

How to use Sampling to Reduce the Volume of Participant Evaluation Surveys You Collect and/or Enter

A brief introduction to Simple Random and Stratified Cluster Sampling

A sample is a portion of a larger group. In program evaluation, a sample is usually a group of program participants that participate in an evaluation activity, such as a survey. Sampling is typically used to select a feasible (smaller) number of evaluation participants that will represent all program participants. There are many different kinds of sampling. Refer to the “*Sampling Methods Commonly Used in Program Evaluation*” Tip Sheet for descriptions about various approaches. This tip sheet focuses on cluster sampling because it can be very useful for reducing the number of surveys you need to collect (or enter), particularly for classroom-based programs.

This strategy is applicable in the following situations:

- You are evaluating a program using surveys (pre/post-test, post-only, or retrospective post-test surveys)
- The program has 100 or more participants (If the program has fewer than 100 participants, you should collect/enter surveys for all participants.)
- You have limited resources for data entry and are looking for an efficient way to reduce the volume of surveys you process.

When should I make my sampling plan?

Ideally, you will develop a sampling plan before you collect your surveys. But, if you have already collected your surveys, you can still use sampling to reduce the number of surveys you actually data enter and analyze.

Step 1: Make Your List (develop your sampling frame)

In order to do sampling, you need to have a “sampling frame”—the list you will use to draw your sample. This could be a list of individual participants, or of groups of participants (also known as “elements”). For most prevention education programs, it makes sense to have a list of groups of participants, such as classrooms or implementation sites (these groups are known as “clusters”). It is preferable to have a larger number of small clusters (such as 20 classroom groups) than to have a small number of larger clusters (such as 4 schools). Number your list (numbering is necessary if you are using random number table or random number generator). Here’s an example of classroom groups numbered 1-5:

Step 1 Example: A Sampling Frame of Classroom Groups

Safe Dates classroom groups	Number of participants
Acme High School	
1. Mrs. Walters	31
2. Ms. Moreno, 2 nd period	27
3. Ms. Moreno, 4 th period	32
Acme Middle School	
4. Mr. Smith	29
5. Mrs. Jackson	27
Etc.	

Step 2: Determine how many surveys you need (calculate your sample size)

The easiest way to calculate your sample size is to use an online calculator such as the one provided on the Creative Research Systems website: <http://www.surveysystem.com/sscalc.htm#one>

Use the “Determine Sample Size” box:

- Check the “95%” button (95% confidence level)
- Enter “5” into the “confidence interval” box
- Enter your total number of program participants in the “population” box (in the context of program evaluation, the “population” refers to program participants)

The calculator will then provide you with a number in the “sample size needed” box. That’s the total number of *completed* surveys you will need to have to adequately analyze your data. In the example below, we have 500 participants, but only need to collect or enter surveys for 217 of them.

Step 2 Example: Calculating sample size online

Determine Sample Size		
Confidence Level:	<input checked="" type="radio"/> 95%	<input type="radio"/> 99%
Confidence Interval:	5	
Population:	<input type="text" value="500"/>	
Sample size needed:	217	

Step 3: Pick your groups (draw a random sample of clusters)

Now that you have your list of participant groups and the target number of surveys you need to collect or enter, you are ready to actually draw your sample. In order to minimize bias, you should draw a *random* sample—meaning that all the groups on your list have an equal chance of being picked. The traditional way to do this is by using a random number table, or by using an online random number generator, such as <http://stattrek.com/Tables/Random.aspx>. Instructions for using a random number table are provided at the end of this document.

The fastest and easiest way to do draw your sample is to use the List Randomizer Function on Random.org: <http://www.random.org/lists/>. To use this website function:

- Enter your list in the “Part 1” box, with each item/group on a separate line
- Click on the “Randomize” button
- Start at the top of the list and work your way down until you have enough surveys to meet your Sample Size (the result of Step 2).

In the example below, you would plan to survey students in the first 8 randomly selected classroom groups only—Mrs. Valle through Mr. Hamilton. If you had already collected surveys from all classroom groups, then you would enter the surveys from the first 8 groups, and set aside surveys from Ms. Cline onward.

Step 3 Example: Using List Randomizer to Select Classroom Groups

List Randomizer

There were 16 items on your list. Here they are in random order:

1. Mrs. Valle (27)
2. Ms. Moreno, 4th period (32)
3. Ms. Moreno, 2nd period (27)
4. Mrs. Jackson (31)
5. Mr. Smith (29)
6. Mrs. Crouch (35)
7. Mr. Hamilton, Health Class (30)
8. Mr. Hamilton, afterschool program (6)
- >>> **Stop here. Groups 1-8 add up to 217 participants.**
9. Ms. Cline
10. Etc.

Special considerations and additional information

If there are specific groups you want to make sure you include...

Sometimes you may want to make sure that you are analyzing information about a specific group, such as a specific school or grade level or gender. When you do random sampling, it is possible that you may not get adequate representation from some key groups that you are particularly interested in. In order to avoid this issue, you can use “**stratified cluster sampling**”—meaning that you first categorize (or “stratify”) your groups of participants (“clusters”) by the characteristic you are interested in (or, the “stratification variable”).

For example, next year you have to decide if you are going to implement Safe Dates in middle or high schools. This year you have surveys from both levels, and you are very interested in comparing results for 8th graders and 11th graders. In the “Step 3 List Randomizer Example” above, all but one of the 8 classroom groups you selected are 11th grade classrooms. In order to avoid this issue, and to ensure that you have enough surveys from 8th graders, you could stratify your classroom groups by grade level. To do this, you would create two separate sampling frame lists—one for 8th grade classrooms and one for 11th grade classrooms. You would then repeat Steps 1-3 for each list separately.

Example: Sampling Frames for Stratified Cluster Sample

Cluster Strata #1: 8 th grade classroom groups	Cluster Strata #2: 11 th grade classroom groups
1. Mr. Smith	1. Mrs. Jackson
2. Mrs. Valle	2. Ms. Moreno, 2 nd period
3. Ms. Cline	3. Ms. Moreno, 4 th period
4. Ms. Stevens	4. Mrs. Crouch
5. Ms. Williams	5. Mr. Hamilton, Health Class
6. Mr. Porter	6. Mr. Hamilton, afterschool program
	7. Dr. Ortega
Population= 200 students	Population= 300 students
Sample size needed=132	Sample size needed=169

If you are developing your sampling plan before collecting your surveys...

Ideally you will know ahead of time how many program participants you will have. In the case of classroom presentations, school staff can tell you how many students are in each classroom, section, or group. In reality, however, there may not be as many participants as expected. It's important to keep this in mind when developing your sampling plan so that you will make sure you end up with enough completed surveys. This is called "**anticipating non-response.**" To compensate for this, estimate ahead of time the percent of participants you think will not complete the survey. For example, you may assume that 10% of students will be absent from the classroom on the day of the post-test (90% present). If your desired sample size is 200, then you would actually attempt to get surveys from 222 participants (200/.90; desired sample # divided by anticipated percent present).

If you would like to learn how to do more sophisticated sampling techniques...

There are many different kinds of sampling techniques. If you have large numbers of participants, complex evaluation questions, and/or a high capacity for conducting research in your organization, you may want to explore additional concepts, such as proportionate and disproportionate stratified sampling (for focusing in on special populations), multi-stage sampling (for several different stratification variables), and random sampling of individual participants (instead of clusters). You can also choose to use different confidence intervals and levels. The following resources are available online and provide additional information about sampling:

Research Methods Knowledge Base: Sampling:

<http://www.socialresearchmethods.net/kb/sampling.php>

Stat Trek Statistics Tutorial: Survey Sampling Methods:

<http://stattrek.com/AP-Statistics-2/Survey-Sampling-Methods.aspx?Tutorial=Stat>

For a general overview of sampling, including probability and non-probability techniques, see University of Wisconsin-Extension site:

<http://learningstore.uwex.edu/Sampling-P1029C237.aspx>

References: Above websites, and Wolfer, Lauren. *Real Research: Conducting and evaluating research in the social sciences*. Pearson Education, Inc. 2007.