GLOBAL EDITION

EXPLORING RESEARCH

TENTH EDITION

NEIL J. SALKIND

Exploring Research

Tenth Global Edition

Neil J. Salkind
University of Kansas

Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • Sao Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Table 4.1 Group of 50 names constituting a population for our purposes. Notice that each one is numbered and is ready to be selected (also, realize that populations are often much larger).

1. Jane	11. Susie	21. Ed T.	31. Dana	41. Nathan
2. Bill	12. Nona	22. Jerry	32. Bruce	42. Peggy
3. Harriet	13. Doug	23. Chitra	33. Daphne	43. Heather
4. Leni	14. John S.	24. Glenna	34. Phil	44. Debbie
5. Micah	15. Bruce	25. Misty	35. Fred	45. Cheryl
6. Sara	16. Larry	26. Cindy	36. Mike	46. Wes
7. Terri	17. Bob	27. Sy	37. Doug	47. Genna
8. Joan	18. Steve	28. Phyllis	38. Ed M.	48. Ellie
9. Jim	19. Sam	29. Jerry	39. Tom	49. Alex
10. Terrill	20. Marvin	30. Harry	40. Mike G.	50. John D.

4 or a 5 is equal. This means that when names are attached to the numbers, the likelihood of selecting any particular name is equal as well.

With that fact in mind, we will select one group of 10 names using the table of random numbers in Table 4.2. Follow these steps:

1. Select a starting point somewhere in the table by closing your eyes and placing your finger (or a pencil point) anywhere in the table. Selecting your starting point in this way ensures that no particular starting point (or name) is selected. For this example, the starting point was the first column of numbers, last row (36,768), with the pencil point falling on the fourth digit, the number 6.
2. The first two-digit number, then, is 68 (in boldface type in Table 4.3).

   Because the population goes up to 50, and there is no number 68, this number is skipped and the next two-digit number is considered. Because you cannot go down in the table (no place to go), go to the top of the next column and read down, once again selecting the first two digits. For your convenience, each pair of two-digit numbers in the second column of Table 4.3 is separated.
3. The next number available is 48. Success! Person 48 on the list is Ellie, and she becomes the first of the 10-member sample.
4. If you continue to select two-digit numbers until 10 values between 01 and 50 are found, the names of the people that correspond in Table 4.1 with the numbers in boldface type in Table 4.4 are selected. Here is a breakdown of which numbers worked and which did not for the purposes of selecting a random sample of 10 people from the population of 50.

Table 4.2 Partial table of random numbers. In such a table, you can expect there to be an equal number of single digits that are randomly distributed throughout all the numbers.

23157	48559	01837	25993
05545	50430	10537	43508
14871	03650	32404	36223
38976	49751	94051	75853
97312	17618	99755	30870
11742	69183	44339	47512
43361	82859	11016	45623
93806	04338	38268	04491
49540	31181	08429	84187
36768	76233	37948	21569

Table 4.3 Starting point in selecting 10 cases using the table of random numbers. You can begin anywhere, as long as the place you begin is determined by chance and is not intentionally chosen.

23157	48 55 9	01837	25993
05545	50 43 0	10537	43508
14871	03 65 0	32404	36223
38976	49 75 1	94051	75853
97312	17 61 8	99755	30870
11742	69 18 3	44339	47512
43361	82 85 9	11016	45623
93806	04 33 8	38268	04491
49540	31 18 1	08429	84187
367**68**	76 23 3	37948	21569

Randomly determined starting point

Reading down the first column of two-digit numbers, 48, 50, 03, 49, and 17 are fine because they fall within the range of 50 (the size of the population) and they have not been selected before:

- 04 and 31 are fine
- 76 is out of the range

Because you cannot read farther down the column, it is time to go up to the next set of two digits (in the same

Table 4.4 Ten two-digit numbers (each one appearing in bold) selected from the population.

23157	**48** 55 9	01837	25993
05545	**50 43** 0	10537	43508
14871	**03** 65 0	32404	36223
38976	**49** 75 1	94051	75853
97312	**17** 61 8	99755	30870
11742	69 **18** 3	44339	47512
43361	82 85 9	11016	45623
93806	**04 33** 8	38268	04491
49540	31 18 1	08429	84187
36768	76 23 3	37948	21569

five-digit column) at the top of the column, which begins with the number 55.

- 55 is not within the range
- 43 is fine
- 65, 75, and 61 are not acceptable
- 18 is
- 85 is not
- 33 is

And there you have the 10 people:

Number	Name
48	Ellie
50	John D.
03	Harriet
49	Alex
17	Bob
04	Leni
31	Dana
43	Heather
18	Steve
33	Daphne

Now you have a sample of 10 names from a population of 50 selected entirely by chance. Remember, the probability of any one of these people being selected from the population is the same as the probability of any other person from the population being selected.

Your sample is selected by chance because the distribution of the numbers in the partial table of random numbers in Table 4.2 was generated by chance. Is it just a coincidence that three of the first five numbers (48, 50, 03, 49, and 17) in the partial table of random numbers are grouped together? Absolutely yes. This group of five is the best approximation and the most representative of any sample of five from the entire population, given that each member of the population has an equal and independent likelihood of being chosen (1/50 or .02 or 2% in this case).

A further assumption is that the names in the population (Table 4.1) were listed in a random fashion. In other words, names 01–20 were not listed as the first 20 of 50 because they come from a different neighborhood, are very wealthy, or have no siblings, or some other characteristic that might get in the way of an unbiased selection.

The general rule (and this may be the most important point in the entire chapter) is to *use a criterion that is unrelated to that which you are studying*. For example, if you are doing a study on volunteering, you do not want to ask for volunteers!

USING THE COMPUTER TO GENERATE RANDOM SAMPLES You should always do new things at least once manually, so you understand how a process works, such as selecting a random sample from a population using a table of random numbers as you were shown earlier. After you are comfortable with the technique, it is time to turn to the computer.

The computer can be used to quickly generate values that can be used to determine a random selection of participants from a population.

As you can see in Figure 4.1, 10 participants have been selected. Those who have not been selected have a diagonal line (e.g., case) through the cell, and subsequent analyses will only use the selected cases. This figure shows how IBM® SPSS® Statistics software (SPSS)* works, but any capable data analysis tool can do the same.

Systematic Sampling

In another type of sampling, called **systematic sampling**, every kth name on the list is chosen. The term *kth* stands for a number between 0 and the size of the sample that you want to select. For example, here is how to use systematic sampling to select 10 names from the list of 50 (although these steps apply to any population size and sample) shown in Table 4.1. To do this, follow these steps:

1. Divide the size of the population by the size of the desired sample. In this case, 50 divided by 10 is 5. Therefore, you will select every fifth name from the list. In other words,

   $$\begin{matrix}\text{Size of population} \longrightarrow \\ \text{Size of sample} \longrightarrow\end{matrix} \frac{50}{10} = 5 \longleftarrow \text{Size of step}$$

2. As the starting point, choose one name from the list at random. Do this by the "eyes closed, pointing method" or, if the names are numbered, use any one or two digits from the serial number on a dollar bill. The dollar bill used in this example has as its first two digits 43, which will be the starting point.
3. Once the starting point has been determined, select every fifth name. In this example, using the names in Table 4.1 and starting with Heather (#43), the sample will consist of Ellie (#48), Harriet (#3), Joan (#8), Doug (#13), Steve (#18), Chitra (#23), Phyllis (#28), Daphne (#33), and Ed M. (#38).

Systematic sampling reduces the chances of certain participants being selected; therefore, it is less unbiased than simple random sampling.

Because systematic sampling is easier and less trouble than random sampling, it is often the preferred technique. However, it is also less precise. Clearly, the assumption of each member of the population having an equal chance to be selected is violated. For example, given that the starting point is Heather (#43), it would be impossible to select Debbie (#44).

---

*SPSS Inc. was acquired by IBM in October, 2009.

Figure 4.1 Selecting a sample.

	id_stud	id_prof	sex_stud	teacher1	teacher2	teacher3	teacher4	teacher5	filter_$
1	80761	1	Male	4.00	4.00	2.00	1.00	3.00	0
2	22920	1	Male	5.00	4.00	2.00	1.00	3.00	0
3	84768	2	Male	4.00	3.00	3.00	2.00	3.00	1
4	89225	2	Male	3.00	3.00	3.00	1.00	4.00	0
5	94354	3	Male	3.00	5.00	1.00	2.00	3.00	0
6	40351	4	Male	2.00	4.00	2.00	2.00	2.00	1
7	62034	4	Male	2.00	4.00	2.00	2.00	4.00	0
8	26841	5	Male	2.00	4.00	2.00	1.00	3.00	0
9	13691	5	Male	2.00	3.00	1.00	3.00	2.00	0
10	70386	6	Male	3.00	4.00	2.00	3.00	5.00	1
11	45176	6	Male	5.00	4.00	1.00	1.00	4.00	1
12	29627	6	Male	2.00	4.00	2.00	1.00	4.00	0
13	24824	6	Male	1.00	5.00	2.00	2.00	4.00	0
14	39986	7	Male	5.00	5.00	2.00	1.00	5.00	0
15	77974	7	Male	5.00	4.00	2.00	1.00	4.00	1
16	38396	7	Male	5.00	5.00	2.00	1.00	5.00	1
17	78665	7	Male	4.00	3.00	2.00	1.00	3.00	0
18	20227	7	Male	4.00	5.00	1.00	1.00	5.00	0
19	66298	8	Male	5.00	5.00	1.00	1.00	3.00	0
20	27423	8	Male	5.00	5.00	1.00	1.00	4.00	1
21	79359	8	Male	5.00	5.00	1.00	1.00	5.00	0
22	66513	8	Male	5.00	5.00	1.00	1.00	5.00	1
23	28796	9	Male	2.00	4.00	4.00	3.00	1.00	1
24	67564	9	Male	3.00	3.00	4.00	4.00	1.00	0
25	35598	9	Male	3.00	3.00	2.00	1.00	2.00	0
26	77823	10	Male	5.00	5.00	2.00	1.00	5.00	0
27	34850	10	Male	5.00	5.00	1.00	2.00	5.00	0
28	54217	10	Male	3.00	5.00	1.00	1.00	4.00	0

Stratified Sampling

The two types of random sampling that were just discussed work fine if specific characteristics of the population (such as age, gender, ethnicity, and ability group) are of no concern. In other words, if another set of 10 names were selected, one would assume that because both groups were chosen at random, they are, in effect, equal. But what if the individuals in the population are not "equal" to begin with? In that case, you need to ensure that the profile of the sample matches the profile of the population, and this is done by creating what is referred to as **stratified sampling**.

Stratified sampling takes into account the different layers or strata that characterize a population and allow you to replicate those layers in the sample.

The theory behind sampling (and the entire process of inference) goes something like this: If you can select a sample that is as close as possible to being representative of a population, then any observations you can make regarding that sample should also hold somewhat true for the population. So far so good. Sometimes, though, random sampling leaves too much to chance, especially if you have no assurance of equal distributions of population members throughout the sample and, most important, *if the factors that distinguish population members from one another (such as race, gender, social class, or degree of intelligence) are related to what you are studying*. This is a very important point. In that case, stratified sampling is used to ensure that the strata (or layers) in the population are fairly represented in the sample (which ends up being layered as well, right?).

For example, if the population is 82% Catholics, 14% Muslims, and 4% Buddhists, then the sample should have the same characteristics *if* religion is an important variable in the first place. Understanding the last part of the preceding sentence is critical. If a specified characteristic of the population is not related to what is being studied, then there is no reason to be concerned about creating a sample patterned after the population and stratifying on one of those variables.

Let's assume that the list of names in Table 4.1 represents a stratified population (females and males) and that attitudes toward legalizing abortion is the topic of study. Because gender differences may be important, you want a sample that reflects gender differences in the population. The list of 50 names consists of 20 females and 30 males, or 40% females and 60% males. The sample of 10 should mirror that distribution and contain four females and six males. Here is how you would select such a sample using **stratified random sampling**. Once again, the example is the population we created, but these steps apply to all circumstances.

1. All the males and all the females are listed separately.
2. Each member in each group receives a number. In this case, the males would be numbered 01–30 and the females 01–20.

Figure 4.2 Selecting a sample from a population that is stratified on two factors or layers: grade and location. Here the sample size is shown in brackets below the population size.

	Grade			
Location	**1**	**3**	**5**	**Total**
Rural	1,200 [120]	1,200 [120]	600 [60]	3,000 [300]
Urban	2,800 [280]	2,800 [280]	1,400 [140]	7,000 [700]
Total	4,000 [400]	4,000 [400]	2,000 [200]	10,000 [1,000]

3. From a table of random numbers, four females are selected at random from the list of 20 using the procedures outlined earlier.
4. From a table of random numbers, six males are selected at random from the list of 30 using the procedures outlined earlier.

Although simple examples (with only one stratum or layer) such as this often occur, you may have to stratify on more than one variable. For example, in Figure 4.2, a population of 10,000 children is stratified on the variables of grade (40% first grade, 40% third grade, and 20% fifth grade) and location of residence (30% rural and 70% urban). The same strategy is used: Select 10% (1,000 is 10% of 10,000) of each of the stratified layers to produce the sample size shown in Figure 4.1. For example, of the 1,200 rural children in the first grade, 10% (or 120) were randomly selected. Likewise, 140 urban children in fifth grade were selected.

Cluster Sampling

The last type of probability sampling is **cluster sampling**, in which units of individuals are selected rather than individuals themselves. For example, you might be doing a survey of parents' attitudes toward vaccination. Rather than randomly assigning individual parents to two groups (say, e.g., those who will be sent informational material and those who will not), you could just identify 30 pediatricians' offices in the city and then, using a table of random numbers, select 15 for one group and designate 15 for the second group. Another example can be found in large cities where police stations are divided into districts, and each district becomes one entry, as a cluster of stations.

Clusters are groups of occurrences that occur together.

Cluster sampling is a great time saver, but you must be sure that the units (in this case, the people who visit each pediatrician) are homogeneous enough such that any differences in the unit itself might not contribute to a bias. For example, if one pediatrician refuses to vaccinate children before a certain age, that would introduce a bias you would want to avoid.

Test Yourself

Why is it critically important that the criterion used to assign people to groups not be related to the focus of the study or to the topic of interest?

What are the four steps involved in simple random sampling?

Nonprobability Sampling Strategies

In the second general category of sampling strategies, nonprobability sampling, the probability of selecting a single individual is not known. Because this is the case, you must assume that potential members of the sample do not have an equal and independent chance of being selected. Some of these sampling methods are discussed in the following.

Convenience Sampling

Convenience sampling is just what it says. A football coach gives each team member a questionnaire. The audience (the team) is a captive one, and it is a very convenient way to generate a sample. Easy? Yes. Random? No. Representative? Perhaps, but to a limited extent.

You might recognize this method of sampling as the reason why so many experiments in psychology are based on results using university students; these students are a captive audience and often must participate for credit.

Quota Sampling

You might be in a situation where you need to create a sample that is stratified on certain variables, yet for some reason proportional stratified sampling is not possible. In this case, quota sampling might be what you want.

Quota sampling selects people with the characteristics you want (such as first-grade, rural children) but does not randomly select from the population a subset of all such children, as would occur in **proportional stratified sampling**. Rather, the researcher would continue to enlist children until the quota of 120 is reached. The 176th rural kid in first grade never has a chance, and that is primarily why this is a nonprobability sampling technique.

Here is another example of a quota system. You have to interview 20 first year university students of both genders. First, you might interview 10 men and, knowing that the distribution of males and females is approximately a 50/50 split, you interview the next 10 women who come along, and then you call it quits. Whereas quota sampling is far easier than stratified sampling, it is also less precise. Imagine how much easier it is to find any 10 men, rather than a specific 10 men, which is what you would have to do in the case of stratified sampling.

Table 4.5 provides a summary of probability and nonprobability sampling methods.

Samples, Sample Size, and Sampling Error

No matter how hard a researcher tries, it is impossible to select a sample that perfectly represents the population. The researcher could, of course, select the entire population as the sample, but that defeats the purpose of sampling—making an inference to a population based on a smaller sample.

One way that the lack of fit between the sample and the population is expressed is as **sampling error**, which is the difference between a measure of the characteristics of the sample and a measure of the characteristics of the population from which the sample was selected. For example, the average height of 10,000 fifth graders is 40 inches. If you take 25 samples of 100 fifth graders and compute the average height for each set of 100 children, you will end up with an average height for each group, or 25 averages. If all those averages are exactly 40 inches, there is no sampling error at all. This result, however, is surely not likely to be the case.

Reducing sampling error is the goal of any sampling technique.

Life is not that easy nor is the selection of samples that perfect. Instead, you will find the values to be something like 40.3 inches, 41.2 inches, 39.7 inches, 38.9 inches, and so on. The amount of variability or the spread of these values gives you some idea of the amount of sampling error. The larger the diversity of sample values, the larger the error and the less precise and representative your sample. Think for a moment what would happen if the entire population of 10,000 fifth graders were the sample. You would find the average height to be 40! Perfect! No error! The lesson? The larger the sample, the smaller the sampling error, because larger samples approach the size of the population and thus are more representative of the population. But, as you already know, studying too large a sample is expensive and inefficient, and often not necessary.

Table 4.5 Summary of the different types of probability and nonprobability strategies.

Type of Sampling	When to Use It	Advantages	Disadvantages
Probability Strategies			
Simple random sampling	When the population members are similar to one another on important variables	Ensures a high degree of representatives	Time consuming and tedious
Systematic sampling	When the population members are similar to one another on important variables	Ensures a high degree of representatives and no need to use a table of random numbers	Less random than simple random sampling
Stratified random sampling	When the population is heterogeneous and contains several different groups, some of which are related to the topic of study	Ensures a high degree of representatives of all the strata or layers in the population	Time consuming and tedious
Cluster sampling	When the population consists of units rather than individuals	Easy and convenient	Possibly members of units are different from one another, decreasing the technique's effectiveness
Nonprobability Sampling Strategies			
Convenience sampling	When the members of the population are convenient to sample	Convenient and inexpensive	Degree of generalizability is questionable
Quota sampling	When strata are present and stratified sampling is not possible	Ensures some degree of representativeness of all the strata in the population	Degree of generalizability is questionable

The exact process for computing the sampling error, which is expressed as a numerical value, is beyond the scope of this book, but you should recognize that your purpose in selecting a good sample is to minimize that value. The smaller the value, the less discrepancy there is between the sample and the population.

But there's more. You already know that the larger a sample is, the more representative the sample is of the population. And, in general, the better that the samples represent their respective populations, the more accurate any test of differences, for example, will be. In other words, better sampling leads to more accurate, more valid tests of population differences.

How do you minimize sampling error? Use good selection procedures as described earlier in this chapter and increase the sample size as much as possible and reasonable. The next question you are ready to ask (I hope) is, "How big should the sample size be?" Glad you asked. Let's look at the last section in this chapter for more insight into the answer to that question.

How Big Is Big?

Now that you know something about sampling, just how many of those high school students do you need to select from the population of 4,500? If 50 is good, is not 500 better? And why not 1,500, if you have the time and resources to commit to the project?

A big enough sample is one that is big enough to answer the question being asked.

You already know that too small a sample is not representative of the population and too large is overkill. Sampling too many high school students would be self-defeating because you are no longer taking advantage of the power of inference. Some people believe that the larger the sample the better, but this strategy does not make economic or scientific sense.

Remember, the less representative the sample is of the population, the more sampling error is present. In addition, the larger the sampling error, the less generalizable the results will be to the population and the less precise your test of the null hypothesis.

A more advanced way of dealing with sample size is through a consideration of effect size. This concept was made popular with the pioneering work of Jacob Cohen (1988) and the notion that the stronger the effects of a treatment (such as the larger the expected difference between samples), then the smaller the sample size need be. Now this is pretty advanced stuff, but you can use a set of tables and, given the expected effect (or the magnitude of the difference you expect between two groups, e.g.), you can get a pretty clear estimate of the number of participants you need in each group.

ESTIMATING SAMPLE SIZE Every situation is different. Let's assume that you are examining the difference between two groups. How would you go about determining what the "correct" sample size might be? There are several numerical formulas for this, but you should at least be aware of what the important factors are that figure into your decision. Keep in mind that 30 is the general magic number of how many participants should be in each group.

In general, you need a larger sample to represent the population, accurately when

- The amount of variability within groups is greater and
- The difference between the two groups gets smaller.

Why is this the case? First, as variability increases within groups, it means that the data points (perhaps representing test scores) are more diverse, and you need a larger number of data points to represent all of them. For example, if you test two groups of college sophomores to determine whether their grade point averages differ and each group is highly variable, then it is likely that you will need a larger number of data points to represent the population fairly and show any difference between the groups (if a difference exists).

Second, as the difference between groups gets smaller, you need a larger number of participants to reach the critical mass where the groups can differ. For example, if you were to compare a first grader and a sixth grader on height, you would need only one participant in each group to say fairly confidently that there is a difference in height. In fact, there are very few (if any) short sixth graders who are shorter than the tallest first grader. But, if you examined a first grader and a third grader, the differences become much less noticeable, and a larger number of participants would be necessary to reveal those differences (if they are even there).

Do you want the real scoop on sample size? Keep the following in mind:

- In general, the larger the sample is (within reason), the smaller the sampling error will be and the better job you can do.
- If you are going to use several subgroups in your work (such as males and females who are 10 years of age, and healthy and unhealthy rural residents), be sure that your initial selection of subjects is large enough to account for the eventual breaking down of subject groups.
- If you are mailing out surveys or questionnaires (and you know what can happen to many of them), count on increasing your sample size by 40% to 50% to account for lost mail and nonresponders.
- Finally, remember that big is good, but accurate and appropriate are better. Do not waste your hard-earned money or valuable time generating samples that are larger than you need. In Chapter 13, we make specific