THE ANALYSIS OF SAMPLING ERRORS

K.P. Krotki, Statistics Canada; L. Kish, The University of Michigan;
R.M. Groves, The University of Michigan

1. Introduction

This investigation is based on eight fertility surveys from five countries (South Korea, Taiwan, Malaysia, Peru and the United States), all of them conducted before 1974. The unique aspect of this investigation is the large number and variety of sampling error results that are calculated and analyzed. We suggest methods for the analysis and presentation of sampling errors for future surveys. Continued work in this field will hopefully lead to a type of data bank containing sampling errors for a large number of statistics originating from a vareity of sample designs.

2. Methodology

2.1 Formulas and calculations of deft and roh values

Deft (the square root of deff, the design effect) and roh (the synthetic intra-class correlation coefficient) are presented for approximately 40 means on the total sample and on 24 subgroups from each survey. We will refer to these means as "characteristics" and the subgroups as "subclasses." The choice of these characteristics was a subjective process guided by a desire to achieve a wide variety of substantive issues and some variation in the sensitivity of the statistic to clustering effects.

The formulas used, in their most basic form, are:

$deft^2 = var(r) / (s^2/n)$ where r is the ratio mean for a characteristic, var(r) is the computed sampling variance, and s^2/n is the simple random sample variance (estimatable by (pq)/n in the case of a proportion p).

$roh = (deft^2 - 1) / (\bar{b} - 1)$ where $\bar{b}$ is the average cluster size measured as the sample size, n, divided by the number of clusters, a.

The sample mean, r, a ratio mean, is of the form (y/x) where, because of clustering, x (as well as y) is a random variable because of variation in cluster size. In order to calculate the variance or r we use the approximate formula:

$$var(r) \doteq (1/x^2)\left[var(y) + r^2 var(x) - 2r cov(x,y)\right] .$$

Stratification and clustering are introduced into the calculation of var (r) in the standard fashion. The paired difference calculation was deemed appropriate in all the surveys. The samples on which the surveys were based were stratified, clustered areal probability samples. The sampling elements were women of child-bearing ages, and the primary sampling units (PSU's or clusters) were geographical units (e.g., counties, townships, city blocks).

Sampling errors were calculated for means and proportions of both the total sample, subclasses, and differences between subclass means. These consisted of differences $(y/x - y'/x')$ for the same characteristic in two categories of the same variable; the computations of these variances contain two variances and a covariance term. To compute a "synthetic roh," the value of $\bar{b}$ for the difference of means uses the harmonic mean of the sample sizes for the two subclasses.

2.2 Portability

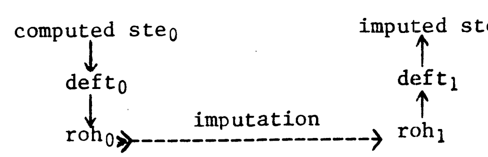
Our goal is to compute and present estimates of design parameters that can be used both simply and generally for diverse multipurpose designs. We think that portable estimates conveys the meaning we need. Portability refers to properties of the estimate that facilitate its use far from its source.

To illustrate, let us begin with the standard error, $ste(\bar{y})$, one computes for making inferential statements like $\bar{y} \pm t_p ste(\bar{y})$. Standard errors computed for one statistic can be imputed directly only to essentially similar survey designs. They are specific to the estimate $\bar{y}$ and depend on: a) the nature of the variables, b) their units of measurement, c) the nature and design of statistics derived from variables, d) sizes of the sample bases, which can vary greatly for subclasses, e) sizes of selections from sample clusters, f) nature and size of sampling units.

Design effects are considerably more portable than standard errors. They are widely used to modify simple random estimates $ste_{srs}(\bar{y})$ to guess at some ste(r) as $\left[deft \times ste_{srs}(\bar{y})\right]$. When we compute $deft = ste(r)/ste_{srs}(\bar{y})$, we remove the effects of the units of measurement and of the sample's aggregate size.

However, design effects for most subclasses diminish along with sample size, and using values of deft computed from the entire sample grossly exaggerates the actual effect of the design on subclasses. Also, deft values depend heavily on the sizes of sample clusters used.

We need portability to make inferences from one set of results to a set of variates with different values of $\bar{b}$. Values of roh are more portable for this purpose than deft or ste. We found usable stable relationships of roh for subclass means to roh for total sample means-much more stable than for values of deft or ste. Also we found relative stability of roh values across diverse subclasses for each characteristic from a sample; and similarities for similar characteristics across samples. Thus we propose the following indirect method of imputation from a computed standard error (ste_0) to an unknown one (ste_1):



We must, however, remain aware of factors that interfere with complete portability. The computed values of roh are also functions of the kind of sampling units used and of the selection procedures in several stages.

2.3 The use of roh and deft for imputation

We need to impute roh for subclasses from values computed for the entire sample or for similar type subclasses. Thus we need stability (portability) for roh values and we seem to find that for crossclasses. This type seems to cover most subclasses used in survey analysis. Crossclasses is a term we coined for subclasses that cut across clusters and strata used in the selection process. The sizes of sample clusters for each subclass are roughly $\bar{b}_s = \bar{b} M_s$, where M_s is the proportion of the subclass in the sample and $\bar{b}$ is for the entire sample. Design effects tend to decrease linearly almost to 1 as the crossclass size decreases and roh remains relatively constant. We must first impute some value $roh_1 = \lambda_1 roh_0$ from computed values of roh_0 and a correction factor λ_1. Then we estimate the unknown $deft_1$ from $deft_1^2 = 1 + \lambda_1 roh_0 (\bar{b}_1 - 1)$. We computed values of roh_0 based on means for the entire sample for each of 40 characteristics on each survey. We then computed and found values near (and slightly over) $\lambda_1 = 1$ for the diverse subclasses.

2.4 Summarizing sampling error results

Sampling errors computed from survey samples are themselves usually subject to great sampling variability. Many samples are not based on a large enough number of PSU's to yield sufficient precision for individual estimates for sampling errors. In addition, most surveys are highly multipurpose in nature and we must combine results from diverse statistics for joint decisions and designs. Some form for combining them must be sought, because combining their results is preferable to its alternatives. We argue against following the common practice of choosing a single variable among many for making inferences about the design and planning future designs.

Several methods were applied to the sampling error results in this investigation in order to identify underlying trends and relationships. Much of what was done was on an ad hoc basis as each survey presented its own idiosyncracies. Thus the methods shown here should be viewed more as a progress report than as final optimal techniques. Hopefully we have pointed out some approaches that may be applicable on a more general scale.

First, characteristics were listed by order of magnitude of roh. Another approach to arrive at the same information is to group supposedly "similar" characteristics and to calculate the average roh for each group. The mean and range of roh values for the characteristics within each group can serve as summary statistics. Measurements on the same characteristics at different points in time or under different survey conditions provide further data on the sampling behavior of these characteristics.

The study of sampling errors for subclasses is an important need because much survey analysis involves comparisons of subclasses. It is difficult to give guides for how the choice of subclasses should be made, but using measures which are candidates for independent variables in analysis of the data may be desirable. In this view, the characteristics would be analogous to the dependent variables. Comparison of sampling errors for the total sample and for the subclasses can give the survey designer an idea of how to impute in general from total results to subclasses. This is a common requirement since sampling errors cannot be calculated for all possible subclasses for each characteristic.

3. Empirical Results

The above described methodology was applied to the sampling errors calculated for eight fertility surveys in five countries. In this section we discuss in detail the results for one of these surveys.

Detailed analysis of sampling error results for Taiwan: General Fertility Survey (1973 KAP-4)

3.1 Sample design

The universe of 331 townships was divided into 27 strata using level of urbanization, education, and fertility. Within strata, townships were geographically ordered and 56 were selected systematically. Within selected townships the sample had three stages, yielding 5588 married women aged 20-39. The coefficient of variation of size among the 56 ultimate clusters is 0.03 for the entire sample; within the 24 subclasses used it ranges from 0.02 to 0.08.

3.2 Results for the total sample

Results for 40 characteristics are presented in Table 1. The characteristics are ordered from highest to lowest values of roh. Deft values follow this trend closely with minor exceptions due to slight differences in sample bases (n), hence cluster size (n/a). Note the large range of roh values (col. 4) for the 40 characteristics, essentially from 0 to 0.3. The quartiles are about 0.075, 0.025 and 0.015. These correspond to deff values of about 8.4, 3.2, and

Table 1

Taiwan Fertility Study (KAP), 1970, Means, Ste's, Deft's and Roh's for 40 Characteristics Together with Summary Roh Values for Subclasses and Differences[a]

Char. Type	Characteristic	1 Mean	2 Std. Error	3 Total Sample Deft	4 Total Sample roh	5 Sub-Class roh	6 (5)/(4)	7 Ave. rohd (r-r')
3	Sex preference scale	5.23	.053	5.41	.290	.334	1.15	.012
4	Approve contraception strongly	0.38	.034	5.28	.273	.350	1.28	.020
4	Approve sterilization	0.72	.029	4.75	.219	.251	1.15	.007
4	Should have many children	.037	.029	4.49	.194	.241	1.24	.015
4	Ideal first birth interval	20.86	.478	3.82	.140	.181	1.29	.006
3	Number preference scale	4.70	.053	3.59	.122	.186	1.52	.016
3	Husbands not wanted last pregnancy	0.24	.019	3.39	.106	.125	1.18	.010
4	Ideal marriage age	23.10	.076	3.23	.096	.115	1.19	.012
4	Expect sterilization	0.33	.020	2.98	.088	.107	1.22	.003
4	Approve abortion	0.24	.017	2.94	.078	.134	1.72	.014
2	Visited Health Station	0.47	.019	2.88	.074	.105	1.42	.009
4	Others should have < 3 children	0.66	.018	2.87	.074	.088	1.19	.007
3	Desired children < expected	0.06	.008	2.50	.057	.079	1.39	.002
2	Contraception from private MD	0.47	.018	1.96	.055	.090	1.63	.018
3	Ideal number of children	1.37	.018	2.42	.051	.063	1.23	.006
3	Husband's ideal number of children	3.24	.028	2.26	.048	.075	1.55	.014
2	Visited by health worker	0.37	.015	2.37	.047	.072	1.55	.005
3	Ideal number of boys	1.89	.014	2.08	.036	.043	1.22	.005
2	Plan no future contraception	0.10	.008	1.92	.028	.042	1.47	.007
6	Age at marriage	20.31	.072	1.86	.025	.041	1.62	.008
3	Wife-husband want same number of children	0.19	.010	1.83	.024	.037	1.55	.006
1	Able to have children	0.86	.008	1.81	.023	.028	1.22	.003
3	Desired number of children	3.54	.031	1.79	.023	.038	1.68	.005
2	Contraception started after pregnancy number	3.57	.042	1.55	.022	.040	1.86	.006
1	Husband's mother's number children	6.05	.059	1.72	.021	.036	1.74	.005
3	Expected total births	3.58	.030	1.68	.020	.040	2.06	.006
5	Literate wife	0.75	.010	1.67	.018	.042	2.31	.008
1	Number of live births	3.20	.037	1.65	.017	.032	1.86	.008
1	Wife's mother's number children	6.45	.051	1.62	.016	.020	1.25	.004
2	Ever used contraception	0.67	.010	1.61	.016	.020	1.28	.001
3	Want no more children	0.67	.010	1.56	.014	.014	1.01	-.003
1	First birth interval	15.14	.236	1.49	.013	.017	1.29	-.002
1	Open birth interval	45.22	.836	1.52	.013	.025	1.93	.003
5	Literate husband	0.92	.005	1.50	.013	.024	1.89	.007
2	Contraception before 1st pregnancy	0.02	.003	1.35	.011	.006	.050	.000
2	Currently using contraception	0.45	.010	1.45	.011	.006	0.57	-.002
1	Living sons number	1.54	.021	1.43	.011	.012	1.08	.002
1	Living children number	3.04	.029	1.39	.010	.017	1.75	.006
1	Pregnant now	0.12	.005	1.21	.005	.005	1.11	-.001
2	Induced abortions number	0.31	.012	1.19	.004	.012	2.72	.004
	Averages				.0592	.0790	1.436	.00652
	Ratios of means of col. 5/col. 4 and col. 7/col. 5					1.334		.083

[a] The characteristic type denotes: 1) fertility experience, 2) contraceptive practice, 3) birth preferences and desires, 4) attitudes, 5) socio-economic background, 6) demographic background.

2.5; these large factors arise because of the large number of elements, almost 100, per cluster. The mean roh on the total sample is 0.0592.

It is useful to observe the clear differences in roh values between the 6 classes of characteristics. Attitudinal variables are all in the first quartile, with roh value over 0.075. Birth preferences and desires are mostly in the top two quartiles, with roh values over 0.025. Contraceptive practice is spread evenly between the second quartile (0.075 - 0.025) and the second half under 0.025. Fertility experience variables are all in the lower half with roh values under 0.025. They are evenly spread among socio-economic (which, in this survey, only indicates literacy) and demographic variables. These three classes of variables (codes 1, 5 and 6) are contained in the lower half, with roh values under 0.025, while classes 3 and 4 are above that.

If roh values were unusually high for all variables, we should look either into causes for unusual segregation in the population or into the choice of small and homogeneous sampling units. However, roh's for demographic variables are not high. Their spread under 0.025 is similar to values found in other populations. Two explanations are possible for the high roh values for the subjective variables of attitides and birth preferences and desires. First, is is sociologically reasonable to think that when attitudes change rapidly, the spread of the change takes place unevenly and is clustered in areas. Second, clustering of the measured values can be caused by interviewer effects which are not separable from the effects of clusters themselves.

3.3 Results for subclasses

Clustering of values for subgroups of the sample was investigated for the 24 subclasses in Table 2 for each of the 40 characteristics. This vast amount of data is summarized in Column 5 of Table 1. Each entry is the mean of the rohs over the same 24 subclasses of Table 2. This mean subclass roh is shown as the ratio to the roh for the total sample (col. 6). Note that the mean subclass roh values parallel closely the total roh values. The ratios of subclass/total roh values do not vary greatly around their mean of 1.436. A more useful average is .0790/.0592 = 1.334, the ratio of the two mean values. This gives greater weight to the larger roh's where more fluctuations can be observed. A quick rule of thumb would guide the researcher to use the total roh times 1.33 to obtain subclass roh's. This yields

$$deff_{subclass} = \left[1 + 1.33\, roh_{Total}\,(b_{subclass} - 1)\right]$$

Column 4 of Table 2 presents values of roh for each subclass averaged over all 40 characteristics. Column 5 notes the ratios of these averages to the mean roh value of 0.0592 when the total sample is the base. For these values of subclass bases there exists no clear separation between socio-economic and demographic subclasses that we found for them as characteristics. Though the former tend to be a little higher, most of the variation is within the groups. The

Table 2

Taiwan Fertility Study (KAP), 1970, Deft's and Roh's for Twenty-four Subclass Variables Treated as Characteristics and Subclass Base.

		1	2	3	4	5	6	7
		Population Base			Subclass base		Differences	
		Prop.	Deft	Roh	Ave. Roh	Ratio to .0592[1]	Ave. Rohd	(6)/(4)[2]
Education of husband	None	.255	1.684	.0186	.1212	2.05	.0101	.111
	Primary	.548	1.727	.0201	.0615	1.04		
	Junior High	.081	1.453	.0112	.0410	0.69	.0053	.077
	Senior High +	.070	1.739	.0205	.0969	1.64		
Occupation of husband	Farmer	.219	2.437	.0509	.1474	2.49	.0208	.189
	Labr.&Opertv.	.202	2.002	.0310	.0726	1.23		
	Skilled	.149	1.951	.0289	.0733	1.24	.0041	.065
	White Collar +	.359	1.872	.0258	.0525	0.89		
Income of family (1000 NT)	0-23.9	.154	4.171	.1987	.1765	2.98	.0211	.160
	24.-35.9	.172	1.445	.0132	.0868	1.47		
	35.-47.9	.172	1.807	.0274	.0639	1.08	.0044	.067
	48. +	.303	2.476	.0621	.0671	1.13		
Ave. for 12 classes			2.064	.0424	.0884	1.494	.0110	.112
Children ever born	0-1	.147	1.221	.0050	.0671	1.13	.0036	.054
	2	.172	1.122	.0026	.0667	1.13		
	3	.239	0.987	-.0002	.0613	1.04	.0025	.036
	4 or more	.396	1.429	.0105	.0766	1.29		
Marriage duration	0-4	.228	1.139	.0031	.0622	1.05	.0031	.049
	5-9	.267	0.874	-.0024	.0647	1.09		
	10-19	.386	1.038	.0008	.0741	1.25	-.0001	-.001
	20 +	.058	1.037	.0008	.0936	1.58		
Age of wife	19-24	.189	1.150	.0032	.0554	0.94	.0014	.022
	25-29	.252	1.187	.0041	.0715	1.21		
	30-34	.260	1.169	.0037	.0678	1.14	.0006	.008
	35-42	.255	0.892	-.0021	.0733	1.24		
Ave. for 12 classes			1.104	.0024	.0694	1.174	.0019	.028
Ave. for 24 classes					.0790	1.334	.0064	.070

1. 0.0592 is the average roh for the 40 characteristics on the total sample (see bottom of Col. 4 of Table 1).
2. In calculating the ratio, the mean of the two entries in col. 4 is used.

average roh for the 24 subclasses is 0.0790, and the ratio 0.0790/0.0592 1.334 measures the average increase over the roh value based on the total sample.

3.4 Results for differences between subclass means

We have computed roh values for the difference of each of 2 pairs in each set of 4 subclasses, for each of the 40 characteristics. The averages over the 12 values are shown in col. 7 of Table 1, where rohd is the roh for the difference. These rohd values are substantially lower than the corresponding subclass values. The individual ratios (not shown) of values in column 6 to column 4 vary considerably around their average of .095. A better average is the ratio of means: .00652/.0790 = .083. The individual ratios range most from 0.30 to 0.00, except from some trivial cases near the bottom of the table, where negative values appear. We have also found in many other studies positive but smaller effects for differences than for the corresponding subclasses. The effects of covariance between subclasses seem unusually strong in this design. Consequently, the effects of clustering of differences though still present, are considerably reduced. In column 6 of Table 2 are shown roh values for differences of pairs of subclass means. Each of the 12 entries represents an average over the 40 variables of Table 1. Note the great reductions in design effects due to positive covariances in clusters. The ratios of the average rohs is .0064/.0790 = 0.081.

4. Highlights from other surveys

The 1971 and 1973 South Korea fertility studies provided an opportunity to study sampling errors for the same characteristics at two points in time. At first glance it seemed that the roh values in 1973 were considerably smaller than those in 1971. The average roh value for some 40 characteristics was 0.049 in 1971 and 0.033 in 1973. However, when we examined only the subset of characteristics which were common to both surveys the average roh values were 0.037 in 1971 and 0.030 in 1973. In this subset the design effects are 3.85 and 2.02 respectively because the average cluster size in 1973 was much smaller than in 1971. This is an example of why we argue for portability in terms of roh rather than deft. The range of roh values in the South Korean fertility surveys was 0 to 0.2.

A fertility survey of Malaysia was conducted in 1969 and yielded 2,950 interviews with women involved in two large family planning programs. The sample was drawn after stratification into rural and urban areas. It was found that the design effects were far larger in the rural than in the urban areas. For 29 variables, the average deft's for the rural and urban areas were 1.92 and 0.99 respectively. The average roh for rural areas was 0.046. In the urban areas there was no clustering since the respondents were selected individually from lists of names. The range of roh values for the total sample was 0.02 to 0.05.

Arranging the characteristics by size of roh revealed two striking results. The characteristics "proportion using NFPB clinic," "proportion Malay" and "proportion with farmer husband" produced abnormally large sampling errors (deft's of 4.06, 2.65 and 2.58 and roh's of 0.36, 0.14 and 0.13 respectively). The first is explained by the fact that women in a given cluster either attended one type of clinic or the other. (This variable could have been an appropriate stratification variable.) The second result suggests that ethnicity is a highly clustered variable in Malaysia. The third result is due to the fact that clusters follow geographical boundaries with diverse densities of farmers.

Another result gleaned form the Malaysia survey is that subclasses that approximate crossclasses produce different sampling errors than do subclasses that are segregation classes. Over 5 pairs of crossclasses (e.g., income, age, marital status) the average roh across 14 characteristics was 0.0318, which has a ratio of 1.15 to the average roh for these characteristics on the total sample. On the other hand, if we consider the segregation classes (e.g., type of clinic, ethnicity, rural-urban birth and farmer-nonfarmer occupation) the average roh is 0.0750.

5. Summary of Results from Eight Surveys

For each survey sampling errors were computed for about 30 to 40 characteristics. This was done in each survey for means based on the entire sample and on about 24 subclasses and for differences between about 12 pairs of subclass means. The great range across different variables in values of roh in each of the surveys is the most important result. The roh values have an effective hundredfold range in each survey from about 0.001 to 0.002 to about 0.1 or 0.2.

Some differences between types of variables can be detected on each survey in Table 3. However these differences are not consistent and are also marked by considerable sampling variability. Socio-economic variables appear noticeably high for Korea and Peru. Demographic background variables tend to be near the lower end for all surveys. Attitudes and birth preferences appear high though more often in the lower half with roh values mostly from 0.005 to 0.05. The ranges within types (not shown) seem to be factors of about 5 to 10. They are considerably less than the range of 50 or 100 for rohs of all variables within surveys. Thus the typing of variables seems an effective and simple way to reduce our level of ignorance.

The individual computations of rohs for each characteristic/subclass combination are subject to great variability. But the average roh for each characteristic computed over several subclasses is quite stable. We refer to subclasses that are approximately crossclasses (more or less evenly distributed in the sample clusters). Other kinds of subclasses, those that are very unevenly distributed in sample clusters, need special considerations.

Table 3 summarizes a vast body of computations over the eight surveys. Since the variables included had not been coordinated initially, it is comforting that some very useful stabilities may nevertheless be drawn from them. The average values of overall rohs (first row) varies from .024 to .063. This stability is quite good, considering the diversity of variables and sample designs. It is helpful for choice of sample designs, since accepting .04 or .05 for roh would not badly mislead one. For fertility experience and demographic background variables, the roh values are lower and more stable, .011 to .038. For general attitudinal variables the roh values are very high for Taiwan and Peru and fertility preferences are also high in Taiwan. It would be interesting to investigate how much

TABLE 3

Summary of Average Roh's for Eight Surveys [a]

STATISTIC	SAMPLE SURVEY South Korea 1971	South Korea 1973	Taiwan	Peru	Malaysia	United States 1960 Whites	United States 1970 Whites
A. ROH'S FOR TYPES OF VARIABLES FOR TOTAL SAMPLE (Number of characteristics below roh values)							
1. All Characteristics	.050	.033	.059	.063	.045	.024	.037
	40	39	40	29	29	9	36
2. Fertility Experience	.016	.009	.014	.034	.025	.011	.019
	11	6	9	8	3	4	6
3. Contraceptive Practice	.047	.021	.030	.054	.022	.043	.029
	9	11	9	8	3	2	8
4. Fertility Preferences	.023	.024	.072	---	.028	.025	.019
	6	8	11	0	3	2	6
5. Attitudes	.028	.026	.145	.094	.017	---	.061
	2	3	8	1	2	0	16
6. Socio-economic Variables	.128	.081	.016	.126	.045	---	---
	9	8	2	7	12	0	0
7. Age, Marriage (demographic background)	.014	.025	.025	.024	.010	.039	.105[b]
	3	3	1	5	2	1	1
B. ROH'S FOR SUBCLASSES AND FOR DIFFERENCES							
Number of Characteristics	40	39	40	20	14	9	36
Number of Subclasses	23	22	24	10	20	8	24
8. Roh's for Total Sample	.050	.033	.059	.056	.028	.024	.037
9. Roh's for Subclasses	.059	.044	.079	.065	.032[c]	.048	.052
10. Ratio of Subclass/Total (9)/(8)	1.19	1.36	1.33	1.15	1.15[c]	2.00[d]	1.41
11. Differences of Means	.0060	.0000	.0065	.0170	.0200[c]	.0130	.0050
12. Ratio of Difference/Subclass (11)/(9)	.100	.000	.083	.026	.210[c]	.270	.096
C. COMPARISONS OF SUBCLASSES: SOCIO-ECONOMIC (SE) VERSUS CROSSCLASSES							
13. SE as Characteristics	.076	.092	.042	.105	---	---	.122
14. Others as Characteristics	.006	.007	.002	.015	.037	---	.020
15. SE Subclass Base	.063	.040	.088	.073	---	---	.063
16. Others as Subclass Base	.057	.038	.069	.063	.032	---	.047

[a] The eighth survey pertaining to blacks in 1970 was unreliable due to sample design and small sample size.

[b] Results unreliably high for unknown reasons.

[c] Results are for 10 crossclasses only.

[d] This result is based on 8 subclasses and removing one of them reduces the ratio to 1.15.

of these high roh values are due to homogeneity of the respondents in compact clusters, or how much of the effects of interviewer variance of response from large workloads. The high roh values for socio-economic variables in Peru and South Korea have implications for sample designs, as well as for sociological studies of their sources.

When we separate socio-economic subclasses from others we regularly note considerable differences between the two groups, when these are computed as characteristics based on the entire sample (rows 13 and 14). However, when used as subclasses (rows 15 and 16) the differences between the two sets of subclass roh's (averaged over all characteristics) are not great, say 1.2 versus 1.4. It is the characteristics, much more than the subclass, that are the sources of variability in sampling errors.

The ratio of the rohd's for difference to the average roh's for subclass means (rows 11 and 12) is not stable. In all cases the reductions due to covariances between clusters are substantial. The central value may be 0.1 and 0.2.

6. Strategies for Large-Scale Calculation, Summarization and Presentation of Sampling Errors

(1) Paired selection considerably simplifies sampling error calculations.

(2) The coefficient of variation of cluster size should always be calculated and inspected before the results of sampling error calculations are published, since the approximate formula for var(r) requires $cv(x) < 0.2$.

(3) Codes identifying the primary sampling units and the strata must be included together with the data. Our experience has been that these codes are seldom readily available.

(4) Sampling errors should be calculated for the entire sample for many variables. We think it inadequate to single out a few critical survey variables or several categories of one variable. Rather than exhausting all categories for a few variables, more variables should be used, each one for one or a few categories. Variability between variables is generally greater than between categories within variables. This is especially true for characteristics, but also for subclass variables. The range of variables should parallel the aims of the survey, of its analysts and of its users. Also, it should aim to cover the range of design effects.

(5) The variables should be separated into a few groups within which the sampling errors are expected to be relatively similar.

(6) Sampling errors should be computed for many characteristics each based on a moderate number of subclasses. Sampling errors, particularly roh's, were found subject to greater diversity across characteristics than across subclasses. Subclass results should be compared to the results obtained for the total sample.

(7) Most of the needed subclasses tend to approximate crossclasses. However, partially segregated subclasses, if important, should also be investigated.

(8) In choosing subclass categories a range of subclass sizes should be selected to obtain empirical evidence of the effect of subclass size on deft and roh.

(9) All chosen characteristics should be analyzed by all chosen subclasses (rather than using different subclasses for each characteristics). This yields a symmetrical table and averaging can be done over both subclasses and characteristics. However, other designs may be used, especially for a larger number of subclasses.

(10) Sampling errors should be computed for the difference of means of pairs of subclasses. For many subclass variables one or two pairs usually suffice. These results should be compared with the individual results for each of the two subclasses.

(11) Sampling error results should be preserved and publicized for the use of survey designers who would find such data useful in the design of future surveys.

In addition to the 40 characteristics that we treated as "dependent," we also computed roh values for 24 variables later used for subclass analysis. Here a clear dichotomy emerged. The 12 characteristics based on demographic variables had roh values under 0.005 (Table 2, col. 3). However, the 12 socioeconomic characteristics had roh values 0.01 to 0.20. Within the two classes of characteristics there is variation, but much of it is too haphazard to be of general use.