

SUMMARY OF THE 4-STEP SAMPLING PROCESS

I. INTRODUCTION

The quality and usefulness of results from a survey depend largely on the procedures used to select the clients. Because surveying every client in an agency, region, or state is usually impossible, impractical, and unnecessary, a sample of the entire population should be selected. Samples can be selected in many ways. The results from a good sample can be generalized to the entire client population from which the sample was drawn. The results from a poor sample only refer to the clients who participated.

The two main factors contributing to a good sample are sample size and sample selection. Sample size is directly related to the magnitude of the error expected in the results. Any estimate that is based on a sample will have some degree of error in it. The magnitude of the error will reflect your confidence in the statements that you make about the population based on your sample results. Typically prior to conducting a survey, an agency would decide on the maximum amount of error allowed in their estimate of the population. Deciding on the margin of error involves marking off the limits within which the population mean probably lies, and ascertaining what that probability is. In technical terms, you would state the level of confidence of your hypothesis that the population mean resides within the specified confidence interval.

The second issue of sample selection is concerned with the method for selecting the clients to be surveyed. To be able to generalize findings to the population, a random sample of clients should be selected. A random sample is one where every member of the population has an equal chance of being included in the sample and each selection is made independently of all the others. A representative sample is not a random sample. An example of a representative sample may involve selecting within a local area agency at least one small, medium, and large provider and then randomly selecting clients from only these selected providers. Since all clients served by the local area agency did not have an equal chance of being selected, the findings can only be used to describe the clients served by the selected providers and not the entire population served by the local area agency.

A summary of the procedure for selecting a random sample is discussed here and is preceded by an explanation of tools provided for this project to assist in sampling. A two phase sample, which may involve randomly sampling a subset of providers and then administering the survey to a random sample of their clients, is a more complex sampling plan that is not discussed here and would not use the procedures discussed here.

II. PROCEDURE

To select a sample of clients from the population the following steps are involved.

1. **Construct a master client list.** Construct a list of all eligible clients in the population from which your sample will be randomly selected.
2. **Decide on the sample size.** The size of your sample is related directly to the size of the eligible population, the estimated response rate, and the desired precision of your results.

The size of the eligible population is determined by the purpose of your survey. For example, if the purpose of the survey is to describe the satisfaction of clients using transportation services, the eligible population may be defined as clients who are 60 years of age or older and have been enrolled in transportation services for at least six months.

The desired precision of your results is another important consideration. Not every eligible client is surveyed, so survey results provide an estimate rather than an exact measure of the population. Thus, you will need to define within what range you can be reasonably sure that the population mean resides. The level of precision is conveyed by providing the survey estimate plus or minus its margin of error (i.e., also referred to as the confidence interval). The exact result should fall within this range for a specified level of confidence.

In administering a survey, it has to be anticipated that a certain percentage of the clients will refuse to participate, cannot be located, or other reasons. To insure that the number of completed questionnaires matches the number needed for the desired level of precision, you should adjust the initial sample size estimate by the percentage of clients you expect will refuse or cannot participate for other reasons. To make this adjustment, you need to estimate the response rates for the survey, which is the percent of clients in a sample that will complete the survey. The response rate usually is estimated from previous surveys conducted by your local agency. For example, the results from a previous survey may show that 80% of the clients called completed the questionnaire and that 20% did not complete the questionnaire. The initial sample size estimate would have increased to accommodate for the 20% that were not completed.

A Sample Size Calculator program has been developed to assist you in determining your sample size requirements. This program applies a population correction factor to the sample size estimate, since it is assumed that most agencies are going to be sampling from a small finite population. When working with relatively large populations, this correction factor is not needed. Table 1 highlights how sample sizes vary by population size and precision levels. It is recommended that at a minimum the sample size should be at least 30.

Table 1. Sample Size Estimates Based on Population Size and Confidence Interval*

	Population Size										
Confidence Interval	50	75	100	150	200	250	300	350	400	450	500
+/- 1%	50	74	99	148	196	244	292	339	385	432	477
+/- 2%	49	73	96	142	186	228	269	309	347	384	420
+/- 3%	48	71	92	133	171	206	239	270	298	325	350
+/- 4%	47	67	87	122	154	181	206	229	249	267	284
+/- 5%	45	64	81	111	136	157	175	191	205	218	229
+/- 6%	43	60	75	99	119	135	148	160	169	177	185
+/- 7%	41	56	68	89	104	116	125	133	140	146	150
+/- 8%	39	52	62	79	91	99	106	112	117	121	124
+/- 9%	36	48	57	70	79	86	91	95	98	101	103
+/- 10%	34	44	52	62	69	74	78	81	83	85	87

	Population Size										
Confidence Interval	550	600	650	700	750	1,000	1,500	2,000	5,000	10,000	15,000
+/- 1%	523	568	612	656	700	913	1,313	1,681	3,390	5,128	6,186
+/- 2%	455	489	521	553	584	725	956	1,137	1,724	2,083	2,239
+/- 3%	374	397	418	438	457	539	657	738	948	1,047	1,085
+/- 4%	300	314	327	339	351	397	458	495	581	617	630
+/- 5%	239	248	256	263	270	297	329	348	388	404	410
+/- 6%	191	197	202	206	211	226	245	255	276	284	287
+/- 7%	155	158	162	165	167	177	188	194	206	210	212
+/- 8%	127	129	131	133	135	141	148	152	159	162	163
+/- 9%	105	107	108	110	111	115	120	122	127	128	129
+/- 10%	88	90	91	92	92	95	98	100	103	104	105

* Table based on a 95% confidence level and a 95% responses rate.

3. **Selecting Clients.**

There are two popular methods for randomly selecting a sample of clients. The simple random sampling method selects clients from a list of all clients based on a random number assignment. The systematic sampling method selects clients based on a sampling rate, such as, every fifth person who enrolls in the program will be included in the sample and given the questionnaire.

The method selected will depend on several factors such as how the information on the clients is stored (i.e., paper logbooks versus a computer database) and whether the population of clients is known prior to data collection. In some instances, no random selection method may be needed; instead, all clients in the population of interest would be administered the questionnaire.

Example of Simple Random Sampling in Microsoft Excel

The simple random sampling method would work best for agencies who want to sample from a known list of clients. This list would contain all the names of the clients in the population of interest, e.g., all transportation clients who are 60 years of age or older and enrolled for at least 6 months. The steps for selecting clients to be included in the sample would involve (1) assigning a random number to every client in the list, (2) sorting the list in ascending order by the random number, (3) selecting the first “*N*” records, where “*N*” is the sample size estimate, and (4) assigning these records to the sample list. These selected records in the sample list will be the clients who are contacted and administered the questionnaire.

An example of using Microsoft Excel to select a sample is given here. We'll assume that we have a population of 600 transportation clients and had to obtain a sample of 90 from this population. The first step would be to import your list of clients into Excel, such that each client is assigned to one row in the spreadsheet. Next would be to assign a random number to each client by using the RAND function, i.e., create a new column and set each cell in this column to the value “=RAND()”.¹ Next, place your cursor on the first name in the list and select the sort function under the data menu. The sort function will automatically highlight all 600 names in your list, just verify that it contains all the records you want to sort. Under the sort menu, it will ask you which column you want to sort by, select the column with the random number. You can sort in ascending or descending order, it doesn't matter. Once you resort the data, select the first 90 clients in the list for your sample. Since the whole list is now randomly sorted, you could select the first 90 clients on the list or the last 90 or a group of 90 in the middle for your sample.

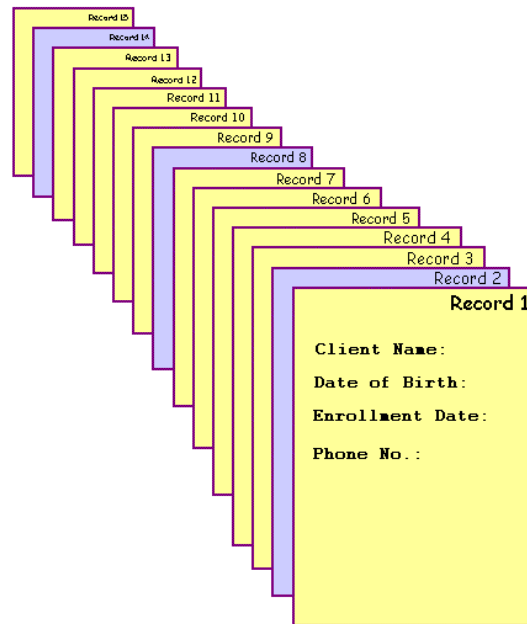
¹ Once you assign a random number to every client, you might want to hard code the value, since the RAND() will change its value every time you make a change or edit to the workbook. To hard code the value, highlight the random number column, copy it to your clipboard, keep the random number column highlighted, and perform a paste special command that will only paste the value of the cells.

Systematic Sampling Method

The systematic sampling method or the sampling rate method is useful for sampling historical records that are in paper logbooks or for sampling clients from an unknown population, such as persons calling for I&R service during the next two months. This method determines a sampling rate that will be used to select clients for the sample. For example, the sampling rate may indicate that every fifth person calling for information should be included in the sample.

Systematic sampling is a fairly easy procedure. The first step involves determining the sampling rate by dividing the number of clients in your population by the required sample size. For example, if the population is 600 transportation clients and if the sample size is 90, then the sampling rate is 600 divided by 90 or 6.7, which should be rounded down to every 6th client. Once the sampling rate is determined, you need to determine a starting point. Starting with the first client on the list is not recommended; instead, randomly select a number between one and the sampling rate number (e.g., a number from 1 to 6). This will be the starting point and the first client selected for the sample. To select the next client add the sampling rate to the previous record number. For example, if 2 is the starting record, then the next client selected would be 2 + 6 or the eight record and the following would be 8 + 6 or the 14th record. This process is continued until the end of the list of clients. The following illustrations highlight this process for selecting names on a list or out of a stack of records.

Transportation Clients		
Client ID	Client Name	Phone No
1101	Cathy Crispin	212-8268
✓1103	Charles Irwin	212-4967
1104	Billie Bob	214-5188
1106	Sherlie Temple	214-6299
1109	John Rogers	213-0210
1110	Gloria Stevens	212-2211
1111	Victor Hugo	212-1661
✓1112	Paul Thomas	212-7898
1113	Neil Carter	214-5090
1114	Clara Barton	212-1231



4. **Document your sample.** The final characteristic of a good sample is adequate documentation of the sampling process. It should summarize the three previous steps, and include information on the size of the population of interest, the size of the sample, and if possible the demographics of the population and the sample.