were defined in terms of Standard Metropolitan Statistical Areas (SMSAs) and included residential codes 1 through 3. Rural PSUs were defined as non-SMSA areas and contained residential codes 4 or 5. A summary of residential code definitions follows:

1, one of the 16 SMSAs containing the largest number of people;

2, an SMSA containing 500,000 people or more and not previously classified as code 1;

3, an SMSA containing fewer than 500,000 people;

4, non-SMSA area which is less than 60 percent rural;

5, non-SMSA area which is rural (i.e., 60 percent or more rural)

The seven urban PSUs with only a small number of individuals were combined with others in the same region and residential area code so that

Table 2.1 Comparison of stratum representation in total and limited population

			TOTAL			LIMITED		
Stratum	Region	Urbanization	No. PSU's	No. Indiv.	% Indiv.	Realloc. PSU's	No. Indiv.	% Indiv
1	Northeast(NE)	Urban	31	7,275	18.0	26	1,300	18.3
2	Northcentral(NC)	Urban	26	6,268	15.5	22	1,100	15.5
3	Combined NE and NC	Rural	16	5,077	12.6	18	900	12.7
4	Southeast (SE)	Urban	24	7,670	19.0	27	1,350	19.0
5	West (W)	Urban	20	5,533	13.7	19	950	13.4
6	Combined SE and W	Rural	25	8,497	21.2	30	1,500	21.1
			142	40,320	100.0	142	7,100	100.0

the equal cluster sample of 50 individuals per PSU would not overrepresent the urban area in the limited population. Second, six strata were determined by combining region and urban or rural categories. The strata are listed in Table 2.1 along with the number of PSUs, number of individuals, and percentage individuals per stratum. Third, the 142 PSUs were reallocated in proportion to the percent individuals per stratum (6th column of Table 2.1) in the limited population for the dual purposes of equalizing the average cluster sizes in the total population prior to selection and thus, minimizing small or empty domains per PSU; and of maximizing the similarity between the total and limited populations. The reallocated PSUs, number of individuals, and percent individuals for the limited population are specified in the right hand portion of the Table 2.1. Close agreement between the total and limited population with respect to stratum representation can be seen by examining columns 6 and 9 in Table 2.1. Lastly, the limited population was identified by randomly ordering the households within each of the 142

PSUs, and then selecting the first 50 individuals from intact households on the ordered list. If the last household on the list was not the correct size to yield 50 individuals, then the next acceptable one on the randomly ordered list was taken.

Frequency distributions for the proportion of respondents in several domain and analysis categories are compared for the total and limited populations in Table 2.2. The agreement is very good implying the composition of these variables is indeed similar in the NMCES file and the limited population. Domain variables of age, sex, marital status, race, and veteran disability status were examined, as well as analysis variables of perceived health status, limitation of activity, and usual source of care.

2.1.2 Select Ten Independent Samples From Limited Population

Ten independently replicated samples were selected from the limited population so that variances of sample estimates could be calculated for hypothesis testing. In subsequent discussions, these replicated samples are called simply replicates. Each replicate was selected to be representative of the limited population, as well as being independent and approximately self-weighting. The sample size per replicate was somewhat arbitrarily set at 500 so the total of ten replicates would not yield an untenable number of observations. Small domains will be studied in each replicate. The sample design objectives are enumerated as follows:

1. Each replicate will be selected independently, and the total number of individuals per replicate will be 500.

Table 2.2 Comparison of response frequencies for total and limited population

AGE

Category	TOTAL No.	TOTAL Percent	LIMITED No.	LIMITED Percent
0-5	3,361	8.3	589	8.3
6-18	9,824	24.4	1,699	23.9
19-24	4,580	11.4	836	11.8
25-54	14,048	34.8	2,434	34.3
55-64	3,947	9.8	718	10.1
65-102	4,560	11.3	824	11.6
Total	40,320	100.0	7,100	100.0

SEX

Category	TOTAL No.	TOTAL Percent	LIMITED No.	LIMITED Percent
Male	19,461	48.3	3,432	48.3
Female	20,859	51.7	3.668	51.7
Total	40,320	100.0	7,100	100.0

MARITAL STATUS

Category	TOTAL No.	TOTAL Percent	LIMITED No.	LIMITED Percent
Under 17	11,377	28.2	1,979	27.9
Nev. married	6,068	15.0	1,077	15.2
Married	16,832	41.8	2,953	41.6
Widowed	2,321	5.8	420	5.9
Separated	874	2.2	158	2.2
Divorced	1,346	3.3	257	3.6
Unkown	1,502	3.7	256	3.6
Total	40,320	100.0	7,100	100.0

RACE

Category	TOTAL No.	TOTAL Percent	LIMITED No.	LIMITED Percent
White Non-Hisp.	26,167	65.0	4,627	65.2
Black Non-Hisp.	5,340	13.2	908	12.8
Hispanic	1,830	4.5	283	4.0
Other Non-Hisp.	493	1.2	94	1.3
No Answer	6,490	16.1	1,188	16.7
Total	40,320	100.0	7,100	100.0

VETERAN DISABILITY STATUS

Category	TOTAL No.	TOTAL Percent	LIMITED No.	LIMITED Percent
Dis.Male Vet.	488	1.2	86	1.2
Other Male Vet.	4,647	11.5	811	11.4
Other Male 20+	7,170	17.8	1,296	18.3
All Other	28,015	69.5	4,907	69.1
Total	40,320	100.0	7,100	100.0

PERCEIVED HEALTH STATUS

Category	TOTAL No.	TOTAL Percent	LIMITED No.	LIMITED Percent
Excellent	16,220	40.2	2,793	39.3
Good	14,812	36.7	2,596	36.6
Fair	4,510	11.2	829	11.7
Poor	1,489	3.7	266	3.7
Unknown	3,289	8.2	616	8.7
Total	40,320	100.0	7,100	100.0

LIMITATION OF ACTIVITY

Category	TOTAL No.	TOTAL Percent	LIMITED No.	LIMITED Percent
Ltd. Major Act.	1,331	3.3	247	3.5
Ltd. Amt/Kind	1,634	4.1	300	4.2
Ltd. Not Maj.	674	1.7	110	1.5
Not Limited	30,709	76.1	5,323	75.0
Unknown/NoAns.	5,972	14.8	1,120	15.8
Total	40,320	100.0	7,100	100.0

USUAL SOURCE OF CARE

Category	TOTAL No.	TOTAL Percent	LIMITED No.	LIMITED Percent
Has Usual Sour.	28,844	71.6	5,154	72.6
No Usual Source	5,483	13.6	921	13.0
Don't Know	129	0.3	17	0.2
No Answer	5 864	14.5	1,008	14.2
Total	40,320	100.0	7,100	100.0

2. Within each replicate, the frame will be stratified and the number of sample clusters selected to be representative of the limited population.
3. In order to study small domains, an equal number of elements will be selected per cluster; additionally, the secondary sampling rate within clusters will be large.
4. Each replicate will be self-weighting or very near to self-weighting.

Mindful of these objectives, it was proposed that each replicate consist of two stages with 20 clusters (a) selected from the total of 142 (A) and 25 individuals (b) from each total of 50(B) in the frame. Thus, each sample consisted of 500 individuals (20 X 25). Furthermore, an equal number of elements was selected from each cluster (25), and the sampling rate per cluster was large (f_b = b/B = 25/50 = 1/2). Thus, a very small domain of five percent still contained an average of approximately one individual per selected cluster. One of these samples is described in Table 2.3. The six region and level of urbanization strata defined for the limited population were used again, and the 20 sample PSU's were allocated to the strata in proportion to the number of clusters in the limited population (column 4 of Table 2.3) to obtain the sample PSU allocation (column 7). The number of sample PSU's per stratum was multiplied by 25 individuals per sample PSU to obtain the number of sample individuals per stratum (column 8). We see that the percentage individuals in the population and sample were very close

Table 2.3 Description of one independent replicate from limited population

Stratum			POPULATION			SAMPLE				
Stratum	Region	Urban-ization	No. PSU's	No. Indiv.	(%)	No. PSU's	No. Indiv.	(%)	Sampling Rate	Weight
1	Northeast(NE)	Urban	26	1,300	(18.3)	3	75	(15)	75/1300	17.33
2	Northcentral(NC)	Urban	22	1,100	(15.5)	3	75	(15)	75/1100	14.67
3	Combined NE and NC	Rural	18	900	(12.7)	3	75	(15)	75/900	12.00
4	Southeast (SE)	Urban	27	1,350	(19.0)	4	100	(20)	100/1350	13.50
5	West (W)	Urban	19	950	(13.4)	3	75	(15)	75/950	12.67
6	Combined SE and W	Rural	30	1,500	(21.1)	4	100	(20)	100/1500	15.00
			142	7,100	(100.0)	20	500	(100)		

to each other implying each replicate was representative of the limited population. The sampling rate is the ratio of sample to population individuals, and the weight is the inverse of the sampling rate.

While the samples were not Equal Probability SElection Method (EPSEM) or equal weighting, the effect of the weights on variance computation was measured and found to be minimal. Thus, the design was approximately EPSEM as are designs in many surveys. To measure the weight effect, one forms the ratio of the design effect from a sample with variable weights and complex design to the design effect from a complex design only (Kish, 1965, p. 427).

$$\text{Weight Effect} = \frac{\Sigma_h^6 \; W_h \; K_h^2}{\left[\Sigma_h^6 \; W_h \; K_h\right]^2}$$

where h refers to the 6 strata defined in Table 2.3; W_h is the proportion of sample elements contained in stratum -h; and K_h is the weight associated with the stratum. If there were no weight effect, this ratio would be one (1.00). For the sample, the weight effect is negligible (1.0139).

<u>Selection Of First-Stage Units:</u>

To insure an adequate representation of veterans in each replicate, the set of 142 primary sampling units was ordered by number of veterans; PSUs were then systematically selected. The procedure was as follows. First, the total number of veterans was determined for each PSU. Second, within each stratum, the PSUs were selected with equal probability by ordering alternately for each stratum from low to high and then high to low number of veterans. Then, within each stratum, systematic selection was applied to identify selected clusters (Kish, 1965, p. 230). Separate independently chosen random numbers were used to identify the systematic sample within each stratum; thus, each replicate was truly independent. Since the sampling interval for the systematic sample within the stratum was not always an integer, the selection probabilities for each cluster were not strictly equal. To insure each cluster an equal chance of selection, circular systematic sampling was used (Kish, 1965, p. 116). By this procedure, a random number between one and the total number of clusters in the stratum was identified; this unit was then selected and the sampling interval applied continuously; when the last cluster in the stratum was encountered, the list skipped back to the first cluster. By this procedure, clusters containing small, medium, and large numbers of veterans were selected in proportion to their representation in the population. We were less likely to select PSU's containing too few veterans.

Selection of Second-State Units: The goal here was to randomly select households containing a total of 25 individuals in each PSU. This result was accomplished by randomly ordering households within each PSU and individuals within each household. The first 25 individuals in each selected PSU were then chosen. In some cases, all members from the last household on the ordered list may not have been selected. No substitutions were made to insure complete households were selected.

2.2 Selection Of Domain And Analysis Variables

Actual rates of homogeneity for the limited population were calculated for 42 analysis variables and 15 domain variables using equation 1.9. There were approximately 200 analysis variables and 25 domain variables in total for the limited population. Based on a careful examination of these ranges of roh, specific domain and analysis variables were selected for further study in the dissertation. The selection of these particular variables is discussed in the following sections.

A Statistical Analysis System (SAS) macro or computer software subroutine was developed to calculate actual intracluster homogeneity in the limited population for unequal clusters and for any domain and analysis variable. This subroutine and its flowchart are included as Appendix A. It can be shown that Roh for unequal clusters (equation 1.9) reduces to Rho (equation 1.3) when clusters are equal (see Appendix B). The subroutine was carefully checked to insure it was working properly by (1) verifying the results from several textbook examples and (2) independently verifying the correct computation of between and within sum of squares by means of SAS's General Linear Models (GLM) procedure. Further specific checks to insure the subroutine was accurate are enumerated in Appendix C.

The subroutine was then used to calculate intracluster homogeneity for the continuous analysis variable of average time in minutes to get to usual source of care for all domain categories; it was similarly used to calculate Roh for the categorical analysis variable of perceived health status for all domain variables. The domain variables included age, sex, region, size of community, urban or rural status, education, race, income, occupation, veteran status, and veteran disability status.

Table 2.4. Some summary measures of intracluster homogeneity for analysis variables in limited population

Description	No. of analysis variables	Range	Median	Mean
Continuous variables	23	-0.001 to 0.070	0.012	0.016
Categorical variables	19	-0.001 to 0.179	0.060	0.070
	42			

For the continuous analysis variable, of average time in minutes to get to usual sources of care, a total of 50 Rohs were computed (1 total + 49 domain categories). For the categorical variables of perceived health status, a total of 250 Roh's were computed (50 domain and total categories X 5 analysis categories). Additionally, Roh was computed for all other analysis variables and domain categories on the file or 101 Rohs (40 analysis variables X 2.525 categories/analysis variable). The analysis variable results for the total study population are summarized in Table 2.4. Both continuous and categorical variables were found to have small rates of homogeneity (< 0.05); however, larger rates of homogeneity (> 0.10) were confined to categorical variables.

2.2.1 Determination Of Analysis Variables

Based on these results and other empirical results from section 1.2.6.1, four ranges for intracluster homogeneity were determined for analysis variables as: (1) Very Small, approximately 0.001; (2) Small, from 0.01 to 0.03; (3) Medium, from 0.05 to 0.08; and (4) Large, from 0.15 to 0.20. The analysis variables selected to represent each category are described in Table 2.5. Continuous and categorical variables are present for both the low and high range of Roh. For the two

Table 2.5 Selected analysis variables

Category	Roh	Description	Type
Very small	0.001	Average dollars for all hospital admissions	Continuous
Small	0.021	Perceived health status "good"	Categorical
Medium	0.073	Average time in minutes to get to usual source of care	Continuous
Large	0.179	Insured by Health Maintenance Organization "Yes"	Categorical

categorical variables, only one response category was selected; the response of good was chosen for perceived health status, and the response of yes was identified for whether insured by a Health Maintenance Organization (HMO) or not.

2.2.2 Determination Of Domain Variables

There were four average domain sample sizes which were of interest for cross-class variables. They are enumerated in Table 2.6. Because SESUDAAN computer software (Shah, 1981) did not print standard errors and design effects for clusters smaller than two, there was interest in

Table 2.6 Four interesting average domain sample sizes

Class	Average domain size (b_d)	Fraction of sample	Percentage
1	1	1/25	4%
2	2 - 3	2/25 - 3/25	8 - 12%
3	5 - 6	5/25 - 6/25	20 - 24%
4	$\geq$ 10	$\geq$ 10/25	$\geq$ 40%

Table 2.7 Selected domain variables

Description	Categories and percentages	Total cat.	Type
Race	White non-Hispanic (65%), Black non-Hispanic (13%), Hispanic or other (5%)	3	Cross-class
Age	0-18 (32%), 19-59 (52%), 60-64 (4%), 65-102(12%)	4	Cross-class
Disabled veteran status	Disabled male vet (1%), Other male vet (11%), Other male 20+(18%), All other (70%)	4	Cross-class
Urbanization	Urban, rural	2	Segregated
		13	

the first two classes. CLUSTERS software (Verma, 1978) censored results for clusters of size 6 or smaller. This fact accounted for the interest in class 3. Lastly, results needed to be compared for small and large domains; thus, the selection of the last class which is a large domain. These categories were not inclusive but represented a range felt to be appropriate for this research.

All domain variables were then examined to see which included some or all of these interesting categories. The three cross-class domains listed in Table 2.7 appeared to be good candidates. Age categories were artifically formed to include the four interesting size categories. The fourth variable was a segregated domain variable selected because the rural category produced a relatively large Roh (0.29) for average time to get to usual source of care.

2.3 Primary Purpose: *To Determine Under What Circumstances The Design Effect And Proportion Variation Estimation Methods Are Appropriate To Calculate The Rate Of Homogeneity*

2.3.1 Estimation of Roh Using Design Effect Method

Using equation 1.10, Roh was estimated from the design effect (deff) and average sample cluster size ($\overline{b}$) for the four analysis variables in each of the 13 domain categories. For each sample, a total of 52 rates of homogeneity (4 X 13) were thus calculated by the design effect method. Furthermore, these rohs were calculated in each of ten replicates; the total number of rohs determined by the design effect method was 520 (52 X 10).

The design effect and average sample cluster size was determined using SESUDAAN, a program developed at Research Triangle Institute (RTI), Research Triangle Park, North Carolina to compute survey estimates and their standard errors from complex multistage samples (Shah, 1981). The program produced, in addition to the estimate and standard error, the deff and sample size. The program also counted the number of PSUs used to calculate the estimate. The average sample cluster size ($\overline{b}$) was then calculated by dividing the sample size by the number of PSUs in the sample. For a cross-class domain, the average sample cluster size was the ratio of the number of sample elements in the domain category to the total number of clusters in the sample ($a = 20$). For a segregated domain, the average sample cluster size was the number of sample elements in the domain category divided by the number of sample clusters containing nonzero elements for that domain category. A SAS computer program was written to compute roh using equation 1.10 and given the design effect and average cluster size. To calculate the variance of the design, SESUDAAN used a first-order Taylor

Series approximation of the squared deviations of estimates from their expected values. The variance of this linear approximation was obtained using the squared deviations in the ultimate cluster variance estimator (Shah, 1981). SESUDAAN estimated deff by forming the ratio of the variance of the estimate in the given design to the variance of the estimate in a simple random sample of the same size. SESUDAAN used the following formula for the estimated variance of a ratio-type estimate (r).

$$\mathrm{var}(r) = \frac{\Sigma_h^H \; \Sigma_i^{a_h} \; \Sigma_j^{b_{hi}} (W_{hij}(Z_{hij} - Z)^2)}{\Sigma_h^H \; \Sigma_i^{a_h} \; \Sigma_j^{b_{hi}} \; W_{hij}},$$

where $Z_{hij} = Y_{hij} - rX_{hij}$;

$$r = \frac{\Sigma_h^H \; \Sigma_i^{a_h} \; \Sigma_j^{b_{hi}} \; W_{hij} \; Y_{hij}}{\Sigma_h^H \; \Sigma_i^{a_h} \; \Sigma_j^{b_{hi}} \; W_{hij} \; X_{hij}} ;$$

and $$Z = \frac{\Sigma_h^H \; \Sigma_i^{a_h} \; \Sigma_j^{b_{hi}} \; W_{hij} \; Z_{hij}}{\Sigma_h^H \; \Sigma_i^{a_h} \; \Sigma_j^{b_{hi}} \; W_{hij}} .$$

In the preceding formulas, it was assumed that the sample was composed of H strata; that a_h primary selections (PSUs) were made from stratum-h; and that b_{hi} secondary selections were made within PSU-i from stratum-h. It was further assumed that for stratum-h, PSU-i, and secondary units-j, W_{hij}, Y_{hij}, and X_{hij} represented the weight, numerator, and denominator for the unit, respectively.

2.3.2 Estimation of Roh Using Proportion Variation Method

As estimate for roh from a stratified random sample was derived in Chapter 1 and specified in equation 1.12. This estimator is called the proportion variation method of estimation. The equation was derived from the population definition for Roh specified in Hansen, Hurwitz, and Madow (1953) using two-stage, unequal clusters. The estimator can also be shown to be consistent according to the definition provided in Hansen, Hurwitz, and Madow (1953), and this is done in Appendix F. We think the proportion variation method or equation 1.12 will be superior to the design effect method or, equation 1.10 for estimating large roh, and it was this hypothesis which would be tested as the primary purpose of the dissertation.

As with the design effect method, 52 point estimates for roh by the proportion variation method were obtained for each sample by computing roh for each analysis variable in each domain category. A computer software macro was written to execute equation 1.12 (proportion variation method). For the ten samples, a total of 520 point estimates for roh were obtained.

2.3.3 Comparison Of Methods

Sample rates of homogeneity were calculated by the design effect and proportion variation estimation methods for each of the 52 different categories formed by the four analysis variables in each of the 13 domain categories. The sample rohs calculated by each estimation method were determined separately for each of the ten samples. Additionally, the true rate of homogeneity was calculated from the limited population for each of the 52 categories. A relative deviation was determined for each estimation method in each of the 52 categories and ten samples by comparing the sample roh with the true value and dividing by the true value. The differences in the absolute values of the relative deviations for each estimation method were calculated for each of the 52 domain categories and ten samples. For each domain category, these differences were averaged across the ten samples. A paired data t test was used to examine these average differences in absolute relative deviations for each of the 52 domain categories. The testing procedure is detailed in the remainder of this section.

A relative deviation was calculated for each estimation method in each of the 52 categories and ten samples as

$$RD_{ijk} = \frac{r_{ijk} - \rho_i}{\rho_i},$$

where i = domain category from 1 to 52;

j = estimation method and 1 is design effect, while 2 is proportion variation;

k = replicate from 1 to 10;

r_{ijk} = sample rate of homogeneity for ith domain category using jth estimation method from kth sample;

ρ_i = true rate of homogeneity for domain category -i.

A difference in the absolute values of the relative deviations for each estimation method was calculated for each domain category (i) and sample (k) as

$$d_{i.k} = |RD_{i1k}| - |RD_{i2k}|.$$

Absolute values were used to determine which of the two relative deviations was closer to zero (i.e., the difference of absolute values) or which of the two methods actually came closer to estimating the true rate of homogeneity.

An average difference was then calculated for each domain category across the ten samples as

$$\overline{d}_{i..} = \frac{\Sigma_k^{10} d_{i.k}}{10} .$$

It is shown it Appendix E that these average differences in absolute values, minus their expected values, and divided by their standard errors follow Student's t distribution with 9 degrees of freedom. The standard error for $\overline{d}_i$, is defined as

$$\text{ste}\,(\overline{d}_{i..}) = \sqrt{\frac{\Sigma_k^{10} d_{i.k}^2 - \frac{\{\Sigma_k^{10} d_{i.k}\}^2}{10}}{9(10)}}$$

where $d_{i.k} = |RD_{i1k}| - |RD_{i2k}|$.

For a cross-class domain, the design effect method was conjectured to be superior to the proportion variation method (i.e., $|RD_{i1k}|$ was closer to zero than $|RD_{i2k}|$ and $d_{i.k}$ was, thus, negative.) The true average difference, or Δ_i.., was thus negative.

In a similar manner, the proportion variation method was speculated to be superior to the design effect method for segregated domains. That is, relative deviations from method 2 were conjectured to be small and Δ_i.., the true average difference in absolute values, to be positive.

Suppose we examined a confidence interval for the average difference associated with each of the 52 domain categories (i)

$$\bar{d}_i.. \pm t_{9,.95} \text{ ste } (\bar{d}_i..)$$

where $t_{9,.95}$ was the t distribution critical value associated with 9 degrees of freedom and a single tail probability of $\alpha/2 = .025$. If the confidence interval was mostly positive, then the true average difference ($\Delta_i..$) was greater than zero and the proportion variation method was generally better than the design effect method. If the confidence interval was entirely positive, then the proportion variation method was significantly preferred at the five percent level. Alternately, if the confidence interval was mostly negative, then the true average difference was less than zero and the design effect method was probably superior to the proportion variation method. The design effect method was significantly preferred at the five percent level, if the confidence interval was entirely negative.

By this procedure, it could be determined whether the design effect or proportion variation method was superior to estimate roh for domain size, domain type, size of roh for analysis variable, and whether the analysis variable was continuous or categorical. The confidence interval might not be significant in every case; however, by examining its general direction, we gained some feeling for whether the design effect method was superior (i.e., the confidence interval was generally negative) or the proportion variation method was preferred (i.e., the confidence interval was usually positive). By examining these confidence intervals, we gained some insight into when each method was appropriate to estimate roh for very small, small, medium, and large domain sizes;

for segregated and cross-class domain types; for very small, small, medium, and large rohs for analysis variables; and for categorical and continuous analysis variables.

2.4 Secondary Purpose: *To Investigate How The Relationship Between The Analysis Variable And The Nature Of The Domain May Determine The Appropriate Formula*

The secondary purpose was investigated in two parts. First, how the estimate of average domain ratio was affected by estimation method was explored. Secondly, the relationship between the proportion of the domain contained in a particular category ($\overline{b}_d/\overline{b}_t$) and the ratio of roh for domain to roh for the total population (roh_d/roh_t) was investigated.

2.4.1 Evaluate How Average Domain Ratio May Determine Appropriate Formula For Roh

The average domain ratio (ADR) as suggested by Kalsbeek (1983) was defined for a particular analysis variable and domain variable as the ratio of the arithmetic average of the rates of homogeneity for the analysis variable across all domain categories to the analysis variable roh for the total population estimate. For the first analysis variable of average dollars for all hospital admissions and the first domain variable of urban or rural, the first population ADR was defined as

$$\Lambda_1 = \frac{\frac{\rho_1 + \rho_2}{2}}{\rho_{t1}} = \frac{\rho_1 + \rho_2}{2\rho_{t1}}$$

where ρ_1 was the rate of homogeneity for average dollars on all hospital admissions for urban members of the population; ρ_2 was the rate of homogeneity for average dollars on all hospital admissions for rural members of the population; and ρ_{t1} was roh for average dollars on all hospital admissions for all members of the population.

In a similar manner, the true population ADRs were calculated for the other 15 analysis and domain variables by averaging the true rates of homogeneity across domain categories and dividing by the true rate of homogeneity for the total population estimate. There were a total of 16 analysis and domain variables, obtained from the four analysis variables and four domain variables.

Using sample rates of homogeneity calculated by the design effect and proportion variation methods, sample average domain ratios were computed for each of the 16 analysis and domain variables by averaging sample rohs across domain categories and dividing by sample roh for the total population estimate. These average domain ratio estimates were made separately for each of the ten replicates. A relative deviation was calculated for each estimation method in each of the ten replicates and for each of the 16 analysis and domain variable categories as:

$$RD_{ijk} = \frac{\lambda_{ijk} - \Lambda_i}{\Lambda_i},$$

where i = analysis and domain variable categories from 1 to 16;

j = estimation method where 1 is design effect and 2 is proportion variation;

k = sample from 1 to 10;

λ_{ijk} = sample average domain ratio for analysis and domain category -i, estimation method -j, and sample -k;

Λ_i = true population average domain ratio for ith analysis and domain variable category.

For each of the 16 analysis and domain variable categories (i), a difference in the absolute values of the relative deviations for each estimation method was made separately for each replicate (k) as

$$d_{i.k} = |RD_{i1k}| - |RD_{i2k}| .$$

An average difference across the ten replicates was obtained for each of the 16 analysis and domain variable categories (i) as

$$\overline{d}_{i..} = \frac{\Sigma_k^{10} d_{i.k}}{10} .$$

As described earlier, confidence intervals based on the the paired data t test were used to compare average differences in absolute values of relative deviations by estimation method. If the confidence interval was generally positive, then the proportion variation method was better than the design effect method for estimating average domain ratio. Conversely, if the confidence interval was mostly negative, then the design effect method was superior to the proportion variation method.

Thus, whether the design effect or proportion variation method was generally superior to estimate the average domain ratio could be determined. Furthermore, this differentiation could be made for

1. Segregated or cross-class domain variable type;
2. Continuous or categorical analysis variable type;
3. Very small, small, medium, and large rohs for analysis variables.

We conjectured that the proportion variation method would be superior for segregated domains and the design effect method preferred for cross-class domains.

2.4.2 Investigate How The Proportion Of The Domain Contained In A Particular Category May Affect The Appropriate Formula

The ultimate goal of this portion of the dissertation was to model the proportion of the domain contained in a particular category ($\bar{b}_d/\bar{b}_t$) as a function of the ratio of roh for domain to roh for the total population estimate ($roh(y_d)/roh(y_t)$). Verma et al (1980) have proposed such a model. Their model was first reviewed; then the model was verified by substituting true population values. Finally, it was determined whether estimation method affected the performance of the Verma et al model by means of the confidence interval procedure described earlier.

Verma et al (1980) noted that the ratio $roh(y_d)/roh(y_t)$ was composed of two small numbers; $roh(y_d)$ and $roh(y_t)$ were the rates of homogeneity for the estimate, y, in the domain category, d, and in the total population, t, respectively. In the majority of cases roh was less than 0.1 (Kish, 1965). They also noted that this ratio was often close to one (1.00). Thus, Verma et al used the ratio $(deff\ (y_d) -1)/(deff\ (y_t) -1)$

to model the proportion of the domain contained in a particular category; deff (y_d) and deff(y_t) were the design effects for the estimate, y, in the domain category, d, and in the total population, t.

Their approximation was obtained as follows. From the basic equation (Kish, 1965)

$$\text{deff}(y) = 1 + \text{roh}(y)\ (\bar{b}_y - 1)$$

they obtained that

$$\text{deff}(y) - 1 + \text{roh}(y) = \text{roh}(y)\ \bar{b}_y .$$

In the expression above, $\bar{b}_y$ was the average sample cluster size for the estimate y. Therefore, the ratio of the product of roh and the average sample cluster size for domain and total was

$$\frac{\text{roh}(y_d)\bar{b}_{y_d}}{\text{roh}(y_t)\bar{b}_{y_t}} = \frac{\text{deff}(y_d) \ \ -1 + \text{roh}(y_d)}{\text{deff}(y_t) \ \ -1 + \text{roh}(y_t)} .$$

Assuming roh (y_d) and roh(y_t) were both small and their ratio was near one (1.00), then the expression above was nearly

$$\frac{\bar{b}_{y_d}}{\bar{b}_{y_t}} = \frac{\text{deff}(y_d) \ \ -1}{\text{deff}(y_t) \ \ -1} = M_d .$$

Verma et al called this ratio M_d, the proportion in the domain category, d. More generally, this proportion was expressed as

$$\left[\frac{\bar{b}_{y_d}}{\bar{b}_{y_t}}\right]^{\alpha} = (M_d)^{\alpha} = \frac{\text{deff}(y_d) \ \ -1}{\text{deff}(y_t) \ \ -1} \qquad (2.2)$$

where α was empirically determined by least squares, as explained below, and expected to range from zero (0.00) to one (1.00); zero corresponded to segregated domains and one was associated with cross-class domains.

Verma et al suggested that separate estimates of α be made for similar analysis variables in domains with similar proportions. The procedure for estimating α by least squares follows.

First, goodness of fit was found to improve considerably when indivi-dual points were weighted by their relative sample sizes, M_d; thus, equation 2.2 was multiplied by M_d to obtain

$$(M_d) \frac{\mathrm{deff}(y_d) \quad -1}{\mathrm{deff}(y_t) \quad -1} = M_d^{\alpha + 1}.$$

Secondly, the above expression was inverted to fulfill the basic linear model requirement of homoschedasticity or equal variance for each independent observation of M_d

$$\left[M_d \quad \frac{\mathrm{deff}(y_d) \quad -1}{\mathrm{deff}(y_t) \quad -1} \right]^{-1} = M_d^{\quad - (\alpha + 1)}.$$

Lastly, the expression was linearized by taking natural logarithms

$$- \ln \left[M_d \quad \frac{\mathrm{deff}(y_d) \quad -1}{\mathrm{deff}(y_t) \quad -1} \right] = - (\alpha + 1) \ln M_d$$

or $y_d = (1 + \alpha)x_d$

where $y_d = -\ln (M_d D_d)$; (2.3)

$$D_d = \frac{\mathrm{deff}(y_d) - 1}{\mathrm{deff}(y_t) - 1}; \text{ and}$$

$$x_d = -\ln (M_d) .$$

Equation 2.3 was now in the form of a linear model with the line forced through the origin ($b_o = 0$)

$$y = bx$$

with $b = (1 + \alpha)$. The least squares estimate of b was

$$\check{b} = (\Sigma xy)/(\Sigma x^2).$$

Similarly, the least squares estimate of $(1 + \alpha)$ was

$$(1 + \alpha) = (\sum_d x_d y_d)/(\sum_d x_d^2) .$$

Forcing the line through the origin corresponded to the total sample or $M_d = 1$. Alpha was then estimated as

$$\check{\alpha} = \frac{\sum_d x_d y_d}{\sum_d x_d^2} - 1 .$$

The goodness of fit of the model was examined by the coefficient of determination, R^2, which was the ratio of regression sum of squares to total sum of squares. R^2 was determined as follows.

$$R^2 = \frac{\text{Regression sum of squares}}{\text{Total sum of squares}}$$

$$= \frac{\text{Total sum of squares - Error sum of squares}}{\text{Total sum of squares}}$$

$$= 1 - \frac{\text{Error sum of squares}}{\text{Total sum of squares}}$$

$$= 1 - \frac{\sum_d (y_{\text{predicted}} - y_{\text{observed}})^2}{\sum_d (y_{\text{observed}} - \bar{y})^2}$$

$$= 1 - \frac{\sum_d ((1 + \alpha)x_d - y_d)^2}{\sum_d (y_d - \bar{y})^2}$$

where $\bar{y}$ was the arithmetic mean across all analysis variables and all domains considered.

In keeping with Verma et al's suggestion that separate estimates be made for similar analysis variables in domains with similar proportions, the model degree of segregation (or cross-classness), α, was determined and the coefficient of determination, R^2, estimated separately for five domain types which are explained below. Estimates for α and R^2 were pooled across the four analysis variables.

The five domain types consisted of

1. Segregated domains;
2. Very small cross-class domains where M_d is approximately 0.04;
3. Small cross-class domains where M_d is between 0.08 and 0.12;
4. Medium cross-class domains where M_d ranges from 0.20 to 0.24;
5. Large cross-class domains where M_d is 0.40 or greater.

For each of the five domain types, the domain categories (d) which were summed over to estimate α and R^2 were

1. Segregated - Urban and rural;
2. Very small cross-class - Race of Hispanic or other, age of 60 to 64 years, and disabled male veteran;
3. Small cross-class - Race of Black non-Hispanic, age of 65 to 102 years, and other male veteran;
4. Medium cross-class - Age of 0 to 18 years and other males at least age 20 who are not veterans;
5. Large cross-class - Race of White non-Hispanic, age of 19 to 59 years, and females or males whose veteran status was not determined.

The domain categories were taken from the domain size categories and domain variables defined in Tables 2.6 and 2.7.

The degree of segregation (or cross-classness), α, was modeled, and the coefficient of determination, R^2, estimated separately for the population and for each sample using the design effect and proportion variation methods to estimate roh and in turn, the design effect (deff). First, for the limited population, α_i and R_i^2 were determined for each of the five domain types.

α_i = degree of segregation (or cross-classness) for for domain type -i (i = 1,...5);

R_i^2 = coefficient of determination for domain type -i (i = 1,...,5).

By examining R_i^2, it was determined whether the Verma et al model worked. If it worked, we would expect large R_i^2, say near 0.75 or 0.80. Secondly, assuming the model worked, α_{i1k} and R_{i1k}^2 would be determined separately for each of the 10 replicates using the rates of homogeneity calculated from the design effect method.

α_{i1k} = degree of segregation (or cross-classness) for domain type -i (i=1, ...,5), design effect estimation method(j = 1), and sample -k (k = 1,...,10);

R_{i1k}^2 = coefficient of determination for domain type -i (i = 1,...,5), design effect estimation method (j = 1), and sample -k (k = 1,...,10).

We expect R_{i1k}^2 to be similar to those obtained by Verma et al (i.e., 0.32 to 0.79). Thirdly, if the model works, α_{i2k} and R_{i2k}^2 would be calculated separately for each of the ten samples using rohs determined by the proportion variation method.

α_{i2k} = degree of segregation (or cross classness) for domain type -i (i = 1,...,5), proportion variation estimation method (j = 2), and sample -k(k = 1,...,10);

R_{i2k}^2 = coefficient of determination for domain type -i (i = 1,...,5), proportion variation estimation method (j = 2), and sample -k(k = 1,...10).

Comparing differences in degree of segregation and coefficient of determination for the two estimation methods, we may notice higher R^2 for certain domain categories. This might imply the proportion variation estimation method is superior to the design effect method under certain conditions.

To further investigate this phenomenon for the degree of segregation, one-sided t tests on the average differences in absolute values of relative deviations are suggested. The relative deviation is first calculated for each of the two estimation methods (j), in each of the five domain types (i), and for each of the ten samples (k) as

$$RD_{ijk} = \frac{\alpha_{ijk} - \alpha_i}{\alpha_i},$$

where α_{ijk} = degree of segregation for domain type -i (i = 1,...,5), estimation method -j (j = 1,2), and sample -k (k = 1,...,10);

α_i = true degree of segregation for domain type -i (i = 1,...5).

Secondly, the difference in the absolute values of the relative deviations are determined for each domain type (i) and replicate (k) as

$$d_{i.k} = | RD_{i1k} | - | RD_{i2k} | .$$

By averaging across the 10 replicates, an average difference was obtained for each of the five domain types (i) as

$$\bar{d}_{i..} = \frac{\Sigma_k^{10} d_{i.k}}{10} .$$

Confidence intervals based on the average differences in absolute values of the relative deviations will be examined separately for the five categories. If the confidence interval for each category was predominantly negative, then the design effect method was superior to the proportion variation method for estimating degree of segregation. Conversely, if the confidence interval was generally positive, then the proportion variation method was better. In the same manner, the methods of estimation will be compared for the computation of the coefficient of determination with ${R_{ijk}}^2$ and ${R_i}^2$ substituted for α_{ijk} and α_i in the preceding formulas.

By these methods, a determination was made as to (1) whether the domain proportion could be modeled as a function of the ratio of roh for domain to roh for the total population estimate; and (2) if the model held, whether the design effect or proportion variation method yielded the superior model for domain type (segregated and cross-class) and domain size (very small, small, medium, and large). Assuming the model fit, the proportion variation method was conjectured to be superior for segregated domains and the design effect method preferred for cross-class domains.

2.5 Analyses Of Interest To Veterans

Suggested domain and analysis variables for the two veteran oriented analyses specified in section 1.3.3 are described in Tables 2.8 and 2.9, respectively. The total NMCES file was used for these analyses rather than the limited population. This decision was made because the Veterans Administration was interested in a specific subpopulation on.

the NMCES file called veteran users of medical services and defined as (1) veterans with (2) a physician hospital outpatient visit or hospital admission in 1977. There were found to be 1,110 of these veteran users on the total file. Thus, in the 1/6 sample limited population there would be about 185(1/6 X 1,110) veteran users. This number was felt to be too small for any meaningful analysis when it was further subdivided by domains such as urbanization and region. The possibility of including all veteran users in the limited population and applying differential weights was considered but rejected as being (1) complicated and contrived; and (2) not in keeping with the primary purpose of this research. Thus, it was decided to use the total NMCES file for veteran-related analyses. The results of these analyses are discussed in Chapter 3.

Table 2.8 Recommended domain and analysis variables to calculate rates of homogeneity

Domain:	Urban veteran users of medical services Rural veteran users of medical services
	Northeastern veteran users of medical services Northcentral veteran users of medical services Southeastern veteran users of medical services Western veteran users of medical services
Analysis:	Average number of visits Average cost of visits Average time in minutes to get to usual source Average number of prescribed medicines Average cost of prescribed medicines Average number of dental visits Average cost of dental visits

Table 2.9 Suggested domain and analysis variables to compare differences in domain proportions

Domain:	Veterans Nonveterans (males, over age 20)
	Veteran users of medical services Other Veteran
	Urban veteran users of medical services Rural veteran users of medical services
Analysis:	Average age Average income Average years of education Average number of visits Average cost of visits Average time in minutes to get to usual source of care Average number of prescribed medicines Average cost of prescribed medicines Average number of dental visits Average cost of dental visits

Specific output for the first Veterans Administration (VA) analysis, defined in section 1.3.3, is specified below:

1. A total of 42 rates of homogeneity will be calculated from the total NMCES file using the NMCES weight and the appropriate formula determined from the dissertation. Each of the six domain variables in Table 2.8 is a segregated domain; the first two involve urban or rural status, and the last four are regional classifications. The 42 rohs will be determined from the six domain variables and seven analysis variables in Table 2.8.
2. The rates of homogeneity will be interpreted and recommendations made as to the generalizability of findings based on the magnitude of roh.

The second VA analysis may be summarized as follows:

1. For each of the three domain difference categories defined in Table 2.9 a difference in the proportion in each category will be calculated for each analysis variable. The three domain difference categories are:
   a. Veterans and nonveterans;
   b. Veteran users of medical services and other veterans; and
   c. Urban and rural veteran users of medical services.

   SESUDAAN calculates these differences as well as the standard errors of the differences. A total of 30 differences and standard errors will be determined from the three domain difference categories and the ten analysis variables of Table 2.9.

2. Using large sample theory, the null hypothesis of no difference in the two proportions for each analysis and domain category -i ($i = 1, \ldots, 30$) may be tested as

$$Z = \frac{\check{p}_{i1} - \check{p}_{i2}}{\text{ste}(\check{p}_{i1} - \check{p}_{i2})} \tag{2.4}$$

where $(p_{i1} - p_{i2})$ is the difference in proportions for analysis and domain category -i and ste $(p_{i1} - p_{i2})$ is the standard error of that difference. Z is distributed as a normal random variable with mean 0 and variance 1.

3. The Z calculated from equation 2.4 will be compared with a critical value associated with a single tail probability of $\alpha/2$ from the normal distribution with mean 0, variance 1, and Type I error rate of α. For example, to test the alternative hypothesis of a difference in the proportions, Z calculated from equation 2.4 will be compared with the critical Z values containing .05/2 probability in a single tail of the distribution. If the tabular value exceeds the critical value in absolute magnitude, then we assume a significant difference in the two proportions. Otherwise, we assume no difference between the proportions.

CHAPTER 3

Findings

3.1 Results From Primary Objective

The primary purpose of this dissertation was to quantify when the design effect and proportion variation estimation methods were appropriate to estimate roh. Sampling rates of homogeneity were computed by the design effect and proportion variation estimation methods for each of the four analysis variables and 13 domain categories in each of the ten replicates. These sampling rates of homogeneity are summarized by analysis variable, estimation method, and type of domain in Table 3.1. Of particular interest are the four domain size categories; size 1 is a small domain containing about five percent of the population, and size 4 is a large domain composing approximately 50 percent of the population. These domain size categories are not inclusive and were subjectively chosen to represent a range of values felt to be appropriate for this research. Their selection and description are detailed in Table 2.6. For each domain class, estimates for mean roh were obtained by averaging across replicates for all categories in the domain class. For example, the total class included ten estimates, one from each replicate; urbanization contained 20 estimates or 10 replicates from each category of urban and rural. Ideally, the number of estimates for each size class should have been 30, 30, 20, and 30, respectively. However, for some replicates, sample observations were not present for the particular classification.

After examining Table 3.1, it can be seen, first, that the mean across replicates and domain categories for each class for both the proportion variation and design effect methods are generally larger than the true value implying that both methods are overestimating the true parameter. Secondly, we observe that means for the design effect method are generally larger than the proportion variation method implying that, although neither method is particularly accurate, the proportion variation method is generally closer approximating the true value. The design effect method occasionally generated estimates which exceeded (1.000) as evidenced by the mean value of 1.666 for the last analysis variable in domain size class 1 of Table 3.1. This situation occurred in size class 1 when a large design effect (deff $(\bar{\bar{y}})$ in equation 1.10) was larger than a small average sample cluster size $(\bar{b})$. Thirdly, variation among estimates obtained by the proportion variation method as measured by the standard deviation among estimates are generally smaller than for the design effect method; thus, the proportion variation method generates a more compact distribution than the design effect method. However, for the two continuous analysis variables in segregated domains, we observed that variability among estimates was greater for the proportion variation method than the design effect method. We also noted that the mean for the proportion variation method was larger than the design effect method. This occurrence was again observed in the computation of rates of homogeneity for veteran-related continuous analysis variables in segregated domains discussed in section 3.3.1. And lastly, for each estimation method, variation among estimates is generally larger for small domains than large ones implying both methods

Table 3.1 Comparison of sample rates of homogeneity by estimation method and analysis variable

Categorical analysis variable: Perceived Health Status "Good"

Domain			Design effect		Proportion variation		
Type	Class	No. est.[1]	Mean	Std. dev.[3]	Mean	Std. dev.[3]	True Mean Roh[2]
-	Total	10	0.059	0.057	0.054	0.051	0.021
Segregated	Urbanization	20	0.057	0.063	0.050	0.055	0.021
Cross-class	Size 1	25	0.122	3.999	0.091	0.414	0.139
Cross-class	Size 2	30	0.262	0.515	0.065	0.206	0.052
Cross-class	Size 3	20	0.165	0.150	0.095	0.099	0.034
Cross-class	Size 4	30	0.065	0.065	0.056	0.053	0.027

Categorical analysis variable:
Insured by Health Maintenance Organization "Yes"

-	Total	10	0.152	0.112	0.141	0.090	0.179
Segregated	Urbanization	10	0.085	0.114	0.143	0.090	0.125
Cross-class	Size 1	6	0.629	3.393	0.349	0.473	0.315
Cross-class	Size 2	16	0.122	0.346	0.142	0.374	0.225
Cross-class	Size 3	18	0.090	0.261	0.058	0.208	0.172
Cross-class	Size 4	29	0.140	0.142	0.139	0.111	0.191

Continuous analysis variable:
Average dollars for all hospital admissions

-	Total	10	-0.001	0.014	0.002	0.008	0.001
Segregated	Urbanization	20	-0.001	0.018	0.005	0.020	0.001
Cross-class	Size 1	19	-0.050	1.846	0.169	0.438	0.493
Cross-class	Size 2	30	-0.019	0.165	-0.048	0.096	0.001
Cross-class	Size 3	20	0.012	0.070	0.009	0.055	-0.005
Cross-class	Size 4	30	-0.005	0.020	-0.005	0.018	0.003

Continuous analysis variable:
Average time in minutes to get to usual source of care

-	Total	10	0.061	0.028	0.135	0.058	0.077
Segregated	Urbanization	20	0.059	0.045	0.141	0.085	0.134
Cross-class	Size 1	25	1.666	2.378	0.547	0.437	0.290
Cross-class	Size 2	30	0.248	0.582	0.278	0.175	0.118
Cross-class	Size 3	20	0.119	0.163	0.292	0.130	0.067
Cross-class	Size 4	30	0.068	0.045	0.157	0.071	0.077

[1]Estimates were obtained by averaging across replicates and domain categories within each class.
[2]Obtained by averaging true Rohs from study population across domain categories for the particular domain class.
[3]Standard deviation defined as square root of variance among sample estimates of roh across replicates and domain categories within each class.

are more variable for estimating roh in small domains than large ones. This last observation provides an indication that both estimators may be consistent; i.e., smaller sample size implies greater instability of estimates. We show in Appendix F that the proportion variation estimator is consistent using the definition of consistency provided in Hansen, Hurwitz, and Madow (1953, Volume II).

Now comparing in detail the means from the two estimation methods in Table 3.1 with the true mean Roh, it appears generally that neither method accurately estimates the true value. True mean Rohs were obtained by averaging true Roh from the study population across domain categories for the particular domain class. We see, as observed earlier, that both the design effect and proportion variation method generally overestimate the true value.

Mean sample rohs are compared with true Rohs separately by estimation method and domain category for categorical and continuous analysis variables in Tables 3.2 and 3.3, respectively. Means were obtained by averaging sample estimates for roh across replicates. First, for categorical analysis variables in Table 3.2, we generally notice that the design effect estimate exceeded the estimate from the proportion variation method. We also notice for the first categorical analysis variable where true Roh is small that both methods provided overestimates; whereas, for the second analysis variable, where true Roh is larger, both methods supplied underestimates. Second, examining continuous analysis variables in Table 3.3, we generally observe for the second analysis variable where Roh is large that the proportion variation estimate is larger than the design effect estimate. Third, we notice mean rohs estimated by the design effect method which exceeded

Table 3.2 Sample rates of homogeneity by estimation method and domain variables for categorical analysis variables.

Categorical analysis variable: Perceived health status "Good"

Domain			No.	Design effect		Proportion variation		True
Type	Size	Category	Est.	Mean[2]	Std. dev.[3]	Mean[2]	Std. dev.[3]	Roh
-	-	Total	10	0.059	0.057	0.054	0.051	.021
Segregated	-	Urban	10	0.062	0.060	0.063	0.060	.022
Segregated	-	Rural	10	0.051	0.069	0.037	0.050	.019
Cross-class	1	Hisp. & Other	10	2.644	3.663	0.147	0.287	.149
Cross-class	2	Black	10	0.683	0.614	0.116	0.224	.117
Cross-class	4	White	10	0.083	0.075	0.062	0.060	.036
Cross-class	1	60-64	10	-2.004	5.337	0.140	0.340	.077
Cross-class	2	65+	10	0.175	0.366	0.126	0.227	.052
Cross-class	3	0-18	10	0.175	0.105	0.122	0.061	.074
Cross-class	4	19-59	10	0.053	0.069	0.052	0.059	.013
Cross-class	1	Dis. Male Vet.	5	0.017	0.514	-0.169	0.781	.190
Cross-class	2	Other Male Vet.	10	-0.073	0.116	-0.046	0.117	-.012
Cross-class	3	Other Male 20+	10	0.156	0.191	0.069	0.124	-.006
Cross-class	4	All Other	10	0.060	0.050	0.054	0.045	.031

Categorical analysis variable: Insured By Health Maintenance Organization "Yes"

Type	Size	Category	No. Est.	Design effect Mean[2]	Design effect Std. dev.[3]	Proportion variation Mean[2]	Proportion variation Std. dev.[3]	True Roh
-	-	Total	10	0.152	0.112	0.141	0.090	.179
Segregated	-	Urban	10	0.085	0.114	0.143	0.090	.176
Segregated	-	Rural	-[1]	-[1]	-[1]	-[1]	-[1]	.073
Cross-class	1	Hisp. & Other	6	0.629	3.393	0.299	0.448	.551
Cross-class	2	Black	3	0.088	0.182	0.080	0.052	.243
Cross-class	4	White	9	0.167	0.166	0.174	0.115	.212
Cross-class	1	60-64	-[1]	-[1]	-[1]	-[1]	-[1]	.428
Cross-class	2	65+	4	0.184	0.546	0.192	0.552	.293
Cross-class	3	0-18	10	0.222	0.301	0.165	0.202	.209
Cross-class	4	19-59	10	0.103	0.109	0.109	0.101	.175
Cross-class	1	Dis. Male Vet.	-[1]	-[1]	-[1]	-[1]	-[1]	-.035
Cross-class	2	Other Male Vet.	9	0.094	0.217	0.129	0.333	.139
Cross-class	3	Other Male 20+	8	-0.042	0.120	-0.075	0.126	.135
Cross-class	4	All Other	10	0.151	0.153	0.138	0.118	.185

[1]Sample observations were not present for this classification.

[2]Mean was obtained by averaging sample estimates for roh across replicates.

[3]Standard deviation defined as square root of variance among sample estimates of roh across replicates.

Table 3.3 Sample rates of homogeneity by estimation method and domain variable for continuous analysis variables.

Continuous analysis variable: Average dollars for all hospital admissions

Domain Type	Domain Size	Domain Category	No. Est.	Design Effect Mean[1]	Design Effect Std. dev.[2]	Proportion variation Mean[1]	Proportion variation Std. dev.[2]	True Roh
-	-	Total	10	-.001	0.014	0.002	0.008	0.00100
Segregated	-	Urban	10	-.004	0.010	-0.003	0.009	0.00300
Segregated	-	Rural	10	.003	0.024	0.013	0.024	-0.00200
Cross-class	1	Hisp. & Other	6	.236	2.540	-0.098	0.105	0.02100
Cross-class	2	Black	10	-.058	0.136	-0.060	0.103	0.00100
Cross-class	4	White	10	.011	0.017	0.003	0.019	0.00600
Cross-class	1	60-64	9	-.504	2.228	0.322	0.377	0.56400
Cross-class	2	65+	10	.054	0.224	0.001	0.099	0.00700
Cross-class	3	0-18	10	.006	0.066	0.005	0.034	0.00200
Cross-class	4	19-59	10	-.011	0.020	-0.015	0.016	0.00005
Cross-class	1	Dis. Male Vet.	4	.129	0.321	0.206	1.123	0.89500
Cross-class	2	Other Male Vet.	10	-.053	0.101	-0.085	0.073	-0.00800
Cross-class	3	Other Male 20+	10	.019	0.077	0.013	0.071	-0.01200
Cross-class	4	All Other	10	-.015	0.014	-0.001	0.013	0.00400

Continuous analysis variable: Average time in minutes to get to usual source of care

Domain Type	Domain Size	Domain Category	No. Est.	Design Effect Mean[1]	Design Effect Std. dev.[2]	Proportion variation Mean[1]	Proportion variation Std. dev.[2]	True Roh
-	-	Total	10	0.061	0.028	0.135	0.058	0.077
Segregated	-	Urban	10	0.059	0.033	0.123	0.064	0.033
Segregated	-	Rural	10	0.059	0.056	0.160	0.102	0.234
Cross-class	1	Hisp. & Other	8	2.942	2.604	0.522	0.502	0.076
Cross-class	2	Black	10	0.648	0.849	0.257	0.186	0.216
Cross-class	4	White	10	0.089	0.042	0.130	0.066	0.091
Cross-class	1	60-64	9	2.342	2.549	0.551	0.307	0.181
Cross-class	2	65+	10	0.066	0.175	0.284	0.186	0.113
Cross-class	3	0-18	10	0.233	0.151	0.302	0.128	0.122
Cross-class	4	19-59	10	0.037	0.037	0.178	0.082	0.053
Cross-class	1	Dis. Male Vet.	8	0.037	0.659	0.591	0.600	0.613
Cross-class	2	Other Male Vet.	10	0.030	0.270	0.292	0.170	0.025
Cross-class	3	Other Male 20+	10	0.004	0.066	0.281	0.138	0.011
Cross-class	4	All Other	10	0.078	0.042	0.163	0.061	0.087

[1]Mean was obtained by averaging sample estimates for roh across replicates

[2]Standard deviation defined as square root of variance among sample estimates of roh across replicates.

1.00 in both tables. This situation occurred in size 1 domains and is indicative of the design effect exceeding the average sample cluster size. Fourth, in both Tables 3.2 and 3.3, we generally notice that there is less variability among estimates for the proportion variation method than the design effect method as exemplified by a smaller standard deviation among proportion variation estimates. Lastly, for both estimation methods in Tables 3.2 and 3.3, we see more variability among estimates for small domains than large ones as exhibited by larger standard deviations for small domains than large ones. Thus, we may say in summary that neither estimation method provided an accurate point estimate of the true value; however, in terms of relative performance, the proportion variation method did appear to come closer to the true value more often than the design effect method. The design effect method appeared to breakdown especially for small domains as evidenced by estimates of roh which were greater than 1.00 and thus, out-of-range. The proportion variation method also generated a more compact sampling distribution than the design effect method.

In Tables 3.4 and 3.5, sample relative deviations are examined. A relative deviation was defined as the difference in the sample estimate for roh and the study population true Roh and divided by the true Roh for the study population. Specifically, in Table 3.4, means and standard deviations for the relative deviations are listed by estimation method, analysis variable, and domain class. Also included in the table are true mean Rohs obtained by averaging true Rohs from the study population across domain categories for the particular class. Means and standard deviations were obtained across replicates and domain categories within each domain class. A positive mean signified an overestimate, while a negative mean denoted an underestimate. First, we see mean relative deviations for each estimation method are very large implying again that neither method accurately estimated true Roh. Ideally, we would like all of the means for one estimation method to be zero. The mean relative deviation for the first analysis variable in the total population class implied that the relative deviation for total roh estimated by the design effect method and averaged across the ten replicates was 180 percent above the true value. Secondly, we see that the magnitude of the relative deviations is influenced by the size of true Roh. When the true parameter was small, as in average dollars for all hospital admissions, where the true Roh was 0.001, the relative deviations were larger because we were dividing by a small true Roh. On the other hand, relative deviations were generally small for the response of "yes" to whether insured by a Health Maintenance Organization because of division by a relatively large true Roh (0.179) parameter. Lastly, we noticed variability of the relative deviations was usually less for the proportion variation method as evidenced by smaller standard deviations. This finding was consistent with the earlier observation that the proportion variation method generated a more compact distribution.

We generally conclude that neither the design effect nor the proportion variation method accurately estimated roh as evidenced by the large magnitude of mean relative deviations in Table 3.4. We did notice that mean relative deviations for the proportion variation method were generally smaller than for the design effect method implying the proportion

Table 3.4 Comparison of mean relative deviations by estimation method and analysis variable

Categorical analysis variable: Perceived health status "Good"

Domain			Design effect		Proportion variation		
Type	Class	No. est.[1]	Mean	Std. dev.[2]	Mean	Std. dev.[2]	True Mean Roh[3]
-	Total	10	1.800	2.725	1.569	2.449	0.021
Segregated	Urbanization	20	1.764	3.118	1.388	2.631	0.021
Cross-class	Size 1	25	-4.382	44.306	0.021	3.489	0.139
Cross-class	Size 2	30	4.100	7.391	1.422	6.158	0.052
Cross-class	Size 3	20	-12.777	26.303	-5.883	15.774	0.034
Cross-class	Size 4	30	1.762	3.424	1.494	3.021	0.027
Categorical analysis variable: Insured by Health Maintenance Organization "Yes"							
-	Total	10	-0.153	0.626	-0.214	0.501	0.179
Segregated	Urbanization	10	-0.461	0.729	-0.098	0.576	0.125
Cross-class	Size 1	6	0.106	6.177	-0.301	0.973	0.315
Cross-class	Size 2	16	-0.444	1.424	-0.221	2.041	0.225
Cross-class	Size 3	18	-0.626	1.360	-0.810	1.153	0.172
Cross-class	Size 4	29	-0.269	0.730	-0.273	0.573	0.191
Continuous analysis variable: Average dollars for all hospital admissions							
-	Total	10	-1.728	13.652	1.420	8.111	0.001
Segregated	Urbanization	20	-2.382	8.346	-4.635	9.127	0.001
Cross-class	Size 1	19	2.081	63.020	-2.207	3.741	0.493
Cross-class	Size 2	30	-15.432	83.892	17.413	66.229	0.001
Cross-class	Size 3	20	-0.255	23.286	-0.285	12.655	-0.005
Cross-class	Size 4	30	-73.906	243.281	103.986	230.351	0.003
Continuous analysis variable: Average time in minutes to get to usual source of care							
-	Total	10	-0.204	0.369	0.758	0.748	0.077
Segregated	Urbanization	20	0.012	1.058	1.208	2.074	0.134
Cross-class	Size 1	25	14.805	25.242	3.159	4.828	0.290
Cross-class	Size 2	30	0.594	6.535	4.131	6.161	0.118
Cross-class	Size 3	20	0.152	4.274	13.010	14.651	0.067
Cross-class	Size 4	30	-0.147	0.555	1.216	1.329	0.077

[1]Estimates were averaged across replicates and domain categories within each class.

[2]Standard deviation was obtained as square root of variance among relative deviations across replicates and domain categories within each class.

[3]Obtained by averaging true Rohs from study populations across domain categories for the particular domain class.

variation method closer approximated the true value. Another method must be developed which more accurately estimates true Roh.

In Table 3.5, the differences in the absolute value of the relative deviations are compared by domain category and analysis variable. For each domain category and analysis variable, the mean and standard deviation of the differences is listed in Table 3.5. These differences are the $d_{i.k}$ of section 2.3.4. We would interpret the mean difference of -0.052 for the Perceived Health Status "Good" analysis variable from the urban domain as signifying that the design effect method was 5.2 percent better than the proportion variation method based on average differences in absolute values of relative deviations for the two methods. The differences were compared to ensure that the distribution for a particular domain category -i across the ten replicates (k = 1,..., 10) was approximately normal. This verification was necessary to ensure that Student's t distribution held in the Confidence Interval procedure (see Appendix E).

Both Lilliefors test (Conover, 1980) and the Kolmogorov-Smirnov test (Conover, 1980) were used to check for normality. First, Lilliefors test enabled us to determine whether the distribution of differences across all domain categories and analysis variables was normally distributed. The null hypothesis was that the distribution of differences followed a normal distribution and the alternative was that the distribution was nonnormal. We compared the empirical distribution function for the distribution of differences from all of the 53 domain categories and analysis variables to a normal cumulative distribution function with mean equal to the grand mean of all differences in Table 3.5 (4.3) and

Table 3.5 Comparison of mean differences in absolute values of relative deviations by domain category and analysis variable

Domain			Categorical and analysis variable						Continuous analysis variables					
			Perc. health stat. "Good"			"Yes" to Insured by HMO			Ave. dollars for hosp. adm.			Ave. time to usu. source of care		
Type	Size	Class	#Est.[2]	Mean	Std. dev.[3]	#Est.[2]	Mean	Std. dev.[3]	#Est.[2]	Mean	Std. dev.[3]	#Est.[2]	Mean	Std. dev.[3]
	-	Total	10	0.179	0.449	10	0.072	0.281	10	4.126	7.274	10	-0.521	0.874
Seg.		Urban	10	-0.052	0.274	10	0.073	0.291	10	0.846	1.717	10	-1.632	1.674
Seg.		Rural	10	0.731	1.309	-[4]	-[4]	-[4]	10	-1.185	7.208	10	0.297	0.201
Cross	1	Hisp. & Other	8	17.538	22.398	5	7.329	12.970	5	113.366[1]	85.561	8	31.429	29.524
Cross	2	Black	10	3.879	4.212	2	-0.011	0.106	10	6.257	114.775	10	2.513	2.760
Cross	4	White	10	0.288	0.856	9	0.031	0.424	10	-0.574	2.023	10	-0.248	0.783
Cross	1	60-64	9	33.632	62.188	-[4]	-[4]	-[4]	9	2.720	3.143	9	13.245	8.915
Cross	2	65+	10	0.902	4.103	5	0.175	1.446	10	10.649	20.160	10	-0.522	1.299
Cross	3	0-18	10	0.694	0.676	10	0.158	0.840	10	10.503	17.292	10	-0.351	0.798
Cross	4	19-59	10	0.405	1.890	10	0.067	0.214	10	-6.871[1]	308.015	10	-1.731	1.578
Cross	1	Dis. Male Vet.	4	-1.120	1.962	-[4]	-[4]	-[4]	2	0.266	1.750	5	0.633	1.568
Cross	2	Other Male Vet.	10	1.346	4.281	10	-0.348	1.617	10	-0.152	10.560	10	-2.518	7.873
Cross	3	Other Male 20+	10	13.529	14.145	8	-0.164	0.511	10	1.214	5.787	10	-19.726	13.290
Cross	4	All Other	10	0.115	0.338	10	0.060	0.366	10	2.259	2.674	10	-0.551	0.789

[1]Denotes sample which did not appear to come from hypothesized normal distribution based on Kolmogorov-Smirnov test.
[2]Estimates were obtained by averaging across replicates.
[3]Standard deviation defined as square root of variance between difference in absolute values of relative deviations across replicates.
[4]Sample observations were not present for this classification.

standard deviation equal to the square root of the variance for all differences (48.882). The critical value T = 0.65 had a probability value (p-value), i.e., the probability of rejecting a critical value at least as big as the one determined, of approximately 0.15. Thus, we did not reject the null hypothesis; we assumed the distribution of differences was normally distributed. Secondly, for the Kolmogorov-Smirnov test, we assumed that the distribution of differences across all domain categories and analysis variables was normally distributed with mean equal to the grand mean (4.3) and standard deviation equal to the square root of the variance (48.882) across all domain categories and analysis variables of Table 3.5. This was the hypothesized distribution for the test. Each set of ten differences from the 53 domain and analysis variable categories of Table 3.5 was compared with the hypothesized distribution. The null hypothesis was that both samples came from the same normal distribution and the alternative hypothesis was that they came from different distributions. In two out of 53 comparisons, the null hypothesis was rejected, i.e., the sample distribution of differences was found to be different from that hypothesized. These samples are noted in the Table 3.5 footnote. The p-value associated with one of the rejected samples, the Hispanic and Other category, was $p = 0.015$, and the p-value for the other rejected sample, age 19 to 59, was $p = 0.025$. In the remaining 51 samples, the differences appeared normally distributed. Based on the outcome of both this test and the Lillifors test, we concluded that the differences in absolute values of the relative deviations within each domain category and analysis variable were approximately normally distributed.

Now moving on to examine the outcome of our speculation that the proportion variation method more accurately estimates roh than the design effect method in segregated domains. The results of the confidence interval procedure are presented in Table 3.6 for the two categorical analysis variables and in Table 3.7 for the two continuous analysis variables. The type and size of domain are specified, as well as the mean difference in the absolute values of the relative deviations, the standard error of this difference, the confidence interval, and lastly, the preferred method and true rate of homogeneity. Significance is noted at the five percent level. We would interpret the mean of -0.052 for the first analysis variable of perceived health status "good" from the urban domain category in Table 3.6 as signifying that the design effect method was 5.2 percent better than the proportion variation method based on the average differences in absolute values of the relative deviations. The preferred method and significance level were determined from the confidence interval. If the confidence interval was mostly positive, then the proportion variation method was preferred, and if the confidence interval was entirely positive, then the proportion variation method was significantly preferred over the design effect method at the five percent level. Similarly, a mostly negative confidence interval signified that the design effect method was preferred, and an entirely negative interval implied the design effect method was significantly preferred over the proportion variation method at the five percent level. Focusing our attention on segregated domains in Table 3.6 and 3.7 it does not appear, as hypothesized, that the proportion variation method was preferred over the design effect method for

Table 3.6 Comparison of estimation methods for categorical analysis variables

Perceived Health Status "Good"

Domain Type	Size	Category	Mean[1]	Std. error[2]	Confidence Interval	Preferred method	True Roh
-	Total	-	0.179	0.142	(-0.1, 0.5)	PV	0.021
SEG	Urbani-	Urban	-0.052	0.087	(-0.2, 0.1)	DE	0.022
SEG	zation:	Rural	0.731	0.414	(-0.2, 1.7)	PV	0.019
CROSS	1 Race:	Hisp.&Other	17.538	7.919	(-1.2, 6.3)	PV	0.149
CROSS	2	Black	3.879	1.332	(0.9, 6.9)	PV*	0.117
CROSS	4	White	0.288	0.271	(-0.3, 0.9)	PV	0.036
CROSS	1 Age:	60-64	33.632	20.729	(-14.2, 81.4)	PV	0.077
CROSS	2	65+	0.902	1.298	(-2.0, 3.8)	PV	0.052
CROSS	3	0-18	0.694	0.214	(0.2, 1.2)	PV*	0.074
CROSS	4	19-59	0.405	0.598	(-0.9, 1.8)	PV	0.013
CROSS	1 Vet.	Dis. Male Vet.	-1.120	0.982	(-4.2, 2.0)	DE	0.190
CROSS	2 Status:	Other Male Vet	1.346	1.354	(-1.7, 4.4)	PV	-0.012
CROSS	3	Other Male 20+	13.529	4.473	(3.4, 23.6)	PV*	-0.006
CROSS	4	All Other	0.115	0.107	(-0.1, 0.4)	PV	0.031

Insured by Health Maintenance Organization "Yes"

Domain Type	Size	Category	Mean[1]	Std. error[2]	Confidence Interval	Preferred method	True Roh
-	Total	-	0.072	0.089	(-0.1, 0.4)	PV	0.179
SEG	Urbani-	Urban	0.073	0.092	(-0.1, 0.3)	PV	0.176
SEG	zation:	Rural	-[3]	-[3]	-[3]	-[3]	0.073
CROSS	1 Race:	Hisp.&Other	7.329	5.800	(-8.8, 23.4)	PV	0.551
CROSS	2	Black	-0.011	0.077	(-1.0, 1.0)	DE	0.243
CROSS	4	White	0.031	0.141	(-0.3, 0.4)	PV	0.212
CROSS	1 Age:	60-64	-[3]	-[3]	-[3]	-[3]	0.428
CROSS	2	65+	0.175	0.647	(-1.1, 2.0)	PV	0.293
CROSS	3	0-18	0.158	0.266	(-0.4, 0.8)	PV	0.209
CROSS	4	19-59	0.067	0.068	(-0.1, 0.2)	PV	0.175
CROSS	1 Vet.	Dis. Male Vet.	-[3]	-[3]	-[3]	-[3]	-0.035
CROSS	2 Status:	Other Male Vet	-0.348	0.511	(-1.5, 0.8)	DE	0.139
CROSS	3	Other Male 20+	-0.164	0.181	(-0.6, 0.3)	DE	0.135
CROSS	4	All Other	0.060	0.116	(-0.2, 0.3)	PV	0.185

*Denotes significance at the five percent level.
[1]Mean is average difference in absolute values of relative deviations across the ten replicates.
[2]Standard error is the square root of the variance for the mean difference in absolute values of relative deviations.
[3]Dash signifies no sample observations were present for this classification.

estimating roh in segregated domains. On the contrary, both methods appeared to do about equally well. Across all seven cases where a segregated domain comparison could be made, the proportion variation method was preferred in four cases and the design effect in three. The difference between the two methods was significant in two of these seven cases. In one case, the design effect method was significant; the proportion variation method was significant in the other case.

For cross-class domains, we conjectured that the design effect method would be superior. After examining Tables 3.6 and 3.7, that conjecture also appeared to be false. Restricting our attention to cross-class domains and categorical analysis variables of Table 3.6, we see the proportion variation method is uniformly preferred over the design effect method. The proportion variation method was preferred in 16 out of 20 comparisons. A significant difference in favor of the proportion variation method was observed for three of those 16 cases. The strong showing by the proportion variation method probably stems from earlier observations on Table 3.2 that both the design effect and proportion variation methods overestimate Roh for categorical analysis variables when Roh is small. However, the proportion variation method generally provides less of an overestimate. Similarly, when Roh is large, both methods underestimate, but the proportion variation method provides more of an underestimate.

Now examining cross-class domains among continuous analysis variables of Table 3.7, we see the design effect method is generally preferred in larger domains, whereas the proportion variation method is superior for smaller domains. The design effect method was the method

Table 3.7 Comparison of estimation methods for continuous analysis variables

Average dollars for all hospital admissions

Domain Type	Size	Category	Mean[1]	Std. error[2]	Confidence Interval	Preferred method	True Roh
-	Total	-	4.126	2.300	(-1.1, 9.3)	PV	0.00100
SEG	Urbani-	Urban	0.846	0.543	(-0.4, 2.1)	PV	0.00300
SEG	zation:	Rural	-1.185	2.280	(-6.3, 4.0)	DE	-0.00200
CROSS	1 Race:	Hisp.&Other	13.366	38.264	(7.1,219.6)	PV*	0.02100
CROSS	2	Black	6.257	36.295	(-75.8, 88.4)	PV	0.00100
CROSS	4	White	-0.574	0.640	(-2.0, 0.9)	DE	0.00600
CROSS	1 Age:	60-64	2.720	1.050	(0.3, 5.1)	PV*	0.56400
CROSS	2	65+	10.649	6.375	(-3.8, 25.1)	PV	0.00700
CROSS	3	0-18	10.503	5.468	(-1.9, 22.9)	PV	0.00200
CROSS	4	19-59	-6.871	97.403	(-227.2,213.5)	DE	0.00005
CROSS	1 Vet.	Dis. Male Vet.	0.266	1.196	(-14.9, 15.5)	PV	0.89500
CROSS	2 Status:	Other Male Vet	-0.150	3.340	(-7.7, 7.4)	DE	-0.00800
CROSS	3	Other Male 20+	1.214	1.830	(-2.9, 5.4)	PV	-0.01200
CROSS	4	All Other	2.259	0.846	(0.4, 4.2)	PV*	0.00400

Average time in minutes to get to usual source of care

Domain Type	Size	Category	Mean[1]	Std. error[2]	Confidence Interval	Preferred method	True Roh
-	Total	-	-0.521	0.276	(-1.1, 0.1)	DE	0.077
SEG	Urbani-	Urban	-1.632	0.529	(-2.8, -0.4)	DE*	0.033
SEG	zation:	Rural	0.297	0.063	(0.2, 0.4)	PV*	0.234
CROSS	1 Race:	Hisp.&Other	31.429	10.438	(6.7, 56.1)	PV*	0.076
CROSS	2	Black	2.513	0.873	(0.5, 4.5)	PV*	0.216
CROSS	4	White	-0.248	0.248	(-0.8, 0.3)	DE	0.091
CROSS	1 Age:	60-64	13.245	2.972	(6.4, 20.1)	PV*	0.181
CROSS	2	65+	-0.522	0.411	(-1.5, 0.4)	DE	0.113
CROSS	3	0-18	-0.351	0.252	(-0.9, 0.2)	DE	0.122
CROSS	4	19-59	-1.731	0.499	(-2.9, -0.6)	DE*	0.053
CROSS	1 Vet.	Dis. Male Vet.	0.633	0.700	(-1.3, 2.6)	PV	0.613
CROSS	2 Status:	Other Male Vet	-2.518	2.490	(-8.1, 3.1)	DE	0.025
CROSS	3	Other Male 20+	-19.726	4.203	(-29.2,-10.2)	DE*	0.011
CROSS	4	All Other	-0.551	0.250	(-1.1, 0.0)	DE	0.087

*Denotes significance at the five percent level.
[1]Mean is average difference in absolute values of relative deviations across the ten replicates.
[2]Standard error is the square root of the variance for the mean difference in absolute values of values of relative deviations across the ten replicates.

of choice in ten of 22 cases, and it was significantly better in two of those ten cases. The proportion variation method was preferred in 12 of 22 cases, and it was significantly better in six of the 12 cases. This outcome is an illustration of the fact that the design effect method occasionally falls apart for estimating roh in small cross-class domains among continuous analysis variables. Out-of-range rohs in excess of 1.00 were sometimes generated by the design effect method. In fact, in Table 3.3, we see that mean sample rohs by the design effect method for domain size category 1 often exceeded 1.00.

It is interesting that the proportion variation method was preferred to estimate roh for the total analysis variable in three of the four cases. We expected the proportion variation method to perform better when we had equal-sized clusters as was the case with the total population and the segregated domains of urban and rural. We observed that the design effect and proportion variation methods performed about equally well for segregated domains. Apparently this also happened for the total population.

We conclude from these data that for segregated domains, either the design effect or the proportion variation estimation method appear to perform about equally well. For cross-class domains and a categorical analysis variable, the proportion variation method appears to be superior; whereas for cross-class domains and a continuous analysis variable, the design effect method appears to be better for large domains and the proportion variation method preferred for smaller domains. The design effect method particularly breaks down for small cross-class domains where out-of-range rates of homogeneity in excess of

1.00 were sometimes generated. We must also keep in mind our earlier findings that neither method provides an accurate point estimate of roh. We also need to recall that this study is an empirical study. The population of inference is the study population of the dissertation. To generalize the results, they must be verified by applying the same methodology to different populations using alternate sampling designs.

3.2 Secondary Objective Findings

The reader will recall that the secondary objective was divided into two parts. The first part consisted of determining the magnitude of the average domain ratio (ADR) used to impute roh for a domain category from roh for the population total. The ADR can be pictured as the ratio of an average roh across all categories in the domain to roh when domains are ignored (i.e., the total population). So we can think of the ADR as a ratio of average roh to total roh. Kish, Groves, and Krotki (1976) using World Fertility Survey data found the ADR to be about 1.2; Kalsbeek (1983) suggested an ADR of 2.5 using POPLAB data. The second part of the secondary objective concerns modeling the ratio of roh for a domain category to roh for the population total separately for each domain category via the Verma et al model (1980). Thus, we can think of this portion of the secondary objective as decomposing the ADR into its component parts with respect to the number of domain categories. The results from the first and second parts are summarized in the two subsequent sections. The global purpose of this secondary objective is to clarify the magnitude of the constant λ_y= used in the imputation of Figure 1.1 when we desire to know the total sample size required for a particular domain category.

Table 3.8 Comparison of true average domain ratios (ADR) for study population

Analysis type	Analysis description	Domain type	Domain description	No. of domain categories	True population ADR	True ADR omitting small domains
Categorical	Perceived health status "Good"	Segregated	Urbanization	2	0.976	0.976
		Cross-class	Race	3	4.794	1.714
		Cross-class	Age	4	2.571	0.619
		Cross-class	Vet. status	4	2.417	1.476
			Total	13	2.791	1.152
Categorical	Insured by HMO "Yes"	Segregated	Urbanization	2	0.696	0.696
		Cross-class	Race	3	1.873	1.184
		Cross-class	Age	4	1.543	0.978
		Cross-class	Vet. status	4	0.592	1.034
			Total	13	1.196	0.917
Continuous	Average dollars for all hosp. admissions	Segregated	Urbanization	2	0.500	0.500
		Cross-class	Race	3	9.333	6.000
		Cross-class	Age	4	143.260	0.050
		Cross-class	Vet. status	4	219.750	4.000
			Total	13	113.927	2.210
Continuous	Average time to get to usual source of care	Segregated	Urbanization	2	1.734	1.734
		Cross-class	Race	3	1.658	1.182
		Cross-class	Age	4	1.523	0.688
		Cross-class	Vet. status	4	2.390	1.130
			Total	13	1.853	1.294

3.2.1 Determination of Average Domain Ratio

Table 3.8 lists true study population average domain ratios for each analysis and domain variable. For a particular analysis and domain variable, the study population ADRs were obtained from the true population rates of homogeneity specified in Tables 3.6 and 3.7 by averaging true Roh for the analysis variable across all domain variable categories and dividing by true Roh for the population total. We notice first in Table 3.8, the great variability in magnitude of the ADR. For segregated domains, the ADRs were less variable than for cross-class domains. ADRs for segregated domains ranged from 0.500 to 1.734, whereas for cross-class domains they varied from 0.592 to 219.75.

The majority of the ADRs were within the average ranges of 1.2, previously reported by Kish, Groves, and Krotki (1976), and 2.5 noted earlier by Kalsbeek (1983). By examining the total ADR for each analysis variable, we saw that an unusually large ADR was generated from division by a very small population total Roh (0.001) for the continuous analysis variable of average dollars for all hospital admissions. This analysis variable represented the very small rate of homogeneity category in Table 2.5. Total ADR for a particular analysis variable was calculated by forming the ratio of average Roh across all 13 domain variable categories to total population Roh.

In the first chapter, we speculated that the larger ADR of 2.5 observed by Kalsbeek (1983) over the 1.2 of Kish et al. (1976) might have been caused by the omission of domains of size 6 or smaller from the Kish et al (1976) data and the inclusion of these small domains in Kalsbeek's (1983) data. We can test this conjecture by computing true

ADRs for the study population omitting size categories one through three (i.e., domains of size 6 or smaller). These true ADRs appear in the rightmost column of Table 3.8. By comparing this column with the true ADR column, we see, as speculated, that smaller ADRs were obtained by omitting small domains. This would explain why Kish et al (1976) observed smaller ADRs than Kalsbeek or than we did in this dissertation. We also notice much steadier ADRs when small domains are omitted as did Kish et al (1976).

We conclude that the ADRs for our study population are generally not steady. Extremely large ADRs can occur when an analysis variable rate of homogeneity for the population total is small. Thus, for our study population, we would not feel particularly comfortable using the Kish, Groves, and Krotki imputation procedure because the true lambda or true ADR exhibited great variability. Kalsbeek (1983) also observed variability among estimates of the ADR. For Bolivia, estimates ranged from -0.75 to 3.72 with an average of 1.47; for Indonesia, the range was from 2.19 to 7.14 and the average value was 4.65; values ranged from 0.60 to 18.17 for Somalia with an average of 4.79. The less extreme values for ADR observed by Kish, Groves, and Krotki (1976) may be primarily explained by the omission of small domains. It may also be partially explained by noting that World Fertility Survey (WFS) analysis variables pertaining to fertility experience, contraceptive practice, fertility preferences, and attitudes in developing countries generally exhibited larger rates of homogeneity than the 0.001 selected for the very small roh category.

Sample ADRs were determined for each of the replicates using rohs computed by the design effect and proportion variation estimation methods. Relative deviations were then computed for each estimation method as the sample ADR minus the study population true value and divided by the true value. The average difference in absolute values of the relative deviations were compared via the confidence interval procedure. The results are summarized in Table 3.9. For the first analysis variable or response of "good" to perceived health status in the urbanization domain we observed that the proportion variation method was 45.2 percent better than the design effect method for estimating ADR based on the average difference in absolute values of the relative deviations. Examining segregated domain types, we see that the design effect and proportion variation estimation methods performed about equally well to estimate the ADR. The design effect method was preferred in two of the four cases, and the proportion variation method in the other two. The design effect method was significantly preferred to the proportion variation method in one case. For cross-class domains, the proportion variation method generally appeared to be superior to the design effect method. The proportion variation method was preferred in ten of the 12 cases or 83 percent. It was significantly preferred in one of the ten cases. The proportion variation method probably made such a strong showing because many small cross-class domains were studied in this research and the proportion variation method was the method of choice for small cross-class domains. Again, it must be stressed that neither the design effect nor the proportion variation method provided an accurate point estimate for roh as noted in the findings from the

Table 3.9 Comparison of average domain ratios by estimation method and analysis variable

Categorical analysis variable: Perceived Health Status "Good"

Domain variable Type	Category	Mean[1]	Std. Error[2]	Confidence Int.	Preferred method	True ADR
Segregated	Urbanization	0.452	0.791	(-1.3, 2.2)	PV	0.976
Cross-Class	Race	46.411	44.149	(-53.5, 146.3)	PV	4.794
Cross-Class	Age	116.079	111.697	(-136.6, 368.7)	PV	2.571
Cross-Class	Veteran Status	1.444	2.738	(-4.8, 7.6)	PV	2.417

Categorical analysis variable: Insured by Health Maintenance Organization "Yes"

Domain variable Type	Category	Mean[1]	Std. Error[2]	Confidence Int.	Preferred method	True ADR
Segregated	Urbanization	-0.147	0.005	(-0.2, -0.1)	DE*	0.696
Cross-Class	Race	7.190	5.308	(-4.9, 19.2)	PV	1.873
Cross-Class	Age	2.510	2.656	(-3.5, 8.5)	PV	1.543
Cross-Class	Veteran Status	-0.135	0.627	(-1.6, 1.3)	DE	0.592

Continuous analysis variable: Average dollars for all hospital admissions

Domain variable Type	Category	Mean[1]	Std. Error[2]	Confidence Int.	Preferred method	True ADR
Segregated	Urbanization	-1.300	2.076	(-6.0, 3.4)	DE	0.500
Cross-Class	Race	12.410	6.625	(-2.6, 27.4)	PV	9.333
Cross-Class	Age	0.546	0.739	(-1.1, 2.2)	PV	43.260
Cross-Class	Veteran Status	-0.105	0.138	(-0.4, 0.2)	DE	19.750

Continuous analysis variable: Average time in minutes to get to usual source of care

Domain variable Type	Category	Mean[1]	Std. Error[2]	Confidence Int.	Preferred method	True ADR
Segregated	Urbanization	0.467	0.384	(-0.4, 1.3)	PV	1.734
Cross-Class	Race	33.185	27.434	(-28.9, 95.2)	PV	1.658
Cross-Class	Age	6.164	1.636	(2.5, 9.9)	PV*	1.523
Cross-Class	Veteran Status	2.519	1.971	(-1.9, 7.0)	PV	2.390

*Denotes significance at the five percent level.
[1]Mean is average difference in absolute values of relative deviations across the ten replicates.
[2]Standard error is the square root of the variance for the mean difference in absolute values of relative deviations across the ten replicates.

primary objective of this research. Thus, we would tend to lack confidence in an estimate of the ADR produced by one of these procedures.

We conclude from these results that for this study population, it does not appear feasible to obtain a simple estimate of the ADR for use in the Kish, Groves, and Krotki (1976) imputation for the following reasons: (1) the study population ADR varied considerably; (2) neither the design effect nor proportion variation estimation methods provided good point estimates for roh.

3.2.2 Model Proportion Of Domain Contained In A Particular Category

Verification of the Verma et al (1980) model is the first question to be addressed in this section. The Verma et al (1980) model is summarized in Table 3.10. The model did not work when applied to the study population.

1. The model fit (R^2) was extremely poor ranging from 0.035 to 0.592.

2. The model did not accommodate negative rates of homogeneity which were occasionally generated for small domains. For one domain, the domain degree of segregation, α, was out-of-range because negative rohs were not accepted.

The degree of segregation parameter, α, was expected to range from 0, for segregated domains, to 1, for cross-class domains. The magnitude of values, for α were, as expected, generally small for segregated domains and large for cross-class domains.

Table 3.10 Estimates of study population degree of segregation and model fit for Verma et al model

Domain variable		No. Categories	$\hat{\alpha}$	$\hat{R}^2$	Actual Obs.	Expected Obs.
SEG	Urbanization	2	0.372	0.045	7	8
CROSS	Size 1	3	0.320	0.035	11	12
CROSS	Size 2	3	0.822	0.437	10	12
CROSS	Size 3	2	1.159	0.592	6	8
CROSS	Size 4	3	0.551	0.292	12	12

The estimates for α, ranged from 0.320 to 1.159; the latter value exceeded the range for α because negative rohs were not accepted by the model. It is interesting that Verma et al (1980) also obtained two out-of-range values for α; both were for analysis variables pertaining to contraceptive knowledge in cross-class domains; a value of 1.14 was obtained for Thailand and 1.07 for Indonesia. For a given group of similar analysis variables, the estimate of α as determined by Verma et al (1980) was generally higher for cross-class domains than segregated domains, as expected. However, for cross-class domains, α varied widely for similar analysis variables between countries. For segregated domains, α was considerably more stable.

Examining the coefficient of determination (R^2) in Table 3.10, we see a better fit was obtained for larger cross-class domains than segregated domains. This result is consistent with Verma et al (1980) observations. However, our R^2 values were considerably lower, particularly for segregated domains. We must remember that in this dissertation we were using from 6 to 12 observations to estimate each R^2, whereas Verma et al were using from 24 to 125. Their estimates were based on much larger sample sizes than ours.

The number of observations expected and actually used to fit each α are also specified in Table 3.10. The expected number of observations was obtained as the product of the number of domain categories and analysis variables in each domain class. The number of actual observations was smaller than expected because the model did not accommodate negative rates of homogeneity. Negative rohs for a domain made y_d undefined in equation 2.3, because the negative roh generated a domain design effect less than one, which in turn yielded a negative D_d; y_d or the natural logarithm of a negative number, D_d, was thus, undefined. Since CLUSTERS computer software (1978) censors calculation of roh for small domains, negative rohs were generally not a problem to Verma et al (1980) when their model was developed. Examining the true Rates of Homogeneity from Tables 3.6 and 3.7, we find these negative Rohs occurring among both continuous and categorical analysis variables in the smaller cross-class domains particularly the three smallest veteran status categories. The negative Rohs signify little variability among population means per cluster. The mean per cluster was examined for those analysis and domain variables yielding a negative Roh and found to be extremely stable from cluster to cluster.

For the limited population of this dissertation, the Verma et al (1980) model did not work as exemplified by the bad fit of Table 3.10. Thus, for this study population, modeling the ratio of roh for domain to roh for population total separately for each category in the domain did not appear to be a viable alternative.

Since the Verma et al (1980) model did not hold for the study population, the earlier-specified comparisons to determine whether estimation method affected the model were not made. Some reasons as to why the model worked for Verma et al and did not work for the study population of this dissertation are included in the discussion of section 3.4.

From these secondary objective findings, we conclude that for this study population no simple estimate of lambda appeared to be available for use in the Kish, Groves, and Krotki (1976) imputation procedure. We found in the first part of the secondary objective that the study population ADR varied considerably for different domain and analysis variables. Thus, we would feel uneasy using one global estimate for lambda across several different analysis variables. In the second part of the secondary objective, we observed that the Verma et al model did not hold for our study population implying we could not form separate estimates of lambda for each domain category in the domain variable. Since this study was an empirical one, the results must be verified by applying the same methodology to other study populations.

3.3 Findings Of Interest To Veterans

The Veterans Administration was interested in a specific subpopulation on the NMCES file called veteran users of medical services, and defined as (1) veterans with (2) a physician hospital outpatient visit or hospital admission in 1977. This subpopulation, consisting of 1,110 veteran users, was relatively small. Furthermore, the size of the subpopulation would be reduced to 185 (1/6 X 1,110) in the study population. When these 185 veteran users were further subdivided by urbanization and region, the sample size would be too small for any meaningful analysis. Therefore, it was decided to use the total weighted NMCES file for veteran-related analyses. Two separate analyses were performed and are detailed in section 2.5. Briefly, the first analysis consisted of calculating rates of homogeneity using the appropriate formula determined from the dissertation, interpreting these rohs, and making a recommendation as to the generalizability of findings based on the magnitude of roh. For the second analysis, differences between veteran users categories were compared for several analysis variables. These two analyses are discussed in the subsequent sections.

3.3.1 Calculation of Roh For Veteran-Related Domains

Estimated rates of homogeneity for urban and rural veteran users of medical services are reported in Table 3.11. Similar estimates are provided in Table 3.12 for veteran users of medical services by region. Since we determined in the dissertation that neither the design effect nor the proportion variation estimation methods provided accurate point estimates for roh, we decided to make no recommendations as to the generalizability of findings and to calculate roh by each method to better estimate the possible order of magnitude of true Roh.

Estimates of roh from the design effect and proportion variation estimation methods in Tables 3.11 and 3.12 were calculated using the total weighted NMCES file. Since urban and rural domains in Table 3.11 and regional domains in Table 3.12 were segregated, the number of non-zero sample clusters for each domain was used to determine the average sample cluster size in the design effect estimation method. Average sample cluster sizes ($\overline{b}$) are reported in the table footnotes. All domain variables in Tables 3.11 and 3.12 are segregated, and all analysis variables are continuous.

In the following discussion, we must keep in mind that neither the design effect nor the proportion variation estimation method provided an accurate point estimate for roh, as determined in the empirical study. However, results from this epidemiologic application do replicate the empirical study results for continuous analysis variables in segregated domains; thus, more credence is lent to the empirical study results, and more confidence is provided for the ranges of rohs associated with these Veteran Administration results.

Examining rates of homogeneity for urban and rural veteran users of medical services in Table 3.11, we observe small differences in estimated rohs by each method. This result is consistent with the general empirical study conclusion that for segregated domains, the design effect and proportion variation estimation methods performed about equally well. Across both urban and rural categories, design effect rohs ranged from -0.0275 to 0.0801 with a mean of 0.0252; and the proportion variation rohs from -0.0210 to 0.2483 with an average roh of 0.0441. We noticed that the mean for the proportion variation method was slightly larger than the design effect method. This result agreed with the results from Table 3.1 where for the two continuous analysis variables in segregated domains, the mean roh for the proportion variation method was slightly larger than the design effect method. We also observed that estimates obtained for roh in Table 3.11 by the

Table 3.11 Estimated rates of homogeneity by design effect and proportion variation methods for urban and rural veteran users of medical services from entire NMCES sample

Analysis variable	Urban veteran users of medical services[1] Design effect roh	Urban veteran users of medical services[1] Proportion variation roh	Rural veteran users of medical services[2] Design effect roh	Rural veteran users of medical services[2] Proportion variation roh
Average number of visits	-0.0275	-0.0110	-0.0062	0.0244
Average cost of visits	0.0718	0.1204	0.0553	0.0266
Average time in minutes to get to usual source of care	0.0403	0.2483	0.0801	0.0242
Average number of prescribed medicines	0.0701	0.0943	0.0047	0.0317
Average cost of prescribed medicines	0.0301	0.0095	-0.0157	0.0281
Average number of dental visits	0.0384	0.0133	-0.0136	0.0024
Average cost of dental visits	-0.0096	-0.0210	0.0339	0.0262

[1]Urban veteran user estimates were based on 711 responses per analysis variable and an average sample cluster size ($\bar{b}$) of 7.4 for all analysis variables except average time where the total responses were 550 and $\bar{b}$ was 5.7.

[2]For rural veterans users, estimates were based on 399 responses per analysisvariable and a $\bar{b}$ of 9.7 for all variables except average time where the responses were 338 and $\bar{b}$ was 8.2.

proportion variation method were slightly more variable than the design effect method. This result also was in agreement with Table 3.1.

Two negative rates of homogeneity were generated by the proportion variation method in Table 3.11. Negative rohs by definition imply little variability between the estimated means for a cluster. Estimated means per cluster were examined for these negative rohs and found to be stable. Five negative rohs were generated by the design effect method. These negative rohs were caused by the design effect being less than one (1.00). Negative design effect rohs were also generated in the empirical study as evidenced in Table 3.1 by a negative design effect mean roh for one of the two continuous analysis variables in segregated domains.

We also noticed larger rates of homogeneity for the analysis variable of average time in minutes to get to usual source of care. This was the only analysis variable exhibiting item nonresponse. Nonrespondents were of two types: (1) those who did not state a usual source of care and (2) those who failed to respond with number of minutes to usual source of care. The total item nonresponse to this question was about 27 percent, and it was nearly equally divided between the two categories (Cohen and Kalsbeek, 1981). Thus, in addition to the expected variability in response to this question, the average sample cluster sizes were also smaller which was found in the empirical study to lead to larger rohs. These two factors are suggested as possible contributors to the larger rohs observed for the average time analysis variable.

Rates of homogeneity for veteran users of medical services by region in Table 3.12 exhibited these same characteristics previously noted for

Table 3.12 Estimated rates of homogeneity by design effect and proportion variation methods for veteran users of medical services by region from entire NMCES sample

Analysis variable	Northeast region[1] Design effect roh	Northeast region[1] Proportion variation roh	North Central region[1] Design effect roh	North Central region[1] Proportion variation roh	Southern region[1] Design effect roh	Southern region[1] Proportion variation roh	Western region[1] Design effect roh	Western region[1] Proportion variation roh
Average number of visits	-0.0450	-0.0195	-0.0383	0.1069	-0.0030	0.0957	-0.0069	0.0224
Average cost of visits	0.0317	0.0939	0.0405	0.1104	0.0934	0.0897	-0.0159	0.0193
Average time in minutes to get to usual source of care	0.1299	0.2149	0.0342	0.1438	0.1514	0.1098	0.0527	0.0206
Average number of prescribed medicines	-0.0026	-0.0984	0.0207	0.0139	-0.0045	-0.0123	0.0991	-0.0650
Average cost of prescribed medicines	0.0459	0.0457	0.0420	0.1152	-0.0288	0.1206	-0.0238	0.0753
Average number of dental visits	-0.0037	0.0755	0.0394	0.0607	0.0143	0.0135	0.0534	-0.0016
Average cost of dental visits	-0.0038	0.1016	0.0493	0.0490	0.0228	0.0973	-0.0070	0.0126

[1]Estimates were based on 224, 868, 388, and 230 responses per analysis variable for each region, respectively, and a $\bar{b}$ of 6.8, 7.4, 9.2, and 8.8 for all analysis variables except average time where the number of responses were 183, 222, 301, and 182 and $\bar{b}$ was 5.5, 6.2, 7.2, and 7.0, for each region, respectively.

urban and rural veteran users. Estimated rohs by each method were fairly similar agreeing with the empirical study finding of little difference in the two estimation methods for segregated domains. Across all regions, the design effect rohs ranged from -0.0450 to 0.1514 with a mean of 0.0263; while the proportion variation rohs ranged from -0.0984 to 0.2149 with an average of 0.0576. We observed the mean for the proportion variation method to be slightly larger than the design effect method, and the estimates for the proportion variation method to be slightly more variable than the design effect method. Both of these results also agreed with empirical study findings as exhibited in Table 3.1. We noted eleven negative rates of homogeneity for the design effect method and five for the proportion variation method. It appeared from the empirical study, as shown in Table 3.1, that more negative rohs were generated by the design effect method than the proportion variation method for continuous analysis variables in segregated domains. Lastly, larger rohs were observed for the analysis variable of average time in minutes to get to usual source of care. An explanation for this occurrence was offered in the preceding paragraph.

3.3.2 Compare Differences Between Veteran-Related Categories

Differences between veterans and nonveterans with respect to certain demographic, medical utilization, and medical cost variables are summarized in Table 3.13. The differences between the two categories as well as the standard errors of the differences and the associated Z values are provided. The standard errors of the differences were estimated using the computer software SESUDAAN (Shah, 1980) which included the covariances in the differences. Significant differences between veterans and nonveterans are noted with an asterisk in Table 3.13. Significance is noted at the five percent level.

Significant differences were observed between veterans and nonveterans with respect to age, income, years of schooling, and number and cost of dental visits. In each case, the figure was larger for veterans. That is, veterans were significantly older than nonveterans by about 4.5 years; the average age for veterans was 46.5 years and for nonveterans, 42.0 years. Nonveterans were defined as males who were at least 20 years of age and who had never participated in the military service. We also see from Table 3.13 that veterans earned a significantly higher average annual income than nonveterans; the figures were $15,288 and $11,298, respectively; average income directly adjusted for age by pooling the two groups was also significantly higher for veterans ($13,401) than nonveterans ($12,080); thus, we concluded veterans indeed had higher average incomes than nonveterans after adjusting for age. Veterans also had completed significantly more schooling than nonveterans, 12.1 years versus 11.4 years; average income was again directly adjusted for years of schools by pooling the two groups; adjusted average income for veterans ($14,591) was significantly higher than for nonveterans ($11,270) after accounting for educational differences; thus, we concluded veterans again had higher average incomes than nonveterans after adjusting for years of education. Finally, average income directly adjusted for both age and education was still significantly higher for veterans ($12,922) than nonveterans ($12,281). We observe in Table 3.13 that veterans visited dentists significantly more often than nonveterans; veterans average cost of dental visits were also significantly higher than for nonveterans; average number and cost of dental visits were directly adjusted for age by pooling the two groups; both figures were still significantly higher for veterans (1.2, $58) than nonveterans (1.0, $43) after adjustment. Thus, we

Table 3.13 Comparisons between veterans and nonveterans in the total NMCES sample*

Analysis variable	Veteran	Nonveteran	Difference	ste(Diff.)	Z
Average age	46.5	42.0	4.5	0.33	13.64*
Average income	$15,288	$11,298	$3,990	269.68	14.80*
Average years of education	12.1	11.4	0.7	0.08	8.75*
Average number of visits	3.6	3.5	0.1	0.12	0.83
Average cost of visits	$108	$95	$13	6.97	1.87
Average time in minutes to get to usual source of care	19.6	18.9	0.7	0.87	0.80
Average number of prescribed medicines	3.99	4.01	-0.02	0.18	-0.11
Average cost of prescribed medicine	$26.22	$26.26	$-0.04	0.04	-1.00
Average number of dental visits	1.3	1.0	0.3	0.05	6.00*
Average cost of dental visits	$60	$42	$18	3.90	4.62*

*Based on following sample sizes: Veterans (5,135) and Nonveterans (7,170) concluded that veterans had more and costlier dental visits than non-veterans after adjusting for age.

Several of the estimates were compared with results from the 1981 National Health Interview Survey (NHIS). Comparisons were favorable. The average annual number of physician visits reported by males was 4.0 closely agreeing with the 3.6 and 3.5 reported by veterans and nonveterans in the NMCES survey. The average number of dental visits per year for males from NHIS was 1.5 again paralleling the 1.3 and 1.0 visits for veterans and nonveterans from NMCES.

Veteran users of medical services were compared with other veterans in Table 3.14. We noted significant differences, at the five percent level, between all analysis variables except average cost of dental visits. In particular, we saw that veteran users of medical services were significantly older than other veterans and had lower incomes than other veterans. A significantly higher number and cost of physician visits were also seen among veteran users of medical services than other veterans. Lastly, a significantly larger number and higher cost of prescribed medicines were associated with veteran users of medical services. Veteran users of medical services were defined as veterans with a physician hospital outpatient visit or hospital admission in 1977. The group was created to be similar to potential Veteran Administration users. It is interesting that the findings from these comparisons were consistent with previously published results for applicants to Veteran Administration facilities (U.S. Senate, 1977). The report found that Veteran Administration applicants were older and poorer than veterans in general as we did; it also found that applicants had more chronic health

Table 3.14 Comparisons between veteran users of medical services and other veterans in the total NMCES sample

Analysis variable	Veteran Users of Medical Services	Other Veteran	Differences	ste(Diff.)	Z
Average age	49.9	45.6	4.3	0.53	8.11*
Average income	$14,082	$15,607	$-1,525	490.35	-3.11*
Average years of education	11.7	12.2	-0.5	0.13	-3.88*
Average number of visits	8.1	2.3	5.8	0.29	20.35*
Average cost of visits	$279	$63	$216	27.34	7.90*
Average time in minutes to get to usual source of care	23.7	18.3	5.4	1.64	3.29*
Average number of prescribed medicines	8.2	2.9	5.3	0.43	12.45*
Average cost of prescribed medicine	$52	$19	$33	2.67	12.38*
Average number of dental visits	1.5	1.2	0.3	0.10	3.03*
Average cost of dental visits	$58	$61	$-3	6.00	-0.50

*Based on following sample sizes: Veterans users of medical services (1,110) and other veterans (4,025) problems associated with more physician visits and larger numbers of prescribed medicines as we found.

Urban and rural veteran users of medical services were compared in Table 3.15. All comparisons between urban and rural veteran users of medical services were significantly different except age, average number of prescribed medicines, and average cost of prescribed medicines. We observed significant differences for average number and average cost of visits as well as average number and average cost of dental visits. In each case, the figure was larger for urban than rural veteran users of medical services. We also noted that urban veteran users of medical services exhibited a significantly lower average time to usual source of care than rural users. Thus, urban veteran users appeared to have better and more frequent access to medical and dental care than rural veteran users.

3.4 Discussion Of Results

This discussion is speculation. Very little research has been done on estimation of rates of homogeneity in multistage samples. It is hoped that this discussion will generate further research in the area of estimation, particularly a theoretical approach.

The findings of this study which are of greatest practical significance to research concerned the fact that neither the design effect nor the proportion variation estimation method provided accurate point estimates for roh. However, comparing the two methods, the proportion variation method appeared to come closer to the true value more often. The proportion variation estimate was also less variable than the design effect estimate as evidenced by the smaller standard deviations in

Table 3.15 Comparisons between urban and rural veteran users of medical services in the total NMCES sample.

Analysis variable	Urban Veteran Users	Rural Veteran Users	Differences	ste(Diff.)	Z
Average age	49.9	49.8	0.1	1.00	0.10
Average income	$15,042	$12,048	$ 2,994	907.27	3.30*
Average years of education	12.0	11.0	1.0	0.25	4.07*
Average number of visits	8.7	6.8	1.9	0.51	3.69*
Average cost of visits	$313	$206	$107	42.13	2.54*
Average time in minutes to get to usual source of care	20.7	29.5	-8.8	3.70	-2.38*
Average number of prescribed medicines	8.3	8.0	0.3	0.81	0.37
Average cost of prescribed medicine	$52.5	$51.0	$1.5	5.00	0.30
Average number of dental visits	1.7	1.0	0.7	0.16	4.36*
Average cost of dental visits	$68	$36	$32	8.53	3.75*

*Based on following sample sizes: Urban veteran users of medical services (711) and rural veteran users of medical services (399).

empirical sampling distributions of Tables 3.1 through 3.3. These findings suggest a more accurate estimator than the proportion variation method exists. In view of the importance and utility of roh as cited in Chapter 1, we would strongly encourage the search for such an estimator.

It appears that this new estimator should be patterned after the proportion variation method since it appeared stronger; however, the new estimator must also incorporate changes to capitalize on the strengths of the design effect method. For example, the design effect method worked best for continuous analysis variables in large cross-class domains; it may be possible to improve the new proportion variation estimator to better estimate roh for continuous analysis variables in large cross-class domains. It would also be helpful to find a simpler estimator than the proportion variation method. Other specific ideas for improving the proportion variation method are provided in the future research section 4.2.

A second important practical consideration from this research concerns when it is advisable to use the design effect method to estimate roh. We saw in Chapter 1 that the design effect method was generally used to estimate roh. We determined in the first part of Chapter 3 that the design effect method sometimes provided a poor estimate for roh. The method particularly appeared to breakdown for continuous analysis variables in small cross-class domains. We may generally regard roh estimated by the design effect method as a crude or rough estimate. Thus, the design effect method might be used to estimate roh for determining the total sample size for a particular survey. However, we would not want to use it to determine the optimum cluster size or optimum allocation. These more detailed calculations would require a more refined estimate of roh such as that obtained from a pretest or previous similar survey.

Another important research finding from this study was the fact that the average domain ratio (ADR) was not constant for the study population and the Verma, Scott, and O'Murcheartaigh (1980) model provided a poor fit for the study population. Thus, it does not appear advisable to use a single estimate for lambda in the Kish, Groves, and Krotki (1976) imputation procedure of Figure 1.1 to obtain domain roh from total roh. Average domain ratios (ADRs) for the study population were computed in Table 3.8. The ADR for a particular analysis variable was defined as the ratio of average roh across all domain categories to roh for the population total. If the population total estimate for roh was small, the ADR could become sizable because of division by a small number. This phenomenon was observed in Table 3.8 for average dollars for all hospital admissions. Kalsbeek (1983) also observed variability among estimates of ADR. For Bolivia, he observed values ranging from -0.75 to 3.72 with an average of 1.47; the range was from 2.19 to 7.14 with an average of 4.65 for Indonesia; and values ranged from 0.60 to 18.17 for Somalia with an average of 4.79. Thus, estimating lambda of the Kish, Groves, and Krotki (1976) imputation by means of the ADR does not appear feasible. We also observed in Table 3.8 that the Kish et al (1976) estimate of 1.2 for lambda was lower than the Kalsbeek (1983) estimate of 2.5 primarily because small domains (i.e., clusters containing six elements or less) were omitted from the Kish et al (1976) estimate and included in Kalsbeek's (1983). Omission of small domains also helped to explain why the Kish et al (1976) ADRs were more constant or stable than those observed by Kalsbeek (1983) or seen by us in this dissertation.

Verma, Scott and O'Murcheartaigh (1980) attempted to model the Kish, Groves, and Krotki (1976) lambda separately for each domain category. When we applied their methodology to the study population, we found the model supplied a very poor fit. The model coefficient or determination or R^2 in Table 3.9 ranged from 0.035 to 0.592. The following reasons are suggested as to why the Verma et al (1980) model worked on World Fertility Survey (WFS) data but failed to work on the study population of this dissertation. First, CLUSTERS computer software (1978) was used by WFS to estimate roh; estimates of roh were therefore censored for clusters of six elements or less. Thus, many large rohs and also many negative rohs, generated by small domains, were excluded from the WFS data. The WFS rohs were therefore less variable than those from the study population. Secondly, the number of observations on which each estimate of model fit was based was larger for WFS than for our study population. For WFS data, the number of observations per estimate ranged from 24 to 125, whereas ours ranged from 6 to 12. The WFS estimates of the degree of segregation, α, and the coefficient of determination, R^2, were based on larger sample sizes than those from the study population. Third, the coefficients of determination for WFS data, while good, were not spectacular. Fifty percent were greater than 0.65; 30 percent were greater than or equal to 0.50 and less than or equal to 0.65; and 20 percent were less than 0.50. Lastly, the Verma et al model assumed that (1) both roh for the domain and roh for the population total were small and (2) the ratio of roh for domain to roh for population total was 1.00. These assumptions may have been true for the WFS data, but they definitely were not true for the study population. In Tables 3.6 and 3.7, true Roh for the study population was greater than 0.1 in 23 of 56 cases or 41 percent which is not particularly small. Furthermore, the ratio of roh for domain to roh for total was greater than or equal to 2.00 for 20 out of 52 cases or 38 percent. Recall that these small domains were not a problem for WFS data since CLUSTERS (Verma, 1978) computer software censors the data for small domains. Since the study population ADR was not steady and the Verma et al (1980) model failed to work on the study population, it does not appear wise to use a single global estimate for lambda in the Kish, Groves, and Krotki (1976) imputation of Figure 1.1 to obtain the sample size for a particular domain category when roh for the population total is known.

CHAPTER 4

Summary Of Results And Suggestions For Future Research

4.1 Summary of Results

The primary purpose of this research was to quantify when each of two methods was appropriate to estimate the rate of homogeneity or roh. Roh is a measure of how much elements are alike within sampling clusters and is used in sample size determination. It was speculated that the proportion variation method would perform better for segregated domains, which contain only one category per cluster such as region or urban status; and the design effect method better for cross-class domains, those containing more than one category per cluster like age, race, or sex. To investigate the primary issue, two-stage cluster samples were selected from the National Medical Care Expenditure Survey (NMCES) household file which thus served as the population to which inference would be made; sample values for roh were calculated using each formula and compared with the population values from the entire file.

As a secondary issue, how the relationship between the domain and analysis variable may affect the appropriate formula for roh was also investigated. This secondary issue was explored by first, calculating sample values for the average domain ratio using each of the two estimation procedures and comparing with the true population value; and secondly, by modeling the domain degree of segregation or cross-classness for each estimation procedure and evaluating the model fit.

Results from the primary objective showed that neither the proportion variation nor design effect method provided accurate point estimates for roh; however, the proportion variation method, in general, more closely approximated the true population value. For segregated domains, the proportion variation method and the design effect method appeared to perform about equally well. For cross-class domains and a categorical analysis variable, the proportion variation method seemed to be superior; whereas for cross-class domains and a continuous analysis variable, the design effect method looked better for estimating roh in large domains, while the proportion variation method was superior for small domains.

It was found that the average domain ratio (ADR) for the study population was not constant. Also, for segregated domains there appeared to be no difference in the design effect or proportion variation estimation methods for estimating the average domain ratio. For cross-class domains, the proportion variation method was usually preferred, although not significantly so, to estimate the average domain ratio. The Verma et al (1980) model was found to fit very poorly when applied to the population. The model was, therefore, not investigated further to see whether it was affected by estimation method.

Since the study population ADR was not constant and the Verma et al (1980) model failed to work on the study population, it did not seem advisable to use a single estimate for lambda in the Kish, Groves, and Krotki (1976) imputation procedure to obtain the sample size for a particular domain category when roh for the population total was known. It must be emphasized that this study was an empirical one, and the population of inference was the population for the dissertation or the NMCES household file. To generalize the results, they must be verified by applying the same methodology to different populations using alternate sampling designs.

4.2 Suggestions For Further Research

These results have probably generated more questions about estimation of rates of homogeneity, than they have answered. Because so little was known about these estimators at present, this empirical study was basically exploratory in nature; in order to validate the results, the same estimators must be compared using different populations. This dissertation was limited by time, level of effort, and amount of computer funds. For these reasons, a study population was selected from the total NMCES household file; additionally, a subset of the analysis and domain variables on the NMCES household file were also identified for intensive study. Ideally, we would like to conduct a study similar to that undertaken by Kish, Groves, and Krotki (1976) for the World Fertility Survey (WFS) or Kalsbeek (1983) for the POPulation LABoratory (POPLAB) Survey. In this study, we would recommend that first, data from several different countries be analyzed; data from both developing and developed countries should be included as well as sampling designs employing both equal and unequal selection probabilities; comparing results from self-weighting and non-self-weighting sampling designs allows the effects of sampling weights on the estimates to be assessed.

Secondly, we would recommend that a large number and great variety of domain and analysis variables be studied. Kish, Groves, and Krotki (1976) strongly recommend this also. In fact, they suggest that in order to conserve funds, one study more variables omitting some of the categories rather than studying all categories for a few variables. Specifically for domain variables, we would recommend the inclusion of socioeconomic, race, age, and urbanization variables. Many domain categories within these domain variables may lead to large rates of homogeneity, i.e., race of black, age of 60 to 64, and rural categories.

Thirdly, we would recommend that results be compared (1) for clusters containing six or fewer elements; (2) for clusters comprising more than six elements; and (3) for moderate and large clusters. The comparison between six or fewer and greater than six elements would assess differences in computer software criteria for censoring results between SESUDAAN (Shaw, 1980) and CLUSTERS (Verma, 1978); additional comparisons would evaluate estimator results from small domains with those from moderately-sized and large domains.

Fourth, it is recommended that results be compared for the design effect estimation method by defining the average sample cluster size for a domain in several ways. In general, the average sample cluster size for a domain ($\bar{b}_d$) is defined as

$$\bar{b}_d = \frac{n_d}{a_d}$$

where n_d is the number of sample elements in the domain and a_d is the number of sample clusters in the domain. The problem arises when one tries to specify how to count a_d. Three methods are recommended: (1) a_d is the number of sample clusters containing at least one element from any category of the domain; this method was used in the dissertation; for cross-class domains; (2) a_d is the number of sample clusters containing at least one element from any category of the domain in a stratum. If in any stratum some clusters contain domain elements while others contain none, then all clusters in that stratum are counted in obtaining a_d. This definition should be very similar to (1) and is the definition used by CLUSTERS (Verma, 1978) computer software; (3) a_d is the number of nonzero sample clusters in a specific domain category; this definition computes a_d separately for each category -1 (a_{d1}) as the number of nonzero sample clusters for that specific category; this method was used in the dissertation for segregated domains. It is interesting that all three definitions appear to be consistent with Yates' definition (1981, p. 289) when definitions 1 and 2 are applied to cross-class domains and 3 to segregated domains. Yates definition states that the number of clusters must be taken to be the number of first-stage unites for which the number of elements is not zero.

Fifth, since neither the design effect nor proportion variation method provided accurate point estimates for roh, we would strongly suggest using a theoretical approach to develop a better point estimate. The estimator should be consistent using the definition in Hansen, Hurwitz, and Madow (1953). It should also exhibit a small variance and be less biased than the two methods considered in this dissertation. It appears that the new estimator should be patterned after the proportion variation estimator since it generally performed better in this study.

However, it may be possible to improve the estimator to better estimate roh for continuous analysis variables in large cross-class domains. This is the area where the current proportion variation estimator did not perform well.

Lastly, we would suggest that the same formulas for the design effect and proportion variation estimation methods as well as others be applied to the data to see if the results of this dissertation are duplicated.

BIBLIOGRAPHY

Alexander, C. H. and Kobilarcik, E. L. (1981), Selection of Stratifiers for the National Crime Survey Redesign, Proceedings of the Survey Research Methods Section of the ASA, 291-296.

Cochran, W. (1977), Sampling Techniques (Third Edition), New York: John Wiley and Sons, Inc.

Cohen, S. B. and Kalsbeek, W. D. (1981), NMCES Estimation and Sampling Variances in the Household Survey, National Center for Health Services Research, Hyattsville, Maryland.

Conover, W. G. (1980), Practical Nonparametric Statistics (Second Edition), New York: John Wiley and Sons, Inc.

Fellegi, I. P. (1980), Approximate Tests of Independence and Goodness of Fit Based on Stratified Multistage Samples, Journal of the American Statistical Association, 75, 261-268.

Fisher, R. A. (1970), Statistical Methods for Research Workers (Fourteenth Edition), New York: Hafner Press.

Folsom, R. E. (1984), Probability Sample U-Statistics: Theory and Applications for Complex Sample Designs, Unpublished Ph.D. dissertation, University of North Carolina at Chapel Hill.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953), Sample Survey Methods and Theory, Vol. I and II, New York: John Wiley and Sons, Inc.

Hansen, M. H., Madow, W. G., and Tepping, B. J. (1983), An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys, Journal of the American Statistical Association. 78, 776-807.

Harris, J. A. (1913), Biometrika, 9, 446-472.

Hogg, R. V. and Craig, A. T. (1978), Introduction to Mathematical Statistics (Fourth Edition), New York: Macmillan Publishing Co., Inc.

Holt, D., Smith, T.M.F., and Winter, P.D. (1980), Regression Analysis of Data from Complex Surveys, Journal of the Royal Statistical Society A, 143, 474-487.

Johnson, Norman L. and Leone, Fred C. (1977), Statistics and Experimental Design in Engineering and the Physical Sciences (Second Edition), Vol. I, New York: John Wiley and Sons, Inc.

Kalsbeek, W. D. (1984), Biostatistics 264 Class Notes, Unpublished notes for Advanced Sampling class, University of North Carolina at Chapel Hill.

Kalsbeek, W. D. (1983), Measures of Sampling Error in POPLAB Demographic Surveys, Laboratories for Population Statistics, University of North Carolina, Chapel Hill, N. C. 27514.

Kish, L. (1965), Survey Sampling, New York: John Wiley and Sons, Inc.

Kish, L. and Frankel, M. R. (1974), Inference From Complex Samples, Journal of the Royal Statistical Society B, 36, 1-37.

Kish, L., Groves, R. M., and Krotki, K. P. (1976), Sampling Errors for Fertility Surveys, Occasional Paper No. 17. London: International Statistical Institute, World Fertility Survey.

Kotz, S., and Johnson, N. L. (1984), Encyclopedia of Statistical Sciences, Vol. 4, New York: John Wiley and Sons, Inc.

Krewski, D., Platek, R., and Rao, J. N. K. (1981), Current Topics in Survey Sampling, New York: Academic Press.

Lorah, D., Singh, R., and Tegels, R. (1981), Efforts to Improve the Reliability of Minority Estimates in the Redesign of the Current Population Survey, Proceedings of the Survey Research Methods Section of the ASA, 297-302.

Marks, E. S., Seltzer W., and Krotki, K. J. (1974), Population Growth Estimation: A Handbook of Vital Statistics Measurements. New York: The Population Council.

Nathan, G. and Holt, D. (1980), The Effect of Survey Design on Regression Analysis, Journal of the Royal Statistical Society B, 42, 377-386.

Neter, J., Wasserman, W., and Kutner, M. H. (1983), Applied Linear Regression Models, Homewood, Illinois: Richard D. Irwin, Inc.

Rao, J. N. K. and Scott, A. J. (1981), The Analysis of Categorical Data From Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables, Journal of the American Statistical Association, 76, 221-230.

Shah, B. V. (1981), SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data, RTI/5250/00-015. Research Triangle Park, N. C.: Research Triangle Institute.

Sudman, S. (1976), Applied Sampling, New York: Academic Press.

Sukhatme, P. V. and Sukhatme, B. V. (1970). Sampling Theory of Surveys with Applications (Revised Second Edition), Ames, Iowa: Iowa State University Press.

United States Senate (1977), Study of Health Care for American Veterans, Report prepared by the National Academy of Sciences National Research Council for the Committeee on Veterans Affairs. Washington, D.C.: U. S. Government Printing Office.

Verma, V. and Pearce, M. C. (1978), User's Manual for CLUSTERS, London: International Statistical Institute, World Fertility Survey.

Verma, V., Scott, C., and O'Muircheartaigh, C. (1980), Sample Designs and Sampling Errors for the World Fertility Survey, Journal of the Royal Statistical Society A, 143, 431-473.

Waksberg, J., Hanson, R. H., and Jacobs, C. A. (1973), Reliability of Estimates With Alternative Cluster Sizes in the Health Interview Survey, U. S. Department of Health Education and Welfare, Rockville, Maryland.

Williams, R. L., Budescu, D. V., and Chromy, J. R. (1981), NAEP Year 11 Design Efficiency Study Final Report, RTI/1969/01-01 F, Research Triangle Park, N. C.: Research Triangle Institute.

Yates, F. (1981), Sampling Methods for Censuses and Surveys (Fourth Edition), New York: Macmillan Publishing Co., Inc.

APPENDIX A

Flowchart And Computer Software

To Calculate Rate Of Homogeneity

MACRO TO COMPUTE RATE OF HOMOGENEITY

FLOWCHART

OUT = PSUSTAT

Sum PSU-level variables

$X_\alpha, Y_\alpha, B_\alpha$

OUT = TOTSTAT

Sum total variables

X, Y, A

$$N = \sum_{\alpha}^{A} B_\alpha = X = \sum_{\alpha}^{A} X_\alpha$$

$$\bar{B} = N/A$$

DATA TWO and THREE

Obtain PSU-level variables dependent on R:

$$\bar{\bar{X}}_\alpha = X_\alpha / B_\alpha \; ; \; \bar{\bar{Y}}_\alpha = Y_\alpha / B_\alpha$$

$$\bar{\bar{Z}}_\alpha = \bar{\bar{Y}}_\alpha - R\,\bar{\bar{X}}_\alpha \; ; \; Z_\alpha = Y_\alpha - R\,X_\alpha$$

$$R = Y/X$$

$$\bar{\bar{X}} = X/N$$

$$\bar{X}_{.} = X/A$$

BETCOMP

$$\mathcal{B}^2 = \frac{\sum_{\alpha}^{A} Z_\alpha^2}{\bar{X}^2 (A - 1)}$$

WITHIN

$$\mathcal{W}^2 = \frac{\sum_{\alpha}^{A} \frac{B_\alpha}{B_\alpha - 1} \sum_{\beta}^{B_\alpha} (Z_{\alpha\beta} - \bar{\bar{Z}}_\alpha)^2}{A\,\bar{B}\,\bar{\bar{X}}^2}$$

where

$$Z_{\alpha\beta} = Y_{\alpha\beta} - RX_{\alpha\beta}$$

V

$$\mathcal{V}^2 = \frac{A - 1}{A}\mathcal{B}^2 + \frac{\bar{B} - 1}{\bar{B}}\mathcal{W}^2$$

ROH

$$\text{Roh}^{*} = \frac{\frac{A - 1}{A}\mathcal{B}^2 - \frac{\mathcal{V}^2}{\bar{B}}}{\frac{\bar{B} - 1}{\bar{B}}\;{}^2}$$

*This formula for Roh can be found in Hansen, Hurwitz, and Madow (1953, Volume II, page 161.)

```
%MACRO ROH(NE,XAB,YAB,A1,ONE);
PROC SUMMARY DATA=&ONE; BY &NE; VAR &XAB &YAB &A1;
OUTPUT OUT=PSUSTAT SUM=XA YA A2 ;
DATA ONEA; SET PSUSTAT;
IF XA=0 THEN DO; A2=0; END;
PROC PRINT DATA=ONEA; VAR &NE XA YA A2;
PROC SUMMARY DATA=ONEA;     VAR XA YA A2 ;
  OUTPUT OUT=TOTSTAT SUM=X Y A ;
PROC PRINT DATA=TOTSTAT; VAR X Y A ;
DATA TWO; SET TOTSTAT;
R=Y/X;
XDBAR=X/X;
BBAR1=X/A;
XBAR1=X/A;
PROC PRINT;VAR R  XDBAR BBAR1       XBAR1      ;
DATA THREE;
IF _N_=1 THEN SET TWO(KEEP=R A);
SET ONEA;
YDBARA=YA/XA;  XDBARA=XA/XA;
ZDBARA=YDBARA-R*XDBARA;
ZA=YA-R*XA; ZSQ=ZA**2;
PROC PRINT; VAR &NE YDBARA XDBARA ZDBARA ZA ZSQ;
PROC SUMMARY DATA=THREE; VAR ZSQ    ; OUTPUT OUT=BET SUM=SZSQ     ;
PROC PRINT DATA=BET; VAR SZSQ;
DATA FIVE;
MERGE BET(KEEP=SZSQ      ) TWO(KEEP=A XBAR1     );
BETCOMP1=SZSQ /((XBAR1**2)*(A-1));
PROC PRINT; VAR BETCOMP1     ;
DATA SIX;
MERGE THREE(KEEP=&NE R ZDBARA) &ONE(KEEP=&NE &XAB &YAB); BY &NE;
ZAB=&YAB-R*&XAB;
IF (ZAB=0 OR ZAB=.) THEN ZDEV=.;
ELSE ZDEV=(ZAB-ZDBARA)**2;
PROC SUMMARY DATA=SIX;BY &NE;VAR ZDEV;
OUTPUT OUT=WITH2 SUM=ZSQA;
PROC PRINT DATA=WITH2; VAR ZSQA;
DATA SEVEN;
MERGE WITH2    ONEA(KEEP=&NE XA);
IF (XA=0 OR XA=1) THEN WITH=.;
ELSE WITH=(XA/(XA-1))*ZSQA;
PROC PRINT DATA=SEVEN; VAR &NE ZSQA XA WITH;
PROC SUMMARY DATA=SEVEN; VAR WITH; OUTPUT OUT=WITH1 SUM=SWITH;
PROC PRINT DATA=WITH1; VAR SWITH;
DATA EIGHT;
MERGE WITH1 TWO;
WITHIN1=SWITH/(A*BBAR1*XDBAR**2);
PROC PRINT; VAR WITHIN1    ;
DATA NINE;
MERGE TWO FIVE EIGHT;
V1=(((A-1)*BETCOMP1)/A)+(((BBAR1-1)*WITHIN1)/BBAR1);
ROH1=(((A-1)*BETCOMP1/A)-V1/BBAR1)/(((BBAR1-1)*V1)/BBAR1);
PROC PRINT D ; VAR V1    ROH1    ;
%MEND ROH;
```

APPENDIX B

Show Algebraically That ρ^* For Unequal Clusters Reduces To ρ For Equal Clusters

Show ρ^* for unequal clusters reduces to ρ for equal clusters.

$$\rho^* = \frac{\frac{A-1}{A}\mathcal{B}^2 - \frac{\mathcal{V}^{*2}}{\bar{B}}}{\frac{\bar{B}-1}{\bar{B}}\mathcal{V}^{*2}}$$

where A = number of clusters in frame;
$\bar{B}$ = average number of elements per cluster in frame.

$$\mathcal{B}^2 = \frac{\Sigma_i^A \; Z_i^2}{\bar{X}^2 \; (A-1)} = \frac{\frac{A}{A-1} \; \Sigma_i^A \; Z_i^2}{A \; \bar{B}^2 \; \bar{\bar{X}}^2} = \text{Between Component of Variance}$$

$$Z_i = Y_i - RX_i; \; R = Y/X; \; \bar{X} = X/A; \; \bar{\bar{X}} = X/N;$$

$$\mathcal{W}^2 = \frac{\Sigma_i^A \left[\frac{B_i}{B_i - 1}\right] \Sigma_j^{B_i} \; (Z_{ij} - \bar{Z}_i)^2}{A \; \bar{B} \; \bar{\bar{X}}^2} = \text{Within Component Of Variance}$$

$$\bar{\bar{Z}}_i = \bar{\bar{Y}}_i - R\bar{\bar{X}}_i \; ;$$

$$\mathcal{V}^2 = \frac{A-1}{A}\mathcal{B}^2 + \frac{\bar{B}-1}{\bar{B}}\mathcal{W}^2 = \text{Total Variance.}$$

For equal clusters (B) and estimate mean $\bar{\bar{Y}}$ ($X_{ij} = 1$ is a count variable and $R = \bar{\bar{Y}}$),

$$\frac{A-1}{A}\mathcal{B}^2 = \frac{\Sigma_i^A \; Z_i^2}{A \; B^2 \; \bar{\bar{X}}^2} = \frac{\Sigma_i^A \; (Y_i - \bar{Y})^2}{A \; B^2} = \frac{\Sigma_i^A \; (\bar{Y}_i - \bar{Y})^2}{A} = \sigma_A^2 \, .$$

$$\mathcal{V}^{*2} = \frac{N - 1}{N}\mathcal{V}^{2} = \frac{\Sigma_i^A \; \Sigma_j^{B_i} \; Z_{ij}^2}{N \bar{\bar{X}}^{2}} = \frac{N - 1}{N} \; S^2$$

$$= \frac{\Sigma_i^A \; \Sigma_j^{B_i} \; (Y_{ij} - \bar{\bar{Y}})^2}{N} = \sigma^2 .$$

Therefore,

$$\rho^* = \frac{\frac{A - 1}{A}\mathcal{B}^{2} - \frac{\mathcal{V}^{*2}}{\bar{B}}}{\frac{\bar{B} - 1}{\bar{B}}\mathcal{V}^{*2}} = \frac{\sigma_A^2 - \frac{\sigma^2}{B}}{\frac{B - 1}{B}\sigma^2} = \rho .$$

APPENDIX C

Summary Of Checks To Insure Computer Software To Calculate Roh Is Accurate

Insure computer software to calculate population Roh is accurate

PURPOSE	DOMAIN DESCRIPTION	VALUE	ROH	COMMENTS
1. Verify large Roh.	1. Proportion Urban 2. Proportion Northeast	1. 66.1972% 2. 18.3099%	1. 1.0000 2. 1.0000	1. Computer results were verified by hand. 2. Verification of all computer output by hand.
2. Verify small Roh.	1. Proportion 6- to 18-years-old 2. Proportion less than 6-years-old	1. 23.926% 2. 8.29577%	1. 0.0078 2. 0.00097	1. Verification of computer results by hand. 2. Computer output recalculated by hand.
3. Verify negative Roh.	1. Proportion other veterans who responded "Good" to perceived health status	1. 36.1282%	1. -0.0121	1. Computer output independently recalculated by hand.
4. Verify Roh for domain within a domain.	1. Proportion less than 6-year-old in Northeast	1. 6.19355%	1. 0.0050	1. Computer results verified by hand.
5. Verify Roh for domain average instead of proportion.	1. Average income for those over 17-years-old	1. $8,025.3	1. 0.009	1. Recalculation of computer output by hand.
6. Verify previously published roh from Kish (1965).	1. Proportion newspaper subscribers who own home	1. 46.25%	1. 0.34	1. Computer output agreed with results published in Kish (1965).

Insure computer software to calculate population Roh is accurate (continued)

PURPOSE	DOMAIN DESCRIPTION	MSR=$B\mathcal{B}^2$	MSE=$\mathcal{W}^2$	COMMENTS
7. Independently verify between and within variance components for large Roh.	1. Proportion urban	1. 11.2676 = 50(0.2254)	1.0.0000	1. Between and within variance components for Roh agreed with Statistical Analysis System (SAS) General Linear Model (GLM) output.
8. Verify between and within variance components for small Roh using SAS GLM procedure.	1. Proportion less than 6-years-old	1. 0.102396 = 50(0.00204)	1. 0.07555	1. Mean Square Regression (MSR) and Mean Square Error (MSE) output from GLM duplicated between and within variance components from Roh computer macro.
9. Verify between and within variance components for negative Roh.	1. Proportion other veterans who responded "Good" to perceived health status.	1. 0.2096 ≐ (5.7) (0.0385)	1. 0.234	1. Between and within variance components were independently verified using SAS's GLM procedure.
10. Independently verify variance components for Roh for domain within a domain.	1. Proportion less than 6-years-old in Northeast	1. 0.0750 = 50(0.0015)	1. 0.0578	1. Verify MSR and MSE calculated from GLM agree with between and within variance components from Roh macro.

APPENDIX D

Derivation Of Another Estimator For Roh Designed Specifically For Proportion And Assuming Approximately Equal Sampling Clusters

Derivation Of Another Estimator For Roh

Designed Specifically For Proportion And Assuming

Approximately Equal Sampling Clusters

Suppose we are interested in the proportion of population elements possessing some attribute, P. Suppose further that the population is divided into A unequal clusters with B_i elements in each cluster. If we assume two-stage sampling and simple random sampling at each stage without stratification, then a measure of intracluster homogeneity is approximately

$$\text{Roh}(P) \doteq 1 - \frac{\Sigma_i^A \; B_i \; P_i \; (1 - P_i)}{A \; \overline{B} \; P \; (1 - P)} \; . \tag{D.1}$$

The approximation comes about because we assume the average cluster size $\overline{B} \doteq \overline{B}\text{-}1$ where

$$\overline{B} = \frac{\Sigma_i^A \; B_i}{A} \; ; \text{ and}$$

P_i = the proportion of all elements in cluster -i possessing the attribute;

$$= \overline{Y}_i \; = \; \frac{\Sigma_j^{B_i} \; Y_{ij}}{\overline{B}} \; .$$

The proof is as follows. In the notation of equations 1.6 through 1.8, we have

$$N = A\bar{B};$$

$$P = \bar{\bar{Y}};$$

and $$\sigma^2 = \sigma_A^2 + \sigma_B^2,$$

where N is the total number of elements in the population; $\bar{\bar{Y}}$ is the general mean defined in equation 1.6; and σ^2 is the total population variance which can be partitioned into the variance between clusters, σ_A^2, and the variance within clusters, σ_B^2. These variances are defined in equations 1.6, 1.7, and 1.8, respectively.

We know that (Hansen, Hurwitz, and Madow, Volume II, 1953, p. 165)

$$\mathrm{Roh}(\bar{\bar{Y}}) = \frac{\sigma_A^2 - \dfrac{\sigma_B^2}{\bar{B} - 1}}{\sigma^2} = \frac{\sigma^2 - \dfrac{\bar{B}}{\bar{B} - 1}\sigma_B^2}{\sigma^2} \doteq 1 - \frac{\sigma_B^2}{\sigma^2} \qquad (D.2)$$

assuming $\bar{B} \doteq \bar{B} - 1$ and $\bar{B} \doteq B_i$.

Let $Y_{ij} = 1$, if element -j from cluster -i has the attribute;

we know there are NP of these elements in the population.

$Y_{ij} =$, if element-j from cluster -i does not have the attribute;

we know there are N(1-P) of these elements in the population.

Therefore, $(Y_{ij} - P)^2 = (1 - P)^2$, if element -j from cluster -i has the attribute and there are NP of these elements;

$(Y_{ij} - P)^2 = P^2$, if element -j from cluster -i does not have the attribute and there are N(1 - P) of these elements.

$$\sigma^2 = \frac{\Sigma_i^A \ \Sigma_j^{B_i} \ (Y_{ij} - P)^2}{A \ \bar{B}} = \frac{N \ P(1 - P)^2 + N(1 - P)P^2}{N}$$

$$= \frac{N \ P(1 - P)}{N} = P(1 - P). \tag{D.3}$$

Similarly, $(Y_{ij} - P_i)^2 = (1 - P_i)^2$, if element -j from cluster -i has the attribute and there are $B_i(P_i)$ of these in cluster -i; $(Y_{ij} - P_i)^2 = P_i^{\ 2}$, if element -j from cluster -i does not have the attribute and there are $B_i(1 - P_i)$ of these in cluster -i. Thus,

$$\sigma_B^2 = \frac{\Sigma_i^A \ \Sigma_j^{B_i} \ (Y_{ij} - P_i)^2}{A \ \overline{B}} = \frac{\Sigma_i^A \ (B_i P_i \ (1 - P_i)^2 + B_i(1 - P_i)P_i^2)}{A \ \overline{B}}$$

$$= \frac{\Sigma_i^A \ (B_i \ P_i(1 - P_i))}{A \ \overline{B}} \ . \tag{D.4}$$

From equations D.2 through D.4 we have

$$\text{Roh}(P) \doteq 1 - \frac{\sigma_B^2}{\sigma^2} = 1 - \frac{\Sigma_i^A \ B_i \ P_i \ (1 - P_i)}{A \ \overline{B} \ P \ (1 - P)} \text{, as desired.}$$

APPENDIX E

Proof That Mean Of Difference In Absolute Values Of Relative Deviations Minus Its Expected Value And Divided By Its Standard Error Follows Student's t Distribution

We want to show that

$$t = \frac{\bar{d} - E(\bar{d})}{S_d n^{-1/2}}$$

follows Student's t distribution where

$$d_i = |a_i| - |b_i| \; ; \; \bar{d} = \frac{\Sigma_i^n d_i}{n} \; ; \text{ and}$$

$$S_d^2 = \frac{\Sigma_i^n (d_i - \bar{d})^2}{n - 1} .$$

We need to show each d_i has the same normal distribution and the d_i values are independent (Johnson and Leone, 1977). For if the d_i's are all normally distributed

$$\bar{d} = \frac{\Sigma_i^n d_i}{n}$$

is also normally distributed. Then

$$\frac{\bar{d} - E(\bar{d})}{\sigma_{\bar{d}}}$$

is distributed as a normal random variable with mean 0 and variance 1 where $E(\bar{d})$ signifies the expected value of $\bar{d}$ and $\sigma_{\bar{d}}$ denotes the standard error or square root of the variance of $\bar{d}$. Further, if $\frac{(n-1) S_d^2}{\sigma_d^2}$ is distributed as a chi-squared random variable with (n-1) degrees of freedom independently of $\bar{d}$, then

$$\frac{\dfrac{\bar{d} - E(\bar{d})}{\sigma_{\bar{d}}}}{\left(\dfrac{(n-1) S_d^2}{\dfrac{\sigma_d^2}{(n-1)}}\right)^{1/2}} = \frac{\bar{d} - E(\bar{d})}{\sigma_d \, n^{-1/2}} \cdot \frac{\sigma_d}{S_d} = \frac{\bar{d} - E(\bar{d})}{S_d \, n^{-1/2}} = t$$

follows Student's t distribution.

PROOF: First, the d_i's are independent since each comes from separate, independently selected random samples or replicates. Secondly, we must show that the d_i's are normally distributed. Since the replicates are independent and consist of 500 observations each, we may assume that a_i or the relative deviation from the design effect method are distributed normally with mean μ_a and variance σ_a^2. Similarly, we assume the relative deviations from the proportion variation method or b_i are normally distributed with mean μ_b and variance σ_b^2. Therefore, $d_i = |a_i| - |b_i|$ is also normally distributed since it is a linear function of normally distributed random variables. In fact, each d_i is distributed normally with mean $|\mu_a| - |\mu_b|$ and variance

$\sigma^2_{|a|} + \sigma^2_{|b|} - 2\rho_{|a||b|}\sigma_{|a|}\sigma_{|b|}$ because

$$E(d_i) = E(|a_i| - |b_i|) = E|a_i| - E|b_i| = |\mu_a| - |\mu_b|;$$

and, $\sigma^2(d_i) = \sigma^2(|a_i| - |b_i|) = \sigma^2(|a_i|) + \sigma^2(|b_i|) - 2\sigma(|a_i|, |b_i|)$

$$= \sigma^2_{|a|} + \sigma^2_{|b|} - 2\rho_{|a||b|}\sigma_{|a|}\sigma_{|b|}.$$

Therefore,

$$\bar{d} = \frac{\Sigma_i^n d_i}{n}$$

is distributed normally with mean $|\mu_a| - |\mu_b| = E(\bar{d})$ and variance

$$\frac{\sigma^2_{|a|} + \sigma^2_{|b|} - 2\rho_{|a||b|}\sigma_{|a|}\sigma_{|b|}}{n} = \sigma^2(\bar{d}).$$

So,

$$Z = \frac{\bar{d} - E(\bar{d})}{\sigma(\bar{d})}$$

is distributed normally with mean 0 and variance 1.

We must now show that $\dfrac{(n - 1)\, S_d^2}{\sigma_d^2}$ is distributed as a chi-squared variable and that S_d^2 is independent of $\bar{d}$.

Consider the joint distribution of $Y_1 = \overline{D}$, $Y_2 = D_2, \ldots, Y_n = D_n$. The corresponding Helmert Transformation (Hogg and Craig, 1978)

$$\begin{aligned} d_1 &= n\, y_1 - y_2 - \cdots - y_n \\ d_2 &= \quad\quad\; y_2 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ d_n &= \quad\quad\quad\quad\quad\; y_n \end{aligned}$$

had Jacobian n.

The sum $\Sigma_i^n \left(d_i - \{\mu_{|a|} - \mu_{|b|}\}\right)^2 = \Sigma_i^n \left(d_i - \bar{d} + \bar{d} - (\mu_{|a|} - \mu_{|b|})\right)^2$

$$= \Sigma_i^n (d_i - \bar{d})^2 + n\left(\bar{d} - (\mu_{|a|} - \mu_{|b|})\right)^2$$

because $\quad 2\left(\bar{d} - (\mu_{|a|} - \mu_{|b|})\right) \Sigma_i^n (d_i - \bar{d}) = 0$.

So the joint probability density function (pdf) of $D_1, D_2, \ldots, D_n$ which are normally distributed can be written as

$$\left[(2\pi)^{-1/2}\left[\sigma_{|a|}^2 + \sigma_{|b|}^2 - 2\rho_{|a||b|}\,\sigma_{|a|}\sigma_{|b|}\right]^{-1/2}\right]^n$$

$$\exp\left[\frac{-\Sigma_i^n (d_i - \bar{d})^2 - n\left(\bar{d} - (\mu_{|a|} - \mu_{|b|})\right)^2}{2\left(\sigma_{|a|}^2 + \sigma_{|b|}^2 - 2\rho_{|a||b|}\,\sigma_{|a|}\sigma_{|b|}\right)}\right]$$

where $-\infty < d_i < +\infty$, $i = 1, 2, \ldots n$. With $y_1 = \overline{d}$, we find via the transformation that the joint pdf of $Y_1, Y_2, \ldots, Y_n$ is

$$n\left[(2\pi)^{-1/2}\left(\sigma^2_{|a|}+\sigma^2_{|b|}-2\rho_{|a||b|}\,\sigma_{|a|}\sigma_{|b|}\right)^{-1/2}\right]^n$$

$$\exp\left[\frac{-(ny_1-\dots-y_n-y_1)^2-\Sigma_{i=2}^{n}\,(y_i-y_1)^2-n(y_1-(|\mu_a|-|\mu_b|))^2}{2(\sigma^2_{|a|}+\sigma^2_{|b|}-2\rho_{|a||b|}\,\sigma_{|a|}\sigma_{|b|})}\right]$$

$-\infty < y_i < +\infty$, $i = 1, 2, \dots, n$. The pdf of $Y_1 = \bar{d}$ is

$$\frac{n^{1/2}}{(2\pi)^{1/2}\left(\sigma^2_{|a|}+\sigma^2_{|b|}-2\rho_{|a||b|}\,\sigma_{|a|}\sigma_{|b|}\right)^{1/2}}$$

$$\exp\left[\frac{-n(y_1-(|\mu_a|-|\mu_b|))^2}{2(\sigma^2_{|a|}+\sigma^2_{|b|}-2\rho_{|a||b|}\,\sigma_{|a|}\sigma_{|b|})}\right].$$

The quotient of the joint pdf and the pdf of $Y_1 = \bar{d}$ is the conditional pdf of $Y_2, \dots, Y_n$ given $Y_1 = y_1$,

$$n^{1/2}\left[(2\pi)^{-1/2}\left(\sigma^2_{|a|}+\sigma^2_{|b|}-2\rho_{|a||b|}\,\sigma_{|a|}\sigma_{|b|}\right)^{-1/2}\right]^{n-1}$$

$$\exp\left[-\frac{q}{2(\sigma^2_{|a|}+\sigma^2_{|b|}-2\rho_{|a||b|}\,\sigma_{|a|}\sigma_{|b|})}\right]$$

where $q = (ny_1 - y_2 - \dots - y_n - y_1)^2 + \Sigma_{i=2}^{n} (y_i - y_1)^2$. Since this is a joint conditional pdf, it must be for all $[\sigma_{|a}^2 + \sigma_{|b}^2 - 2\rho_{|a|b}\,\sigma_{|a}\sigma_{|b}]^{1/2} > 0$ that

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} n^{1/2}[(2\pi)^{-1/2}(\sigma_{|a}^2 + \sigma_{|b}^2 - 2\rho_{|a|b}\,\sigma_{|a}\sigma_{|b})^{-1/2}]^{n-1}$$

$$\exp\left[-\frac{q}{2(\sigma_{|a}^2 + \sigma_{|b}^2 - 2\rho_{|a|b}\,\sigma_{|a}\sigma_{|b})}\right]dy_2\dots dy_n = 1.$$

Now consider

$$(n-1)\,S_d^2 = \Sigma_i^n (d_i - \bar{d})^2 = (n\,Y_1 - Y_2 - \dots - Y_n - Y_1)^2 + \Sigma_{i=2}^{n} (Y_i - Y_1)^2 = Q.$$

The conditional moment-generating function of $(n-1)\,S_d^2 / \sigma_d^2 = Q/\sigma_d^2$, given $Y_1 = y_1$ is

$$E(e^{tQ/\sigma_d^2} | y_1) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} n^{1/2}[(2\pi)^{-1/2}(\sigma_d)^{-1}]^{n-1}$$

$$\exp\left[-\frac{q}{2\sigma_d^2} + \frac{2tq}{2\sigma_d^2}\right]dy_2 \dots dy_n$$

$$= \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} n^{1/2}[(2\pi)^{-1/2}\,\sigma d^{-1}]^{n-1} \quad \exp\left[-\frac{(1-2t)q}{2\sigma_d^2}\right] dy_2\dots dy_n$$

$$= \left(\frac{1}{1 - 2t}\right)^{\frac{n-1}{2}} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} n^{1/2} \left[\frac{1 - 2t}{2\pi\sigma_d^2}\right]^{\frac{n-1}{2}} \exp\left[-\frac{(1 - 2t)q}{2\sigma_d^2}\right] dy_2 \ldots dy_n$$

where $0 < (1 - 2t)$, or $t < 1/2$. However, this latter integral is exactly the same as that of the conditional pdf of $Y_2, \ldots, Y_n$ given $Y_1 = y_1$ with σ_d^2 replaced by $\frac{\sigma_d^2}{1 - 2t} > 0$, and thus, must equal 1. Hence, the conditional moment-generating function of $(n-1)\, S_d^2/\sigma_d^2$, given $Y_1 = y_1$ or equivalently $\overline{D} = \overline{d}$, is

$$E(e^{t(n-1)S_d^2/\sigma_d^2} \mid \overline{d}) = (1 - 2t)^{-\left(\frac{n-1}{2}\right)} \quad \text{for} \quad t < 1/2.$$

That is, the conditional distribution of $(n-1)S_d^2/\sigma_d^2$, given $\overline{D} = \overline{d}$, is chi-squared with $(n-1)$ degrees of freedom. Morever, since it is clear that this conditional distribution does not depend upon $\overline{d}$, $\overline{D}$ and $(n-1)$ S_d^2/σ_d^2 must be stochastically independent or, equivalently, $\overline{D}$ and S_d^2 are stochastically independent.

QED

APPENDIX F

Proof That Proportion Variation Estimator Provides A Consistent Estimate For Roh According To The Definition Of Consistency In Hansen, Hurwitz, and Madow

Proof That Proportion Variation Estimator Provides A Consistent Estimate For Roh According To The Definition Of Consistency In Hansen, Hurwitz, and Madow

A consistent estimate is defined as follows (Hansen, Hurwitz, and Madow, 1953, Volume II, page 72). Let $u_1, u_2, u_3, \ldots$ be a sequence of random variables where u_n generally refers to the estimator when the sample size is n, then u_n is said to be a consistent estimator of U if, for every $\epsilon > 0$,

$$\lim_{n \to \infty} \Pr \left(\left| u_n - U \right| > \epsilon \right) = 0$$

or equivalently

$$\lim_{n \to \infty} \Pr \left(\left| u_n - U \right| \leq \epsilon \right) = 1.$$

The sequence of random variables $u_1, u_2, u_3, \ldots$ is also said to converge in probability to U. This definition also allows that the population size N increases so that $n < N$ always. We also assume that U remains unchanged as N increases. To establish consistency for u_n, it is sufficient to show

$$\lim_{n \to \infty} \left| E(u_n) - U \right| = 0$$

and

$$\lim_{n \to \infty} \mathrm{Var}\,(u_n) = 0.$$

Additionally, if u_n is an unbiased estimator of U and $\lim_{n \to \infty} \text{Var}(u_n) = 0$, then u_n is also a consistent estimator of U (Theorem 19, Hansen, Hurwitz, and Madow, 1953, Volume II, pages 74 to 75). If $\Phi(U_{1u}, U_{2n} \ldots, U_{kn})$ is some continuous function of u_{jn} $(j = 1,2,\ldots,k)$ each of which is a consistent estimator of U_j, then $\Phi(u_{1n}, u_{2n},\ldots, u_{kn})$ is a consistent estimator of $\Phi(U_1, U_2,\ldots, U_k)$ (Theorem 20, Hansen, Hurwitz, and Madow, 1953, Volume II, pages 75 to 76).

Using these theorems, we may show that the proportion variation estimator (equation 1.12) provides a consistent estimate of Roh (equation 1.11) because the expected value of s_{ha}^2, s_{hb}^2, and $\bar{b}_h$ are respectively, S_{ha}^2, S_{hb}^2, and $\bar{B}_h$; additionally, Var (s_{ha}^2), Var (s_{hb}^2), and Var $(\bar{b}_h)$ all approach 0 as n increases. Therefore, s_{ha}^2, s_{hb}^2, and $\bar{b}_h$ are consistent for S_{ha}^2, S_{hb}^2, and $\bar{B}_h$, and equation 1.12 provides a consistent estimate for equation 1.11 (Theorems 19 and 20, Hansen, Hurwitz, and Madow, 1953, Volume II, pages 74 to 76).