ON THE MEASURE OF HOMOGENEITY IN SUBCLASSES

G. Cicchitelli and G. E. Montanari, University of Perugia
Dipartimento di Scienze Statistiche, Università di Perugia, 06100 Perugia, Italy.

1 Introduction

As is well known, the design effect, *Deff*, is the ratio of the sampling variance for a given design to that for a simple random sample of the same size. A useful model for investigating design effects in multistage samples is

$$deff = 1 + roh(\bar{m} - 1), \quad (1)$$

where *deff* is an estimate of the design effect, $\bar{m}$ denotes the average number of elements sampled per selected PSU (Primary Sampling Unit) and roh is the so called *synthetic ratio of homogeneity*.

The usefulness of model (1) lies in the fact that empirical results suggest a relative closeness of roh in different subclasses to that of the whole sample. This holds true particularly for crossclasses, that is, for subclasses spread throughout the survey clusters.

This empirical evidence allows one to estimate the design effect for a particular subclass, δ, by means of the formula

$$deff_\delta = 1 + roh(\bar{m}_\delta - 1), \quad (2)$$

where $\bar{m}_\delta$ is the average number of elements sampled per PSU for the subclass. Hence, the process of imputation of *deff* to subclasses is as follows (see Kish et al., 1976):

(i) From formula (1), we compute

$$roh = \frac{deff - 1}{\bar{m} - 1}.$$

(ii) Then we substitute this quantity in the right-hand side of (2) to obtain $deff_\delta$.

In this paper we aim to throw a light on how the distribution of the variables over the population, on the one hand, and the features of the sample design, on the other, have effects on the value of the synthetic measure of homogeneity. Our study will be based also on an empirical study carried out on a real population.

The discussion will be at the "parameter level": we will deal with the true design effect, *Deff*, and the true synthetic ratio of homogeneity, Roh, and not with the estimated ones. As a consequence, instead of $\bar{m}$ and $\bar{m}_\delta$ in formulas (1) and (2), we will consider the corresponding expected values.

2 On the measure of homogeneity

Consider a population of M elements grouped into N clusters. Let $\mathcal{Y}$ be the survey variable. Denote by Y_{ij} $(i = 1, 2, \ldots, N; j = 1, 2, \ldots, M_i)$ and by $\bar{Y}_i$, respectively, the value of $\mathcal{Y}$ in unit j of cluster i and the mean of $\mathcal{Y}$ in the same cluster.

First of all, we want to define, at the population level, a measure of homogeneity of elements within clusters. For this purpose, consider the correlation ratio

$$\eta^2 = \frac{\sum_{i=1}^{N}(\bar{Y}_i - \bar{Y})^2 M_i}{M\sigma^2}, \quad (3)$$

where $\bar{Y}$ and σ^2 are the population mean and variance of $\mathcal{Y}$. This index measures the proportion of the variation in the variable $\mathcal{Y}$ explained by the association of $\mathcal{Y}$ with the cluster. It takes on values in the interval [0,1]. The minimum is reached when $\bar{Y}_i = \bar{Y}$, $\forall i$, which means extreme heterogeneity within clusters; the maximum is reached when $Y_{i1} = Y_{i2} = \ldots = Y_{iM_i} = \bar{Y}_i$, $\forall i$, that is, in the case of complete homogeneity within clusters.

We note that, if cluster sizes are equal, $M_i = \bar{M}$, $\forall i$, the following relation holds

$$\rho = \frac{\bar{M}}{\bar{M} - 1}\eta^2 - \frac{1}{\bar{M} - 1}, \quad (4)$$

where ρ is the classical intraclass correlation coefficient. Formula (4) suggests a possible generalization of ρ for unequal cluster sizes. For this purpose, it is sufficient to substitute, in the right-hand side of (4), the average cluster size $\bar{M} = \sum_{i=1}^{N} M_i/N$.

Obviously, we can define formula (3) also for a particular domain δ of the population

$$\eta_\delta^2 = \frac{\sum_{i=1}^{N}(\bar{Y}_{\delta i} - \bar{Y}_\delta)^2 M_{\delta i}}{M_\delta \sigma_\delta^2}$$

(the suffix δ means that the quantities refer to domain δ), and then measure the domain homogeneity by ρ_δ, obtained from (4) replacing η^2 and $\bar{M}$ by η_δ^2 and $\bar{M}_\delta$.

Using ρ and ρ_δ respectively as measures of homogeneity in the whole population and in the domain, we can say that, at the population level, the assumption of $\rho_\delta = \rho$ implies that the patterns of variation of $\bar{Y}_i$ and $\bar{Y}_{\delta i}$ across clusters are similar.

2.1 The case of a partitioned population

When the N clusters are partitioned into H non-overlapping subpopulations, in each subpopulation h ($h = 1, 2, ..., H$) we can measure the homogeneity by formula (3), and obtain η_h^2, say. The issue is then to find a summary measure of homogeneity as a function of η_h^2. Henceforth, for the sake of brevity, formulas will be given only for domain δ, on account that the whole population is a particular domain.

Observe that the following decomposition holds

$$\eta_\delta^2 = \bar{\eta}_\delta^2(1 - \omega_\delta) + \omega_\delta,$$

where

$$\bar{\eta}_\delta^2 = \frac{\sum_{h=1}^{H} M_{\delta h}\sigma_{\delta h}^2 \eta_{\delta h}^2}{\sum_{h=1}^{H} M_{\delta h}\sigma_{\delta h}^2},$$

($M_{\delta h}$ and $\sigma_{\delta h}^2$ being the number of elements and the variance in domain δ of subpopulation h) and

$$\omega_\delta = \frac{\sum_{h=1}^{H}(\bar{Y}_{\delta h} - \bar{Y}_\delta)^2 M_{\delta h}}{M_\delta \sigma_\delta^2},$$

in which $\bar{Y}_{\delta h}$ is the domain mean of subpopulation h.

Now, it is natural to define the quantity $\bar{\eta}_\delta^2$ as the *mean correlation ratio* within the H subpopulations, and then measure the mean homogeneity by formula

$$\bar{\rho}_\delta = \frac{\bar{M}_\delta}{\bar{M}_\delta - 1}\bar{\eta}_\delta^2 - \frac{1}{\bar{M}_\delta - 1}.$$

It is easily seen that $\bar{\rho}_\delta$ takes on values in the interval $[-1/(\bar{M}_\delta - 1), 1]$, the minimum being reached when $\rho_{\delta h} = -1/(\bar{M}_{\delta h} - 1)$, $\forall h$, the maximum when $\rho_{\delta h} = 1$, $\forall h$.

3 Homogeneity and sample design

Our aim is to obtain an explicit expression for Roh, which could account for the impact of homogeneity within clusters and of the sample design on its value.

Consider a general multistage sample design where the clusters (PSU) are stratified and selected with varying probabilities and with replacement. Let N_h and n_h be, respectively, the number of PSU's in stratum h ($h = 1, 2, ..., H$) and the number of PSU's selected from the same stratum ($\sum_{h=1}^{H} N_h = N$ and $\sum_{h=1}^{H} n_h = n$, n being the first stage sample size); let p_{hi} be the probability of selection of PSU i in stratum h.

Consider now the ratio estimator of the mean of $\mathcal{Y}$ in a given domain δ

$$\hat{R}_\delta = \frac{\hat{Y'}_\delta}{\hat{X'}_\delta},$$

where $\hat{Y'}_\delta$ and $\hat{X'}_\delta$ are the unbiased estimators of the totals of the variables

$$Y'_{hij} = \begin{cases} Y_{hij} & \text{if the element belongs to } \delta \\ 0 & \text{otherwise} \end{cases}$$

and

$$X'_{hij} = \begin{cases} 1 & \text{if the element belongs to } \delta \\ 0 & \text{otherwise} \end{cases}$$

The large sample variance of $\hat{R}_\delta$ can be obtained through the Taylor linearization method and it can be always written in the following manner

$$V(\hat{R}_\delta) = \frac{\sigma_{\delta 1}^{*2}}{n} + \frac{\bar{M}_\delta - \mu_\delta}{\bar{M}_\delta - 1}\frac{\sigma_{\delta 2}^{*2}}{n\mu_\delta},$$

where μ_δ is the expected average number of elements of the domain selected per PSU; $\sigma_{\delta 1}^{*2}$ and $\sigma_{\delta 2}^{*2}$ are quantities proportional, respectively, to the first stage variance and to that of the subsequent stages assuming negligible finite population corrections when they appear. We note that $\sigma_{\delta 1}^{*2}$ and $\sigma_{\delta 2}^{*2}$ depend on the sample design features, but not on n and μ_δ.

Dividing $V(\hat{R}_\delta)$ by the variance of the sample mean of a simple random sample of size $n\mu_\delta$, we get

$$Deff(\hat{R}_\delta) = \mu_\delta \frac{\sigma_{\delta 1}^{*2}}{\sigma_\delta^2} + \frac{\bar{M}_\delta - \mu_\delta}{\bar{M}_\delta - 1}\frac{\sigma_{\delta 2}^{*2}}{\sigma_\delta^2}.$$

Since the synthetic ratio of homogeneity for domain δ is given by

$$Roh_\delta = \frac{Deff(\hat{R}_\delta) - 1}{\mu_\delta - 1}, \tag{5}$$

substituting in (5) the expression of $Deff(\hat{R}_\delta)$, after some algebra, we obtain

$$Roh_\delta = \bar{\rho}_\delta + \frac{\bar{M}_\delta}{\bar{M}_\delta - 1}\left(\frac{\sigma_{\delta 1}^{*2}}{\sigma_\delta^2} - \bar{\eta}_\delta^2\right) + \frac{1}{\mu_\delta - 1}\frac{\bar{M}_\delta - \mu_\delta}{\bar{M}_\delta - 1}\left(\frac{\sigma_\delta^{*2}}{\sigma_\delta^2} - 1\right), \tag{6}$$

where $\sigma_\delta^{*2} = \sigma_{\delta 1}^{*2} + \sigma_{\delta 2}^{*2}$ and the parameter $\bar{\rho}_\delta$ is the mean correlation ratio within the strata considered as subpopulations (see section 2.1). Note that this relation holds also for the whole population.

Formula (6) is a meaningful representation of Roh_δ, expressed as the sum of three terms. It is easily seen that the sample design affects the second and third term but not the first one; observe that only the third term depends on μ_δ.

It is interesting to establish when Roh_δ is equal (or approximately equal) to $\bar{\rho}_\delta$. Consider the following conditions:

(i) $n_h = nM_{\delta h}/M_\delta$ (i.e. the number of sampled PSU's in each stratum is proportional to the stratum size);

(ii) $p_{hi} = M_{\delta hi}/M_{\delta h}$ (i.e. the selection probabilities of the PSU's are proportional to their sizes);

(iii) the sample is self-weighting;

(iv) samples within PSU's are equivalent to simple random samples of elements.

It can be shown (Montanari, 1990) that, if the previous conditions hold, then

$$Roh_\delta \approx (1-\omega_\delta)\bar{\rho}_\delta - \frac{\omega_\delta}{\bar{M}_\delta - 1} + \frac{1}{\mu_\delta - 1}\frac{\bar{M}_\delta - \mu_\delta}{\bar{M}_\delta - 1} \times$$
$$\left[-P_\delta\omega_\delta + (1-P_\delta)(1-\omega_\delta)(\bar{\rho}_\delta + \bar{M}_\delta^{-1})\right],$$

where $P_\delta = M_\delta/M$ is the relative size of the domain. We note that the last relation holds exactly if $M_{\delta hi} \propto M_{hi}$. Therefore, when ω_δ is negligible compared to 1 – as normally is for binary variables (Cochran, 1977; 5.11) –, if μ_δ is not small then $Roh_\delta \approx \bar{\rho}_\delta$.

Conditions (i) and (ii) yield

$$\frac{\bar{M}_\delta}{\bar{M}_\delta - 1}\left(\frac{\sigma_{\delta 1}^{*2}}{\sigma_\delta^2} - \bar{\eta}_\delta^2\right) = -\omega_{\delta 0}\left(\bar{\rho}_\delta + \frac{1}{\bar{M}_\delta - 1}\right)$$

which is normally negligible. The first condition is normally approximately satisfied for the total population (it is often suggested to build equal size strata when all n_h's are equal to one or two), and in such a case it is true also for domains evenly (or nearly so) distributed across strata. Condition (ii) is rarely satisfied (usually only a certain measure of PSU sizes is known); however, the ratio estimator is likely to have the same impact on $\sigma_{\delta 1}^{*2}$ that would be achieved if condition (ii) were true (see Hansen et al., 1953; I p. 267). Condition (iii) is necessary, even though not sufficient, to make σ_δ^{*2} approximately equal to σ_δ^2. The effect of departures from selfweighting on Roh_δ is considered in the literature and appropriate corrections to take into account unequal weights are proposed (Verma et al., 1980).

The last condition is the most problematic. If it does not hold, again, σ_δ^{*2} may be considerably different from σ_δ^2, because of the homogeneity of elements within sampling units other than primaries (see next section). As far as we know, this aspect has not been fully considered in the literature on the design effect.

In practice, if the previous conditions are not far from being satisfied, the behaviour of Roh_δ mirrors that of the parameter $\bar{\rho}_\delta$, regardless of the sample design. In this respect, some evidence can be drawn from the tables presented in the following section.

4 An empirical study

We considered the population of the region of Umbria (Italy), as given in the 1981 Census with Townships as natural clusters of population elements (persons). We examined some binary variables concerning the occupational status for the whole population and for some domains.

We computed:

(a) The value of ρ_δ given by formula (4).

(b) The difference $Roh_\delta - \rho_\delta$ for an unstratified two-stage self-weighting design where at the first stage Townships were selected with probability proportional to their sizes and with replacement and at the second stage elements were drawn by simple random sampling.

(c) The same quantity as in (b) for a sampling design which differs from the previous one in that at the second stage households are selected, instead of elements.

The analysis was made excluding the larger Townships (those with a population greater or equal to 20,000; the so called self-representing Townships in the labour force survey carried out by the Italian Statistical Office). The results are shown in table 1.

As we can see from the table, Roh_δ values obtained for the first design (sampling of elements) are always larger but close to ρ_δ, except where μ_δ is below 10, say. As regards the second design (sampling of households), the difference $Roh_\delta - \rho_\delta$ presents a more complex behaviour: it is always positive and larger than that observed in the previous case for "agriculture", sometimes assuming negative values for the other characteristics. This can be explained by looking at formula (6). Since the second term in the right-hand side is the same in the two designs, the difference between columns (b) and (c) must be attributed to the degree of homogeneity of elements within households which affects the third term. Note that such difference decreases in absolute value as μ_δ increases.

Table 2 contains results obtained under a stratified design. Townships were stratified according to the method used by the Italian Statistical Office in the labour force survey. From each stratum a Township was selected with probability proportional to its population. We note that the size of strata presents a large variability, the ratio of the largest to the smallest being 10 to 1. The two biggest values of ω_δ were 0.025 and 0.022.

The structure of table 2 is similar to that of table 1, but now $\bar{\rho}_\delta$ is reported in column (a). The pattern of variation in the two tables are very close. However we note a greater positive contribution of the term $(\sigma_{\delta 1}^{*2}/\sigma_\delta^2 - \bar{\eta}_\delta^2)$, probably because of the large variability of stratum sizes (condition (i) is not fulfilled).

It is worthwhile to pointing out that the differences between columns (c) and (b) are quite similar in tables 1 and 2. In this regard we must keep in mind that the contribution of households as clusters of elements is not affected by the stratification of PSU's.

To sketch the behaviour of synthetic ratios of homogeneity, plots of Roh_δ against $\bar{\rho}_\delta$ (using data from table 2) are shown in figure 1 for element sampling and in figure 2 for household sampling. The correlation coefficients values between Roh_δ and $\bar{\rho}_\delta$ are $r = 0.96$ and $r = 0.86$, respectively. In the latter case the synthetic measures are less predictable because of the effect of homogeneity of elements within households. At last, the average percentage increases of Roh_δ with respect to $\bar{\rho}_\delta$ are 38% for elements sampling and 58% for household sampling. In table 1, such average percentages are 7% and 17%, respectively.

5 The imputation of design effect to subclasses

As illustrated in the introduction, the process of imputation of design effects to subclasses requires the assumption that the synthetic ratio of homogeneity is *portable*, i.e. $Roh = Roh_\delta$. One may wonder about the error produced by this process.

To give some insight into the problem, denoting by $Deff^*(\hat{R}_\delta)$ the design effect imputed to domain δ, consider the difference

$$Deff^*(\hat{R}_\delta) - Deff(\hat{R}_\delta) = (Roh - Roh_\delta)(\mu_\delta - 1),$$

which is proportional to the difference between the synthetic ratios of homogeneity of the whole population and of the domain under study.

We can also write

$$Roh - Roh_\delta = (\bar{\rho} - \bar{\rho}_\delta) + [(Roh - \bar{\rho}) - (Roh_\delta - \bar{\rho}_\delta)].$$

The second term in the right-hand side is the portion of that error which, in essence, can be connected to the different features of the design for the domain and for the whole population. It is worthwhile noting that, even if conditions (i) - (iv) of section 3 were satisfied for the whole population, conditions (i) and (ii) are seldom fulfilled for the domain, furthermore μ_δ is smaller than μ and we expect a greater value of the third term in the right-hand side of (6).

As an illustration, for the stratified sample design table 3 shows the difference $\bar{\rho} - \bar{\rho}_\delta$ in column (a), the difference $(Rho - \bar{\rho}) - (Roh_\delta - \bar{\rho}_\delta)$ in column (b) for the case where elements are selected, and the same difference in column (c) for the case where households are selected. The variables and domains are those examined so far. Data were computed by subtracting, in table 2, the value of each row from the value of the first row.

Looking at table 3 we can observe the clear prevalence of negative quantities in column (a), which indicates that the homogeneity in the domains is often greater than in the whole population. But the design itself contributes to make $Deff^*(\hat{R}_\delta) - Deff(\hat{R}_\delta)$ negative as shown by columns (b) and (c); this occurs particularly when μ_δ is very small.

Generally speaking, the sample design yields an important portion of the "imputation error". However, to a certain extent, one can neutralize these disturbing factors through appropriate corrective strategies. In this context, for example, Verma et al. (1980) propose the correcting factor L to take into account the departures from self-weighting and Skinner (1986) suggests inflating the quantity $\bar{m}_\delta$ in model (2) to control the variability of the number of subclass elements in the sample.

For the error due to the difference between $\bar{\rho}$ and $\bar{\rho}_\delta$, one possibility is to analyse the behaviour of $\bar{\rho}_\delta$ using census data, if available, and then use this information in the process.

References

Cochran W. (1977): *Sampling Techniques.* Wiley, New York.

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

Kish L., Groves R.M., Krotki P.M. (1976): *Sampling Errors for Fertility Surveys.* Occasional Paper, n. 17. World Fertility Survey, London.

Montanari G.E. (1990): *Design effects and ratios of homogeneity in complex sampling designs.* To appear.

Skinner C.J. (1986): *Design effects in two stage sampling.* Jour. Royal Stat. Soc. 48 B, p. 89-99.

Verma V., Scott C. O'Muircheartaigh C. (1980): *Sample design and sampling errors for the World Fertility Survey.* Jour. Royal Stat. Soc. 143 A, p. 431-473.

Table 1. - Ratios of homogeneity within Umbria Townships (Italy) with a population smaller than 20,000 for some survey variables and domains. Values × 1,000 of ρ_δ (column a) and of Roh_δ - ρ_δ for sampling elements (column b) and for sampling households (column c)

Domain	Cluster sample size μ_δ	Employed in Agriculture (a)	(b)	(c)	Employed in Industry (a)	(b)	(c)	Employed in Services (a)	(b)	(c)	Employed in (a)	(b)	(c)
Total Population	121.5	16.8	0.0	3.0	13.0	0.0	0.1	3.8	0.0	1.2	6.4	0.0	-1.7
Age:													
14-29	25.8	15.2	0.8	5.8	25.4	1.0	7.0	6.8	0.3	3.7	17.2	0.7	5.5
30-54	39.8	32.5	0.4	7.1	24.9	0.8	0.2	9.2	0.1	4.6	13.2	0.4	-7.0
55-70	24.0	22.7	1.7	7.5	5.5	0.1	-1.3	3.2	0.2	5.3	9.3	1.1	1.3
Education:													
8 years	108.8	17.7	0.1	3.4	13.3	0.3	0.1	2.9	0.0	1.0	7.0	0.0	-1.5
13 years	10.9	11.5	2.0	8.0	20.0	1.6	2.6	10.7	0.6	6.0	7.0	1.0	1.7
Male	60.3	20.6	0.2	3.1	12.1	0.1	-0.8	6.1	0.1	-0.5	2.5	0.0	-4.7
Male aged:													
14-29	13.1	20.7	2.5	11.4	14.3	1.3	7.5	10.1	0.8	3.5	10.0	1.2	7.3
30-54	20.2	41.1	2.1	7.8	30.9	1.5	1.8	16.7	1.0	2.2	6.0	0.3	0.6
55-70	11.6	26.8	3.4	7.2	11.2	1.1	0.9	5.4	0.7	1.4	13.2	2.4	3.5
Male with:													
8 years ed.	53.4	21.6	0.4	3.4	11.5	0.2	-0.9	5.2	0.1	-0.4	3.0	0.0	-4.7
13 years ed.	5.8	14.6	5.3	12.3	31.5	6.9	11.9	16.5	4.4	7.0	6.2	1.9	7.7
Female	61.2	20.1	0.2	1.6	20.1	0.2	-0.4	3.0	0.0	0.1	17.6	0.1	-1.0

Table 2. - Ratios of homogeneity within strata of Umbria Townships (Italy) with a population smaller than 20,000 for some survey variables and domains. Values × 1,000 of $\bar{\rho}_\delta$ (column a) and of Roh_δ - $\bar{\rho}_\delta$ for sampling elements (column b) and for sampling households (column c)

Domain	Cluster Sample size μ_δ	Employed in Agriculture (a)	(b)	(c)	Employed in Industry (a)	(b)	(c)	Employed in Services (a)	(b)	(c)	Employed (a)	(b)	(c)
Total Population	121.5	10.0	2.5	5.8	3.0	1.2	1.6	2.0	0.7	1.8	2.1	0.5	-1.1
Age:													
14-29	25.8	8.1	3.4	9.2	5.7	1.1	7.1	2.4	1.0	4.4	4.8	1.4	6.0
30-54	39.8	20.2	7.1	14.0	8.6	5.0	4.6	4.4	1.5	5.9	4.8	2.0	-5.3
55-70	24.0	13.5	2.3	8.6	2.2	0.4	0.9	2.3	0.8	6.0	4.3	0.6	1.1
Education:													
8 years	108.8	10.6	2.6	6.1	3.0	1.3	1.5	1.5	0.4	1.4	2.4	0.4	-1.1
13 years	10.9	7.1	5.1	11.3	6.1	1.9	3.2	5.0	2.4	8.0	3.8	3.4	4.0
Male	60.3	11.9	3.2	6.5	4.0	1.0	0.3	2.8	0.9	0.4	1.0	0.1	-4.5
Male aged:													
14-29	13.1	12.2	6.7	16.7	4.7	1.7	7.7	2.6	0.8	3.8	3.3	1.6	7.7
30-54	20.2	24.6	9.2	15.4	12.6	3.4	4.5	7.6	2.6	3.8	3.3	0.7	1.2
55-70	11.6	14.7	3.8	8.2	4.9	1.3	1.4	3.7	2.2	2.9	5.2	1.6	3.2
Male with:													
8 years ed.	53.4	12.5	3.2	6.7	4.0	1.1	0.2	2.5	0.6	0.2	1.1	0.1	-4.6
13 years ed.	5.8	9.0	11.0	18.3	9.6	4.0	9.6	6.0	6.6	9.5	3.0	3.7	9.6
Female	61.2	13.3	2.9	4.6	5.9	2.5	2.2	1.6	0.5	0.6	5.2	1.5	0.5

Figure 1: Plot of synthetic against population ratios of homogeneity (sampling elements)

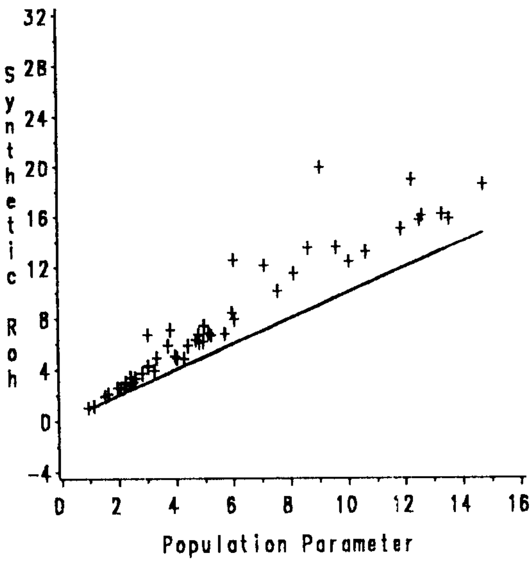

Figure 2: Plot of synthetic against population ratios of homogeneity (sampling households)

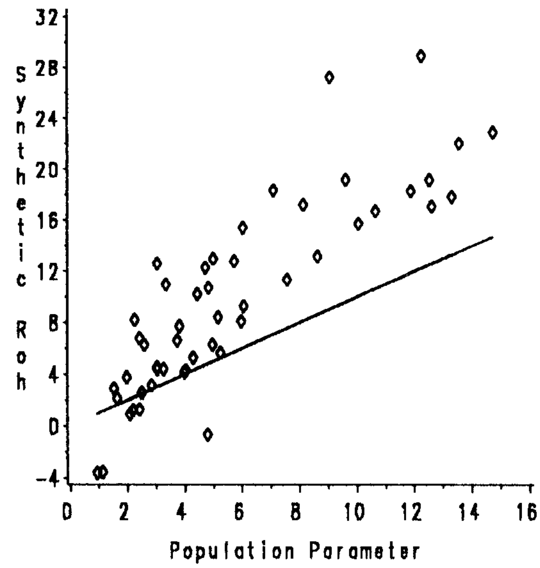

Table 3. - Comparisons between ratios of homogeneity for the whole population and for some domains within strata of Umbria Townships (Italy) with a population smaller than 20,000. Values × 1,000 of $\bar{\rho} - \bar{\rho}_\delta$ (column a) and of $(Roh - \bar{\rho}) - (Roh_\delta - \bar{\rho}_\delta)$ for sampling elements (column b) and for sampling households (column c)

Domain	Cluster Sample size μ_δ	Survey Variable: Employed in Agriculture (a)	(b)	(c)	Employed in Industry (a)	(b)	(c)	Employed in Services (a)	(b)	(c)	Employed (a)	(b)	(c)
Total Population	121.5	–	–	–	–	–	–	–	–	–	–	–	–
Age:													
14-29	25.8	1.9	-0.9	-3.4	-2.7	0.1	-5.6	-0.4	-0.4	-2.6	-2.7	-0.8	-7.1
30-54	39.8	-10.3	-4.6	-8.1	-5.6	-3.7	-3.1	-2.4	-0.8	-4.1	-2.7	-1.5	4.2
55-70	24.0	-3.6	0.0	-2.7	0.8	0.8	2.5	-0.3	-0.1	-4.2	-2.2	-0.1	-2.2
Education:													
8 years	108.8	-0.7	-0.1	-0.3	-0.0	-0.1	0.1	0.5	0.3	0.4	-0.3	0.1	-0.0
13 years	10.9	2.9	-2.6	-5.5	-3.1	-0.7	-1.7	-3.1	-1.7	-6.1	-1.7	-2.9	-5.1
Male	60.3	-1.9	-0.7	-0.7	-1.0	0.3	1.2	-0.9	-0.2	1.5	1.1	0.4	3.4
Male aged:													
14-29	13.1	-2.3	-4.2	-10.9	-1.7	-0.4	-6.1	-0.6	-0.1	-1.9	-1.3	-1.1	-8.8
30-54	20.2	-14.6	-6.7	-9.6	-9.7	-2.2	-2.9	-5.6	-1.9	-2.0	-1.2	-0.2	-2.3
55-70	11.6	-4.8	-1.3	-2.4	-1.9	-0.0	0.2	-1.7	-1.5	-1.1	-3.1	-1.1	-4.3
Male with:													
8 years ed.	53.4	-2.5	-0.7	-0.9	-1.0	0.1	1.3	-0.5	0.1	1.7	0.9	0.5	3.5
13 years ed.	5.8	1.0	-8.5	-12.5	-6.6	-2.7	-8.1	-4.0	-5.9	-7.7	-1.0	-3.2	-10.7
Female	61.2	-3.3	-0.4	1.0	-2.9	-1.3	-0.7	0.3	0.2	1.3	-3.1	-1.0	-1.6