A CRITIQUE OF SURVEY SAMPLING PRACTICE AND USE OF SURVEY DATA IN SOCIAL SCIENCE RESEARCH[1]

by

I. P. DAVID[2]

I. Introductory Remarks

First I wish to have on record my thanks for the invitation to write a paper for this conference, which arrived six weeks ago. How I wished it came six months earlier.

But it was difficult to turn down an invitation which very wisely left the choice of topic open as long as it was about survey sampling. This is a timely subject for discussion in view of the proliferation of surveys in government, research centers and market research organizations. We should be pleased that a tool, shaped and honed to what it is (or could be) today by statistical science, become an important part of the modern scientific method.

But some of us are not entirely happy about the way many surveys are done and put to use. We view with great concern the pedestrian approach to conducting sample surveys and the haphazard use of survey data that seem to pervade research today.

I thought it would be preferable to substantiate the last sentence via a case study approach, using a sample survey I was working on when the invitation came. One drawback is that this particular survey's sampling design (discussed in section II) is hardly pedestrian.

---

[1] Read at the Philippine Statistical Association Annual Conference, Manila, 16 July 1982.

[2] Statistician, Asian Development Bank. The views expressed here are those of the author and not of ADB. The computing assistance of Mss. F. Fernandez and D. Samarita-Maligalig are gratefully acknowledged. I also wish to thank Ms. Ellen de la Cruz and her PBME staff at the National Irrigation Administration for the fruitful discussions and use of their data.

It had the imprints of the collective wisdom of a (statistical) consultative committee that provided advice during its formative stage. It is therefore more of a counter-example. Nevertheless, it had two features which caused some anxiety about the survey's actual usefulness vis-a-vis its original objectives.

Section III presents the 'normal' way a 'typical' researcher would summarize and draw inferences from the data. These are compared with statistically sound procedures. Of the three topics in the section, the reader's attention is directed particularly to III.C which discusses the problem of estimating frequency distributions. The social science research literature is strewn with 'normally done' one-way and multiway frequency tables from survey data.

Section IV attempts to summarize the reasons for this casual approach to designing sample surveys and the widespread use of incorrect inference procedures from survey data. Many of these can be traced back at our own doorstep which, on the brighter side, means that we can do something about them.

II. The Survey

A. Purpose

Here is a story that is replicated frequently in many parts of the developing world. There is a small area in Davao province which the Government, through the National Irrigation Administration (NIA), had identified as potentially suitable for irrigation development. An intention for a loan was filed with a funding agency; this triggered a series of studies including a feasibility report and, finally, an appraisal report which would render a verdict whether or not a loan is justified. The area referred to has been given the identity tag Davao III, for which a feasibility report was completed in mid-1981 and an appraisal mission was fielded a few months ago.

It takes some months and sometimes up to a year for a small team of subject matter specialists to complete a feasibility study. An appraisal team normally is given only a few weeks to finish its report. Although the latter draws heavily on the feasibility study, both require detailed data on the engineering aspects of the project, as well as project area – specific statistics on population, crop yield, land use, prices, production costs and other agronomic-demographic data. Where to get these is always a problem because national statistical systems generally provide aggregate statistics at the national and regional levels only. In the Philippines it is difficult to get reliable and usable data at the provincial level. Thus, although some feasibility study teams manage to do small quick sample surveys to meet part of their data needs, to say that much 'guesstimations' and use of data from judgment samples are part of the tools in project feasibility and appraisal work would not be stretching the truth too far.[3] Some find this disturbing, especially since millions of dollars are on the line and, if approved, the same data very likely will be used again to assess whether or not the project lived up to its billing, such as, at post-evaluation. However, it is difficult to imagine how anybody else can do significantly better, unless some of the old rules are relaxed and innovations are tried.

Davao III was going to try one innovation. To minimize uncertainties about data quality and availability, a benchmark survey of the project area was conducted by NIA in December 1981, the results of which were to be used by the appraisal team around April 1982. Moreover, this survey was to be the first of a series to be undertaken in the framework of a project benefit monitoring and evaluation system (PBMES) for Davao III.[4] The monitoring function of these surveys make it imperative that they be done and analyzed quickly, which means among other things, that they should be smallish both in sample size and subject matter coverage. On the other hand, these should be able to help measure with sufficient accuracy, or detect with high probability, small and varied benefits from the project, e.g. a 0.4 ton change in rice yield, a 10 per cent real increase in median household income, or a 50 per cent reduction of the gap between

[3]Lately, the term 'convenience sampling' had been used to mean judgment or non-probability sampling. At least the first sounds honest, although of course the resulting samples are no more scientifically valid than other judgment samples. Perhaps a lesson that can be drawn here is that users find the class of probability sampling procedures wanting in operational convenience.

[4]For details on PBMES see ADB (1980) Guidelines on Logical Framework Planning and Project Benefit Monitoring and Evaluation; and B.T. Oñate (1982) Benefit Monitoring and Evaluation System in Agricultural and Rural Development Project Design.

measured and recommended daily per capita calorie intake before and after project completion. These requirements call for an efficient sampling design, coupled if possible with precision-increasing estimation techniques.

In order not to leave gaps in the story, it should be mentioned at this point that the PBME benchmark survey returns were not analyzed in time for the appraisal mission. This is not a rare phenomenon. From experience, I cannot recall one sample survey that was analyzed within its original schedule. The reasons are almost always the same – optimistic timetable, data processing problems, and lack of statistical help at the analysis stage. Nevertheless, analysis of the survey data must go on for a *post facto* adjustment of appraisal report statistics (should the need arise) and to provide benchmark information for PBMES. Far from discouraging us, this turn of events only points to the challenge of planning and completing sample surveys on time without sacrifice on scientific validity and statistical accuracy.

B. The Sampling Design and Related Comments

To construct a sampling frame for the benchmark and other future PBME surveys, the NIA PBME Unit listed in August 1981 all households in the proposed project area, along with land use, occupation of members, and a few more items that were required in stratification, sample allocation and actual drawing of samples. This crucial first step in survey sampling work is not so easy to sell because it is costly, time-consuming and its importance is not always convincingly clear to management. "If this is a sample survey, why list everybody? Why not try getting a similar list elsewhere instead of going from house to house?"

The last question needs careful consideration. With so many surveys and censuses that had been going on, it is almost certain that a sampling frame can be assembled with much less trouble and cost. The problems are age (obsolescence) and accessibility of past data. For instance, the sampling frame of the 1981 Census of Agriculture and Fisheries (CAF) of the Philippines covered the same (and more) information as that required in the Davao PBME survey. Furthermore, it would not be surprising if the CAF main questionnaire included many of the items asked in the Davao PBME sample questionnaire which, subject to accessibility, raise the possibility of at least thinning down the latter or, in the extreme case, relying on secondary data completely. However, accessibility to these data is limited by laws that protect the confidentiality of individual returns to government censuses and survey.[5]

The sampling frame identified 6,645 households with farms inside the proposed project's command area[6] (see Chart IA). These were stratified by size of landholding, namely, Small (under 3 hectares), Medium (3 to 7) and Large (over 7). Each of these strata were subdivided further into so-called Unit I and Unit II areas; these are geographically distinct areas identified in the feasibility report as diverse in terms of yield and other farming characteristics. In each Unit the households were again split into Upstream (U) and Downstream (D), signifying the household farm's location relative to the feasibility report's proposed placement of irrigation canals. Up to this point, the stratified population, along with stratum and sample sizes (the latter shown in ()), is as shown in Chart 1A.

There were still two levels of stratification after the (U,D) grouping of households, namely, by major crop grown (rice, corn, coconut, others) and lastly by tenure status (full owner, part owner, share tenant, lessee, amortizing owner, squatter, others). For example, see Chart 1B which shows the complete stratification and final sample allocation for the L-II-D stratum.

The total sample size was initially and more or less arbitrarily set at 210 households. This was to be split equally among the (S, M, L) strata possibly for the following reasons:[7] From the ranges of the

[5]In the U.S. (and possibly other countries), a computer-based system has long been developed whereby the government provides users sub-samples of its huge statistical data banks while still safeguarding the confidentiality of individual respondent identities. If a similar system can be developed in the Philippines, the resources saved by not 'reinventing the wheel' in these times when household surveys have become *de rigueur* among micro-oriented social scientists, is almost inestimable.

[6]It should be mentioned that the target population includes households with residences inside but with farms outside the project area and households with farms inside but residences outside the project area. A 'control' area outside the project's command area was sampled also. In the paper the discussion is limited to households with both residences and farms inside the project area.

[7]These are mere conjectures since the writer was not involved in the designing of the survey.

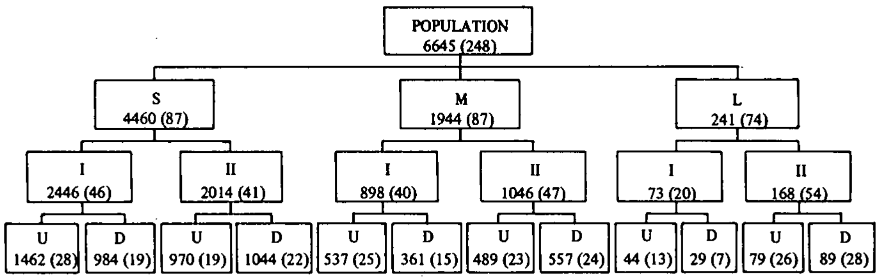

Chart 1A. Partial Stratification and Sample Allocation

Tenure	Rice 18 (6)	Corn 6 (2)	Coconut 56 (17)	Others 9 (3)
Full Owner	13 (4)	2 (2)	47 (13)	8 (3)
Part Owner	2 (0)	0	1 (0)	0
Tenant	3 (2)	0	3 (2)	1 (0)
Lessee	0	0	0	0
Amortizing Owner	0	0	0	0
Squatter	0	3 (2[a])	2 (0)	0
Other	0	1 (0)	3 (2)	0

[a]Missing observations; sample units cannot be located.

Chart 1B. Crop by Tenure Stratification and Sample Allocation, Unit II – Downstream landholdings, it is expected that the variances of the S and M strata will only be marginally different, but that of the L stratum should be many times higher. Proportional allocation will not take these variance deviations into account, but will in point of fact assign a very small sample to L. On the other hand, (Neyman's) allocation formula will likely lead to bigger allocation for L, but more likely less than 70. However, this is an 'analytic survey' and inferences concerning (differences between) stratum parameters may be in the research agenda, in which case a larger sample size in L would be preferable to neutralize the large variances there.[8] Thus, equal allocation may have been resorted to as a compromise (which was possibly a good decision).

The desire to have control of the allocation between Small, Medium and Large farms may be the main reason why first level stratification was by size of landholding; otherwise, one would expect stratification by Unit I and Unit II first.

After assigning equal sample sizes to the (S, M, L) strata, allocation was done proportionately (to the number of farm households) starting from the (I, II) strata and sequentially all the way down to the smallest strata (crop x tenure). Whenever the computed allocation for a crop x tenure stratum was one, a sample of size two was actually drawn; the reason must be so that a valid estimate of the sampling variance can be computed. *However, strata with zero computed allocations remained unsampled.* This raised the total sample size from an intended 210 to an actual 248 households, as seen in Charts 1A and 1B.

Simple systematic sampling was used to draw the sample households independently from each crop x tenure stratum. Data were collected entirely by interview with a one year recall period (two most recent crop seasons) for most variables.

---

[8]Like most sampling topics, sample allocation is almost exclusively seen in light of 'descriptive' surveys, i.e., when the main objective is to estimate parameters (means, totals, ratios) used to describe populations. Neyman's optimum formula, for example, yields an allocation that minimizes the sampling variance of means and totals. There is a need to rethink the allocation problem when the survey is primarily 'analytic', i.e., the principal aim goes beyond point estimation to testing hypothesis, fitting models, etc. In fact, the entire foundation for statistical inference from complex survey data has of late been under scrutiny by the best minds in the statistical profession.

C. Further Comments

This is one of the more elaborate sampling designs I have seen in years. It was obvious that its proponent had a clear understanding of the significance and benefits of investing on a good sampling frame.

It is also one of a few that did not make use of the village (barangay) as a sampling unit. Consequently, I would have preferred simple random sampling over systematic sampling of households for two reasons. Firstly, the practice of treating systematic samples like simple random samples, along with all its implied assumptions, would have been unnecessary. There is enough empirical evidence to support the assumption that these two procedures produce close enough estimates for means or totals. However, the same cannot be said about variance estimates. There is still no way to get around the fact that a systematic sample, no matter how large, is based on one random number, hence there is no valid and practical estimate for the sampling variance.

Secondly, with systematic sampling the households in a village can be listed, the sample identified and the respondents interviewed simultaneously in one pass. These are the three distinct steps to be done sequentially in simple random sampling. Without the villages as sampling units and with prior complete listing of households, this time saving appeal of systematic sampling became irrelevant.

There are two other aspects of the sampling design which were important enough to require a change in the statistical analysis of the data and a rethinking of the inferences to be made from such analysis.

The first is zero sample allocation in some of the strata. There were 220 crop x tenure strata; hence with 248 total sample households allocated proportionately, it was inevitable that some 138 strata will be unsampled. The logical implication of this is a diminution of the target population (the object of inferences) from 6,645 households to a sampled population (the real subject of inferences) of about 6,000 households. This problem could have been circumvented by collapsing tenure classifications from 7 to 3 and allotting a minimum of 2 sample households per stratum. This would result in a slightly higher total sample which is probably a good trade-off for the problems concomitant to zero allocation. As discussed in III.A, all estimation and other inferential procedures will have to be adjusted to reflect this difference between target and sampled populations.

The second has to do with trying to make the survey satisfy the requirements of different users, which is not easy to do especially if the requirements are brought to light after the data had been collected. Just prior to appraisal a conclusion was reached that the project actually is four subprojects – Hijo, Kipaliku, Libuganon and Manat – which would be appraised together for administrative economy and expediency, but whose water sources, hence irrigation construction and development, are completely independent. It follows also that the survey should have been designed to be four PBME benchmark surveys rolled into one. However, inasmuch as this was not clear during the planning stage of the survey (at least not to the NIA PBME Unit), the sampling design was not geared for independent estimation of subproject.[9] This problem is discussed further in III.B.

III. Data Analysis: How It Is, How It Ought to Be

A. Inference for Which Population?

It must be said that zero sample allocation in some subset of the target population is rarely seen or done on purpose. It happens occasionally when defective sampling frames, small sample allocation and high non-response come into play together. The implication is the same: When we define a population in statistics, it is understood that all its units are to be assigned *positive* probabilities of selection or inclusion in the sample. In statistical inference these probabilities serve as the bridge from the sample back to the population. With zero probabilities, there is nothing to build a bridge with. Thus, inferences from the sample relate only to the units with positive probabilities (the sampled population), no longer to the original target population. This is a limitation inherent with the modern scientific method, which governs us all.

---

[9]A small area, Tuganay, which was part of the proposed command area up to the feasibility study stage and from which 18 farm households were included in the sample, was eventually dropped from the project.

However, no one of course can put a sanction against efforts to see more than what the sample can really show. For instance, extraneous (outside the sample) information or assumptions may be added, which together with the sample may be used in an attempt to 'get at' the unsampled subsets. Sometimes semantics alone will do the trick, at least to a few consumers of research reports. For instance we come across titles that give the impression that the inferences pertain to the Philippines, but upon careful reading it turns out the study made use of a survey of a few villages in one province.

A look at the population and sample compositions by crop and by tenure status gives an idea of the extent of the unsampled subpopulations (Table 1). The amortizing owners were completely left out, along with a number of cells of part owners, lessees, squatters and other groups. Similar tabulations by subproject will show more unsampled cells with the result that no systematic estimation by tenure status is possible.

To illustrate the statistically correct estimation procedure, consider the L-II-D-Rice stratum which consists of (see Chart 1B)

	Population	Sample
Full Owners (F.O.)	13	4
Part Owners (P.O.)	2	0
Share Tenants (S.T.)	3	2
	18	6

Estimates are computed separately for each of the two sampled tenure strata. For example, estimates of total rice production (tons) and their corresponding sampling variances are[10]

F.O. total =	266.5;	variance	=	6,133.3
S.T. total =	43.5;	variance	=	992.2
	310.0			7,125.5

[10]The estimated total is the sample total multiplied by the reciprocal of the sampling rate (13/4 for F.O. and 3/2 for S.T.). Ignoring finite corrections and assuming systematic samples to behave like simple random samples, the sampling variance estimate of a tenure total is N^2s^2/n, where n/N is the sampling rate in the tenure class and s^2 is the usual sample variance.

Table 1. Target Population and Sample Composition, Davao III PBME Benchmark Survey

Tenure	Rice	Corn	Coconut	Others	Total
Full Owner	1,295 (55)	281 (16)	403 (42)	209 (22)	2,188(135)
Part Owner	182 (10)	37 (0)	21 (2)	10 (0)	250 (12)
Share Tenant	2,159 (41)	520 (18)	124 (8)	96 (2)	2,899 (69)
Lessee	507 (16)	4 (0)	1 (0)	4 (0)	516 (16)
Amortizing Owner	75 (0)	7 (0)	8 (0)	3 (0)	93 (0)
Squatter	150 (6)	73 (2)	32 (0)	40 (0)	295 (8)
Other	253 (6)	88 (0)	28 (2)	35 (0)	404 (8)
Total	4,621 (134)	1,010 (36)	617 (54)	397 (24)	6,645(248)

Note: Figures in parenthesis are sample sizes.

We cannot infer about the unsampled partly owned farms. Thus, 310 tons pertains to the *sampled population* of full owners and share tenants only; its variance estimate is obtained similarly by simply adding the individual stratum estimates. *Design-unbiased* estimates for bigger subpopulations can be obtained by simple aggregation of the smallest strata (tenure) estimates. This is done with rice area and production separately for all Unit II-Upstream and Unit II-Downstream farms. The results are shown in the first and third columns of figures in Table 2. In the next section we shall see that these areas are Hijo and Manat subprojects, respectively.

Others may try short-cuts. For instance, the Full Owners and Share Tenants samples may be combined to form a single sample of six households. Estimates from this and similarly formed aggregate samples may be used as building blocks for forming higher subpopulation estimates. However, the use of such short-cut procedures should be discouraged (unless sample allocation at the lower strata is strictly proportional to size). The problem is that estimates so obtained are biased, the worse, it is hard to guess in advance the probable magnitude and direction of the bias. For example, see the second and fourth columns of figures in Table 2. Note that the estimates of total area and production are all high compared to their design-unbiases counterparts. However, the biases of the variance estimates seem to be negative for Hijo but positive for Manat.

Table 2. Comparison of Design-Unbiased and Short-Cut (Biased) Estimates, Hijo and Manat Subprojects

Subproject/ Item	Estimate		Variance	
	Unbiased	Short-Cut	Unbiased	Short-Cut
Hijo/				
Rice Area (ha.)	2,711	2,835	143,430	108,160
Rice Production (tons)	7,873	8,430	1,793,074	1,126,547
Manat/				
Rice Area (ha.)	2,297	2,666	61,381	74,580
Rice Production (tons)	6,480	7,858	762,937	1,423,358

B. Consequence of Mismatch in Survey Design and Research Objective

As mentioned in II.C, the need for separate estimates for the individual subprojects – Hijo, Kipaliku, Libuganon and Manat – came to light only after the survey had been carried out. There is no problem for Hijo and Manat since by coincidence, these are the areas covered by Unit II-Upstream and Unit II-Downstream strata, respectively (see Chart 1A). Thus, separate estimates for these two subprojects can be computed directly as in the preceding section III.A. However, Kipaliku and Libuganon are interwoven in both Upstream and Downstream areas of Unit I. This is shown schematically in Chart 3 below which is a rearrangement of Chart 1.

In practice, either one of the following three steps is likely to to be taken. A decision is made that it is not possible to compute separate estimates for the the two subprojects. Or, simple unweighted averages may be used to build separate estimates. If $\bar{y}_{KU}$ and $\bar{y}_{KD}$ are the simple averages of Kipaliku sample observations and N_{KU} and N_{KD} are the total number of sampling units in the U and D strata, respectively, then $N_{KU}\bar{y}_{KU} + N_{KD}\bar{y}_{KD}$ may get proposed to estimate a Kipaliku total. However, the estimate of its sampling error cannot be computed (for roughly the same reasons mentioned in the next section).

A third statistically correct alternative exists.[11] The problem fits that of estimating parameters of subpopulations that cut across stratum boundaries. Suppose the target parameter is the total of Y, say rice production. Define a new variable $Y^{(K)}$ which takes on the actual value of Y for farms in Unit I that are in Kipaliku, otherwise it is zero. Likewise, let $Y^{(L)}$ by Y for all farms in Unit I that belong to Libuganon, or else it is zero. (For convenience, ignore the area Tuganay, which was eventually dropped from the appraisal report.) The identity

[11]Actually, the survey first came to the author's attention when the NIA PBME team asked for advice on the resolution of this problem. For details of the proposed technique, see for example W.G. Cochran (1977) *Sampling Techniques*, Third Edition, John Wiley, 35-38.

$$\Sigma Y = \Sigma Y^{(K)} + \Sigma Y^{(L)}$$

holds when the Σ extends over all farms in Unit I-Upstream, Unit I-Downstream, or the whole of Unit I. $\Sigma Y^{(K)}$ and $\Sigma Y^{(L)}$ are the desired Kipaliku and Libuganon totals. Hence, the estimation procedure in section III.A can be applied to $Y^{(K)}$ in computing the Kipaliku total in Unit I-Upstream $\hat{Y}_U^{(K)}$, and Unit I-Downstream, $\hat{Y}_D^{(K)}$, whose sum, $\hat{Y}^{(K)} = \hat{Y}_U^{(K)} + \hat{Y}_D^{(K)}$ is the estimate for Kipaliku. Similarly, $\hat{Y}^{(L)} = \hat{Y}_U^{(L)} + \hat{Y}_D^{(L)}$ is the estimate for Libuganon.

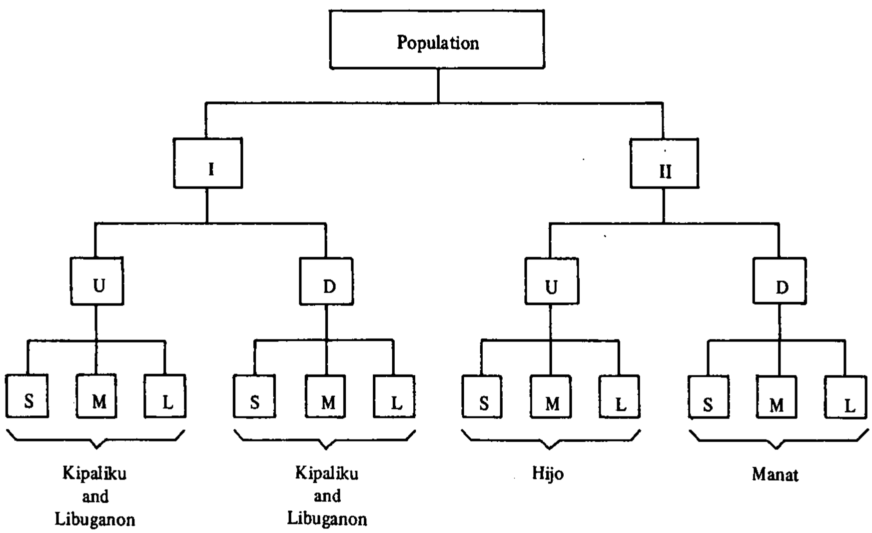

Chart 3. Relationship between Strata and Subprojects

From their definitions, it can be surmised that $Y(K)$ and $Y(L)$ will have considerably higher variances than the original Y. Another way of looking at this is that estimates would have been more precise if, say, Kipaliku = Upstream and Libuganon = Downstream. For example, we compare the direct Y estimates (obtained as in III.A)

$$\begin{aligned} \hat{Y}_U &= \text{Unit I-Upstream total,} \\ \hat{Y}_D &= \text{Unit I-Downstream total, and} \\ \hat{Y}_U + \hat{Y}_D &= \text{Unit I total,} \end{aligned}$$

with their corresponding numerical equivalents

$$\begin{aligned} &\hat{Y}_U(K) + \hat{Y}_U(L), \\ &\hat{Y}_D(K) + \hat{Y}_D(L), \text{ and} \\ &\hat{Y}(K) + \hat{Y}(L). \end{aligned}$$

The results for rice area and production are presented in Table 3. It is seen from the variance ratios that the latter group of estimates are only about 50 to 75 per cent as precise as the former. This is the price paid for failing to match strata with subpopulations of interest.

The above example shows in particular the importance of matching whenever possible, domains of study with strata and, in general, research objectives with survey design. The extra time spent in planning for choosing a more efficient sampling design should be more than compensated by the increase precision of estimates. It could also spell the difference between a simple analysis and a salvage operation.

C. The Effect of Ignoring the Sample Design When Estimating Frequency Distribution

With few exceptions, researchers treat survey samples as 'simple random samples when constructing frequency distribution tables; that is, they assign equal weights to individual observations. The problem with this procedure is that it does not take into account the fact that sampling rates among strata usually differ, sampling units may have

Table 3. Comparison of Variances of Numerically Equal Estimates, Unit I Stratum

Estimator l. h. s. = r. h. s.	Estimate	v(l. h. s.)	v(r. h. s.)	v(l. h. s.)/ v(r. h. s.)
Area (hectares)				
$\hat{Y}_U = \hat{Y}_U(K) + \hat{Y}_U(L)$	3,683	165,348	334,262	0.49
$\hat{Y}_D = \hat{Y}_D(K) + \hat{Y}_D(L)$	2,990	141,627	193,369	0.73
$\hat{Y}_U + \hat{Y}_D = \hat{Y}(K) + \hat{Y}(L)$	6,673	306,975	527,632	0.58
Production (tons)				
$\hat{Y}_U = \hat{Y}_D(K) + \hat{Y}_U(L)$	12,194	3,333,316	5,105,497	0.65
$\hat{Y}_D = \hat{Y}_D(K) + \hat{Y}_D(L)$	9,388	1,409,004	1,875,462	0.75
$\hat{Y}_U + \hat{Y}_D = \hat{Y}(K) + \hat{Y}(L)$	21,582	4,742,320	6,980,959	0.68

been picked with varying probabilities; or in sum, that the sampling design assigned unequal inclusion probabilities, hence unequal weights, to the population units.

Although this practice of invariably assigning equal weights to observations may not be peculiar to social science research, there seems to be greater awareness in the biological and natural sciences of the importance of proper experimental design and the crucial role it plays in the statistical treatment of the resultant data. Thus data from experiments laid out in completely randomized plots, randomized complete blocks, and latin square designs are analyzed differently. However, it appears that the connection between sampling design and proper statistical treatment of data in the social sciences have yet to be seen with comparable logical clarity. There are at least two reasons for this. One has to do with what people think a statistician does or supposed to do. A layman's stereotype statistician is one who goes through life crunching numbers, compiling and computing averages. Statisticians do simple arithmetic on voluminous data. In reality, data hold no lasting interest to a professionally trained statistician. What really matter to him are how you get them (data) and what you can do with them, and that the *what* is logically dependent on the *how*.

The other has to do with the way we ourselves present statistics. We have oversimplified *how* in mainstream classical statistics to mean simple random sampling almost exclusively, or more precisely, independent, identically distributed (i.i.d.) sampling. It is of course perfectly logical to give i.i.d. sample units equal weights. However, we may have neglected to make clear (not in textbooks or even journal articles) that the only time a sample from a finite population is i.i.d. is if the units are drawn with complete replacement, each time with equal probability for all population units. Thus we have failed to spread the truth that in actual practice, survey sampling almost always uses anything but i.i.d. sampling: Populations are stratified, sample allocation among strata is often disproportionate, units are drawn with probability proportional to some measure of size, sampling is done in stages and without replacement, etc. All these are departures from i.i.d. sampling.

Indeed, these two factors are greatly responsible for the current thinking that it takes only a few formal courses to be a statistical expert, hence the growing number of persons with inadequate statistical training who do statistical work. This phenomenon is universal. In the American Statistical Association, it has triggered talk about its threat to the statistician's identity and the need to prevent this by pushing for a strict licensing procedure as prerequisite to statistical practice. In his presidential address to that Association, H. O. Hartley said,[12]

> ". . . we must convince colleagues in other areas that they would be wasting their time dabbling in statistics in an inept do-it-yourself extravaganza . . . ."

R.A. Bradley's 1981 presidential address to that same group dwelt on the same problem.[13]

The computation of weighted or design-unbiased estimates of frequencies follows the usual procedure for estimating point parameters such as totals or means. As an illustration, consider the frequency distribution of area planted to rice among farm households in Hijo and Manat subprojects. Let us use the following intervals (in hectares): Under 2.67,[14] 2.67-6.66, over 6.66 and no area under rice. *In each of the lowest level strata* (tenure x crop), count the number of sample observations, N_1, N_2, N_3, N_4 that fall in these four intervals, where $N_1 + N_2 + N_3 + N_4 = N$ is the sample size in the stratum. With N as the total number of units in the stratum, the design-unbiased estimates of total households in the intervals are N_1 (N/n), N_2 (N/n), N_3 (N/n), N_4 (N/n), respectively. Finally, the design-unbiased estimates of total frequencies in Hijo, for example, are obtained by simply adding the estimates in like intervals through

---

[12] H.O. Hartley (1980) Statistics As a Science and As a Profession. *Jour. Amer. Stat. Associ.* 75. 1-7.

[13] R.A. Bradley (1982) The Future of Statistics As a Discipline. *Jour. Amer. Stat. Assoc.* 77, 1-10.

[14] Integers were not chosen as end-points because of an observed bunching of responses around 1, 2, 3, . . . hectares which can be a source of response error. This will be reported more fully in a forthcoming paper.

all the tenure x crop strata in Hijo.[15] The results for Hijo, Manat and Hijo + Manat are shown in Table 4. It is estimated that 711 out of 2,740 or 26 per cent of farm households in Hijo + Manat have no rice. Of those with rice, only slightly over 5 per cent have areas over 6.66 hectares and over three-fourths have areas less than 2.67 hectares. In addition to being unbiased, the sampling errors of these estimates can be easily computed (see footnote [15]).

In practice, what is commonly done is to simply count the number of sample observations that fall in designated intervals and draw inferences directly from these (Table 5). Note that the percentage of farms with no rice is now 42 per cent; among farms with rice, only one-half have areas under 2.67 hectares and 17 per cent have areas over 6.66 hectares.[16] Compare these with the unbiased 26 per cent, three-fourths and 5 per cent, respectively, shown in Table 4.

In statistical parlance, the design effect (deff) is ignored. Among others, an unweighted analysis does not make adjustment for the differences in sampling rates for the Large (74/241), Medium (87/1944) and Small (87/4460) farms (see Chart 1A). Thus, frequency estimates will be increasingly biased in favor of bigger areas. Furthermore, this bias will be present in all variables correlated with landholding, including household income distribution.

Because of the simplicity of a design-unbiased analysis and the risk of large biases with unweighted estimates, there really is no reason to perpetuate the current practice of using the latter method. All that is required is for the user of survey data to spend time to

---

[15]The estimates can be cast in the form of the usual weighted stratified point estimates. Let n_{ij} be the observed frequency in the interval j out of the n_i sample units in stratum i; $\hat{P}_{ij} = n_{ij}/n_i$ estimates the true proportion $P_{ij} = N_{ij}/N_i$, hence $N_i\hat{P}_{ij}$ estimates the true frequency N_{ij}. The sum through the strata, $\Sigma N_i\hat{P}_{ij}$, estimates the total frequency in the interval j, which when divided by N takes the form $\Sigma(N_i/N)\hat{P}_{ij} = \Sigma w_i\hat{P}_{ij}$ which is the usual stratified sample mean with sampling error given in any standard sampling book.

[16]These are unweighted estimates comparable to the simple average $\Sigma yi/n$ in estimating the mean of a variable Y. The latter is seldom considered in practice, since the unbiased weighted stratified mean $\Sigma w_i\bar{y}_i$ is popularly known, where $\bar{y}_i$ and w_i denote the sample mean and relative size of the i-th stratum, respectively. Sampling books, however, are either silent or not explicit about the problem of design-unbiased estimation of frequency distributions.

Table 4. Design-Unbiased Frequency Distribution of Rice Area, Hijo and Manat

Hectares	Estimated Total Households			Per Cent
	Hijo	Manat	Hijo + Manat	(with rice only)
Under 2.67	764	809	1,571	77.4
2.67 – 6.66	184	171	355	17.5
Over 6.66	62	41	103	5.1
No rice	300	411	711	–
Total	1,310	1432	2,740	–

Table 5. Unweighted Household Distribution of Rice Area, Hijo and Manat Samples

Hectares	Number of Households			Per Cent
	Hijo	Manat	Hijo + Manat	(with rice only)
Under 2.67	20	21	41	50.6
2.67 – 6.66	13	13	26	32.1
Over 6.66	9	5	14	17.3
No rice	26	34	60	–
Total sample size	68	73	141	–

learn how the survey was done. The only occasion that unweighted estimates may be acceptable is when the design is self-weighing – as when sample allocation is strictly proportional to the sizes of the strata and sampling is with equal probability (with or without replacement). Most of the major surveys in the Philippines do not come near this category. Until the early 1970's the Bureau of Agricultural Economics' Rice and Corn Survey (which was earlier called

Crop and Livestock Survey) used varying sampling rates among strata of barrios and the latter were drawn with very unequal probabilities (proportional to farm area). Later, when the survey was redesigned, all barrios with over 500 hectares of rice area were completely enumerated (100% sampling rate) which, if ignored at the analysis stage can be a serious mistake inference-wise.[17] In the National Census and Statistics Office's Integrated Survey of Household and its predecessors, the sample municipalities until the mid-70's were selected with probability proportional to the sizes (pps) of the their populations. In the current design, the household sampling rates for the different strata vary from 1/50 to 1/300.[18]

As mentioned previously, design-unbiased estimation of parameters (total, means, ratios) are well-known and are applied extensively in favor of unweighted estimates. This is not the case with estimating one-way or multi-way frequency distributions. In the preceding numerical example, we have seen the rather serious conssequence of ignoring the effect of disproportionate sampling rates among strata. The effect of pps sampling can be equally, if not more, serious. (This problem along with a numerical example in the context of parameter estimation was discussed in an earlier paper.[19]). For example, a sample of barangays drawn with pps of population will tend to have more urban units, hence unweighted estimates for variables positively correlated with degree of urbanization (e.g. income) will be positively biased. The bias will go in the other direction when units are drawn with pps of farm area since then the sample will tend to have proportionately more rural units than the sampled population.

---

[17] For details about the RCS design, see for example I.P. David (1966) Development of a Statistical Model for Agricultural Surveys in the Philippines. M.S. Thesis, U.P. at Los Baños Library; and B.D. Villanueva (1982) Measurement Errors in Rice Surveys. M.S. Thesis, U.P. at Los Baños Library.

[18] See for example NCSO, NEDA (1977) Integrated Survey of Households Bulletin, Series No. 48.

[19] I.P. David (1977) Analytic Use of Survey Data: Current Issues and Problems. Presented at the Conference on the New Household Economics sponsored by the Agricultural Development Council, Phil. Econ. Society and Phil. Agric. Econ. and Development Assoc., Manila.

IV. Concluding Remarks

Some years ago I was asked to design a sample survey for an evaluation of a government supported farm credit program. The study involved researchers in business administration, demography, economics and sociology. After a series of meetings to clarify the proposed research method and concomitant data requirements of each study leader, we commenced work on the sampling design. We figured we had to *make sure* that the sample had enough respondents with irrigated and non-irrigated farms, borrowers and non-borrowers, some who did and some who did not repay their loans small and large farms, etc., because data from these different groups were called for by the proposed analytical models and hypotheses to be tested. We were also aiming for as near a self-weighing design as possible so that simple unweighted analysis of the data (which social science researchers tend to do most of the time anyway) will provide statistically sound results. All these required a sampling frame which would allow barrios to be grouped into four strata, namely repayment rate (low, high) and by water control (mainly irrigated, mainly rainfed). The farm households within sample barrios were to be classified further according to size of farm (small, large) x (borrower, non-borrower).

The study leaders, who were working within their respective timetables, could not see the need for all these time consuming 'preparations' and judged the proposal to be too elaborate for their requirements. Why not just get a list of barrios in the province, draw a simple random sample from it, then go on and draw a simple systematic sample of households in each sample barrio? Something like this was eventually done.

We can find in this little incident many of the reasons for the prevailing sampling practice and use of survey data.

(1) A researcher's priority is high on testing hypotheses and verifying ideas within his particular field of expertise. If fresh data are required for him to be able to do so, he may be prone to take these in the quickest and most convenient manner. His priorities are expected and understandable; they are directly aligned with his training. If he can be blamed at all, it is in his assumption that

he is adequately trained to collect data also. But even this is partly our own making as explained below.

(2) There is a widening communication gap between survey statisticians and survey practitioners, as alluded to in III.C. As W.G. Cochran noted earlier,[20]

> ". . . sample survey theorists are seldom actively engaged in survey practice, and thus tend to write for other theorists . . ."

Sampling theory courses primarily for statistics majors abound. Thus, in an environment where survey practitioners are left to talk among themselves, it was inevitable that sampling schemes would soon be limited mostly to stratified random sampling and data sets invariably get a simple random sampling treatment. There is a need to talk, in mutually comprehensible language, about survey design being an exercise in *controlled selection:* The research objectives, analytical methods and other constraints that bind the researcher make some samples more preferred than others. Simple random sampling ignores this preference scale and assigns equal chances to each sample. But there is a whole array of sampling schemes that assign higher chances to *preferred* samples and lower likelihood of being stuck with *non-preferred* samples. The search and choice of schemes however, is played under certain probabilistic and statistical rules.

(3) The preceding two reasons have progressively limited the role of a survey statistician to that of a consultant. We can reverse this trend once we demonstrate convincingly, again in mutually understandable terms, that the statistician has as much to contribute to a survey-supported research undertaking if he instead works as collaborator.

---

[20]W.G. Cochran, in his discussion of T.M.F. Smith (1976) The Foundations of Survey Sampling: a Review. *Jour. Royal Stat. Society, Series A, 139,* 183-204.