UNITED NATIONS SECRETARIAT **ESA/STAT/AC.93/9**
Statistics Division **03 November 2003**

English only

Expert Group Meeting to
Review the Draft Handbook on
Designing of Household Sample Surveys
3-5 December 2003

D R A F T

Annexes*

---

* This document is being issued without formal editing.

Table of contents

Annexes

Annex One. Overview of sample survey design

1. Sampling is a technique by which a part of the population is selected and results from this fraction are generalised on the whole population from which the part or sample has been selected. In general there are two types of samples, namely probability and non-probability samples. Our focus in this handbook is on probability samples.

A.I.1. Sample design

2. At the outset it should be stressed that sample design cannot be isolated from other aspects of survey design and implementation. In general, sampling theory is concerned with how, for a given population, the estimates from the survey and the sampling errors associated with them are related to the sample size and structure. In practice sample design involves the determination of sample size, structure and takes into account costs of the survey.

Procedures of selection, implementation and estimation

- Each element in the population should be represented in the frame from which the sample is to be selected.
- The selection of the sample should be based on a random process which gives each unit a specified probability of selection.
- All and only selected units must be enumerated
- In estimating population parameters from the sample, the data from each unit/element must be weighted in accordance with its probability of selection.

Significance of probability sampling to large-scale household surveys

- It permits coverage of the whole target population in sample selection.
- It reduces sampling bias.
- It permits generalization of sample results to the population from which the sample is selected.
- It has been argued that it allows the surveyor to present results without having to apologise for using non-scientific methods (Kish, 1965).
- It allows the calculation of sampling errors, which are reliability measures.

Basic requirements for designing a probability sample

- The target population must be clearly defined.
- There must be a sampling frames or frames in case of multistage samples.
- The objectives of the survey must be unambiguously specified in terms of:

  a. Survey content (what do you want to find out?).
  b. Analytical variables (what estimates are envisaged?).
  c. Level of desegregation (e.g. do you need estimates or data at national, rural-urban, provincial, district, etc. levels?).

- Budget and field constraints should be taken into account.
- Precision requirements must be spelled out in order to determine the sample size.

3. Random selection of units reduces the chance of getting a non-representative sample. Randomisation is a safe way to overcome the effects of unforeseen biasing factors. The method of sample selection used depends to a greater extent on the sampling scheme being used. The more complex the sample designs the more demanding the selection procedures required.

Survey units and concepts

4. ***Elements***: Elements (units) of a population are units for which information is sought. They can be the elementary units comprising the population about which inferences are to be drawn. For example in a survey covering students at a school each student will be an element (unit). To facilitate data collection in a survey it is absolutely essential that elements should be well defined and physically easy to identify.

5. ***Population*** : The population is the aggregate of elements defined above. Elements are, therefore, the basic units that comprise and define the population. It is essential to define the population in terms of:

- Content, this calls for the definition of the type and characteristics of the elements which comprise the population.
- Extent refers to geographic boundaries as they relate to coverage.
- Time would refer to the time period to which the population refers.

6. ***Observational unit***: These are units from which the observations are obtained. In interview surveys they are called respondents. Reporting units are often the elements, as in surveys of attitude. Note that in some cases observational and reporting units may be different. For example in a survey of children under five, parents will normally give, as proxies, information pertaining to their children. In such cases the selected children will be observational units while parents are reporting units.

7. ***Sampling unit***: The sampling units are used for selecting elements into the sample. In element sampling each sampling unit contains one element, while in cluster sampling, for instance, a sampling unit comprises a group of elements called a cluster. For example an enumeration area (EA) which may be a first stage sampling unit contains a cluster of households. It is possible for the same survey to use different sampling units. A good example is multi-stage sampling which uses a hierarchy of sampling units (see chapter 2).

8. ***Sample units***: Selected sampling units may be termed sample units and the values of the characteristics under study for the sample units are known as sample observations.

9. ***Unit of analysis***: This is a unit used at the stage of tabulation and analysis. Such a unit may be an elementary unit or group of elementary units. It should be noted that the unit of analysis and the reporting unit need not necessarily be identical. The reporting unit in a survey of children under five may be parents while the unit of analysis will be children under five.

10. ***A sampling frame*** is used to identify and select sampling units into the sample and for making estimates based on sample data. This implies that the population from which the sample has to be selected has to be represented in a physical form. The frame ideally should have all sampling units belonging to the population under study with proper identification particulars. Frames should be exhaustive and preferably mutually exclusive. The commonly used types of frames in surveys are: list, area and multiple frames.

11. ***List frame*** contains a list of sampling units from which a sample can be directly selected. It is preferable that the frame should have relevant and accurate information on each sampling unit such as size and other characteristics. The additional information helps in designing and/or selecting efficient samples.

12. ***Area Frames*** multi-stage frames are commonly used in household surveys. In this connection the frame consists of one or more stages of area units. In a two stage sample design, for example, the frame will consist of clusters, which can be called primary sampling units (PSUs), in selected PSUs a list of households becomes the second stage frame. In general frames are needed for each stage of selection. The durability of the frame declines as one moves down the hierarchy of the units.

13. ***Area units*** cover specified land areas with clearly defined boundaries which can be physical features such as roads, streets, rivers rail lines, or imaginary lines representing the official boundaries between administrative divisions. Census enumeration areas (EAs) are usually established within the smaller administrative units that exist in a country. This facilitates the cummulation of counts for the administrative units as domains.

14. The frame or frames used for a household survey should be able to provide access to all the sampling units in the survey population such that every unit has a known and non-zero probality of selection in the sample. Access can be achieved by sampling from the frames usually through two or more stages of selection. The frame for the first stage of sampling must include all the designated sampling units. At subsequent stages of sample selection frames are needed only for the sample sample units selected at the preceding stage.

15. Sampling frame can be stored either on hardy copy or electronic media. It is possible to convert the electronic frame into a hardy copy and ***vice-versa***.

A.I.2. Basics of probability sampling strategies

16. The presentation covers simple random, systematic, stratified and cluster sampling.

A. Simple random sampling

17. Simple random sampling (SRS) is a probability sample selection method where each element of the population has an equal chance/probability of selection. Selection of the sample can be with or without replacement. This method is rarely used in large-scale household surveys because it is costly in terms of listing and travel. It can be regarded as the basic form of probability sampling applicable to situations where there is no previous information available on the population structure. SRS is attractive for being simple in terms of selection and estimation procedures (e.g. sampling errors).

18. While SRS is not very much used in practice, it is basic to sampling theory mainly because of its simple mathematical properties. Most statistical theories and techniques, therefore, assume simple random selection of elements. Indeed all other probability sample selections may be seen as restrictions on SRS, which suppress some combinations of population elements. SRS serves two functions:

- Sets a baseline for comparing the relative efficiency of other sampling techniques.
- It can be used as the final method for selecting the elementary units, in the context of the more advanced sampling techniques such as clustering and stratified sampling.

19. Considering a finite population of 100 house holds

$H_{1,}H_{2,-----}H_{i----------}H_{100}$

with income values $X_{1,}X_{2,} -—, X_i, ---, X_{100}$

The probability of any particular unit being selected is $1/100$

20. In drawing the sample, households can be numbered serially and using random numbers a sample, say, of size 25 can be selected. For equal probability selection method (EPSEM) f is the overall sampling fraction for the elements. Thus, $f = n/N$. If n =25 and N=100, the sampling fraction is $25/100 = 1/4$.

A.I.2.1. Sample selection

21. There are two common methods of sample selection, namely:

a. Simple random sampling with replacement (SRSWR).
b. Simple random sampling without replacement (SRSWOR).

22. It is better intuitively to sample without replacement as you get more information because there is no possibility of duplication of sample units.

a. Simple random sampling with replacement

23. SRSWR is based on a random selection from a population by replacing the chosen element in the population after each draw. The probability of selection of an element remains unchanged after each draw, and any selected independent samples are independent of each other. This property explains why SRS is used as the default sampling technique in many theoretical statistical studies. In addition, because the SRS assumption considerably simplifies the formulas for estimators, such as variance estimators, it is used as a reference.

Estimation

24. Given a sample of n units selected using SRSWR for which information on variable x, has been collected the mean, variance and population estimate are given by

Mean

$$\bar{x} = \frac{1}{n}\sum_{i}^{n} x_i = \frac{1}{n}[x_1 + x_2 + --- + x_n]$$

$x_1 = 24,\quad x_2 = 30,\quad x_3 = 27,\quad x_4 = 36,\quad x_5 = 31,\quad x_6 = 38,\quad x_7 = 23,\quad x_8 = 40,$
$x_9 = 25,\quad x_{10} = 32$

$$\bar{x} = \frac{24 + 30 + 27 + --- + 25 + 32}{10} = 30.6$$

$x_4 = 35$

Variance

$$\text{V}\ (\bar{x}) = \frac{S_x^2}{n}$$

where $$s^2 = \frac{1}{n-1}\sum_{i}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left[\sum_{i}^{n} x_i^2 - \frac{x^2}{n}\right]$$

and shows calculation $x^2 = \sum (x_i)^2$

b. Simple random sampling without replacement

25. The simple random sampling without replacement (SRWOR) is the most frequently used form of simple random sampling procedure. In this procedure, the selection process is continued until *n* distinct units are selected and all repetitions are ignored. This is the same as retaining the unit or units selected and selecting a further unit with equal probability from the remaining units in the population.

26. The following are some of the properties SRSWOR:

- It gives a fixed sample size.
- Results in equal probability of selection for every element/unit (EPSEM).
- Like in SRSWR the sample mean and variance are unbiased estimates of population parameters.

▪ ***Example***

27. Total number of primary schools in a region where 275 in 1983. A sample of 55 schools was selected without replacement during a pilot manpower survey. Figures below show number of employees (y_i) in each of the selected schools, using SRSWOR.

5	10	32	6	8	2
15	16	35	7	50	6.
2	6	47	20	20	6
7	6	35	6	16	2
21	2	48	4	15	2
7	5	46	6	7	
4	4	8	2	6	
7	2	7	8	2	
5	12	10	6	2	
2	40	7	7	19	

$$\sum y_i = 688$$

$$\sum y_i^2 = 18{,}182$$

The sample mean is

$$\bar{y} = \sum y_i / n = \ 688/55 \ = \ 12.5$$

The variance of the sample mean is

$$V(\bar{y}) = \ 1\text{-f} \ s_y^2 / n$$

Where *1-f* is the population correction factor and

$$s_y^2 = 1/n-1\left[\sum y_i^2 - \left(\sum y_i\right)^2 / n\right]$$

$$= 1/54\left[18{,}182 - (688)^2/55\right]$$
$$= 177.\ 33$$

i.e. $V(\bar{y}) = (1 - 55/275)177.3/55$

$$= \ 2.579$$

$Se(\ \bar{y}\ ) = \sqrt{2.579} \quad = 1.6$

A.I.2.2. Determination of sample size under SRS

28. At an early stage in the sample design, the size of the sample must be determined. At the outset it should be mentioned that the answer may not be easy. However, there are rational methods to tackle the problem. Let us assume simple random sampling. We want to avoid making the sample so small that the estimate is too inaccurate to be useful. On the

other hand we want to avoid taking a sample that is too large so that the estimate is supouriously more accurate than necessary taking into account costs. We therefore have to keep a balance between accuracy and costs.

29. The ***first step*** is to decide how large an error we can tolerate in the estimate. This demands careful thinking about the use to be made of the estimate and the consequences of a sizable error. The figure reached may be to some extent arbitrary.

30. The ***second step*** is to express the allowable error in terms of confidence limits. Suppose L is the allowable error in the sample mean and we are willing to take a **5% chance** that the error will exceed L. Implying that we want to be reasonably certain that the error will not exceed an amount $\pm L$. The 95% confidence limits computed from a sample mean, based on a large sample, assumed to be approximately normally distributed are:

$$\bar{x} \pm 2 \, {}^{s}\!/\!\sqrt{n}$$

If we put $L = 2 \, {}^{s}\!/\!\sqrt{n.}$, this gives for the required sample size, which is

$$n = {}^{4s^2}\!/\!{}_{L^2}$$

31. To use this relation we must have an estimate of the population standard deviation S. For practical reasons a good guess can be made from the results of the previous sample surveys of the a particular population or similar populations.

- ***Example***

N= 1,004 employees

32. Given that the average $(\bar{x})$ income per month of staff members of a particular firm is

$$\bar{x} = \$4{,}250$$

where $S^2 = \frac{1}{N-1}\sum_i^N (X_i - \bar{X})^2$ Variance of population elements = \$77,690

$$L = 1.5\% \text{ (Allowable error)}$$

What would be the appropriate sample size if we wanted to estimate the true mean emolument with $\pm \$64, (i.e. 0.015 \, x \$4{,}250)$, with a risk that the error will not exceed \$64?

$$n = \frac{4(77{,}690)^2}{(64)^2}$$

$$n = \frac{3{,}107{,}600}{4{,}096} = 758.69 \approx 759$$

If the allowable error is 2%, $n = \frac{3{,}107{,}600}{(85)^2} = 430.12 \approx 431$. It is always safe to round up in determining sample sizes.

A.I.2.3. Coefficient of variation

33. The coefficient of variation (CV) measures the precision of the estimator. In some situations it is useful to consider some relative measures instead of absolute measures of variation. The absolute measures such as standard error appear in the units of measurement of the variable and that can cause difficulties in some comparisons. The common relative measure of variation is the coefficient of variation, in which the unit of measurement is cancelled by dividing the standard error with the mean.

$$cv(\bar{x}) = \frac{se(\bar{x})}{\bar{x}}$$

- ***Example***

Given the mean as 13 and standard error as 3.55, the ***cv*** will be

$$3.55/13 = 0.27 \, or \, 27\%$$

34. Coefficients of variation are useful for variables that are always or mostly positive. Such occur frequently in surveys, especially as count data. Comparison of variability of these items becomes more meaningful when expressed in relative terms. For example the spread of a variable e.g. income between two countries can be best compared by the coefficients of variation rather than standard errors, because the latter will be measured by different currency units. In this case the coefficient of variation may provide a reasonable comparison of average income because of the absence of units of measurement.

A.I.2.4. Confidence interval

35. In general if t is unbiased and normally distributed, the interval $\{t - kd(t), t + kd(t)\}$ is expected to include the population parameter in P $\%$ of the cases.

36. For example a 95 per cent confidence interval strictly means that, if we assert that the true unknown mean q lies in this interval, then we shall be correct (on the assumptions made) for 95 per cent of the time. General expression for the confidence limits for the mean of normal distribution based on a reasonably large sample (n>30) is given by $\bar{x} \pm kse(\bar{x})$. Thus 95 per cent range excludes 5 per cent probability for which k =1.96.

- ***Example***

37. Suppose an estimate of proportion of persons with a particular ailement it is 0.7 and the standard error has been calculated to be 0.02.

The 95% confidence interval is given as

$$.07 \pm (1.96 x 0.02), \quad \text{or} \quad 0.7 \pm 0.039$$

This means that the chances are 95 in 100 that the population proportion of persons with the particular ailment is in the range 0.66 to 0.74.

B. Systematic sampling

38. Systematic sampling is a probability sample selection method in which the sample is obtained by selecting every ***kth*** element of the population where ***k*** is an integer greater than 1. The first number of the sample must be selected randomly from within the first ***k*** elements. The selection is done from an ordered list.

39. This is a popular method of selection especially when units are many and are serially numbered from 1 to N. Suppose that *N* the total number of units is an integral multiple of the required sample size n and k is an integer, such that N = nk, a random number is selected between 1 and k. Let us suppose 2 is the random Start, then the sample will be of size *n*, the units serially numbered as follows:

$$2, 2 + k, 2 + 2k, \ldots, 2 + (n - 1)K$$

40. It will be observed that the sample comprises of the first unit selected randomly and every k^{th} unit, until the required sample size is obtained. The interval k divides the population into clusters or groups. In this procedure we are selecting one cluster of units with probability *1/k*. Since the first number is drawn at random from 1 to *k*, each unit in the supposedly equal clusters gets the same probability of selection l/k.

A.I.2.5. Linear systematic sampling

41. If N, the total number of units, is a multiple of n, thus if N = nk, is the sample size and k is a sampling interval, then the units in each of the possible systematic samples is n. In such a situation the system amounts to categorising the *N* units into *K* samples of n units each and selecting one cluster with probability *l/k*. When $N = nk$, $\bar{y}$ is unbiased estimator of the population mean $\bar{Y}$. On the other hand when N is not a multiple of k, the number of units selected using the systematic technique with the sampling interval equal to the integer nearest to N/n may not necessarily be equal to n. Thus when N is not equal to nk the sample sizes will differ, and the sample mean is a biased estimator of the population mean.

Figure A.1.: Linear systematic sampling

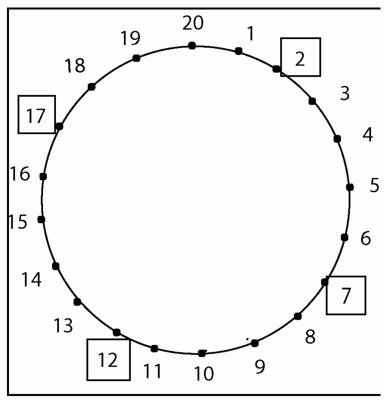
1	2	**3**	4	5	6	7	**8**	9	10
11	12	**13**	14	15	16	17	**18**	19	20

Random start is 3, $N = 20$, $n = 4$, and K=5 thus 3, 8, 13 and 18 are selected.

A.I.2.6. Circular systematic sampling

42. We note that in linear systematic sampling the actual sample size can be different from the desired and the sample mean is biased estimator of the population mean when N is not a multiple of n. However, a technique of circular systematic sampling overcomes the above mentioned limitation. In this method of selection you assume the listings to be in a circle such that the last unit is followed by the first. A random start is chosen from I to N. You then add the intervals K until exactly n elements are chosen. If you come to the end of the list, you continue from the beginning.

Figure A.2.: Circular Systematic Sampling



N = 2
n = 4, k = 5

Random start is 7
In the above case 7, 12, 17, and 2 are selected.

A.I.2.7. Estimation in systematic sampling

43. For estimating the total, the sample total is multiplied by the sampling interval.

$$\hat{Y} = k \sum y_i$$

Estimate of the population mean is

$$\bar{y} = k \sum y_i / N$$

44. Estimation of variance is intricate in that a rigorous estimate cannot be made from a single systematic sample. A way out is to assume that the numbering of the units is random in such a case a systematic sample can be treated as a random sample. The variance estimate for the mean is therefore given by

$$V(\bar{y}) = 1/n(1 - n/N) \sum s^2$$

where $s^2 = 1/n - 1 \sum (y_i - y)^2$ and y = $\sum y_i$

45. A rigorous estimate of unbiased variance from a systematic sample can computed by selecting more than one systematic sample from a particular population.

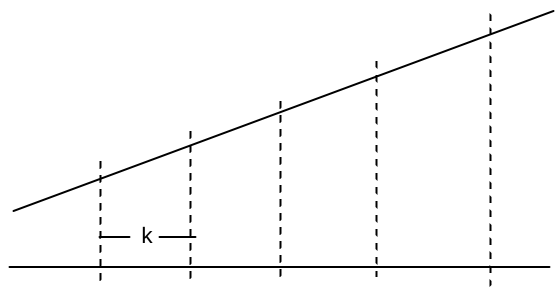
A.I.2.8. Advantages of using systematic sampling

46. The selection of the first unit determines the entire sample. This aurgurs well for field operations as ultimate sampling units can be selected in the field by enumerators as they list the units.

47. The sample is spread evenly over the population when units in the frame are numbered appropriately. However, the sample estimate will be more precise if there is some kind of trend in the population.

48. Systematic sampling provides implicit stratification; it can be regarded as stratification of the population into implicit stratum.

Figure A.3. Monotonic linear trend

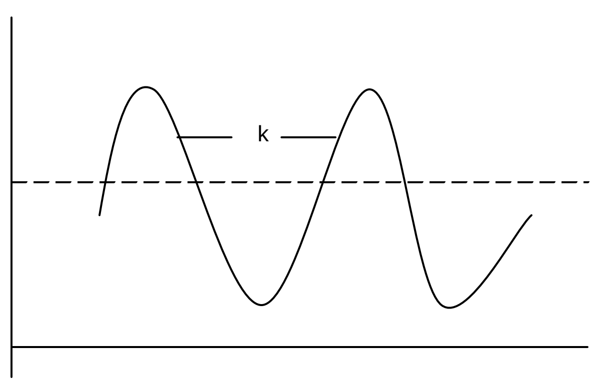


A.I.2.9. Disadvantages of Systematic Sampling

49. If there is periodic variation in the population, systematic sampling can yield results that are either under-estimates or over-estimates. In such a case, the sampling interval falls in line with the data. For example, if you are studying transport flow for 24 hours on a busy street in a cit y, if your interval falls on pick hours, therefore you will consistently get high figures and you will not get representative results.

50. The selection method is prone to abuse by some enumerators/ field staff.

Figure A.4. Periodic fluctuations



51. Strictly speaking, you cannot obtain a rigorous estimate of variance from a single systematic sample.

- ***Example***

52. There are 180 primary schools in a county area having an average of 30 or more people under the age of 21 per class. A sample of 30 schools was drawn using systematic sampling with an interval of k = 6.

Number of people under the age of 21 (yj) in the 30 selected villages.

60	200	45	50	40	79	35	41	30	120
300	65	111	120	200	42	51	67	32	40
46	55	250	100	63	90	47	82	31	50

$$\sum y_i = 2{,}542$$

Estimated number of students

$$\hat{Y} = k\sum y_i = 6 \text{ x } 2542 = 15,\ 252$$

Average number of students per farm

$$\bar{y} = k\sum y_i \big/ N = 6x2542/180 = 84.7$$

The variance of the sample mean will be calculated on the basis of the assumption that the numbering of farms is random.

$$\text{V}(\ \bar{y}\ ) = \text{1-f } \ s_y^2 \big/ n$$

where $s_y^2 = \dfrac{1}{n-1}\left\{\sum y_i^2 - \left(\sum y_i\right)^2 \big/ n\right\}$

$= 1/29(3{,}48700 - 215{,}392.13)$
$= 4{,}596.8$

therefore $\text{V}(\bar{y}) = (0.833)\,(153.227)$
$= 127.64$

$\text{Se}(\bar{y}) = \sqrt{127.64}$
$= 11.30$

C. Stratification

53. Stratified sampling is a method in which the sampling units in the population are divided into groups called strata. Stratification is usually done in such a way that the population is sub-divided into heterogeneous groups which are internally homogeneous. In general when sampling units are homogeneous with respect to the auxiliary variable termed stratification variable, the variability of strata estimators is usually reduced. Further there is considerable flexibility in stratification in the sense that the sampling and estimation procedures can be rightly different from stratum to stratum.

54. In stratified sampling, therefore, we group together units/elements which are more or less similar, so that the variance d_h^2 within each stratum is small, at the same time it is essential that the means $\left(\bar{x}_h\right)$ of the different strata are as different as possible. An appropriate estimate for the population as a whole is obtained by suitably combining stratum-wise estimators of the characteristic under consideration.

A.I.2.10. Advantages of stratified sampling

55. The main advantage of stratified sampling is the possible increase in the precision of estimates and the possibility of using different sampling procedures in different strata. In addition stratification has been found useful in the following situations:

- In case of skewed populations since larger sampling fractions may be necessary for selecting from the few large units, resulting in giving greater weight to few extremely large units for reducing the sampling variability.

- When a survey organization has several field offices in various regions into which the country has been divided for administrative purposes it may be useful to treat the regions as strata, so as to facilitate the organization of fieldwork.
- When estimates are required within specific margins of error, not only for the whole population, but also for certain sub-groups such as provinces, rural or urban, gender, etc. Through stratification such estimates can conveniently be provided.

56. If the sampling frame is available in the form of sub-frames, which may be for regions or specified categories of units, it may be operationally convenient and economical to treat sub-frames as strata for sample selection.

57. Steps followed in stratified sampling:

- The entire population of sampling units is divided into internally homogeneous but externally heterogeneous sub-populations.
- Within each stratum, a separate sample is selected from all sampling units in the stratum.
- From the sample obtained in each stratum, a separate stratum mean (or any other statistic) is computed. The strum means are then properly weighted to form a combined estimate for the population.
- Usually proportionate sampling within strata is used when overall, e.g. national estimates are the objective of the survey and the survey is multipurpose.
- Disproportionate sampling is used when sub-group domain have priority, in cases where estimates for sub-national areas are wanted with equal reliabilities.

Notations

a. Population values

For H strata, total number of elements in each stratum will be denoted by $N_1\ N_2$ _ _ _, N_h _ _ _ N_H such information is usually unknown.

Total population value is $\sum_{h}^{H} N_h = N$

b. $\overline{X}_{hi} = \frac{1}{N}\sum_{i}^{N_h} X_{hi} = \frac{X_h}{N}$

where X_{hi} is the value of the h_{th} element in the h_{th} stratum X_h is the sum of the h_{th} Stratum.

A.I.2.11. Weights

58. The weights generally represent the proportions of the population elements in the strata and $W_h = N_h / N$

$$\text{So, } \sum W_h = 1$$

$$S_h^2 = \frac{1}{N-1}\sum_{i=1}^{N_h}\left(X_{hi} - \overline{X}\right)^2$$

A.I.2.12. Sample values

a. For H strata, the sample sizes in each stratum can be denoted by $n_1, n_2 --- n_n$ where $\sum n_h = n$ the total sample size

b. x_{hi} is the sample element i in stratum h

c. $\overline{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$

d. $\overline{x}_{st} = \sum W_h \overline{x}_h$

e. $f_h = \frac{n_h}{N_h}$ is the sampling fraction for the stratum.

Variance of n_h element in the h_{th} stratum

$$v(\overline{x}_h) = \sum\left[1 - \frac{n_h}{N_h}\right]\frac{s_h^2}{n_h}$$

Where s_h^2 is the element variance for the h_{th} stratum and is given by

$$s_h^2 = \frac{\sum \left(x_{hi} - \bar{x}_h\right)^2}{\left(n_h - 1\right)}$$

The variance of sample mean is given by

$$V\left(\bar{x}_{st}\right) = \sum W_h^2 \left(1 - f_h\right) \frac{s_h^2}{n_h}$$

A.I.2.13. Proportional allocation

59. Proportional allocation in stratified sampling involves the use of a uniform sampling fraction in all strata. This implies that the same proportion of units is selected in each stratum. For example if we decide to select a total sample of 10 percent it means that we shall select 10 percent units from each stratum. Since the sampling rates in all strata are the same, the sample elements selected in the sample will vary from stratum to stratum. Within each stratum the sample size will be proportionate to the number of elements in the stratum.

60. In this case the sampling fraction is given by $f_h = n_h / N_h = \frac{n}{N} = f$ implying an EPSEM design.

Sample mean

$$\bar{x}_{st} = \sum W_h \overline{x_h}$$

Variance of the overall mean is

$$v\left(\bar{x}_{st}\right) = \frac{\left(1 - f\right)}{n} \sum W_h s_h^2$$

A.I.2.14. Optimum allocation

61. The method of disproportionate sampling involves the use of different sampling rates in various strata. The aim is to assign sampling rates to the strata in such a way as to obtain the least variance for the overall mean per unit cost.

62. In using this method the sampling rate in a given stratum is proportional to the standard deviation for that stratum. This means that the number of sampling units to be selected from any stratum will depend not only on the total number of elements but also on the standard deviation of the characteristics used as an auxiliary variable.

In optimum allocation, the notion of a function is also introduced. For example

$$C = C_o + \sum c_h n_h$$

where c_o is the fixed cost.

c_h is the cost of covering the sample in a particular stratum.

63. In many situations we may assume that c_h is a constant in all strata. Therefore, for our purpose we shall consider the Neyman's allocation.

Where c_h is constant and $n = \sum n_h$ noverall sample size is fixed.

The number of units to be selected within a stratum is given by

$$n_h = \frac{W_h s_h n}{\sum W_h s_h} \quad \text{or} \quad n_h = \frac{N_h s_h . n}{\sum N_h s_h}$$

Variance is given by

$$v(\bar{x}_{st}) = \frac{\left(\sum W_h s_h\right)^2}{n} - \frac{1}{N} \sum W_h s_h^2$$

64. The second term on the right is a finite population correction factor which may be dropped if you are sampling from a very large population, thus if the sampling fraction is small.

A.I.2.15. General observations

- Population values S_h and C_h are generally not known, therefore estimates can be made from previous or pilot sample surveys.
- Disproportionate allocation is not very efficient for selecting proportions.
- There may be conflicts on variables to optimize in the case of multi-purpose survey.
- In general, optimum allocation results in the least variance.

- ***Example***

The total number of primary schools in a province is 275. A sample of 55 schools is selected and stratified on the basis of number of employees.

Stratum	Number of employees per selected schools (y_{hi})	Total number of schools in each stratum (N_h)	Selected No. of schools by stratum		W_h	s_h^2	s_h	$W_h s_h$	$W_h s_h^2$
			Proportional allocation (n_h)	Optimum allocation (n_h)					
1	2,4,2,2,4, 2,2,4,2,2, 2,2,2,2,5,5	80	16	8	0.2909	1.663	1.289	0.3750	0.48
2	7,7,7,6,8, 7,7,6,7,6 6,8,6,7,8, 6,7,6,6,6	100	20	6	0.3636	0.537	0.733	0.2665	0.19
3	10,12,10,15, 21,16,20,20, 16,19,15	55	11	18	0.2000	15.564	3.945	0.7890	3.11
4	32,35,35,48, 46,47,50,40	40	8	23	0.1455	48.836	6.989	1.0169	7.10
		275	55	55	1.0000			2.4474	10.8

N = Total number of primary schools

n = total number of primary schools in the whole sample

N_h = Sample size of the h_{th}

A.I.2.16. Determination of within stratum sample sizes

a. Proportional allocation

$\frac{n}{N} = f$ which is the overall sampling fraction applied to the total number of units in the stratum in our example above, $f = \frac{55}{275} = 0.2$ or 20% for the distribution of sample sizes see column 4 in the table e.g. $n_h = 0.2x80 = 16$

b. Optimum allocation

The formula for obtaining sample sizes for different strata is given by

$$n_h = \frac{W_h s_h}{\sum W_h s_h}(n)$$

for example $n_h = \frac{0.3750}{2.4474} x55 = 8$

The rest of the results are given in column (5) in the table.

Example of how to compute variance based on proportional allocation and optimum allocation

a. Proportional allocation:

$$V(\bar{y}_{prop}) = \frac{1-f}{n}\sum w_h s_h^2$$

$$= \frac{(1-0.2)}{55}(10.898) = 0.158$$

b. Optimum allocation:

$$V(\bar{y}_{opt}) = \frac{\left(\sum w_h s_h\right)^2}{n} - \frac{1}{n}\sum w_h s_h^2$$

$$= \frac{(2.4474)^2}{55} - \frac{10.898}{275} = 0.0693$$

c. In general

$$v(\bar{x}_{st})OP \le v(\bar{x}_{st})PROP \le v(\bar{x}_{st}) \le (\bar{x}_{st})SRS$$

D. Cluster sampling

65. The discussions in the previous sections have so far been about sampling methods in which elementary sampling units were considered as arranged in a list from a frame in such a way that individual units could be selected directly from a frame.

66. In Cluster Sampling, the higher units e.g. enumeration areas (see chapter 2) of selection contain more than one elementary unit. In this case, the sampling unit is the cluster.

67. For example, to select a random sample of households in a city a simple method is to have a list of all households. This may not be possible as in practice there may be no complete frame of all households in the city. In order to go round this problem, clusters in the form of blocks could be formed. Then a sample of blocks could be selected, subsequently a list of households in the selected blocks made. If need be, in each block a sample of households say 10% could be drawn.

A.I.2.17. Some reasons for using cluster sampling

a. Clustering reduces travel and other costs of data collection.
b. It can improve supervision, control, follow-up coverage and other aspects that have an impact on the quality of data being collected
c. The construction of the frame is made cheaper as it is done in stages. For instance in multi-stage sampling discussed in chapter 2 a frame covering the entire population is required only for selecting PSUs i.e. clusters at the first stage. At any lower stage, a frame is required only within the units selected at the preceding stage. In addition, frames of larger and higher stage units tend to be more durable and therefore usable over longer period of time. Lists of small units such as households and particularly of people tend to become obsolete within a short period of time.

d. There is administrative convenience in the implementation of the survey.

68. In general we should note that in comparing a cluster sample with an element sample of the same size, we shall find that in cluster sampling the cost per element is lower owing to lower cost of listing and/or locating of elements. On the other hand, the element variance is higher due to irregular homogeneity of elements (intra-class correlation) in the clusters. We illustrate the basic cluster sampling by considering a single stage design (multi-stage designs are presented in chapter 2).

A.I.2.18. Single stage cluster sampling

69. In a particular district, it may not be feasible to obtain a list of all households, and then select a sample from it. However, it may be possible to find a list of villages prepared during a previous survey or kept for administrative purposes. In this case we would obtain a sample of villages, then obtain information about all the households in the selected villages. This is a single-stage cluster sampling design because after a sample of villages has been selected all units in the cluster, in this case households, are canvassed.

70. Sample selection under clustering can be illustrated as follows:

Assume that from a population of EAs (clusters) a sample is selected with equal probability. For a single stage cluster sampling, all households from the EAs would be included in the sample.

Given that
$A =$ Total number of clusters
$B =$ Total number of households in the cluster
$a =$ A sample of clusters

i.e. $aB =$ n elementary units in the total sample
$AB = N$

71. The probability of selecting an element with equal probability is given by

$$\frac{a}{A} \; x \; \frac{B}{B} = \frac{n}{N} = f$$

where N is the number of elementary units and f is the sampling fraction. In this case the probability of selection is simply $\frac{a}{A}$.

A.I.2.19. Sample mean and variance

$$\bar{y} = 1 \Big/ aB = \sum_{a=1}^{a} \sum_{b=1}^{B} \bar{y}_{ab} = 1/a \sum_{a=1}^{a} \bar{y}_a$$

The sample mean is unbiased estimate of the population mean:

$$E(\bar{y}) = 1/A \sum_{a=1}^{a} \bar{y}_a = \bar{Y}$$

72. In fact because the sample size is fixed ($aB = n$) and the selection is of equal probability then the mean ($\bar{y}$) is unbiased estimate of the population mean $\bar{Y}$.

73. If the clusters are selected using a simple random selection the variance can be estimated as follows:

$$V(\bar{y}) = (l\text{-}f)\ s_a^2$$

$$\text{where } s_a^2 = 1/a - 1 \sum_{a=1}^{a} (\bar{y}_a - \bar{y})^2$$

74. It is important to note that the values yare without sampling error as they are based on the values of the entire B elements within the clusters. In this case the variance of the sample mean is due to the variance between the cluster means.

Annex Two: Planning and execution of surveys

A.II.1. Planning of surveys

1. For a survey to yield desired results, there is need to pay particular attention to the preparations that precede the field work. In this regard all surveys require careful and judicious preparations if they have to be successful. However, the amount of planning will vary depending on the type of survey, materials and information required. It is therefore not surprising that many of the problems faced by researchers and survey organisations emanate from poor planning.

A.II.1.1. Objectives of a survey

2. It is helpful and imperative that the objectives of a survey should be clearly spelled out. There should be a clear statistical statement on the desired information, giving a clear description of the population and geographical coverage. It is also necessary at this stage to stipulate how the results are going to be used. The given fixed budget of the survey should guide the survey statistician in tailoring the objectives. Taking due cognizance of the budgetary constraints will facilitate the successful planning and execution of the survey.

3. In some cases objectives of surveys are not explicitly stated. For instance, a survey organization may be called upon to carry out a study on the activities of the informal sector. If the purpose is not clearly stated, it is for the statistician or researchers to define the informal sector, outlining in detail, particular economic activities that may reflect the requirements of the sponsoring agency. It should be mentioned that a survey which has ambiguous and vague objectives is very much susceptible to high nonsampling errors.

4. Some of the surveys conducted by survey organisations have clear objectives. For example, the 1983, Zambian Pilot Manpower Survey had the following objectives:

   a. To collect information on the size and composition of currently working population in the formal sector.
   b. To assess manpower demand and supply.
   c. To serve as a basis for making manpower projections for particular occupations.
   d. To assist in planning for the expansion of education in fields which are crucial to economic development.

5. It should he noted that clearly stated objectives help to form a basis of questions for which statistical answers are required.

A.II.1.2. The universe

6. When planning to carry out a survey, it is necessary to define the geographical areas to be covered and the target population. In a household Income and Expenditure Survey, for instance, the survey may cover the urban areas and perhaps exclude rural areas.

7. In defining the universe, the exact population to be sampled should be identified. In the above survey the universe of first stage units would be enumeration areas (EAs) in urban areas, at the second stage households in selected EAs.

8. We should note that in practice, however, the target population is somewhat smaller than the population forming the universe. This means that the target population may be different from the survey population. It is usual to restrict the target population for a number of reasons. For instance, in some surveys, some military households in barracks may be excluded from the survey. In labour-force surveys, children below a specified age may be shown as members of households surveyed, but would not be part of the labour-force.

9. It is important to note that when the actual population differs from the target population, the results will apply to the particular population from which a sample was drawn. As discussed in chapter 4, comprehensive and mutually exclusive frames should be constructed for every stage of selection.

A.II.1.3. Information to be collected

10. From the list of questions requiring statistical answers, a list of items that could provide factual information bearing on issues under investigation can be produced. It is always important to bear in mind that some of the required data could be available from existing sources. In producing the list of items, provision should be given for the inclusion of supplementary items that are correlated with the main items. In a survey of Employment and Earnings, for example, supplementary information on age, sex and education may be canvassed. Such information would give additional insight into related questions and enrich the analysis.

11. We may add that a tabulation plan should be produced at the time of planning the survey. The blank tables should be circulated for comments and improvement.

A.II.1.4. Survey budget

12. The survey budget indicates the financial requirements of the survey which is to be conducted. The budget shows cost of personnel, equipment and all other items of expense. if there is a pre-determined ceiling (which is usually the case) of funds available, the overall survey budget should be within the pre-determined framework. It is also advisable to follow the general guidelines of the financing agency in preparing the budget. This may facilitate the approval of the budget estimates. If there is need to depart from the Regulations, authority should be sought from the relevant organization. The financial requests of the survey should be prepared at an early stage. In general, the budget will depend largely on the survey design, precision required and geographical coverage.

13. In summary the survey budget may include the following items:

- Personnel costs: ad-hoc enumerators, supervisors and other field **staff** etc.
- Training costs.
- Travel and transportation costs.
- Cost of supplies, printing and data processing.

14. In addition to the items listed above, vehicles and other equipment may be needed, therefore should be budgeted for.

15. The statistical organization should mount an effective and concerted defence of the budget indicating the usefulness of the survey results to users. Such an approach is necessary because there are many competing demands for limited available resources.

16. It is essential that an effective cost control system is established in the organization that is conducting the survey. In most large scale census and survey operations, chances are high of losing control of monitoring the disbursement of funds once field work starts. In such circumstances a large amount of funds tend to be channelled in areas unrelated to the major survey operations. Judicious cost control helps to monitor actual expenditures in relation to estimated costs and actual work accomplished. It is imperative that management responsible for the survey should ensure accountability of funds. This greatly enhances the credibility of the survey organization.

A.II.2. Data collection methods

17. There are a number of methods used in data collection, among them, direct observation and measurement; mail questionnaire; telephone and personal interview.

18. Direct observation and measurement: is the most ideal method as it is usually more objective. It is free from memory lapse and subjectivity of both respondents and enumerators. Examples of areas where direct observation has been used are:

a. Some aspects of food consumption surveys.
b. Price collection exercises, where enumerators can purchase the produce and record prices.

19. This method though useful has a snag of being expensive both in terms of resources and time. In most cases, enumerators have to use some equipment. Experience has shown that the method of direct observation and measurement tends to be useful and practical when the sample sizes or populations are relatively small.

20. Mail questionnaires: the use of mail questionnaires is fairly cheap and quick. The major cost component being postage. After the questionnaire is designed and printed it is mailed to respondents. In this case the respondents are assumed to be literate as they are expected to fill the questionnaire on their own. This may be an erroneous assumption especially in developing countries where literacy levels are still low. The major weakness of this method is the high non-response rates associate with it. This may be due to the complexity of questionnaires used. However, apathy cannot be completely ruled out, in some cases there is good questionnaire response but high item nonresponse.

21. In trying to improve the response rate, some reminders have to be sent to non-respondents. However, it is advisable to select a sub-sample of the nonrespondents and cover them by the personal interview method. This may be necessary because the characteristics of the non-responding units may be completely different from those that responded. In this case the responding and nonresponding units are treated as two strata. In order to increase the response rate, the mailed questionnaires should be attractive, short and as simple as possible. Enclosing stamped and addressed returns may help to improve the response rate.

22. Here is a summary of some of the advantages and disadvantages of mail questionnaire surveys:

Advantages:

a. It is cheaper.
b. Sample can be widely spread.
c. Interviewer bias is eliminated
d. It is quick.

Limitations:

a. Nonresponse is usually high.
b. The answers to the questions are taken at their face value as there is no opportunity to probe.
c. If it is an attitude survey, it is difficult to ascertain whether the respondent answered the questions unaided.
d. The method is useful only when the questionnaires are fairly simple, therefore, it may not be an ideal method for complex surveys.

23. In order to use this method, there must be an up to date frame. The addresses of the respondents should be up-to-date and the survey organization must be convinced that respondents in a particular survey are capable of completing the questionnaires on their own.

24. <u>Personal interview method</u>: this method is the most common in collecting data through large scale sample surveys in developing countries. Apart from the usually high response rate resulting from personal interviews, the method is appropriate because of the prevailing high illiteracy rates in some of these countries. The method entails enumerators going to selected respondents collecting information by asking questions. The main advantage of this method is that the enumerators can persuade (through motivation) respondents to answer questions and can explain the objectives of the survey. Further, in using the personal interview method, there is scope for collecting statistical information on conceptually difficult items which are likely to yield ambiguous answers in a mailed questionnaire.

25. The following are some of the disadvantages in using the personal interview method:

a. Different enumerators may give different interpretations to the questions, thereby introducing bias in the survey results as very few enumerators consistently refer to the instructions manual.
b. In the process of probing, some enumerators may suggest answers to respondents.
c. Personal characteristics of enumerators may influence attitudes of respondents, for example, age, sex at times even race.
d. Enumerators may read questions wrongly because of the divided attention of interviewing and recording.

26. The following points should be taken into consideration when asking questions to respondents:

a. The enumerator should clearly understand the purpose of each question as explained in the enumerators' manual. It is important that enumerators constantly refer to the manual;
b. Experience has shown that it is better for the enumerator to follow the sequence of questions in the questionnaire. In most questionnaires careful thought is given in the ordering of questions; taking into consideration, motivation of respondents, linkage of topics, facilitating memory of the respondent's past events, and taking care of sensitive questions;
c. Enumerators should by all means refrain from suggesting answers to respondents;
d. All questions should be asked. In this way, item nonresponse is minimised. Further, no item in the questionnaire should have a blank space unless it satisfies the skip pattern. If a question is not relevant to a particular respondent, then a comment should be included. Such an approach assures the researcher that all questions included in the questionnaire have been administered.

A.II.3. Questionnaire design

27. It is advisable to design questionnaires at the time of planning for the survey. If the questionnaires have to be mailed to respondents, they have to be attractive and simple. This may increase the response rate. On the other hand a questionnaire to be used in the field for recording responses by enumerators, should be sturdy to survive the field handling.

28. The questionnaire so designed should be ideal in facilitating the collection of relevant and accurate data. In order to enhance accuracy in the survey data, special consideration should be made in ordering the sequence and in the wording of items in the questionnaire. The respondent has to be motivated. We may add that the questionnaire has to be well spread out to facilitate easy reading of questions either by the respondent or the interviewer. We cannot overemphasize that every questionnaire should have clear instructions.

29. Special care, therefore, should be taken by researchers in giving precise definition of the data to be collected and the translation of data requirements and related concepts into operational questions.

30. In summary a good questionnaire should:

a. Enable the collection of accurate information to meet the needs of potential data users in a timely manner.
b. Facilitate the work of data collection, data processing and tabulation.
c. Ensure economy in data collection, i.e. avoid collection of any non-essential information.
d. Permit comprehensive and meaningful analysis and purposeful utilization of the data collected.

31. This implies that survey questionnaires must be developed so as to yield information of the highest quality possible with special emphasis on relevance, timeliness and accuracy.

This must be accomplished efficiently, minimizing the cost and burden involved in the provision of the necessary information.

(a) **Question construction**: Open and closed-ended questions are used in sample survey questionnaires. In an open-ended question, the respondent gives his own answer to a question. In an attitudinal survey we may ask respondents to define what they consider is good quality life. Obviously different respondents will define in their own way what constitutes quality of life. On the other hand, a closed-ended question restricts the respondent to select answers from a list already given by the researcher. The following is an example of a closed-ended question:

Do you have any permanent mental disabilities that limit you in daily activities?

Yes	No
☐	☐

How do you evaluate your capacity to see (even with glasses or contact lenses, if used)?

1. Unable
2. Severe permanent difficulty
3. Some permanent difficulty
4. No difficulty

The advantages of using closed-ended questions is that: (i) they yield more uniform responses and (ii) they are easy to process. The main limitation of such questions is that the designer of the survey has to structure the possible answers. In such a case, important possible responses may be overlooked. In most surveys, complex issues and questions pertaining to attitudes and perceptions that may not be known are best handled by open-ended questions.

(b) **Question wording**: The questions should be clear, precise and unambiguous. The respondent should not be left to guess what the researcher wants out of him/her. The use of definitions and concepts may seem obvious to the researcher while this *may* not be so to the respondent. This way, a respondent may use discretion when answering questions. The end result is a proliferation of nonsampling errors in survey results. Let us consider a simple example. The question, What is your "home" address? such a question is bound to create confusion in many African countries, especially for the urban population, unless "home" is clearly defined. There are respondents who take "home" to mean the village they originally come from.

(c) **Loaded questions**: A loaded question persuades a respondent to answer a question in a certain way. This means that the question tends to be biased in favour of a certain answer. Here is an example of a loaded question in a healthy survey: "How many days in a week do you drink more than two bottles of beer"? This question courts the respondent into admitting that he/she drinks beer, above all, not less than two bottles a day. Such questions tend to bias answers of respondents. It is important to remind researchers to avoid creating data but to simply collect data.

(d) **Ask relevant questions**: The purpose of a questionnaire is to enlist information that would be used in studying the situation. It is therefore, imperative for the researcher to ask relevant questions in order to obtain a true picture of a particular situation under study. The questions included in a questionnaire should be relevant to most respondents. For instance, it is pointless to canvass a questionnaire clustered with questions on individual achievement with regard to higher university education in a typical rural environment of most African countries today. Similarly, it is not appropriate in a fertility survey to include females say of age 10, and ask them questions on number of children ever born, whether married, divorced or widowed. These questions would be relevant to females above a certain age, but certainly not to girls who are less than child bearing age. The danger in asking irrelevant questions is that the researcher can easily be misled.

(e) **Question sequence**: The order of items in a questionnaire should try to motivate and facilitate recall in the respondent. It is suggested that the first questions should be easy, interesting and not sensitive. This builds up the confidence of the respondent to carry through the interview which in most cases he/she provides voluntarily. Questions should assist the respondent in providing accurate answers. We should emphasize that there must be a logical link in questions, especially those that are contingent. Sensitive questions must be among the last questions to ask.

A.II.4. Organization of field work

32. In most developing countries, the organization of field work is seriously constrained by lack of resources. However, if a survey has to be carried out, field work should be properly organised in order to utilise the limited resources at the disposal of the researcher efficiently. For the survey operations to succeed, the theoretical aspects of the survey subject matter should be clearly understood by those involved in the survey operations. Further, enumerators must thoroughly master the practical procedures that may lead to the successful collection of accurate data. In order for the survey operations to be rationalised, there is always a need to have a well organised and effective field organization.

A.II.4.1. Equipment and Materials

33. In many developing countries it is necessary, that well in advance, equipment such as vehicles, bicycles, etc are available and in working condition. It is also necessary to have some spare parts. Vehicles and bicycles facilitate quick mobility of team leaders and supervisors/enumerators, respectively.

34. Adequate materials, like folders, clipboards, pencils, pencil sharpeners, notebooks and fuel (for vehicles) should be available in adequate supplies for use during the survey operations.

Management of survey operations

35. A large scale sample survey is usually a demanding and complex operation therefore the need for judicious, effective and efficient management of activities at various levels cannot be overemphasized.

36. There must be a clear and well defined line of command from the survey manager to the enumerator. It should be noted that control forms for monitoring progress of the survey have been found useful.

Publicity

37. Some surveys have had limited success partly due to high nonresponse owing to refusals. It is, therefore, incumbent upon survey organisers to mount some publicity campaigns for the survey, different approaches can be adopted depending on prevailing circumstances. For example in some countries, in the urban areas radio, television and news paper messages can complement posters. While in the rural areas, radio messages and posters could be used.

38. Further, it may be necessary to arrange meetings with local opinion leaders in selected areas. During such meetings people would be briefed on the objectives of the survey. In addition the leaders should be requested to persuade people in their respective areas to provide requisite information to the enumerators.

39. Before going into the field it is important that the relevant legal provision for conducting the survey is published. The announcement should among other information give the survey objectives, duration and topics to be covered.

40. Experience has shown that publicity plays an important role in soliciting cooperation from respondents during survey operation. However, some funding organizations/age ncies consider expenditures on publicity as a waste of resources.

Selection of enumerators

41. An enumerator is at the interface with the respondents. He/she is the representative of the survey organization who is always in contact with the respondent. This is a clear indication of why an enumerator's job is so crucial to the success of the survey programme. The selection of an enumerator should, therefore, be given great consideration and care. An enumerator should be capable of effectively communicating with the respondent. He/she should have qualities of enlisting all the information with accuracy within a reasonable given time.

42. Depending on the type of survey, an enumerator should have an adequate level of education. In addition, an enumerator should be able to record information honestly, without "cooking *figures*". The selected enumerators should follow instructions and use definitions and concepts as provided for in enumerator's field manual.

43. The following selection procedures may assist in selecting suitable enumerators:

a. The prospective enumerators should preferably fill in a form among others, indicating his/her age, marital status, address, education and employment history. Such a form will provide useful background information.
b. Those initially selected may be subjected to an intelligence test and an additional test in simple numerical calculations.
c. Apart from written tests, there is usually a need to interview the candidates. The interviews should be conducted by a panel of interviewers who will independently rate the candidates. Some of the attributes to be considered in rating the candidates are the following: friendliness, interest in work, expression and alertness.

44. Field work can be tedious, with problems of travel under difficult terrain, therefore an enumerator so selected should be committed and prepared to work under difficult conditions.

Training

45. The selected enumerators should be thoroughly trained before being sent into the field. Qualified instructors should be responsible for the training. Such instructors should be well versed in the aims and objectives of the survey. Preferably, they should be part of the research group carrying out the survey.

46. In order for the enumerators to genuinely be involved in the ideals of the survey, they have to be well trained in the concepts and definitions used in the questionnaire. Indeed the enumerators are entitled to know the purposes of the survey and how the results are going to be used.

47. As part of the training process, the enumerators, in the presence of the instructor, should take turns in explaining to others the various items in the questionnaire. Practical sessions should be arranged both in class and in the actual field situation. For example, enumerators could take turns in asking questions to each other in a class room setting and then a field trip could be arranged in the neighbourhood, where a few households could be interviewed by the trainee enumerators. The instructor should always be present to guide and correct the enumerators. After the field interviews, the trainees should discuss the results under the guidance of the instructor. Depending on the results of such interviewers, a decision can be made as to whether the enumerators were ready to go into the field or not

48. The main purpose of a training programme is to bring about uniformity in the procedures of the survey. As it is obvious that if definitions and concepts are differently interpreted by enumerators, the aggregate results will not represent the true picture of the situation under study.

Field supervision

49. It is generally agreed that training is a precursor to effective and successful field work. However, training without proper supervision may not yield the desired results. The success of field work requires dedicated, continuous and effective supervision by superior staff who should be more experienced and better qualified than enumerators. Supervisors should undergo training in all aspects of the survey. It cannot be overemphasized that the supervisor is an important link between the research organisation and the enumerator. The supervisor is supposed to organise work for enumerators; reviews completed work and maintains level of high commitment to the survey programme by the enumerators. We advocate that if possible, there should be a relatively high ratio between the supervisory staff and the enumerators. The ratio of one supervisor to four or five enumerators has been suggested. However, this is just a guide.

Follow-up of non-respondents

50. In most surveys, there are bound to be cases of non-response. Some respondents refuse to co-operate with the enumerators, while in some cases, certain items in the questionnaire are not attended to. Since the aim, in any survey, is to achieve the highest

possible response rate, it is necessary to collect information from some of the nonrespondents. When a nonresponding unit has been reported to the supervisor, he has to contact the sample unit and try to solicit for the information. After all the supervisor is supposed to be better qualified and more trained and experienced. If some nonrespondents do not give in, the questionnaire may be modified to get information only on key variables. On the other hand, if the nonresponse is considerably high, a frame of non respondents can be prepared from which a sub-sample may be selected. In this case, the survey effort is then directed to the sub-sample preferably using supervisors as enumerators.