Practitioner's Guide To Stratified Random Sampling: Part 1

By **Brian Kriegler**
November 30, 2018, 3:53 PM EST


Brian Kriegler

This is the first of two articles on stratified random sampling. In the first article, I discuss when it might be advantageous to select a random sample that has been divided into multiple subpopulations. The second article covers common misperceptions about this statistical technique.

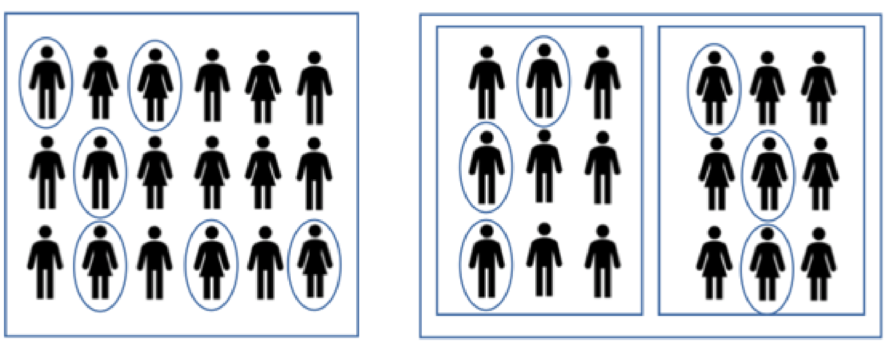
Each article in this two-part series addresses stratified sampling within a legal setting. The case study below is about Medicare reimbursement, an issue that arises in False Claims Act litigation. The subsequent article uses a hypothetical wage and hour example involving on-duty meal period agreements.

These two articles follow **an earlier series** on random sampling that was published in Law360. My objective for this article is to provide practical recommendations and considerations when using this statistical technique.

Four Reasons to Select a Stratified Random Sample

Statistical practitioners frequently are tasked with constructing a sampling design. Two common sampling designs are "simple" or "stratified." Simple random sampling is like placing everyone's name in a hat and selecting a subset of these names. Stratified sampling entails two steps. People are separated into multiple groups, e.g., for men and women, and a simple random sample is selected from each hat. A visual comparison of simple random sampling versus stratified random sampling is shown in Figure 1 below.

Figure 1: Comparison of Simple Random Sampling to Stratified Random Sampling



Visual of Simple Random Sampling: Selection of 6 out of 18 People

Visual of Stratified Random Sampling: Selection of 3 out of 9 Men and 3 out of 9 Women

Both the simple and stratified samples include six people. Each person in Figure 1 has the same probability of selection in a simple random sample (left). The random selection happens to include four women and two men. In a stratified sample (right) each person has the same probability of selection within each stratum. In this case the three men and three women are selected at random.

Virtually any population can be divided into multiple categories. For example, a population of people could be split based on their eye color, hair color, height, age or geographic region. Just because a population can be stratified does not necessarily mean that it should be. Sometimes a simple random sample will suffice. In this article we will demonstrate how a stratified sampling design can be advantageous given one or more of the following goals:

- Mitigating nonresponse bias.
- Lowering the margin of error.
- Addressing certain relevant data questions.
- Reducing the impact of outliers.

We will apply these practical considerations to a "Medicare reimbursement example."[1] Assume that the assignment is to estimate the rate at which a large medical provider submitted ineligible reimbursements over the last 10 years. Medicare reimbursement ineligibility is determined by examining patient files. It is too expensive to examine all 10,000 patient files in the population. Up to 300 patient files can be examined given time and budget constraints.

We will explore how stratified sampling may be worthwhile given the challenges and facts listed below. Each of the aforementioned goals can address these issues.

- ***Mitigating nonresponse bias:*** Some patient files are missing or incomplete.

- ***Lowering the margin of error:*** Reimbursement ineligibility is correlated with the time period during which the patient was treated.

- ***Addressing certain relevant data questions:*** Specific subpopulations are of particular interest.

- ***Reducing the impact of outliers:*** Some extremely large reimbursements may be ineligible.

Reason 1: Stratifying Can Mitigate Nonresponse Bias When Sample Selections Are Missing or Incomplete

Nonresponse bias is a result of missing and/or incomplete sampled data. This may lead to a sample that is not representative, e.g., if patient files begin to get purged after five years. Recall that this hypothetical study covers 10 years.

A stratified sampling design can mitigate potential nonresponse bias. Each patient can be placed into one of two subpopulations: (2) patients from the most recent five years or (2) patients from more than five years ago. Population estimates based on this design will be unbiased if the sample within each stratum is selected at random. Therefore we will look more closely at each stratum.

- ***0-5 years ago:*** Assume that each sampled patient file is located and analyzed. There are no issues concerning missing/incomplete data and nonresponse bias.

- ***5-10 years ago:*** Some sampled patient files may be missing. The goal is to preserve the integrity of the population estimate and margin of error. There are a few options.

  - *Option 1:* Assume that all missing sampled patients are eligible — or alternatively ineligible — for reimbursement. The practitioner's assumption will depend on whether it is better to err on the side of underestimating or overestimating reimbursement.

  - *Option 2:* Demonstrate that missing/incomplete observations are likely no different than the nonmissing sampled observations. The practitioner then can work with a relatively smaller sample than originally intended.

  - *Option 3:* Similar to option 2, except missing patient files are replaced with additional random selections.

Reason 2: Stratifying Into Homogeneous Subpopulations Can Yield a Lower Margin of Error

Another objective may be to identify if there are any characteristics strongly correlated with patient ineligibility for reimbursement. The margin of error may be lower if the population is stratified on these characteristics.

Assume that ineligibility for reimbursement is strongly correlated with time. Stratifying by time period makes the data within each stratum relatively homogeneous in terms of reimbursement ineligibility. This likely leads to less variability within each stratum and a lower margin of error across all strata.

Note that both stratified and simple random sampling designs can be used to calculate unbiased population estimates. However, the stratified design is preferred because it yields a lower margin of error and a more precise population estimate.

Reason 3: Stratifying Allows the Practitioner to Draw Inferences About Both the Entire Population and Specific Subpopulations

Recall in this example that time and cost constraints allow for a sample of up to 300 patient files. A stratified random sample can be allocated so that estimates are reasonably precise for the whole population and within specific subpopulations.

Assume that in one state, patient ineligibility is more costly and therefore important to measure. A stratified sampling design could include 150 patients from this one state and 150 patients from the remainder of the country. The goal is to obtain an informative estimate — and margin of error — for both this specific state and the whole country. A higher nationwide margin of error is exchanged for the ability to reach conclusions about both the whole population and a specific state.

Reason 4: Examining All Outliers May Improve the Precision and Robustness of the Population Estimates

Sometimes potential outliers may be identified before selecting a sample. Giving potential outliers their own stratum may stabilize the population estimates and reduce the margin of error.

Thus far we focused on the percentage of patients that were ineligible for reimbursement. A random sample of patients can also be used to estimate the total amount of ineligible Medicare payments. The goal is to obtain an informative estimate — and margin of error — for the total amount of reimbursement ineligibility.

Assume the patient database reveals that 100 patients account for the majority of all reimbursement dollars. The estimated total amount ineligible for reimbursement and margin of error may change depending on whether these patients were eligible. One way to tackle this issue is to examine the entire subpopulation of outliers. Then, a random sample of 200 patients is selected from the remaining population of 9,900 people. This can be advantageous for two statistical reasons.

- Giving outliers their own stratum limits their impact on the population estimate. This can improve the robustness of the estimated total.

- The outliers' contribution to the margin of error is eliminated because this entire subpopulation is examined. The margin of error for the total ineligible reimbursement amount is attributed to the sample of 200 from 9,900 patients.

Practical Considerations

Each of the aforementioned reasons to stratify requires knowledge about the population and/or what the sample ultimately will reveal. How does the practitioner know which of these will be issues before the sample is analyzed? Oftentimes the sampling design is justified based on one or more of the following:

- Analyzing a pilot sample.
- Collaborating with subject matter experts.
- Reviewing a population's database.

Performing a combination of these steps can be advantageous. A pilot sample may reveal the largest sample size possible. Subject matter experts may have insights on which data characteristics are strongly correlated with the outcome variable(s). Reviewing available data about the population can indicate whether the population includes potentially important subpopulations and/or outliers.

The aforementioned reasons to stratify can also be applied to determine if a simple random sample will suffice. Sometimes a simple random sample provides all of the data needed to address the relevant issues. Stratifying the population may add unnecessary complexity to the design and analysis. Indeed the margin of error may be higher if the population is split on variables that are not strongly correlated with the measurement(s) of interest.

Concluding Remarks

Stratification arms the practitioner with a flexible sampling tool. There is virtually no limit to the number of ways that a population can be stratified. There also are no hard-and-fast rules about how to allocate the sample size across multiple strata. The sample size can be the same across all strata, in proportion with the size of each subpopulation, or in some other fashion. Choosing a sampling design entails both science and discretion. Used properly, stratified sampling allows the practitioner to tackle multiple issues efficiently and offer well-informed interpretations about a defined population.

Brian Kriegler, Ph.D., is a managing director at Econ One Research Inc. in Los Angeles.

[1] This Medicare reimbursement example is similar but not identical to the example discussed in a previous article titled “Making Valid Statistical Inferences When Sample Selections are Missing.”

Practitioner's Guide To Stratified Random Sampling: Part 2

By **Brian Kriegler**
December 3, 2018, 1:55 PM EST


Brian Kriegler

This is the second of two articles on stratified random sampling. In **part one**, I discussed when stratified sampling is a viable option. In part two below, I explore common misperceptions about this statistical technique.

Each article in this two-part series addresses stratified sampling within a legal setting. The case study in the previous article was about Medicare reimbursement, an issue that arises in False Claims Act litigation. This article uses a hypothetical wage and hour example involving on-duty meal period agreements.

These two articles follow an **earlier series** on random sampling that was published in Law360. My objective for this article is to simplify the application of this statistical technique so that it is used correctly in a legal setting.

Three Misconceptions About Stratified Random Sampling

Stratified random samples frequently are used to estimate the population average and corresponding margin of error. Neutral and unbiased calculations require four steps. First, each observation is assigned to a defined stratum, i.e., subpopulation. Second, a random sample is selected from each stratum. Third, sampled observations are analyzed. Finally, results from each stratum are "weighted" based on the size of each respective sub-population. The formulas used to estimate the population average/percentage and margin of error are established and can be found in numerous statistical sampling textbooks. Nevertheless misconceptions about stratified sampling requirements still persist.

Observers, auditors and opposing experts may well ask questions about the sampling design, such as: (1) Why did the practitioner use this particular design? (2) Why did the practitioner select this sample size within each stratum? (3) Why wasn't some other stratification and/or sample size allocation implemented? Additionally some practitioners expect the sampling design to exhibit the following attributes:

- The population needs to be stratified in a way that creates highly homogeneous subpopulations.
- The population must be divided on specific data values.
- The sample must be allocated equally or proportionally to the size of each subpopulation.

It is a common misconception to think that the population estimate will be biased if the above three conditions are not met. Below we discuss why this is not true.

The absence of some or all of these three conditions may lead to a different margin of error — but not a biased estimate.

Case Study Overview

The hypothetical case study represented in this article will refer to a stratified sampling design involving employees who signed an on-duty meal period agreement. The goal is to estimate the percentage of 10,000 employees who signed an agreement based on a random sample and to calculate the corresponding margin of error. For purposes of this case study we will assume the following facts:

- There is a pending class action litigation where plaintiffs allege that the agreement is not in compliance with wage and hour laws.
- Each employee's personnel file reveals whether s/he signed the agreement.
- Given time and cost constraints, between 200 and 250 employees' personnel files can be reviewed.

Details About the Hypothetical Sampling Design

The population of 10,000 employees is divided into three strata. The intention is to stratify by geographic region in order to create relatively homogeneous subpopulations. A random sample of employees — and their personnel files — is selected from each sub-population. The population, sampling design and sample results are shown in Table 1 below. We will use these results to calculate a population estimate and a margin of error.

Table 1: Summary of the Stratified Random Sampling Design and Results

Stratum / Intended Geographic Region	# of Employees (% of population)	# of Sampled Employees (% of sample)	% of Employees That Signed the Agreement	Employees Assigned to the "Wrong" Stratum	
				In the Population	In the Sample
1. Northern California	4,000 (40%)	100 (47.6%)	15/100 = 15%	None	None
2. Central California	3,600 (36%)	60 (28.6%)	10/60 = 16.67%	None	None
3. Southern California	2,400 (24%)	50 (23.8%)	12/50 = 24%	100 from Central CA	5 from Central CA

Calculating the Population Estimate

Three sets of computations are needed to derive an unbiased estimate. The steps below are used to estimate the percentage of people in the population who signed the agreement.

> ***Step 1:*** Calculating the percentage of employees who signed the agreement in the sample within each stratum. Respectively, the percentages are 15 percent, 16.67 percent and 24 percent for Northern California, Central California and Southern

California.

Step 2: Multiplying each percentage by the number of observations in each subpopulation. These three numbers are 15 percent x 4,000, 16.67 percent x 3,600, and 24 percent x 2,400.

Step 3: Adding the numbers from step 2 together and dividing by the number of people in the whole population:

[15 percent x 4,000 + 16.67 percent x 3,600 + 24 percent x 2,400] / 10,000 = 17.8 percent.

Based on this stratified random sample of data, the unbiased estimate for the proportion of employees who signed the agreement is 17.8 percent.

Interpreting the Margin of Error

The true proportion of employees who signed the agreement may well differ from sample proportion. As a result there is a “margin of error” around 17.8 percent. The fact that these data constitute a random sample gives the “green light” to calculate the margin of error using well-documented formulas that can be found in numerous sampling textbooks.[1]

Given the hypothetical results of this case study shown in Table 1 above, the margin of error for the percentage of employees who signed the agreement is approximately 5.2 percent. The steps taken to derive this margin of error are shown in the appendix of this article. The conclusion is that we are 95 percent confident that 17.8 +/- 5.2 — between 12.6 and 23.0 — percent of all employees signed the agreement.[2]

Potential Misconceptions

There are three common ways in which this sampling design might be questioned by observers, auditors and opposing experts.

Table 2: Overview of Common Questions and Misconceptions About the Sampling Design Presented in Table 1

Question	Misconception
Why was only one data characteristic — geographic region — used to stratify the population?	Other data variables should have been used to split the data into multiple subpopulations
Why did one subpopulation include people from Central and Southern California?	Each subpopulation should have included employees from one geographic region
Why does the sample size within each subpopulation appear to be arbitrary?	The sample sizes should be equal or in proportion with the subpopulation sizes

We will challenge these issues by: (1) addressing each of these misconceptions, (2) demonstrating why the population estimate is unbiased, and (3) exploring how alternative sampling designs may yield different confidence intervals and margins of error — which are irrelevant.

Misconception 1: The Population Needed to Be Stratified Into Smaller and More Homogeneous Subpopulations

There are three key points regarding why additional stratification is unnecessary when the objective is to derive an unbiased estimate of a population average/percentage.

1. Population estimates and margins of error are statistically reliable provided that the data are randomly selected and correctly analyzed.[3]

2. Stratifying on additional variables redefines the population. This may lead to a different margin of error — not a biased estimate across the three strata.

3. Splitting these data into more strata is unnecessary in this situation. The stated goal is to reach conclusions about the whole population — not specific subpopulations.

Misconception 2: The Population Was Divided Using the Wrong Data Values

In this hypothetical example, employees come from one of three possible geographic areas. However the third stratum captures people from two regions. Some practitioners may argue that each stratum must contain people from the same area. This is incorrect for three reasons.

1. It is established in the statistical literature that conclusions from a random sample apply to the sampled population — which is defined by the practitioner.[4] Here the objective is to estimate the proportion of 10,000 employees who signed the agreement. This is also the population that gets sampled. Therefore this sampling design can be used to reach quantitative conclusions across all employees.

2. In this hypothetical example, a simple random sample was selected from each of the three subpopulations. The estimated proportion of employees who signed the agreement within each subpopulation is neutral and unbiased. The same is true for the estimate across the entire population provided that each stratum is correctly weighted based on the number of employees in each subpopulation.

3. A population that was more homogeneous in terms of geography may yield a lower margin of error. However the margin of error is not indicative of whether a sample is biased. There is no statistical requirement that the population must be split in a particular way when the objective is to estimate a population average/percentage. Indeed there are likely hundreds of ways to stratify this population of 10,000 people.

Misconception 3: The Sample Size Was Not Properly Allocated to Each Stratum

Critics may assert that statistical extrapolations are biased and unreliable because the sample size within each stratum appears to be arbitrary. Here employees in Northern California comprise 47.6 percent of the sample but only 40 percent of the population. The perception is that the sample is not "representative" of the population. There are three reasons that this criticism is statistically incorrect.

1. Using a stratified sampling design, population estimates are weighted by number of observations in each subpopulation — not the sample size. Additionally the number of employees in each subpopulation is a fixed number that does not depend on the number of observations selected. Therefore a disproportionate stratification will not bias the population estimate.

2. A sampling design allocated in proportion with the size of each subpopulation may be worthwhile if an objective is to measure certain characteristics, e.g., the 10th or 25th percentile. However this allocation is not essential when the goal is to estimate a population average/percentage.

3. Sometimes practitioners are interested in obtaining relatively precise measurements within a specific subpopulation. Perhaps the practitioner simultaneously was interested in studying Northern Californians.[5] Using the sample size allocation in Table 1 above, there is a potential trade-off between a higher margin of error across the entire population and the ability to reach informative conclusions about a specific subpopulation.

Practical Considerations

The misconceptions raised in this article imply that the sample must meet a higher standard that unnecessarily goes beyond the fundamentals of statistical sampling. This may be due to a misunderstanding or an intentional effort to confuse the audience, e.g., by muddling the difference between bias and the margin of error.

Practitioners can put themselves in a strong foundational position by getting out in front of these potential misconceptions. Even if the criticisms/questions raised are inconsequential, this does not mean that these issues cannot be addressed proactively. Anticipating these points can mitigate confusion and make the results easier to process. How the sampling design is presented can be persuasive.

Concluding Remarks

Estimated averages/percentages and the corresponding margin of error are statistically reliable when (1) observations are selected at random; (2) every observation in the sampled population has a known probability of selection; and (3) the sample of data is correctly analyzed. These statistical calculations are permitted even if the sample does not look like the population and/or there is a significant amount of variation in the data. It is imperative for the practitioner to have a thorough understanding of which criteria are needed — and when additional conditions are not grounded in statistics.

Appendix of Calculations

Below is the derivation of the margin of error given the data presented in Table 1 above.

Table A: Derivation of the 95% Confidence Interval Using Hypothetical Results Shown in Table 1 Above

Within Each Stratum

Variable Description	Variable Abbreviation	Formula	Stratum 1	Stratum 2	Stratum 3
# of Employees In Population	N		4,000	3,600	2,400
# of Sampled Employees	n		100	60	50
# of Sampled Employees that Signed the Agreement	x		15	10	12
% of Sampled Employees that Signed the Agreement	p	x/n	0.150	0.167	0.240
Variance in % of Employees that Signed the Agreement	v	$\frac{n \cdot p \cdot (1-p)}{n-1}$	0.129	0.141	0.186
Variance in # that Signed the Agreement	T	$\frac{(N-n) \cdot N \cdot v}{n}$	20,091	30,000	20,995

Across the Three Strata

Variable Description	Variable Abbreviation	Formula	Calculation
Standard Deviation Of % that Signed The Agreement	SD	$\frac{\sqrt{(T_{Northern} + T_{Central} + T_{Southern})}}{N_{Northern} + N_{Central} + N_{Southern}}$	2.67%
Margin of Error for a 95% Confidence Interval	ME	$1.96 \cdot SD$	**5.23%**

Brian Kriegler, Ph.D., is a managing director at Econ One Research Inc. in Los Angeles.

[1] See,e.g., Thompson, Steven K. (2002).Sampling. New York: Wiley. pp. 120-121.

[2] 17.8 - 5.2 = 12.6 and 17.8 + 5.2 = 23.0.

[3] See, e.g., Moore, D.S., McCabe, G.P., and Craig, B.A. (2009). Introduction to the Practice of Statistics. 6th Ed. New York: W.H. Freeman. (Hereinafter, “Mooreet al.”), p. 219. (“The act of randomizing **guarantees** that the results of analyzing our data are subject to the laws of probability.”) (bolded emphasis added)

[4] Cochran, W.G. (1999).Sampling Techniques. 3rd Ed. Wiley. p. 5. ([C]onclusions drawn from the sample apply to the sampled population.”)

[5] In a legal context, selecting more people from Northern California might be worth considering if there was a potential subclass of people from this region.