SAMPLING TECHNIQUES

INTRODUCTION

Many professions (business, government, engineering, science, social research, agriculture, etc.) seek the broadest possible factual basis for decision-making. In the absence of data on the subject, a decision taken is just like *leaping into the dark*. Sampling is a procedure, where in a fraction of the data is taken from a large set of data, and the inference drawn from the sample is extended to whole group. [Raj, p4] The surveyor's (a person or a establishment in charge of collecting and recording data) or researchers initial task is to formulate a rational justification for the use of sampling in his research. If sampling is found appropriate for a research, the researcher, then:

(1) Identifies the target population as precisely as possible, and in a way that makes sense in terms of the purpose of study. [Salant, p58]

(2) Puts together a list of the target population from which the sample will be selected. [Salant, p58] [Raj, p4] This list is termed as ***a frame*** (more appropriately ***list frame***) by many statisticians.

(3) Selects the sample, [Salant, p58] and decide on a sampling technique, and;

(4) Makes an inference about the population. [Raj, p4]

All these four steps are interwoven and cannot be considered isolated from one another. Simple random sampling, systematic sampling, stratified sampling fall into the category of simple sampling techniques. Complex sampling techniques are used, only in the presence of large experimental data sets; when efficiency is required; and, while making precise estimates about relatively small groups within large populations [Salant, p59]

SAMPLING TERMINOLOGY

- A ***population*** is a group of experimental data, persons, etc. A population is built up of ***elementary*** units, which cannot be further decomposed.
- A group of elementary units is called a ***cluster***.
- ***Population Total*** is the sum of all the elements in the sample frame.
- ***Population Mean*** is the average of all elements in a sample frame or population.
- The fraction of the population or data selected in a sample is called the S***ampling Fraction***.
- The reciprocal of the sampling fraction is called the R***aising Factor***.
- A sample, in which every unit has the same probability of selection, is called a ***random sample***. If no repetitions are allowed, it is termed as a simple random sample ***selected without replacement***. If repetitions are permitted, the sample is ***selected with replacement***.

PROBABILITY AND NONPROBABILITY SAMPLING

Probability sampling (a term due to Deming, [Deming]) is a sampling porcess that utilizes some form of random selection. In probability sampling, each unit is drawn with known probability, [Yamane, p3] or has a nonzero chance of being selected in the sample. [Raj, p10] Such samples are usually selected with the help of random numbers. [Cochran, p18] [Trochim] With probability sampling, a measure of sampling variation can be obtained objectively from the sample itself.

Nonprobability sampling or j***udgment*** sampling depends on subjective judgment. [Salant, p62] The nonprobability method of sampling is a process where probabilities cannot be assigned to the units objectively, and hence it becomes difficult to determine the reliability of the sample results in terms of probability. [Yamane, p3] Examples of nonprobability sampling used extensively in 1920's and 1930's are the judgment sample, quota sample, and the mail questionnaire. In nonpraobability sampling, often, the surveyor selects a sample according to his convenience, or generality in nature. Nonprobability sampling is well suited for exploratory research intended to generate new ideas that will be systematically tested later. However, if the goal is to learn about a large population, it is imperative to avoid judgment of nonprobabalistic samples in survey research. [Salant, p64] In contrast to probability sampling techniques, there is no way of knowing the accuracy of a non-probabilistic sample estimate.

SAMPLING ERRORS

Sampling errors occur as a result of calculating the estimate (estimated mean, total, proportion, etc) based on a sample rather than the entire population. This is due to the fact that the estimated figure obtained from the sample may not be exactly equal to the true value of the population. For example, [Raj, p4] if a sample of blocks is used to estimate the total number of persons in the city, and the blocks in the sample are larger than the average — then this sample will overstate the true population of the city.

When results from a sample survey are reported, they are often stated in the form "***plus*** or ***minus***" of the respective units being used. [Salant, p72] This "plus or minus" reflects sampling errors. In [Salant, p73], Salant and Dilman, describe, that the statistics based on samples drawn from the same population always vary from each other (and from the true population value) simply because of ***chance***. This variation is ***sampling error*** and the measure used to estimate the sampling error is the ***standard error***.

se (p) = [(p q)/m] where,
se (p) is the standard error of a proportion,
p and q is the proportion of the sample that do (p) and do not (q) have a particular characteristic, and
n = the number of units in the sample.

Standard errors are usually used to quantify the precision of the estimates. Sample distribution theory, points out that about 68 percentage of the estimates lie within one standard error or standard deviation of the mean, 95 percentages lie within two standard deviations and all estimates lie within three standard deviations. [Cochran] [Sukhatme] [Raj +] [Raj, p16-17] Sampling errors can be minimized by proper selection of samples, and in [Salant, p73], Salant and Dilman state — "Three factors affect sampling errors with respect to the design of samples – the sampling procedure, the variation within the sample with respect to the variate of interest, and the size of the sample. [Yamane] adds that a large sample results in lesser sampling error

NONSAMPLING ERRORS

The accuracy of an estimate is also affected by errors arising from causes such as incomplete coverage and faulty procedures of estimation, and together with observational errors, these make up what are termed nonsampling errors. [Sukhatme, p381] The aim of a survey is always to obtain information on the true population value. The idea is to get as close as possible to the latter within the resources available for survey. The discrepancy between the survey value and the corresponding true value is called the ***observational error*** or ***response error.*** [Hansen +] ***Response Nonsampling errors*** occur as a result of improper records on the variate of interests, careless reporting of the data, or deliberate modification of the data by the data collectors and recorders to suit their interests. [Raj, p96-97] [Sukhatme, p381] ***Nonresponse error*** [Cochran, p355-361] occurs when a significant number of people in a survey sample are either absent; do not respond to the questionnaire; or, are different from those who do in a way that is important to the study. [Salant, p20-21]

BIAS

Although judgment sampling is quicker than probability sampling, it is prone to ***systematic*** errors. For example, if 20 books are to be selected from a total of 200 to estimate the average number of pages in a book, a surveyor might suggest picking out those books which appear to be of average size. The difficulty with such a procedure is that consciously or unconsciously, the sampler will tend to make errors of judgment in the same

.

direction by selecting most of the books which are either bigger than the average of otherwise. [Raj, p9] Such systematic errors lead to what are called ***biases***. [Rosenthal]

BASIC PRINCIPLES OF SAMPLING

SAMPLING FROM A HYPOTHETICAL POPULATION

Consider the following hypothetical population of 10 manufacturing establishments along with the number of paid employees in each (Table 2.1). [Raj, p14] The average employment per establishment is the number of paid employees / number of establishments, which amounts to 27. ([31+ 15+ 67+ 20+ 13+18+9+22+48+27] / 10])

Table 2.1 Hypothetical populations of 10 establishments

Establishment number	Number of paid employees y
0	31
1	15
2	67
3	20
4	13
5	18
6	9
7	22
8	48
9	27

Let us now try to estimate the average employment per establishment from a random sample of two establishments. There are in all 45 samples each containing two establishments. We can calculate the average employment from each sample and use it as an estimate of the population average. It is clear from Table 2.2, that the sample estimates lie within the range of 11 to 57.5. Some samples give a very low figure while some others give a high estimate. But the average of all the sample estimates is 27, which is the true average of the population. We can conclude that that the sample mean is an unbiased estimate of the population mean. But, although unbiased, the sample mean varies considerably around the population mean. It is observed as a universal phenomenon that the concentration of sample estimates around the true mean increase as the sample size is increased. [Cochran] [Raj +] This fact is expressed by saying that the sample mean is a consistent estimate of the population mean. While only 30% of the samples produced a mean between 21 and 33 for sample size 2, the corresponding percentage is 43 for n = 3, 90 for n = 7, and so on….

Table 2.2 All possible samples of size 2

Sample	Average	Sample	Average	Sample	Average	Sample	Average	Sample	Average
0,1	(31+ 15) / 2 = 23	1,2	41	2,4	40	3,7	21	5,7	20
0,2	49	1,3	17.5	2,5	42.5	3,8	34	5,8	33
0,3	22.5	1,4	14	2,6	38	3,9	23.5	5,9	22.5
0,4	22	1,5	16.5	2,7	44.5	4,5	15.5	6,7	15.5
0,5	24.5	1,6	12	2,8	57.5	[illegible]	11	6,8	28.5
0,6	20	1,7	18.5	2,9	47	4,7	17.5	6,9	18
0,7	26.5	1,8	31.5	3,4	16.5	4,8	30.5	7,8	35
0,8	39.5	1,9	21	3,5	19	4,9	20	7,9	24.5
0,9	29	2,3	43.5	3,6	14.5	5,6	13.5	8,9	37.5

The Variance Of Sample Estimates

The variance of estimates provides a measure of the degree of concentration of the sample estimates around the expected value. [Cochran, p15] The deviation of each sample estimate from the expected value is squared, and the sum of the squares is divided by the number of samples. [Sukhatme, p7] The greater the variance the lesser the concentration of sample estimates around the expected value. Actually, it is not necessary to draw all possible samples to get a measure of the extent to which the sample estimates differ from the value aimed at. [Raj, p15-18] The variance of the sample average or the sample mean $\hat{y}$ is given by

$V(\hat{y}) = 1/n\ (1 - n/N)\ S_y^2$, where
$S_y^2 = 1/(N-1) \sum (y_i - \acute{Y})^2$, and
$\acute{Y} = 1/N \sum y_i$, where
N is the number of units in the population

.

BASIC PROBABILISTIC SAMPLING TECHNIQUES

SIMPLE RANDOM SAMPLING

Sample surveys deal with samples drawn from populations, and contain a finite number of N units. If these units can all be distinguished from one another, the number of distinct samples of size n that can be drawn from N units is given by the combinatorial formula [Cochran, p18] -

$$NC_n = N! / ((n!) * ((N - n)!))$$

Objective: To select n units out of N, such that each number of combinations has an equal chance of being selected, i.e., each unit in any given population has the same probability of being selected in the sample [Cochran] [Raj, p32] [Raj +].

Procedure: Use a table random numbers, a computer random number generator [Raj, p32] [Sukhatme, p5-6] (such as the Rand function in Excel or the rand function in the low level programming language C), or a mechanical device [Trochim] to select the sample.

Example Suppose there are N = 850 students in a school from which a sample of n = 10 students is to be taken. The students are numbered from 1 to 850. Since our population runs into three digits we use random numbers that contain three digits. All numbers exceeding 850 are ignored because they do not correspond to any serial number in the population. In case the same number occurs again, the repetition is ignored. Following these rules the following simple random sample of 10 students is obtained when columns 31 and 32 of the random numbers given in Appendix 1 are used.

251	546	214	495	074
800	407	502	513	628

Remark: If repetitions are included, the procedure is termed as selecting a sample with replacements. [Raj p32] In the present example the sample is selected without replacement. A detailed analysis of the comparison of sampling with and without replacements is found in [Sukhatme, p45-77]

Exercise A bookshop has a bundle of sales invoices for the previous year. These invoices are numbered 2,615 to 7,389 and a random sample of 12 invoices is to be taken. Use Appendix 1, to select 12 random samples.

Exercise There are 19 classes in a school, the number of students in each class being given in Table 3.1. The students are numbered from 1 to 574 in the last column of the table. Select a random sample of 10 students using Appendix 1.

Table 3.1 Number of students in a school

Class	Number of students	Cumulated number	Assigned range
1	100	100	1-100
2	83	183	101-183
3	71	254	184-254
4	65	319	255-319
5	57	376	320-376

6	50	426	377-426
7	43	469	427-469
8	40	509	470-509
9	35	544	510-544
10	30	574	545-574

Estimation Based On Simple Random Sampling

The basic rule used for estimating population parameters such as means, totals, and proportions, is that a population mean or proportion is estimated by the corresponding mean, total, or proportion in the sample. [Raj +] [Cochran p21-26] [Sukhatme, p8-16] A population total is estimated by multiplying the sample mean by the number of units in the population [Raj, p33] [Raj +]

$$V(\hat{y}) = 1/n\ (1 - n/N)\ S_y^2$$

$$S_y^2 = 1/(N-1) \sum (y_i - \acute{Y})^2$$

$$\acute{Y} = 1/N \sum y_i,$$

Systematic Sampling

Systematic sampling is a little bit different from simple random sampling. Suppose that N units of the population are numbered 1 to N in some order. To select a sample of n units, we must take a unit at random from the first k units and every kth unit thereafter [Cochran, p206] [Yamane, p159]

Procedure [Trochim]:

1. Number the units in population from 1 to N
2. Decide on the n (sample size) that is required
3. Select an interval size k = N/n
4. Randomly select an integer between 1 to k
5. Finally, take every kth unit

Let's assume that we have a population that only has N=100 people in it and that you want to take a sample of n=20. To use systematic sampling, the population must be listed in a random order. The sampling fraction would be n/N = 20/100 = 20%. In this case, the interval size, k, is equal to N/n = 100/20 = 5. Now, select a random integer from 1 to 5. In our example, imagine that you chose 4. Now, to select the sample, start with the 4th unit in the list and take every k-th unit (every 5th, because k=5). You would be sampling units 4, 9, 14, 19, and so on to 100 and you would wind up with 20 units in your sample.

In order for systematic sampling to work, it is essential that the units in the population be randomly ordered, at least with respect to the characteristics you are measuring. Systematic sampling is fairly easy to do and is widely used for its convenience and time efficiency. [Cochran, p206-20] In many surveys, it is found to provide more precise estimates than simple random sampling. [Raj, p39] [Cochran p206] This happens when there is a trend present in the list with respect to the characteristic of interest. [Trochim] mentions that, there are situations where there is simply no easier way to do sampling. Systematic sampling is at its worst, when there is periodicity in the sampled data and the sampling interval has fallen in line with it [Raj, p143]. When this happens, most of the units in the sample will be either too high or low, which makes the estimate very variable. Taking many random starts can reduce the risk by giving rise to a number of systematic samples each of a small size [Sukhatme] [Cochran]. If a number of random starts is planned, it is useful to take them in complementary pairs of the type I, k + 1 – i. [Raj, p141-143] This is particularly important when the stratum size is not an integral multiple of the sampling interval. [Raj +]

Estimation Based On Systematic Random Sampling

For estimating the population total the sample total is multiplied by the sampling interval. This when divided by N gives an estimate of the population mean. As before, the population proportion is estimated in the same way as the mean. The question of estimating the variance from the sample is more intricate [Yamane, p168 –176] [Cochran, p208– 210] and is described in the following example. If the arrangement of units in the population can be considered to be random, the systematic sample behaves like a simple random sample. [Raj, p39]

Example There are 169 industrial establishments employing 20 or more software testers in IBM. The following are the employment figures based on a 1-in-5 systematic sample

35	88	35	36	156	25	24	237	80
468	22	139	163	37	37	27	25	26
38	24	62	331	28	31	81	121	49
23	34	23	22	53	50	50		

N = 169
n = 34
The average of samples,
$\acute{y} = 1/n\ \Sigma\ y_i$ = (35 + 88 + 35 + …….. 53 + 50 + 50) /34 = 78.83

$[1/ (N - 1)] * \Sigma\ S\ (y_i - \acute{y})^2 = [1/168] * ((35 - 78.83)^2 + (88 - 78.83)^2 + \ldots\ (50 - 78.83)^2)$ = 309,795 /34 = 9387.73

Variance of the estimate, $V(\acute{y}) = 1/n\ (1 - n/N)\ [1/ (N - 1)] * \Sigma\ S\ (y_i - \acute{y})^2$ = 1/34 * (1 – 34/169) * 9387.73 = 220.89

Standard Error = $\sqrt{(V(\acute{y}))} = \sqrt{220.89}$ = 14.86

Coefficient of Variation = Standard Error/ ý * 100 % = 18.9 %

Sampling With Unequal Probabilities

If the sampling units vary considerably in size, a simple random or a systematic sample of units does not produce a good estimate. This is due to the high variability of units for the characteristics under study. [Raj, p42-43] In some situations, it is better to give a higher probability of selection to larger units, and a lower probability to a selection of lower units in a population. Several methods of selecting the sample with probability proportional to size (pps) are available. [Hansen] The method of sampling with probability proportional to size is generally used for the selection of large units such as cities, farms, etc and surveys which employ subsampling [Cochran, p251]

ESTIMATION IN UNEQUAL PROBABILITY `SAMPLING

The problem of estimating population mean and variance is more intricate when the sample is selected with unusual probabilities. When sampling with replacement, the value of y (characteristics under study) for the selected unit is divided by the probability of selection in a sample of one. This when aggregated over the sample and divided by the sample size provides an estimate of the population total for y.

In sampling without replacement, y is divided by the probability with which the unit is selected in the whole sample of size n and the ratio is aggregated over the sample to produce an estimate of the total of y for the population. [Raj, p44-47]

USE OF SUPPLEMENTARY INFORMATION IN SAMPLING

In research, quite often the surveyors use past or supplementary (auxiliary) information (denoted by x) available for various units in the population to calculate an estimate of a current variate (y). Such information is generally based on previous census or large-scale surveys. Supplementary information ***x*** can be used to improve the precision of the estimates, and may be used in a number of ways —

(1) Selecting the samples with probability proportional to ***x***,
(2) Stratifying the population on the basis of ***x***, or,
(3) The auxiliary information (***x***) may be used to form a
    a. Ratio estimate, [Cochran, p29-33]
    b. Difference estimate, [Madow] or,
    c. Regression estimate [Sukhatme, p193-221]

STRATIFIED SAMPLING involves dividing the population into homogeneous non-overlapping groups (i.e., *strata*), selecting a sample from each group, and conducting a simple random sample in each stratum. [Cochran, p87] [Trochim] On the basis of information available from a frame, units are allocated to strata by placing within the same stratum, those units which are more-or-less similar with respect to the characteristics being measured. If this can be reasonably achieved, the strata will become homogenous, i.e., the unit-to-unit variability within a stratum will be small.

Surveyors use various different sample allocation techniques to distribute the samples in the strata. In proportional allocation, the sample size in a stratum is made proportional to the number of units in the stratum. [Yamane, p153] [Raj, p51] [Sukhatme, p86-87] In equal allocation, the same number of units is taken from each stratum irrespective of the size of the stratum. [Raj, p51] In Neyman allocation, the number of units in the sample from a stratum is made proportional to the product of the stratum size and the stratum standard deviation. [Yamane, p154, 156] [Sukhatme, p87-91] In [Raj +], it is proved that Neyman's allocation is the best when a sample of specified size is to be allocated to the strata.

ESTIMATION FROM STRATIFIED SAMPLING

The basic rule is to form an estimate from each stratum making use of the rules from the section above – ***BASIC PRINCIPLES OF SAMPLING***. By adding over the strata, the population total can be estimated; from this estimate the mean is estimated by dividing by the number of units in the population. [Raj, p52] From [Raj +], the variance of the estimate from stratified sampling is shown as

$V\ (N_h\ \acute{y}_h) = N_h^2 * 1/n_h * (1 - n_h / N_h)\ S_{yh}^2$, where n_h is the sample size

RATIO ESTIMATION

Ratio estimation involves estimating from a sample the population ratio ***R*** of the variate of interest (***y***) to the auxiliary variate (***x***). Quite often, instead of current information on ***x***, prior information on ***x*** is used to improve the precision of an estimate for ***y*** This is usually done by first estimating the ratio of ***y*** to ***x*** in the population and then multiplying the estimate obtained by the known population total for ***x***. The estimate obtained in this manner is called a ***ratio estimate***. The rationale of this procedure is that because of the close relationship between ***y*** and ***x***, the ratio of ***y*** to ***x*** may be less variable than ***y***. [Sukhatme, p 193-196] [Raj, p57-61]

SAMPLING IN CLUSTERS

CLUSTER SAMPLING

The smallest units into which a population can be divided are called the elements of the population, and groups of elements the clusters. [Sukhatme, p222] The problem with random sampling methods when sampling a population that's distributed across a wide geographic region lies in covering a lot of ground geographically in order to get to each of the units sampled. [Trochim] This geographic trotting to collect samples is an expensive affair. But, without taking samples from across the whole geographic population, it may become difficult to conclude anything affirmatively about the population. The impasse is to determine the best size of the cluster for a specified cost of the survey [Raj, p144-145] [Cochran, p244-247]. This predicament can be solved if the cost of the survey and the variance of the estimate can be expressed as functions of the size of the cluster. [Raj, p144] According to [Trochim], cluster sampling includes:

1. Divide population into clusters
2. Randomly sample clusters
3. Measure all units within sampled clusters

Cluster sampling is ordinarily conducted in order to reduce costs. The variance of the estimate of the mean in simple random sampling of clusters depends on the sample size, the population variance and on the correlation of the variate of interest between units within the same cluster. If the units within a cluster are more similar than units belonging to different clusters, the estimator is subject to a larger variance; thus the smaller the intra-cluster correlation, the better. [Raj, p144] A detailed study of the efficiency of cluster sampling is provided in [Sukhatme, p220-279].

MULTISTAGE SAMPLING

Multistage sampling involves, combining various probability techniques in the most efficient and effective manner possible. The process of estimation is carried out stage by stage, using the most appropriate methods of estimation at each stage. [Sukhatme, p262] mentions that for a given number of elements, greater precision is attained by distributing the elements over a large number of clusters than by taking a small number of clusters and sampling a large number of elements from each one of them.

Quite often, auxiliary information is used to improve the precision of an estimate. But, in the absence of auxiliary information, it may be advantageous to conduct the enquiry in two phases. In the first phase, auxiliary information is collected on the variate of a fairly large sample. Then a sub-sample is taken, and information collected on the variate of interest. Then the two samples are used in the best possible manner to produce an estimate for the variate of interest. The procedure of first selecting clusters and then choosing a specified number of elements from each selected cluster is known as subsampling. [Raj, p25] It is also known as two-stage sampling or double sampling. [Cochran, p118] [Neyman]

The clusters, which form the units of sampling at the first stage, are called ***first stage units,*** and the elements or group of elements within clusters, which form the units of sampling at the second stage, are called ***sub-units*** or ***second-stage*** units. [Sukhatme, p262] The procedure can be easily generalized to three or more stages and hence known as multi-stage sampling. [Raj, p73]Double sampling can be used to reduce the response bias in survey results. [Neyman]

NOTE: THE FOLLOWING IS A DIRECT REPLICATION OF A PAPER PUBLISHED BY WILLIAM M. TROCHIM AT THE BILL TROCHIM'S CENTER FOR SOCIAL RESEARCH METHODS AT THE CORNELL UNIVERSITY

NON-PROBABILISTIC SAMPLING [Trochim]

INTRODUCTION

The difference between non-probabilistic and probabilistic sampling is that non-probabilistic sampling does not involve *random* selection and probability sampling does. Non-probabilistic samples don't depend upon the rationale of probability theory. At least with a probabilistic sample, we know the odds or probability that we have represented the population well. We are able to estimate confidence intervals for the statistic. With non-probabilistic samples, we may or may not represent the population well, and it will often be hard for us to know how well we've done so. In general, researchers prefer probabilistic or random sampling methods to non-probabilistic ones, and consider them to be more accurate and rigorous. However, in applied social research there may be circumstances where it is not feasible, practical or theoretically sensible to do random sampling. Here, we consider a wide range of non-probabilistic alternatives.

We can divide non non-probabilistic probability sampling methods into two broad types: *accidental* or *purposive*. Most sampling methods are purposive in nature because we usually approach the sampling problem with a specific plan in mind. The most important distinctions among these types of sampling methods are the ones between the different types of purposive sampling approaches.

ACCIDENTAL, HAPHAZARD OR CONVENIENCE SAMPLING

One of the most common methods of sampling goes under the various titles listed here. I would include in this category the traditional "man on the street" (of course, now it's probably the "person on the street") interviews conducted frequently by television news programs to get a quick (although nonrepresentative) reading of public opinion. I would also argue that the typical use of college students in much psychological research is primarily a matter of convenience. (You don't really believe that psychologists use college students because they believe they're representative of the population at large, do you?). In clinical practice, we might use clients who are available to us as our sample. In many research contexts, we sample simply by asking for volunteers. Clearly, the problem with all of these types of samples is that we have no evidence that they are representative of the populations we're interested in generalizing to -- and in many cases we would clearly suspect that they are not.

PURPOSIVE SAMPLING

In purposive sampling, sampling is done with a *purpose* in mind. We usually would have one or more specific predefined groups we are seeking. For instance, have you ever run into people in a mall or on the street who are carrying a clipboard and who are stopping various people and asking if they could interview them? Most likely they are conducting a purposive sample (and most likely they are engaged in market research). They might be looking for Caucasian females between 30-40 years old. They size up the people passing by and anyone who looks to be in that category they stop to ask if they will participate. One of the first things they're likely to do is verify that the respondent does in fact meet the criteria for being in the sample. Purposive sampling can be very useful for situations where you need to reach a targeted sample quickly and where sampling for proportionality is not the primary concern. With a purposive sample, you are likely to get the opinions of your target population, but you are also likely to overweight subgroups in your population that are more readily accessible.

All of the methods that follow can be considered subcategories of purposive sampling methods. We might sample for specific groups or types of people as in modal instance, expert, or quota sampling. We might sample for diversity as in heterogeneity sampling. Or, we might capitalize on informal social networks to identify specific respondents who are hard to locate otherwise, as in snowball sampling. In all of these methods we know what we want -- we are sampling with a purpose.

1. MODAL INSTANCE SAMPLING

In statistics, the *mode* is the most frequently occurring value in a distribution. In sampling, when we do a modal instance sample, we are sampling the most frequent case, or the "typical" case. In a lot of informal public opinion polls, for instance, they interview a "typical" voter. There are a number of problems with this sampling approach. First, how do we know what the "typical" or "modal" case is? We could say that the modal voter is a person who is of average age, educational level, and income in the population. But, it's not clear that using the averages of these is the fairest (consider the skewed distribution of income, for instance). And, how do you know that those three variables -- age, education, income -- are the only or event the most relevant for classifying the typical voter? What if religion or ethnicity is an important discriminator? Clearly, modal instance sampling is only sensible for informal sampling contexts.

2. EXPERT SAMPLING

Expert sampling involves the assembling of a sample of persons with known or demonstrable experience and expertise in some area. Often, we convene such a sample under the auspices of a "panel of experts." There are actually two reasons you might do expert sampling. First, because it would be the best way to elicit the views of persons who have specific expertise. In this case, expert sampling is essentially just a specific sub case of purposive sampling. But the other reason you might use expert sampling is to provide evidence for the validity of another sampling approach you've chosen. For instance, let's say you do modal instance sampling and are concerned that the criteria you used for defining the modal instance are subject to criticism. You might convene an expert panel consisting of persons with acknowledged experience and insight into that field or topic and ask them to examine your modal definitions and comment on their appropriateness and validity. The advantage of doing this is that you aren't out on your own trying to defend your decisions -- you have some acknowledged experts to back you. The disadvantage is that even the experts can be, and often are, wrong.

3. QUOTA SAMPLING

In quota sampling, you select people non-randomly according to some fixed quota. There are two types of quota sampling: *proportional* and *non proportional*. In **proportional quota sampling** you want to represent the major characteristics of the population by sampling a proportional amount of each. For instance, if you know the population has 40% women and 60% men, and that you want a total sample size of 100, you will continue sampling until you get those percentages and then you will stop. So, if you've already got the 40 women for your sample, but not the sixty men, you will continue to sample men but even if legitimate women respondents come along, you will not sample them because you have already "met your quota." The problem here (as in much purposive sampling) is that you have to decide the specific characteristics on which you will base the quota. Will it be by gender, age, education race, religion, etc.?

4.NONPROPORTIONAL QUOTA SAMPLING is a bit less restrictive. In this method, you specify the minimum number of sampled units you want in each category. Here, you're not concerned with having numbers that match the proportions in the population. Instead, you simply want to have enough to assure that you will be able to talk about even small groups in the population. This method is the non-probabilistic analogue of stratified random sampling in that it is typically used to assure that smaller groups are adequately represented in your sample.

5. HETEROGENEITY SAMPLING

We sample for heterogeneity when we want to include all opinions or views, and we aren't concerned about representing these views proportionately. Another term for this is sampling for *diversity*. In many brainstorming or nominal group processes (including concept mapping), we would use some form of heterogeneity sampling because our primary interest is in getting broad spectrum of ideas, not identifying the "average" or "modal instance" ones. In effect, what we would like to be sampling is not people, but ideas. We imagine that there is a universe of all possible ideas relevant to some topic and that we want to sample this population, not the population of people who have the ideas. Clearly, in order to get all of the ideas, and especially the "outlier" or unusual ones, we have to include a broad and diverse range of participants. Heterogeneity sampling is, in this sense, almost the opposite of modal instance sampling.

6. Snowball Sampling

In snowball sampling, you begin by identifying someone who meets the criteria for inclusion in your study. You then ask them to recommend others who they may know who also meet the criteria. Although this method would hardly lead to representative samples, there are times when it may be the best method available. Snowball sampling is especially useful when you are trying to reach populations that are inaccessible or hard to find. For instance, if you are studying the homeless, you are not likely to be able to find good lists of homeless people within a specific geographical area. However, if you go to that area and identify one or two, you may find that they know very well who the other homeless people in their vicinity are and how you can find them.

Reference:

[Cochran] Cochran, W. G., *Sampling Techniques*. Second Edition. John Wiley & Sons, Inc. New York. 1953-1963. Library Of Congress Catalog Card Number: 63-7553

[Yamane] Taro Yamane. *Elementary Sampling Theory*. Prentice-Hall, Inc., Englewood Cliffs, N.J. 1967

[Sukhatme] Sukhatme, P. V., and Sukhatme, B. V. *Sampling Theory Of Surveys With Applications.* ISBN: 0-8138-1370-0. Indian Society of Agricultural Statistics, New Delhi, India and the Iowa State University Press, Ames, Iowa, U.S.A.

[Raj] Des Raj, *The Design Of Sample Surveys*. McGraw-Hill Book Company, New York. 1972

[Raj +] Des Raj, *Sampling Theory*. Mc-Graw-Hill Book Company. New York. 1968

[Salant] Salant, P. A., and Dillman, D. A. *How To Conduct Your Own Survey*. John Wiley & Sons, Inc. New York. 1994

[Rosenthal] Rosenthal, R., and Rosnow, R. L. *The Volunteer Subject*. John Wiley & Sons, Inc. New York. 1975

[Trochim] Trochim, M. W., http://trochim.human.cornell.edu/kb/sampnon.htm Bill Trochim's Center for Social Research Methods, Cornell University

[Madow] Hansen, A. A., Hurwitz, W .N., and Madow, W. G. *Sample Survey Methods And Theory*. John-Wiley & Sons, New York

[Neyman] Neyman, J., Contribution To The Theory Of Sampling Human Populations. American Statistics Association., 33, p101-116. 1938

[Hansen] Hansen, M. H., and Hurwitz, W. N., On Theory Of Sampling From Finite Populations. Ann. Math. Stat. 14, p333-362. 1943

[Deming] Deming, W. E., Some Theory Of Sampling. John Wiley & Sons, Inc. New York. 1950

[Hansen +] Hansen, A. A., Hurwitz, W .N., and Mauldin, W. P., *Response Errors In Surveys*. Amer. Stat. Assoc., 41, p517-529