IPDET

Handbook

Module 9

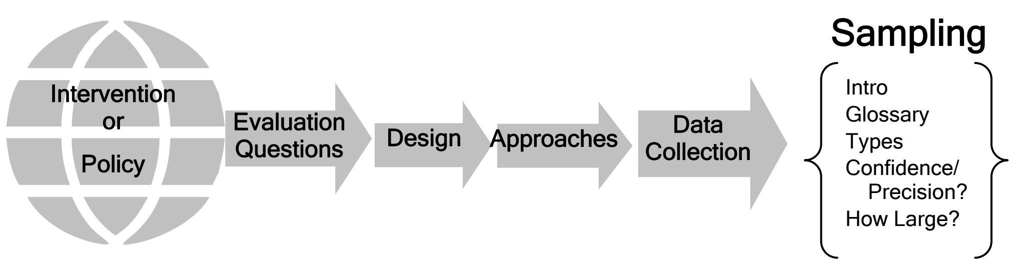
Sampling

Introduction

In the previous three modules, you have learned about writing evaluation questions, and choosing an evaluation design and data collection approach. In this module, you will look at sampling strategies to help you determine how much data you need to collect, and where to collect it so that you can answer your evaluation questions.

This module has five topics. They are:

- Introduction to Sampling
- Sampling Glossary
- Types of Samples: Random and Non-random
- How Confident and Precise Do You Need to Be?
- How Large a Sample Do You Need?



Learning Objectives

By the end of the module, you should be able to:

- define sampling concepts, including: population, sample, sampling frame, sample design, parameter, statistic, and random sample
- discuss the differences between random and non-random samples
- list and define the types of random samples, including: simple, stratified, and cluster
- list and define the types of non-random samples, including: quota, accidental, snowball, judgmental, and convenience
- describe the role of statistics to determine confidence and precision
- describe the guidelines to determine the size of sample that is needed.

Key Words

You will find the following key words or phrases in this module. Watch for these and make sure that you understand what they mean and how they are used in the course.

- census
- sample
- selection bias
- random sampling
- non-random sampling
- inferences
- generalizing to a population
- population
- sampling frame
- sample design
- systematic sampling
- parameter
- statistics
- simple random sample
- stratified random sample
- multi-stage random sample
- cluster random sample
- quota sample
- accidental sample
- snowball sample
- judgmental sample
- convenience sample
- minimum sample size
- confidence level
- confidence interval
- level of precision
- margin of error

Introduction to Sampling

When we begin planning our data collection strategy, we have to decide whether it is possible to collect data from the entire **population** we intend to study: every document, for example, or every farmer, or every mile of a road system. Can we review every file, observe every farmer, examine every road? If we can, we can then accurately report the qualifications of every teacher in our school system, the number of paved miles on all our roads, or the views of all the citizens of the country. If we collect all the data accurately and reliably, then there is little chance of error. The complete coverage of the population in question is called a **census**.

However, most often we are unable to collect data from every file, farmer, or person. It takes too much time and costs too much. Instead, we take a **sample** – a subset of the entire population. If we select a sample, we may be able to draw **inferences** about a population based on our sample results; that is, we can estimate what the population is like based on our sample results. We call this "**generalizing to a population**."

We use samples all the time. For example, when we have a blood test to check on our health, the laboratory takes a sample rather than all our blood. Tests are run using that sample and it is assumed that what they find in the sample is an accurate reflection of what is in all our blood.

Sampling is not just something that applies to large, quantitative studies. Even when conducting a highly qualitative, one-week field visit to assess a program that is spread out over a large geographic region, for example, you still need to be thoughtful about which areas of that region to investigate. Consider, as another example, the biases that might be introduced if program officials select the participants to be studied and how those biases could be avoided with a randomly selected sample. A modest understanding of the basic concepts of systematic sampling can greatly enhance the extent to which your assessment reflects what is really going on in the real world.

Sampling Glossary

population: The total set of units. It could be all the citizens in a country, all farms in a region, or all children under the age of five living without running water in a particular area.

census: A count of (or collection of information from) the entire population.

sample: A subset of units selected from a larger set of the same units.

sampling frame: The list from which you can select your sample.

systematic sampling: a sample drawn from a list using a random start followed by a fixed sampling interval.

sample design: The method of sample selection.

random sample: A sample in which each unit in the population has an equal chance of being selected.

quota sample: A sample in which a specific number of different types of units are selected.

accidental sample: A sample in which the units are selected "by accident."

snowball sample: A type of sampling strategy typically used in interviews, where you ask interviewees who else you should talk to.

judgmental sample: A sample in which selections are made based on pre-determined criteria.

convenience sample: A sample in which selections are based on the convenience to the evaluator (e.g., on easy geographic or organizational access).

parameter: Characteristic of the population.

statistic: Characteristic of a sample.

confidence level: How certain you are (or need to be) that the statistic obtained from your sample is an accurate estimate of the population as a whole.

confidence interval: The calculated range within which the true population value lies and for which we can express 95% confidence (the standard, but may vary).

Types of Samples: Random and Non-random

When we cannot collect data from every country, every person, or every school, we select a smaller subset. As you already know, this subset is called a sample. There are two kinds of samples, random and non-random.

Random Sampling

Random samples are samples in which each unit in the population has an equal chance of being selected. You can take a random sample of files, roads, farms, or people.

One advantage of random sampling is that it eliminates **selection bias**. Since everyone has an equal chance of being selected, you cannot select only those people that look like you or who share a particular viewpoint.

An appropriately sized random sample should be representative of the population as a whole, enabling you to **generalize** to the population from which the sample was drawn.

A complete list of every unit in the population of interest, called a **sampling frame**, is needed to select a random sample. These units are selected using a random schedule; typically, we would use a table of random numbers and select every unit until we reach the sample size we set. Random numbers can be generated fairly simply using any major spreadsheet program. To generate a random whole number between 0 and 100 in Microsoft Excel, for example, enter the formula =RAND()*100 and format the cell as a number with zero decimal places, then copy to as many other cells as you need random numbers (i.e., the size of the sample you need to draw). To create random numbers between 100 and 200, simply add 100 to the formula, i.e., =RAND()*100+100.

An Excel-generated table of two-digit random numbers (the form in which they usually appear in a standard table of random numbers) is included at the end of this module for use in Exercise 9-1. An example would be to use the random numbers in the table near the end of this module to select a sample from an enumerated list of project sites.

Sometimes it is not possible to do a truly random selection. In these cases, a **systematic sampling** technique can be used, in which you make a random start and then select every n^{th} case.

Example of systematic sampling

You want to review records but the records are in boxes and there is no way you can go through and number them all to select a sample. A systematic selection with a random start is acceptable, as long as there is nothing about the original order of the documents that is systematic. For example, you would take a random start and then pick every 20th file until you get the total number of files you want to analyze.

Types of Random Samples

There are four types of random samples. They are:

- simple random samples
- stratified random samples
- multi-stage samples
- cluster samples
- combination random samples.

Simple Random Samples

A simple random sample is the simplest sample. We establish a sample size and then proceed to randomly select units until we reach that number.

Let us say we want to select 100 files from a population of 500. All the files have been consecutively numbered from 001 to 500 and are filed in numerical order from 1-500. We could then use a random numbers table, mentally block it off into three-digit numbers, and then select the first 100 numbers that fall between 001 and 500. These are the files we select for our study.

Stratified Random Samples

Sometimes we want to make sure specific groups are included that might otherwise be missed by using a simple random sample; those groups are usually a small proportion of the population. In this case, we would divide the population into strata based on some meaningful characteristic. This kind of sample is called a **stratified random sample**.

For example, you may want to make sure you have enough people from rural areas in your study. If selected by a simple random sample, you may not get enough people from rural areas if they are a small proportion of all the people in the area. This is especially important if you want to have sufficient numbers in each stratum so you can make meaningful comparisons. For example, we may want to take a stratified sample of farmers at various distances from a major city.

To do a stratified random sample, divide the population into non-overlapping groups (i.e., *strata*) n_1, n_2, n_3, ... $n_{i,}$ such that $n_1 + n_2 + n_3 + ... + n_i = n$. Then do a simple random sample in each stratum. Figure 9.1 illustrates this process.

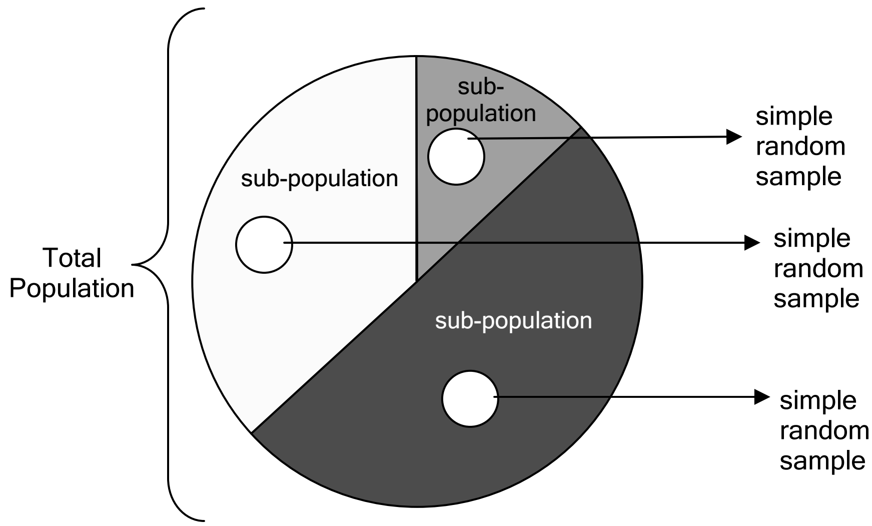


Fig. 9.1: Stratified Random Sample.

Cluster and Multi-stage Samples

Cluster sampling is another form of random sampling. A "cluster" is any naturally occurring aggregate of the units that are to be sampled. Thus households (or homes) are clusters of people, and towns are clusters of households. Cluster samples are most often used when:

- you do not have a complete list of everyone in the population of interest but do have a complete list of the clusters in which they occur, **or**
- you ***have*** a complete list of everyone, but they are so widely disbursed that it would be too time consuming and expensive to send data collectors out to a simple random sample.

In a cluster sample, the cluster is randomly sampled (such as towns or household) and then data is collected on all the target units. For instance, if the evaluation needs to collect data on the height and weight of children ages 2-5 in the program sites scattered across a large rural region, the evaluators might randomly sample 20 villages from the 100 villages receiving the program, and then collect data on all the children ages 2-5 in those villages.

Here is another example. You may want to interview about 200 AIDS patients. There is no compiled list of all AIDS patients, but they are served by 25 clinics scattered over a large region with poor roads. You know that most clinics serve about 50 AIDS patients. Therefore you randomly sample four clinics from the total of 25 clinics, and then study ***all*** AIDS patients in those four clinics, which would provide about 200 patients.

The main drawback of cluster samples is that they are likely to yield somewhat less accurate estimates of the population parameter than simple random samples or stratified random samples of the same size (N). It is possible that the selected clinics will serve clients who differ in economic or religious characteristics from the ones not included in the sample, and if so, the results of the sample will provide biased estimates of the full population of AIDS patients.

Multi-stage random sampling combines two or more forms of random sampling. It applies a second form of random sampling to the results of a first form of random sampling. Most commonly, it begins with random cluster sampling and then applies simple random sampling or stratified random sampling. In the example immediately above of sampling AIDS patients, a multi-stage random sample might initially draw a cluster sample of eight clinics instead of four. Then it might draw a simple random sample of 25 patients from each clinic. That provides a sample of 200 patients, just as in the above example, but they come from a larger number of clinics. This would be somewhat more expensive than collecting data from just four clinics but it is likely to provide less biased estimates of the population parameter.

It is also possible to combine non-random and random sampling in a multi-stage sample. For instance, the clinics might be sampled non-randomly and then the AIDS patients sampled randomly from each selected clinic.

Using the example of collecting data on AIDS patients, for a cluster sample, you would take a random sample of AIDS clinics (just as in the multi-stage sample). From these clinics you would then collect data on ***all patients*** in those selected clinics.

The drawback of multi-stage and cluster samples is that they may not yield an accurate representation of the population.

For example, you may want to interview 200 AIDS patients, but these 200 may be selected from only four randomly sampled clinics because of resource constraints. It is possible that the clinics will serve populations that are too similar in terms of economic background or other characteristics, and therefore may not be representative of all AIDS patients. Likewise, it is possible that the selected clinics are too similar in their level of care to patients to accurately represent the total population of patients.

When you want to interview people living on small, dispersed, and remote farms, it would be very time consuming to sample the people and then travel to all the farms in which they reside. In a cluster sample, you might sample 10 of the 50 farms and then interview all the people at each of those sampled farms.

Combination Random Samples

Sometimes combinations of methods are used. The group may be divided into strata; all the people in one stratum might be selected (as a census) and a random sample selected from the other strata.

The program in Ghana (see Case 9-1) gives an example of a complex use of combination random sample.

Examples of selecting random samples

For example, you want to observe classroom activities to measure the amount of time spent doing hands-on learning activities. You can:

- randomly select classrooms
- randomly select times of day
- randomly select days of the week.

In another example, you might want to observe the amount of traffic on the road from the village to a major town. You can:

- randomly select times and days of the week
- randomly select times of the year
- randomly select observation points or select a single observation point along the road.

Table 9.1 gives you a summary of the random sampling process.

Table 9.1: Summary of Random Sampling Process

Step	Process
1.	Obtain a complete listing of the entire population
2.	Assign each case a unique number.
3.	Randomly select the sample using a random numbers table.
4.	When no numbered listing exists or is not practical to create, use systematic random sampling: • make a random start • select every n^{th} case.

Case 9-1: Impact on Nutrition: Lower Pra Rural Bank Credit with Education Program in Ghana

The intent of the program is to increase the nutritional status and food security of poor households in Ghana. The Credit with Education Program combines: (1) providing credit to participants with (2) education on the basics of health, nutrition, birth timing, and small business skills.

Evaluation Questions: Did the program have an impact on the nutritional status of children, women's economic capacity, women's knowledge of health issues, and ability to offer a healthy diet to their children?

Overall Design: A quasi-experimental design using two surveys. Nineteen communities that did not yet have the Credit with Education Program were the focus of this study. The communities were divided into groups (strata) based on set criteria. Within each of the strata, communities were assigned to either to a treatment group (will receive the Credit with Education Program) or to a control group (who will not receive the program). They were not randomly assigned; three were assigned for political reasons and three were assigned as matched controls.

Sampling Within the Communities: Three groups of women with children were surveyed: those who participated at least one year (all participants were selected); those who did not participate but were in the program communities (random sample); and those in control communities (random sample). In all, ten mother/child pairs with children aged 12-23 months were chosen from each of the small communities; 30 from the large communities.

Non-Random Sampling

When random sampling is not possible, you will need to use a different approach. Non-random sampling approaches enable you to use a group smaller than the total population. While there are many names for different non-random sampling techniques, they all are limited in their ability to generalize your findings back to a larger population. Even when you have a non-random sample (limited to a particular school, for example), you can still use random sampling within it to make your evaluation more credible.

Types of Non-Random Samples

quota: a sample in which a specific number of different types of units are selected. For example, you may want to interview 10 teachers and decide that five will be men and five will be women.

snowball: this type of sampling is used when you do not know who or what you should include. Typically used in interviews, you would ask your interviewees who else you should talk to. You would continue until no new suggestions are obtained.

judgmental: in this kind of sample, selections are made based on pre-determined criteria that, in your judgment, will provide the data you need. For example, you may want to interview primary school principals and decide to interview some from rural areas as well as some from urban areas (but no quota is established).

convenience: in this type, selections are made based on the convenience to the evaluator. Principals from local schools may be selected because they are near where the evaluators are located.

When using a non-random sample, examining the issue of bias is important. Is there something about this particular sample that might be different from the population as a whole? You might want to gather demographic information so you can describe the characteristics of your sample. Ideally, there will be no obvious differences between the sample and the population. But, when you report the demographics of the sample, your audience can then make a judgment as to how similar the sample is to the population.

When using a non-random sample, you need to report your results in terms of the respondents. For example, "Of the mothers interviewed, 70% are satisfied with the quality of the healthcare their children are receiving." Without random sampling, you have to be careful about generalizing to a larger population. However, the data may be very useful and may be the best given your situation. Always make your sample selection criteria and procedures clear.

Combinations

Random and non-random methods can be combined. For example, you may be collecting data on schools. You can select two schools from the poorest communities and two from the wealthiest communities. Then from these four schools, you can randomly select students for your data collection.

How Confident and Precise Do You Need to Be?

Even when you use a random sample, there is some possibility of error. It is possible that your sample will be different from the population. This is where statistics come in (see Module 11, *Data Analysis and Interpretation*). The narrowest definition of **statistics** concerns the validity of data *derived* from random samples. More specifically, it is concerned with estimating the probability that the sample results are representative of the population as a whole. Statisticians have developed theories and formulas for making these estimates and selecting sample size. While we will present some statistics in the next module, we will not present or discuss statistical formulas here. Rather, we will focus on understanding the basic concepts of statistical analysis, and how to apply them to designing evaluations.

You have some options in deciding how accurate and precise you need to be in inferring results to the larger population. The first thing you want to do is decide how **confident** you wish to be that your sample results are an accurate estimate of what is true for the entire population. The standard **confidence level** is 95%. This means you want to be 95% certain that your sample results are an accurate estimate of the population as a whole. If you are willing to be 90% certain, your sample size will be smaller. If you want to be 99% confident (only 1% chance of having the sample be very different from the population as a whole), you will need a larger sample.

Your next choice is about how **precise** you need your estimates to be. This is sometimes called **sampling error** or **margin of error**. We often see this when results from polls are reported. For example, you might read in the paper that 48% favor raising taxes and 52% oppose raising taxes (+/- 3%). What this means is that if everyone in the population were asked, the actual proportions would be somewhere between 45% to 51% (48 +/- 3) favoring raising taxes, and 49% to 55% (52 +/- 3) _ opposing. Most evaluations accept a sampling error of 5%. In the tax example, if we had a 5% margin of error, than the true picture of opinions would be between 43% to 53% favoring raising taxes and between 47% to 57% opposing raising taxes. As you can see, there is more variability (less precision) in our estimates with a +/-5% margin of error as compared to a +/- 3% margin. The more precise you want to be, the larger your sample will need to be. In both examples, however, note that the sampling errors overlap. This means that these results are too close to call.

When working with real numbers, such as age or income, precision is presented in terms of the **confidence interval**. (Note: this is not to be confused with the concept of confidence level explained earlier). We use this when we want to estimate the mean of the population based on our sample results. For example, if the average per capita income of the rural poor in our sample is 2,000 South African Rand per year, the computer might calculate a 95% confidence interval as between R1,800 and R2,200. We can then say that we are 95% certain (this is the *confidence level*) that the true population's average salary is between 1,800 and 2,200 (this is the *confidence interval*).

How Large a Sample Do You Need?

Sample size is a function of size of the population of interest, the desired confidence level, and level of precision. You can calculate a formula to determine the appropriate sample size or you can use a tool such as Table 9.2, a Guide to Minimum Sample Size. This table shows the sample size needed when estimating a population percentage (or proportion) at the 95% confidence level and a + or – 5 percentage point confidence interval. As you can see, the smaller the population, the higher proportion of cases you will need. If your population is 300, you need a sample size of 169 – just over half the total population – to obtain a confidence level of 95%. A population of 900 will require a sample size of 269 – just under a third. When the population is larger than 100,000, 385 will be needed – a much smaller fraction.

Table 9.2: Guide to Minimum Sample Size (95% confidence level, +/- 5% margin of error)

Population Size	Sample Size	Population Size	Sample Size
10	10	550	226
20	19	600	234
40	36	700	248
50	44	800	260
75	63	900	269
100	80	1,000	278
150	108	1,200	291
200	132	1,300	297
250	152	1,500	306
300	169	3,000	341
350	184	6,000	361
400	196	9,000	368
450	207	50,000	381
500	217	100,000+	385

Source: R. V. Krejcie, and D. W. Morgan, "Determining Sample Size for Research Activities", *Educational and Psychological Measurement, Vol. 30:* 607-610, 1970

Note that these are minimum sample sizes. Whenever possible, you should select a larger sample size to compensate for the likelihood of a lower than 100% response rate.

However, low response rates always have the threat of non-response bias. Over-sampling cannot control for this. No amount of over-sampling can control response bias if you have a low response rate (e.g., only 20%). Rather than over-sampling, the evaluator should put extra resources into doing everything possible to obtain a high response rate. Such efforts should include incentives and multiple follow-ups to non-respondents.

While samples are used to keep costs of data collection down, go for as large a sample as you can manage. This will make your estimates of the population as accurate as possible. If you can do the entire population, opt for that, because there will then be no sampling error involved. However, keep mind that censuses can also yield biased data if there are low response rates.

Summary of Sampling Size

- Accuracy and precision can be improved by increasing your sample size. In other words:
  - By increasing sample size, you increase accuracy and reduce margin of error.
- The standard you should aim for is a 95% confidence level and a margin of error of +/- 5%.
- The larger the margin of error, the less precise your results will be.
- The smaller the population, the larger the needed ratio of the sample size to the population size (See Table 9.2)

Table 9.3 gives you a summary of sample sizes for very large populations (those of 1 million or larger). Many national surveys use samples of about 1,100 because that makes the margin of error + or –3 percentage points with a 95% confidence level.

Table 9.3: Sampling Sizes for Large Populations

Precision (margin of error, +/-%)	Confidence Level		
	99%	95%	90%
± 1%	16,576	9,604	6,765
± 2%	4,144	2,401	1,691
± 3%	1,848	1,067	752
± 5%	666	384	271

Tables 9.4 and 9.5[1] show the confidence intervals for two population sizes (100 and 50): a few samples sizes and various proportions found in the sample. Table 9.4 shows confidence levels for populations of 100.

Table 9.4: 95% Confidence Intervals for a Population of 100.

	Proportion Found in the Sample				
Sample Size	.5	.4 or .6	.3 or .7	.2 or .8	.1 or .9
75	± .06	± .06	± .05	± .05	± .03
50	± .10	± .10	± .09	± .08	± .06
30	± .15	± .15	± .14	± .12	± .09

Table 9.4: 95% Confidence Intervals for a Population of 50.

	Proportion Found in the Sample				
Sample Size	.5	.4 or .6	.3 or .7	.2 or .8	.1 or .9
30	± .11	± .11	± .10	± .09	± .07
20	± .17	± .17	± .16	± .14	± .10

[1] Gregg B. Jackson (2005). *Sampling in Development Evaluations*. Presentation at IPDET, July 5 and 6, 2005, p 28.

There are tools available on the Internet that can help you identify the population size you need to achieve your confidence level and margin of error. The following sites are examples of these:

If you want it to determine how your sample size, you can use the Sample Size Calculator from Creative Research Systems at:

> http://www.surveysystem.com/sscalc.htm

The Sample Size Calculator can also help you find your confidence level.

Another calculator for sample size, called the Proportion Difference Power /Sample Size Calculation and can be found at:

> http://statpages.org/proppowr.html

Summary

In this module, you learned about sampling. Review the following checklist. Check those items that you can complete and review those that you cannot.

- ❑ define the sampling concepts, including:
  - population
  - sample
  - sampling frame
  - sample design
  - parameter
  - statistic
  - random sample
- ❑ discuss the differences between random and non-random samples
- ❑ list and define the types of random samples
  - simple random samples
  - stratified random samples
  - multi-stage random samples
  - cluster random samples
- ❑ list and define the types of non-random samples
  - quota
  - accidental
  - snowball
  - judgmental
  - convenience
- ❑ describe the role of statistics to determine confidence and precision
- ❑ describe the guidelines to determine the size of sample that is needed.

Hints for Development Evaluators

Ev

- You can enhance the extent to which your appraisal reflects what is really going on in the real world by using what you know about appropriate sampling techniques and strategies.
- Consider each of the following concepts when deciding upon one or more sampling techniques:
  - random samples
    - simple, stratified, cluster or combination random samples
  - non-random samples
    - quota, accidental, snowball, judgmental, convenience and combination non-random samples
  - confidence and precision
  - sample size.

Hints for Development Evaluation Managers

Mgr

- Go back and consider the evaluation questions again, making sure the evaluator is choosing an appropriate sample strategy that fits with the needed information.
- Consider whether different sample strategies will be needed for different questions.

Quiz Yourself

Answer the following multiple-choice questions to help test your knowledge of the sampling methods evaluators use to collect data, and the guidelines that determine what size sample is required.

You will find the answers to the questions on the last page of this module.

1. Which of the following is a definition of **parameter**?
   a. a subset of the population
   b. the total set of units
   c. a characteristic of a population
   d. a characteristic of a sample
2. Which of the following is a **list of the types of random samples**?
   a. simple, stratified, cluster
   b. quota, stratified, snowball
   c. cluster, accidental, quota
   d. stratified, convenience, cluster
3. Which of the following is a definition of **quota sampling**?
   a. selecting source of data from whoever walks by
   b. selecting source by setting criteria to achieve a specific mix of participants.
   c. selecting source by setting number from each subunit within the population
   d. selecting source from whoever is easiest to contact or whatever is easiest to observe
4. Which of the following is a definition of **snowball sampling**?
   a. selecting source of data from whoever walks by
   b. selecting source by setting criteria to achieve a specific mix of participants.
   c. selecting source by setting number from each subunit within the population
   d. selecting source by asking people chosen, who else you should interview

5. Which of the following is a definition of **judgmental sampling**?
   a. selecting source by asking people chosen, who else you should interview
   b. selecting source by setting criteria to achieve a specific mix of participants.
   c. selecting source by setting number from each subunit within the population
   d. selecting source from whoever is easiest to contact or whatever is easiest to observe
6. List the summary of important information you should know about sampling size.

Reflection

Think back about previous evaluations with which you have been involved and possible evaluation you have now or will in the future.

- What difficulties did you have selecting samples and how would what you have learned help you in the future?
- Which techniques for sampling data have worked best for you and why?
- What ways can you use to improve the confidence and precision of data collection in your situation?

Application Exercise 9-1
Using a Random Number Table

Instructions:

You have been asked to select 20 case files from a group of 300 urban and rural men and women. All the files have been filed according to numbers from 1-300. List three different ways you could use the random number table on the next page to select 20 case files for review. Under what conditions would you use each one?

Strategy 1:

Use if:

Strategy 2:

Use if:

Strategy 3:

Use if:

Finally, list the cases you would select using the simplest random sampling strategy:

____ ____ ____ ____ ____ ____ ____ ____ ____ ___

____ ____ ____ ____ ____ ____ ____ ____ ____ ___

Table of Random Numbers for Exercise 9-1[2]

44	14	12	12	03	12	73	72	62	33	35	62	80	34	77
69	59	54	90	01	50	04	93	76	69	43	95	47	60	80
23	95	24	95	24	55	69	89	41	18	12	94	43	21	43
40	76	50	38	18	05	44	23	72	61	58	67	99	05	75
54	05	51	52	04	34	25	64	90	95	02	86	51	14	37
36	82	03	65	38	93	49	64	06	93	01	30	62	05	68
96	19	97	24	16	26	94	14	17	45	22	51	09	92	16
75	85	18	50	50	60	80	52	42	11	05	70	89	53	38
57	78	12	98	55	51	48	77	54	07	66	15	33	44	64
58	20	10	51	62	06	25	56	63	67	73	73	79	05	65
55	84	17	67	52	38	16	29	05	24	12	05	35	87	31
92	44	84	04	17	47	18	78	54	40	02	59	74	06	73
86	96	79	86	75	67	31	41	40	20	87	17	85	98	70
78	84	03	69	43	38	43	98	90	75	56	49	88	52	78
25	05	76	72	06	59	37	56	24	36	95	05	30	62	02
26	67	04	13	77	37	21	57	77	41	82	30	32	80	09
87	50	88	39	26	92	72	07	07	79	64	30	91	94	34
44	56	72	31	73	50	31	60	61	92	25	83	10	59	12
42	38	35	20	97	80	90	15	96	10	36	26	08	69	40
72	01	08	61	49	52	89	23	29	43	66	48	23	50	79
11	87	48	55	60	23	16	06	63	49	58	12	02	87	92
54	50	07	21	42	60	57	63	64	82	31	00	76	35	49
09	26	43	03	52	49	89	85	48	15	78	10	98	87	90
27	16	06	60	48	24	82	63	16	18	59	66	20	02	70
09	27	37	97	64	95	95	61	84	52	20	96	33	60	12
42	31	92	39	40	06	66	32	71	29	13	48	31	76	43
07	61	35	67	17	45	47	39	40	06	10	85	24	93	91
10	16	62	33	78	48	06	83	22	55	84	21	44	10	67
27	76	80	20	44	27	62	56	95	92	55	31	57	47	74
58	27	64	85	98	91	58	06	70	18	35	35	58	08	41

2 Generated using Microsoft Excel.

Application Exercise 9-2
Sampling Strategy

Instructions: working in small groups if possible, identify an appropriate measure or statistic for each evaluation question, and decide what sampling strategy you would use for each of these situations, and why.

1. How would you determine the quality of the roads in villages in a particular region of Cambodia, immediately after the rainy season?

   Measure:

   Sampling strategy (and reasoning):

2. How would find out what proportion of children contract at least one case of malaria before the age of 10 in a specific region of India?

   Measure:

   Sampling strategy (and reasoning):

Further Reading and Resources:

Guba, E. and Lincoln, Y.S. (1989). *Fourth generation Evaluation*. Beverly Hills: Sage Publications.

Henry, G. T. (1990). *Practical sampling.* Thousand Oaks, CA: Sage Publications.

Kish, L. (1995). *Survey sampling.* New York: John Wiley & Sons.

Kumar, R. (1999). *Research methodology. A step-by-step guide for beginners.* London: Sage Publications.

Laws, S. with Harper, C. and Marcus, R. (2003). *Research for development: A practical guide.* London: Sage Publications.

Levy, P. and Lemeshaw, S. (1999). *Sampling of populations (3rd edition).* New York: John Wiley & Sons.

Lipsey, M.W. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage Publications.

Lohr, S. (1998). *Sampling: Design and analysis.* Pacific Grove, CA: Duxbury Press.

Neuman, W. Lawrence (2006). *Social research methods—qualitative and quantitative approaches, 6th edition.* Boston: Allyn and Bacon.

Patton, M.Q. (2002). *Qualitative research and evaluation methods.* Thousand Oaks, CA: Sage Publications.

Scheyvens, R. and Storey. D. (eds.) (2003). *Development fieldwork: A practical guide.* London: Sage Publications.

Tryfos, P. (1996). *Sampling methods for applied research.* New York: John Wiley & Sons.

Websites:

Easton, V.J., & McColl, J.H. *Statistics glossary: Sampling.*
http://www.cas.lancs.ac.uk/glossary_v1.1/samp.html

Probability Sampling
http://www.socialresearchmethods.net/kb/sampprob.htm

Research Randomizer
www.randomizer.org

Dr. Drott's Random Sampler
http://drott.cis.drexel.edu/sample/content.html

Power Analysis
www.statsoft.com/textbook/stpowan.html

HyperStat Online: Chapter 11: Power
http://davidmlane.com.hyperstat/ch11_contents.html

UCLA Statistics Calculator:
http://calculators.stat.ucla.edu

The Survey System: Sample Size Calculator
www.surveysystem.com/sscalc.htm

Survey Research Methods Section
www.fas.harvard.edu/~stats/survey-soft/survey-soft.html

Web Pages That Perform Statistical Calculations
www.StatPages.net

Answers to Quiz Yourself

1. c

2. a

3. c

4. d

5. b

6. Summary of Sampling Size

   - Accuracy and precision can be improved by increasing your sample size. In other words:
     – By increasing sample size, you increase accuracy and reduce margin of error.
   - The standard you should aim for is a 95% confidence level and a margin of error of +/- 5%.
   - The larger the margin of error, the less precise your results will be.
   - The smaller the population, the larger the needed ratio of the sample size to the population size (See Table 9.2)

To continue on to this
Module's Presentation Slides click here

To continue on to the
Next Module click here

To return to the
Table of Contents click here