FACTSHEET V

This Factsheet addresses the following more advanced topic in the area of sampling:

- Stratified sampling—what it is, when to use it, how to do it.

This Factsheet also provides

- A crosswalk between the customized terminology with regard to customer feedback sampling used in these Guidelines and Factsheets and the more general statistical terminology used by survey statisticians.
- A discussion of some other kinds of errors (beyond sampling error and nonresponse bias) that will be encountered in sampling.
- A discussion of sampling without replacement and how this differs from sampling with replacement. This section also contains a description of how to randomly select a sample without using a computer or a computer spreadsheet.

STRATIFIED SAMPLING

In the section of the Guidelines describing how to analyze the data obtained from a sample of customers, an example is given on pages 39–40 of a simple procedure that can be used to determine if the degree of satisfaction of your customers varies among different kinds of customers. The procedure presented there is simple and useful and will be quite satisfactory for use in conjunction with most customer satisfaction surveys conducted at EPA.

There is, however, one aspect of that procedure that should be noted: The sample results obtained for the different specific kinds of customers (e.g., educators and advocates in the example on pages 39–40) are not as precise as the results obtained for all kinds of customers in the sample, considered all at once. The reason for this is that only a portion of the sample is relevant to each specific kind of customer. Therefore the sampling error determined for the entire sample does not apply to each specific kind of customer. Instead, in effect, we have a smaller sample size for each specific kind of customer, and there will therefore be higher sampling error for each specific kind of customer, when considered separately.

Again, for most customer satisfaction surveys conducted by EPA, the results obtained from the procedure shown on pages 39–40 will still be quite satisfactory and the increased sampling error associated with each specific kind of customer should not be of concern. The results obtained from using those results will still be meaningful and useful.

There may, however, be some situations in which it is so important to track the degree of satisfaction of one or more specific kinds of customers that you decide that the sampling error for each specific kind of customer of concern must be kept within specified limits. In such cases, a specialized procedure known as *stratified sampling* may be used.

The basic principle underlying stratified sampling is very simple. For each subgroup of customers for which it is important to know their degree of satisfaction with a specified level of precision (i.e., with a known maximum level of sampling error), the sample size for that subgroup should be determined separately and a random sample of those customers selected separately. This can be done by applying the procedures presented in the Guidelines and in **Factsheets III** and **IV** *separately* for *each* of these subgroups of customers. It is as though you are no longer conducting one survey but instead are conducting two or three (or more) surveys simultaneously, one for each subgroup of customers who are so important that their degree of satisfaction must be known and tracked with a known maximum level of sampling error for that specific subgroup.

The results are then analyzed separately for each of these different subgroups, again using the methods of analysis that are presented in the Guidelines. The results of these analyses are then used to track separately the degree of satisfaction of each of these subgroups of customers.

The results from each of these subgroups may then also be combined to give an overall result for all the subgroups surveyed taken together. They may be combined by using the following formula:

$$p_{all} = (f_1 \times p_1) + (f_2 \times p_2) + (f_3 \times p_3) + ...$$

Where the equation continues for as many subgroups of customers as were used in the survey and where

p_{all} = the sample result for all customers served

p_1 = the sample result for the first subgroup of customers (i.e., the number of customers in the first subgroup who reported being satisfied)

f_1 = the fraction (percentage) of all customers served who fall in the first subgroup of customers

p_2 = the sample result for the second subgroup of customers

f_2 = the fraction (percentage) of all customers served who fall in the second subgroup of customers, etc.

For the above formula to give an accurate result for all customers, every customer must be included in one (and only one) of the subgroups. If, for example, there are five different kinds of

customers and only two kinds are so important that they have to be tracked separately with a known level of sampling error, then the remaining three kinds of customers can be included in a third subgroup consisting of the remaining three kinds of customers grouped together.

There is one further consideration. Since you have set the sample size separately for each of the subgroups in order to get a known level of sampling error for each of those subgroups, you will know the level of sampling error for each subgroup, but you do not know what the level of sampling error is for any sample results obtained (using the formula given above) for all the customers taken together. There is, however, a second formula that can be used to determine this:

E_{all} = the square root of the following sum

$$\text{sum} = \frac{N_1^2}{N^2} \times \frac{N_1 - n_1}{N_1 - 1} \times \frac{E_1^2}{n_1} + \frac{N_2^2}{N^2} \times \frac{N_2 - n_2}{N_2 - 1} \times \frac{E_2^2}{n_2} + \text{etc.}$$

Where

E_{all} = the sampling error for sample results applicable to all customers taken together

N = the total number of customers served

E_1 = the sampling error for the first subgroup of customers

N_1 = the total number of customers served for the first subgroup of customers

n_1 = the sample size for the first subgroup of customers

E_2 = the sampling error for the second subgroup of customers

N_2 = the total number of customers served for the second subgroup of customers

n_2 = the sample size for the second subgroup of customers

etc. (for each additional subgroup of customers served)

Note that what is referred to as *sampling error* in this section on stratified sampling is actually the *standard error* since it was values obtained from the sample that were used in computing it. The

standard error is used by statisticians to *estimate* the sampling error. (See the next section of this Factsheet for further clarification of this point.)

A GUIDE TO THE APPLICABLE STATISTICAL TERMINOLOGY

In the discussion of surveys and sampling in the Guidelines and Factsheets, we have used terminology customized to the special circumstances of surveys conducted by EPA to assess customer satisfaction and have modified some other statistical terminology to make it easier for nonstatistician to understand. Some users of the Guidelines may, however, wish to consult text books, reference books or journal articles on one or more aspects of sampling. For that reason, we here provide a crosswalk between the terminology used here and the more general terminology used in general works and articles on sampling procedures.

Terminology used in these Guidelines and Factsheets	The more general terminology used by sampling statisticians
1. "The customers served in a specific period of time" (for which information about their level of satisfaction is being sought) or "The customers served" or "The target group"	1. "The target population" or "The population" or "The universe"
2. "The sample of customers" or "The customers in the sample"	2. "The sample"

3. "The sample result" (shown as the percentage of customers who responded who said that they were satisfied in response to a *specific question* about some aspect of the product or service they received)	3. "The sample proportion"
4. The "master list" of customers served	4. "The sampling frame"
5. "The sampling error" values *obtained from the full target*	5. "The sampling error" (when calculated using *population*—which is not possible under normal circumstances because it would be too costly since it would require a census of the target population. And it is to keep costs to a reasonable level that we are using a sampling procedure in the first place.) or "The standard error" (when calculated using values *obtained from the sample*—which are what we must use in most cases. This is so since the corresponding values for the entire target population are not known and cannot be learned without going to great additional expense. Going to such expense would defeat the purpose of using a sampling procedure rather than a census). "The standard error" is also known as "the standard error of the mean."
6. "The unit of analysis"	6. "The unit of analysis" or "The sampling unit"

7. "The 95% confidence level"	7. "A 95% confidence level" By tradition, the article "a" is used, which suggests that there is more than one kind of 95% confidence level. In fact there is only one kind of 95% confidence level, so "the" is the more appropriate article, and "the" is used in these Guidelines to improve clarity for the sake of the nonstatisticians seeking to make use of them.

OTHER KINDS OF ERROR EXPERIENCED IN SAMPLING SURVEYS

We have so far limited our discussion of error encountered in sampling surveys to three kinds of error: 1) *sampling error*, 2) *nonresponse bias*, and 3) the bias associated with use of a poorly chosen master list of the persons in the target group to be sampled. Each of these has been discussed earlier: sampling error was discussed in **Factsheet III**; nonresponse bias and the bias associated with use of a poorly chosen master list were discussed in the first section of this Factsheet. We will now describe briefly two other kinds of error that occur in sampling surveys:

- ***Reporting error.*** This is the kind of error that results when the customer misunderstands the question asked and therefore gives an incorrect answer. It can also occur if the customer misunderstands how to use an interval scale (by thinking, for example, that a high number on the scale means "highly unsatisfied" when in fact it means "highly satisfied"). It can also occur when a customer purposely gives a wrong answer.

  This kind of error can be kept to a minimum by pretesting the questionnaire on people who are similar to those who will be surveyed. Such pretesting can help identify: 1) questions that are confusing, 2) clear instructions on how to respond (e.g., how to use the interval scale to answer a question) to questions that are confusing, and 3) questions that customers may find so intrusive, threatening or offensive that they may prompt some customers to give a false answer.

- ***Recording error.*** This is error on the part of the staff conducting the survey. Even if the customer has responded correctly, the survey staff may misunderstand or misrecord what the customer said (in a telephone or in-person survey) or may misread or misrecord what was written (in a mail survey).

It should, however, be noted that, unlike sampling error, which results from the use of a probabilistic sampling procedure and is an inevitable consequence of using such a procedure, the additional kinds of error identified above will be encountered whenever customers are contacted and asked to respond to specific questions. These kinds of errors would, for example, still occur even if a full census were conducted of all customers served. In other words, use of a census will

eliminate sampling error, but reporting error and recording error will still occur. Similarly, nonresponse bias will occur in the results obtained from a census just as much as in those obtained from a sampling survey. That's what the recent controversy about the Year 2000 Census of the U.S. population has been all about. The concern is that past censuses have had a nonresponse bias that works to the disadvantage of the poor, because those with low incomes or no incomes are disproportionately less likely to participate than those who are more well off economically.

Should the sample be selected with or without replacement?

There is one basic decision that must be made before selecting a random sample of customers from the larger total number of customers served. That decision is, will the selection be made *with replacement* or *without replacement*?

In order to explain what this means, we will describe an alternative procedure for selecting a random sample of customers that tracks closely with that presented in **Factsheet III**. This alternative procedure will, however, be a simpler one for which it is easier to follow the implications of each step. In particular, the process we are about to describe will parallel perfectly the procedures in **Factsheet III**, but the procedure will be described not in terms of entering customer numbers into a spreadsheet on a computer, but rather in terms of putting the customer numbers on slips of paper, putting these slips into a box, and then pulling these slips from the box. This simpler alternative approach will make it clearer what the difference is between sampling *with replacement* and *without replacement*.

Here is the alternative approach for selecting a random sample of customers:

Begin by carrying out the actions called for in steps 1) and 2) of the procedure that begins on page 4 of **Factsheet III**. After having compiled the master list of persons served described in step 2), take the following additional steps:

Here is the alternative approach for selecting a sample of customers from the master list:

a) For each of the customers on the master list, place the number corresponding to that customer on a slip of paper, one customer number per slip of paper. Then fold each slip of paper in a uniform way. All slips of paper used should be identical in every way.

b) Put the folded slips of paper into a box, and shake the box sufficiently that the slips of paper have been well mixed within the box.

c) Have someone begin to remove slips of paper from the box one at a time. While this is being done, the box should be held or positioned in such a way that the person removing the slips of paper cannot see the slips of paper that he or she is picking from.

d) Have a sheet of paper (a recording sheet) ready with numbers running from 1 to a number at least twice the size of the sample size that has been chosen. For example, if the sample size chosen was 65, have numbers on the sheet of paper running up 130. This sheet will be used to record the outcome of the selection process.

e) As each slip of paper is picked from the box, unfold the slip of paper, read the customer number on it, and record that number in order on the recording sheet. Then set that slip of paper aside (or throw it away). The customer number on the first slip picked should be recorded next to the number 1 on the recording sheet. The customer number on the second slip of paper should be recorded next to the number 2 on the recording sheet, and so on. Continue this process until a slip of paper and a corresponding customer number has been picked for each number on the recording sheet.

f) Determine the number of customers to be included in the initial sample by taking into consideration both the desired sample size and the anticipated response rate. Thus, for example, if the desired sample size is 65 and the expected response rate is 85 percent, the number of customers to be included in the initial sample should be 65/85 percent = 76.47. The size of the initial sample should therefore be 77 (since it is always prudent to round fractions up to the next whole number in situations where a certain minimum level must be achieved).

g) The initial sample will then consist of the customers whose customer numbers are recorded next to positions 1 through 77 on the recording sheet. Carry out the survey using these 77 customers.

h) If the response rate in the survey is 85 percent or greater, then nothing further need be done. No further use will need to be made of the recording sheet. If however, the response rate turns out to be less than 85 percent and all reasonable follow-up actions have been taken to increase the response rate and it is still less than 85 percent, then a determination should be made as to how many additional customers need to be added to the sample to bring the number of customers responding up to 65.

   If for example, the number of customers so far responding is 62, then three more responding customers are needed. Since the response rate so far has been 62/77 = 80.5 percent, the minimum additional number of customers who need to be added to the sample is 3/80.5 percent = 3.73 or 4 after rounding up. However, since the there is no guarantee that the response rate for the next set of customers contacted will be identical to that of the customers already contacted, it might be prudent to add not 4 but 5 or 6 additional customers to the sample to ensure that a third survey cycle will not be needed. In this case, the number of additional customers that it is decided to add to the sample is 6.

i) Starting at the point where you left off in taking customer numbers from the recording sheet in step g) above (position 77 in the sample used here), take the customer numbers appearing in

the next 6 positions on the list (i.e., those in positions 78 though 83), and add the customers' names corresponding to these customer numbers to the sample.

The above procedure is an example of sampling *without replacement*. This expression is used because the procedure used called for slips of paper to be picked from the box and after each slip of paper was picked, it was *not replaced* (i.e., put back) in the box.

The alternative procedure would have been to sample *with replacement*. Using the same physical arrangements described above, sampling with replacement would have entailed picking the first slip of paper from the box, unfolding it, reading the customer number off of that slip, then folding the slip back up as it was and putting it back in the box, shaking the box up again, so that all the slips in the box (including the one already removed and put back in into it (i.e., replaced into it) are once again fully mixed. A second slip of paper is then picked from the box, unfolded, the number is read off, the slip is folded again and put back in the box and the box is shaken again. This procedure continues until the same number of slips has been taken from the box as before or, to be more precise, until the same number of picks have been made from the box (in this case, 130).

One of the obvious implications of this new procedure just described (which calls for sampling *with replacement*) is that it is very possible for a single slip of paper to be picked from the box more than once. In fact it could be picked three times or more. If that happens, then that same number is recorded a second time (and if necessary a third time, a fourth time, etc.) on the recording sheet. Let's say for example that the slip with customer number 278 was the fourth slip picked and was also the 36th picked. Then customer number 278 will appear on the recording sheet both at position 4 and at position 36. Slips continue to be picked a total of 77 times as before (i.e., a total of 77 picks are made from the box).

Now, you may well ask, should an additional slip be picked to make up for the duplicate picking of customer number 278? The answer is no. Customer 278 has been picked twice, so he or she now counts as *two customers*. Does this mean that customer 278 will be contacted two different times during the survey and on the second occasion be asked to respond a second time to the survey questions? The answer again is no. Customer 278 will be contacted only once, but his or her response will be used twice in computing the survey results, as though two different customers had responded in exactly the same way to the survey questionnaire (which does in fact sometimes happen).

We have just described a situation in which one slip was picked twice. It's clearly possible for a single slip to be picked three times or four times or more. Similarly, it's possible for two different slips to be picked two or more times each. When the total number of customers served is relatively low and the sample size is relatively large in comparison, multiple picks of two or more slips will be a fairly common occurrence.

The initial reaction of most people to a description of sampling *with replacement* is generally rather negative. Why would anyone ever do it that way? Putting the slip of paper back in the box, knowing full well that it may be picked again, seems bizarre and unreasonable to them.

So why is this procedure used? The answer is that the mathematics that result from using the sampling *with replacement* procedure are much simpler than those that result from use of the sampling *without replacement* procedure.

Why is this so? Because the odds of any one slip being picked are the same throughout the sampling *with replacement* procedure. If for example, there were a total of 534 customers served (and therefore 534 slips of paper in the box) from which the 77 customers to be included in the sample are to be selected, then the odds that any one customer will be selected on the first pick is 1 in 534, the odds of any one customer being selected in round two are 1 in 534, the odds of being selected on the third round are 1 in 534 and so on. The odds of being picked never change from the beginning to the end of the process of picking 77 slips of paper form the box (and thereby picking 77 customers to be surveyed from the total of 534 customers served).

With the sampling *without replacement* procedure, however, the odds of any one customer being picked are constantly changing from the beginning of the process to the end. On the first pick, every customer's odds of having his or her slip picked is 1 in 534. But then, after one slip has been picked, the customer number on that slip has been recorded, and that slip has been set aside, everyone's odds of being picked on the second pick are changed. For the customer already picked, his or her odds of being picked again are now zero. That slip has been set aside so there is no further chance of being picked on that or any further round of picks. But for the remaining customers, the chances of being picked are now 1 in 533. The odds are lower since there is now one fewer slip in the box—only 533 instead of the 534 that were there before the first pick was made. The odds of being picked on the third round then go down to 1 in 532. The odds continue to decrease in this same way throughout the process of picking 77 slips. Because of these changing probabilities, the resulting mathematics get rather complicated.

In summary, we have two different procedures for selecting a random sample of customers from the total number of customers served. Sampling *without replacement* is the process that seems most reasonable to nonstatisticians, but it results in very complicated mathematics that make calculations much more difficult. Sampling *with replacement* seems bizarre to most nonstatisticians (especially counting a customer's response twice if that customer gets picked twice), but results in much simpler mathematics.

Luckily, statisticians have adopted the sampling *without replacement* procedure as the one generally used for selecting the sample in most surveys. Statisticians refer to this procedure as *Simple Random Sampling* (SRS), but it might instead be called the *standard approach for random sampling*. The standard *stratified sampling* procedure described above in the first section of this Factsheet is a slightly more complex variant of Simple Random Sampling. *Stratified sampling* as described above is also based on sampling *without replacement*.

In these Guidelines, we strongly recommend use of *sampling without replacement* because it is a procedure that EPA employees who are not statisticians are much more likely to feel comfortable with and is the procedure most commonly used by statisticians as well. If, however, any unit of EPA determines that it has a compelling reason to use a sampling with replacement procedure instead, it should feel free to do so as long as those who will conduct the survey understand the implications of doing so, will make use of the appropriate formulas (some of which will differ from those presented in these Guidelines and Factsheets), and are willing to double or triple count customers (those that were that were selected more than once) when calculating the survey results.