5-2023

A Monte Carlo Analysis of Nonprobability Sampling & Post Hoc Corrections

Julia Hong
Western Kentucky University, juliahong39@gmail.com

A MONTE CARLO ANALYSIS OF NONPROBABILITY SAMPLING AND POST HOC CORRECTIONS

A Thesis submitted in partial fulfillment of the requirements for the degree
Master of Science, Psychological Sciences

Department of Psychological Sciences
Western Kentucky University
Bowling Green, Kentucky

By
Julia Hong

May, 2023

A Monte Carlo Analysis of Nonprobability Sampling and Post Hoc Corrections

Julia Hong

Date Recommended 4/10/2023

DocuSigned by:
Reagan D. Brown, Ph.D.
C6FF97C84083446...
Chair

DocuSigned by:
Dr. Katrina A. Burch
75BB04A0959F41C...
Committee Member

DocuSigned by:
Dr. Xiaowen Chen
EE91D1B056614FB...
Committee Member

Committee Member

DocuSigned by:
Ranjit Koodali
8F8057903E39448...
Associate Provost for Research and Graduate Education

ABSTRACT

A MONTE CARLO ANALYSIS OF NONPROBABILITY SAMPLING AND POST HOC CORRECTIONS

Nonprobability samples are often used in place of probability samples because the former are less trouble and less expensive. Unfortunately, it is difficult to determine how well a sample represents population parameters when using nonprobability samples. Researchers attempt to mitigate the disadvantages of nonprobability sampling by performing post hoc corrections, but this adjustment may not successfully undo the effects of nonprobability sampling. To examine these effects, a Monte Carlo simulation was conducted to create a pseudo-population from which samples were drawn. Forty-one conditions were replicated 10,000 times each, with each sample consisting of 100 observations. A post-stratification adjustment was made to these sample means as a post hoc correction. Confidence intervals were computed from the sample means before and after the adjustment. It was found that even slight correlations between the dependent variable and the likelihood of being sampled resulted in biased sample means and ineffective confidence intervals. Furthermore, post-stratification adjustments had mixed results but were generally ineffective in correcting for bias.

To my third grade teacher Candace Whites, who believed in me from the start and without whom I would have never gotten here

ACKNOWLEDGEMENTS

I would like to thank my committee members for their invaluable guidance, especially my committee chair, Dr. Reagan Brown. He taught me that I should not be afraid of numbers, even when there are 41 million of them. I would also like to thank everyone in my cohort. I was very fortunate to have had a supportive cohort that never failed to make me laugh even during the most stressful of times. Although no one else in our lives may quite understand what I-O psychology is, at least we have each other to share memes about bad survey items with.

TABLE OF CONTENTS

LIST OF TABLES

A Monte Carlo Analysis of Nonprobability Sampling and Post Hoc Corrections

Nonprobability samples are commonly used in social science research due to practical considerations. However, if sample statistics computed from these samples cannot produce unbiased estimates of population parameters, studies using nonprobability samples will serve to propagate incorrect information. Conclusions drawn from these samples will be incorrect and could be harmful if those findings are applied to practice. If the science is not sound, the practice based on that science will suffer. Various applications of sample-based research (often in the form of employee surveys) include employee theft (Wimbush & Dalton, 1997), sexual harassment (Ilies et al., 2003), workplace incivility (Hershcovis et al., 2018), and job satisfaction (Rogelberg et al., 2003). The present research explores sample bias and the efficacy of post hoc corrections for this bias.

Sampling

A population is the “aggregate from which the sample is chosen” (Cochran, 1977, p. 5). The study population includes every individual who is relevant to a study. As it is infeasible to measure an entire population, researchers settle for measuring samples from populations instead. The sampled population should reflect the target population, but the sampled population is necessarily more restricted than the target population (Cochran, 1977). The very nature of a sample means that some portion of the target population will not be in the sample, the cause of sampling error.

There are different sources of sampling error; two types that are pervasive in surveys are coverage error and nonresponse error. Coverage error is the discrepancy between a population researchers want to measure and the population that is actually sampled, such as a Web survey that only samples people who have access to the Internet (Lee, 2006). Nonresponse error is the discrepancy between people who are selected for a sample and people who participate (Lee, 2006). People may be unreachable, refuse to respond, or be incapable of responding, leading to nonresponse (Lohr, 1999).

Sampling Methods

Sampling methods can be divided into two general categories: probability sampling and nonprobability sampling. With probability sampling, every member of a population has an equal chance of being selected into a sample (Hays, 1988). Although this procedure allows sample statistics to be used as unbiased estimates of population parameters, it can be difficult and expensive to collect a sample in this fashion (Pedhazur & Schmelkin, 1991). Due to the difficulty in collecting probability samples, researchers employ nonprobability sampling, which does not draw from an entire population (i.e., coverage error). As a result, "the incontrovertible fact is that, in nonprobability sampling, it is not possible to estimate sampling errors. Therefore, validity of inferences to a population cannot be ascertained" (Pedhazur & Schmelkin, 1991, p. 321).

Significance Tests and Confidence Intervals

Significance tests and confidence intervals are used to make inferences about a population. Null hypothesis significance testing (NHST) is used to determine whether a difference between means, for example, can be attributed to sampling error (the null hypothesis) or a main effect (Fisher, 1966). Confidence intervals provide a range of values that have a certain probability of a population parameter falling between them and are determined by the sample statistic and the standard error of that statistic. Confidence intervals have been proposed as a substitute for the NHST process (Cohen, 1994; Cumming, 2014). However, a confidence interval similarly structured to a hypothesis tested with NHST yields identical conclusions because they are both based on the same standard error, the same probability distribution, and the same assumptions regarding sampling (Frick, 1996).

The key assumption that significance tests and confidence intervals share is that the data are independent and identically distributed (*iid*). Data are *iid* when they are randomly sampled from a single population (Klenke, 2020). If there is any correlation between the probability of being selected and the variable being measured, the data cannot be *iid*, which means that researchers cannot make valid inferences about the population. One potential remedy for this problem is the post hoc adjustment.

Post Hoc Corrections for Nonprobability Samples

There are many different methods used to correct for sampling bias, but four of the most prominent ones are post-stratification adjustments, raking, generalized regression (GREG) modeling, and propensity scoring. Post-stratification, also known as ratio adjustment and cell weighting, is the adjustment of sample weights so that the sample totals conform to the population cell totals (Kalton & Flores-Cervantes, 2003). Raking is similar to post-stratification, except it adjusts sample weights so that the sample totals conform to the marginal totals instead of the cell totals (Tourangeau et al., 2013). For example, if there is a population of males and females who are college freshmen and sophomores, post-stratification would adjust the sample weights to match the population of male freshmen or female sophomores, but raking would only adjust the sample weights to match the population of males or freshmen. GREG modeling benchmarks sample estimates to the corresponding population figures by assuming a linear relationship between an analysis variable and a set of covariates (Tourangeau et al., 2013). Finally, propensity scoring adjusts for the predicted probability that a participant will end up in a group due to a confound, and it is often used in conjunction with sample matching (Rivers & Bailey, 2009; Tourangeau et al., 2013).

Tourangeau et al. (2013) compared the results of eight different studies that analyzed the effectiveness of these corrections. Across all eight studies, it was found that the adjustments only partially removed the bias, that the adjustments sometimes increased the biases relative to unadjusted estimates, that the relative biases left after adjustment were often substantial, and that there were large differences across variables. Based on this research, post hoc adjustments can be useful but are ultimately fallible.

Monte Carlo Simulation

A Monte Carlo simulation is a strategy used to create a pseudo-population from which samples can be generated (Mooney, 1997). In Monte Carlo studies, a large number of samples are drawn from a population and results are averaged across these samples. In addition, the ready availability of computers allows Monte Carlo studies to explore a variety of conditions for these populations, allowing researchers to explore conditions for which various procedures function poorly. Monte Carlo analyses are primarily used when experiments would be overly time-consuming, costly, or otherwise impractical (Harrison, 2010). However, a Monte Carlo simulation is seldom simple; it can be an involved process. Paxton et al. (2001) outlined nine steps: (1) developing a research question, (2) creating a valid model, (3) designing specific experimental conditions, (4) choosing values of population parameters, (5) choosing an appropriate software package, (6) executing the simulations, (7) file storage, (8) troubleshooting and verification, and (9) summarizing the results. Monte Carlo simulations are used in a variety of disciplines, including engineering, physical sciences, life sciences, investing, and mathematics (Rubinstein & Kroese, 2016). They are used in psychology as well; for example, Brand et al.

(2008) used the Monte Carlo method to simulate psychological experiments and found that effect sizes are overestimated when there is a bias against publishing nonsignificant results.

Although there has been discussion about the fallibility of sampling bias corrections, there is a methodological gap in measuring the efficacy of these corrections in producing representative samples drawn from nonprobability sampling methods. This study will use a Monte Carlo analysis to examine the effect of sampling bias and the effectiveness of post-stratification corrections.

The Present Study

The present study uses a Monte Carlo simulation to first, investigate the effects of non-probability samples and second, investigate whether post-stratification can result in a sample that represents the relevant population. Although it is usually impossible to determine the accuracy of nonprobability sample statistics, the Monte Carlo design of this study allows the researcher to create a population with known values to determine accuracy. I drew samples from a defined population with known values, set a 95% confidence interval around the sample means, and determined whether the population means were within the interval. Confidence intervals for probability samples should contain the population mean in the defined percent (e.g., 95%) of the samples.

I explored the effects of sample bias under a variety of conditions. One benefit of this research is determining how sensitive sample statistics are to sample bias. In addition, this study explores the effectiveness of a post hoc correction for sample bias, post-stratification. Although most statisticians use post-stratification knowing that it will not completely rid a sample of bias (Tourangeau et al., 2013), it can eliminate bias if there is no relationship between the probability of participants being selected and the variable being measured or if the correction perfectly offsets the relationship.

> *Research Questions:* As sampling characteristics depart from simple random sampling, the resultant confidence intervals will successfully locate the population at rates less than the nominal (e.g., 95%) level. Furthermore, post hoc adjustments to sample means will not eliminate sample bias when sample likelihood correlations vary by subgroup.

Method

Procedure

I used the open-source statistical software program R to create two types of population datasets, one homogeneous and the other heterogeneous. The simpler population, homogeneous, consisted of only a single group of observations (i.e., hypothetical participants). The heterogenous population consists of three subgroups of hypothetical participants in equal proportion. In both populations the dependent variable for each observation was a single score on a normally distributed continuous variable. Sample means were computed, and a confidence interval was formed around the sample mean. The outcomes of interest are the sample means and whether the confidence interval contains the population mean.

The population mean was set to zero for all analyses. In the homogeneous population, two variables were manipulated: sample eligibility and the correlation of sample likelihood with dependent variable scores. Sample eligibility refers to the percent of the population eligible to be sampled. For example, a sample eligibility of .20 reflects a situation where 80% of the population has a probability of zero of being sampled. The second variable, the relationship between sample likelihood and dependent variable scores, involved the creation of a continuous variable score for each observation (called sample likelihood), which models how likely an observation is to be sampled. This likelihood variable was generated to have a specified correlation with the dependent variable. Thus, if positively correlated with the dependent variable, then observations with higher dependent variable scores will be more likely to be in the sample eligible group (from which observations are randomly sampled). The combination of these two variables should approximate nonprobability samples in a variety of applications. For example, subjects randomly sampled from students currently enrolled across all American universities (an idealized version of this example) are only a small percent of the adult population in the United States (approximately 8% of the adult population in the United States is currently enrolled in a university). Thus, sample eligibility in this scenario is .08. If the variable of interest happens to be correlated with current college enrollment, then the sample likelihood correlation will be a nonzero value.

Table 1 is a demonstration of the sampling model used in this study. In the table, thirty observations have a score on X and a likelihood score. The correlation between the two variables is.10. Sample eligibility is .5 in the example. Thus, any observation with a likelihood greater than .5 is eligible to be sampled, and the sample collected would be a random sample of those eligible. Finally, the slight positive correlation between scores on X and likelihood indicates that those more likely to be sampled will have greater scores on X.

Sample eligibility, and the low levels of eligibility found in psychological research, can be illustrated in a recent study on delay of gratification in children. Carlson et al. (2018, p. 4) described their sample as consisting of "…540 typically developing children ages 3 to 5 years (see Table 1). Families were recruited from participant pools at two urban universities, University of Washington (n = 296) tested between 2002 and 2007 and University of Minnesota (n = 244) tested between 2008 and 2012." Making some rather generous assumptions about the size and nature of the participant pools (that 10% of all students in a given year were in the participant pool and every one of these students had a child in the 3 to 5 age range), only .26% of the total population of children in America aged 3 to 5 (that is, 55,000 out of 2,143,400) were eligible to be included in the sample. Thus, even if the participants in their sample of 540 were randomly selected from the participant pool, over 99.7% of the population was ineligible for inclusion in the study.

When sample eligibility is low, such as in the Carlson et al. (2018) study, then even weak correlations of sample likelihood with the dependent variable are likely to result in biased sample means. This bias will likely be present even if samples are randomly drawn from the eligible population because it is the "who is eligible?" aspect that is correlated with the dependent variable.

I explored the effects of sample bias through combinations of these two variables. A sample collected via simple random sampling would have a sample eligibility value of 1.0 and a likelihood correlation of 0.0. In such a scenario, the probability that any one person would be sampled equals the sample size divided by the population size, and this probability would be the same for every member of the population. If the eligibility is something less than 1.0 and the likelihood correlation is still 0.0, then the resultant sample mean would still be unbiased; even though some portion of the population has a probability of selection of zero, the characteristics of the eligible portion of the population are the same as the population as regards the measured variable. Thus, sample bias should be expected as the likelihood correlation departs from zero. Furthermore, these effects should be magnified as the eligibility becomes more restrictive. The effect of sample bias was assessed via confidence interval effectiveness (whether the 95% confidence interval correctly locate the population mean in 95% of the samples).

The heterogeneous population has the same variables as the homogeneous population but consists of three subgroup populations, equal in size. The subgroup population means, sample eligibility, and likelihood correlation varied by subgroup. Finally, because there are three subgroup populations of known size, I performed a post hoc adjustment to the sample mean. Confidence interval effectiveness was assessed before and after these adjustments.

Each sample consisted of scores on 100 observations. Each condition was replicated 10,000 times, and results (sample means and confidence interval success) were averaged across the 10,000 samples.

Results

Homogeneous Population

Mean values across 10,000 replications for each condition are listed in Table 2. A few patterns in the results are evident. First, bias in the sample mean (i.e., the difference between the sample mean and the population mean) attenuates the effectiveness of the confidence interval (i.e., the 95% confidence interval fails to contain the population mean in 95% of the samples). Second, if the correlation between sample likelihood and scores on the variable of interest is zero, then no amount of eligibility restriction results in a biased sample mean. In such cases, there is no penalty for collecting a sample from only the sample eligible portion of the population. Third, when the likelihood correlation is not zero, the expected sample mean will depart from the population mean; stronger positive correlations result in a positively skewed sample mean because higher scoring observations are more likely to be selected. Fourth, at low levels of sample eligibility, even very weak correlations (e.g., .05) of dependent variable scores with eligibility status result in a biased mean and an ineffective confidence interval.

Relations between sample eligibility and dependent variable scores can also be described with a point-biserial correlation. A point-biserial correlation indicates the relationship between a dichotomous variable (in the present study, sample eligibility: eligible versus not eligible) and a continuous variable (score on the variable measured). Due to the nature of dichotomized variables, point-biserial correlations are even weaker than the Pearson correlations that were employed to generate the datasets. The point-biserial may better describe the interaction of the variables at work. For example, in one condition shown Table 2, sample eligibility is .01 and the Pearson correlation between likelihood, a continuous variable, and dependent variable scores is .05. Thus, only the top one percent of likelihood scores are eligible to be sampled, and likelihood has a very weak relationship with the dependent variable. However, when the likelihood score is dichotomized into an eligible/not eligible status (samples were randomly drawn from these eligible cases), the resultant point-biserial correlation is a near-zero .0133. Nevertheless, a weak association such as this one is still potent enough to bias the expected sample mean to .134 and cause the 95% confidence interval to successfully locate the population mean only 74% of the time.

Heterogenous Population

The heterogenous group results are more complicated as there are now three group means, three sample eligibilities, and three likelihood correlations. Results for the heterogeneous population analyses are listed in Tables 3-6. An inspection of the results shows that in some conditions the post hoc correction brings the sample closer to the population mean, in other conditions the adjustment skews the sample mean further from the population mean, and in a few conditions the sample mean is adjusted to some value that varies the same amount from the population mean. The conditions that lead to these outcomes are complex and interact. To make matters worse, researchers have no information as to what outcome will occur when making a correction.

Table 3 presents conditions in which all three groups have the same eligibility and a likelihood correlation of zero. Thus, the conditions shown in Table 3 are identical to the first column in Table 2 (homogeneous population results) and the adjustment to the sample mean is no adjustment at all as the groups are sampled in the same proportions.

Table 4 displays conditions where the three groups have different probabilities of being sampled; likelihood correlations are the same for each group. The results shown in Table 4 demonstrate the interaction of the various factors affecting the sample mean and the confidence interval effectiveness. In a pattern counter to that seen in Table 2, as the likelihood correlation increases in strength from 0 to .1, the expected value of the sample mean becomes less biased (and the confidence interval more effective). The fourth column of data (likelihood correlation = .1) features an expected sample mean of -.0093 and a confidence interval with .951 effectiveness. The reason for this unexpected result rests in combination of factors, among them the population means for each group. In all of the heterogeneous population datasets the subgroup population means were set to -.25, 0, and +.25 for Groups 1-3, respectively. In the case of Table 4, Group 1, the group with the lowest population mean, was the group most likely to be sampled (sample eligibility = .5), and Group 3, the group with the highest population mean, was the group least likely to be sampled (sample eligibility = .1). Thus, when the likelihood correlation was set to zero (first column), the sample mean was negatively biased due to overrepresentation of Group 1 (mean $\bar{X}$ = -.111, confidence interval effectiveness = .81). The subsequent adjustment to the sample mean equalized the group representation, brought the mean to a value close to the population mean (mean $\bar{X}$ =.005), and raised the confidence interval effectiveness to .951. However, when the likelihood correlation was set to .1 (fourth column), even though the group with the lowest mean (Group 1) was oversampled, the positive likelihood correlation meant that the observations sampled were the higher scoring ones. Thus, the biased selection within each group (positive likelihood correlations) offset the biased sampling of each group (differential sampling eligibilities). In an ironic conclusion, the subsequent adjustment to the sample mean introduced bias (adjusted mean = +.122) and reduced the effectiveness of the confidence interval (.901).

This confluence of factors that resulted in an unbiased unadjusted sample mean in the fourth column of Table 4 is purely happenstance. To illustrate this point, I repeated the analyses of the fourth column with the means flipped so that Group 1 had the highest mean and Group 3 had the lowest mean. Thus, the group with the highest mean was most likely to be sampled; within all groups higher scores were most likely to be selected. The resultant sample mean was positively biased (mean $\bar{X}$ = .212), and the confidence interval was effective for only 45% of the samples.

Tables 5 and 6 offer more results like those of Table 4. They further demonstrate the complexity of the factors that result in biased sample means and diminished confidence interval effectiveness. They also offer more evidence that post hoc adjustments to the sample mean have unpredictable effects (more bias, less bias, or no change in bias) with no indication for the researcher as regards the likely outcome (more bias, less bias, or no change in bias) of the adjustment.

Discussion

Overview

The results shown in Tables 2-6 indicate a clear conclusion: it is very easy to collect a sample in a way that results in a biased sample mean. In addition, a 95% confidence interval placed around a biased sample mean is frequently unsuccessful at locating the population mean. Moreover, it is difficult to successfully correct for this bias given the available information, and it is impossible to know when any such correction has increased or decreased the amount of bias.

This study modeled selection as a random sample of those eligible. The presence of some degree of random selection (e.g., participants were randomly selected from the university subject pool) may comfort researchers; however, the results presented here show that the damage was already done to the sample mean before any random selection occurred.

It is important to note that researchers are operating in an information-constrained environment. That is, researchers have no knowledge of the population mean, no knowledge of the degree of bias in the sample mean, and no knowledge of the correlation between sample likelihood and the variable measured. This study assumed only that the relative sizes of the subgroups in the population are known. It is on the basis of this knowledge alone that the adjustment was performed.

No Salvation Though Large Sample Sizes

One proposed solution for the ails of sampling error, large sample sizes, offers no benefit for the biased sample. To demonstrate this point, I increased the sample size to 400 and repeated the homogeneous population analyses. The larger sample size did not change the expected sample means; the extent and direction of the bias remained the same. The standard deviation of the sample means (i.e., the standard error of the mean) was reduced, and thus, the width of the confidence interval was reduced as well. The result of placing a narrower confidence interval around the same biased sample mean is a confidence interval that successfully locates the population mean less often. As but one example, I repeated the homogeneous group analyses with a sample eligibility of .01 and likelihood correlation of .04. The effectiveness of the 95% CI dropped from .87 for the $N = 100$ condition to .43 for the $N = 400$ condition.

Illusory Benefits of Post Hoc Adjustments

One curious finding in the heterogeneous population analyses was that some adjustments to the sample mean resulted in more bias (greater deviation from the population mean) but a confidence interval that was no less effective. Apparent improvements in confidence interval effectiveness after adjustments were frequently an artifact of the wider confidence interval that resulted from the reduced sample size associated with the adjustment. For example, a 95% confidence interval formed with a standard deviation of 1.0 and a sample size of 100 would have an interval width of approximately .392, whereas a similar interval from a sample size of 33 results in an interval width of .682 (a 74% wider interval). One such instance can be seen in the third column of result in Table 6 where the correction has changed the expected sample mean from -.0492 to .045. The adjusted sample mean is approximately the same distance from the population mean. Thus, the confidence interval should have the same effectiveness. Because the adjustment lowered the sample size from 100 to 33.4, the wider interval successfully located the population mean more often (92.6% before the adjustment versus 94.6% after the adjustment). This improvement is illusory as a similar sized interval would not display the enhanced effectiveness. Similarly, in situations where the adjustment caused the mean to deviate from the population mean to a greater extent, an interval based on the full sample size would be even less effective than what is shown.

To address the confidence interval width issue, I explored a different type of post-stratification adjustment. This alternate procedure preserved the full sample size by computing an equally weighted composite from the sample means of each group. Thus, I computed an adjusted mean while preserving the full sample size. Results from this analysis demonstrated that unless the adjustment brought the sample mean closer to the population mean, confidence interval effectiveness was negatively affected. Confidence interval effectiveness was even worse than what would be expected given the adjusted mean and interval width; the standard deviation of the sample means also increased indicating that this adjustment method introduced sampling errors not accounted by the standard error of the mean equation.

Limitations

Every Monte Carlo analysis is limited by the conditions the researcher chooses to implement as well as the manner of implementation. In the case of the present study the heterogeneous population is defined by a single factor with three levels. A two-factor categorical variable would enable multiple adjustment procedures, including adjustments based on cell means in addition to row or column means. This study investigated only one possible post hoc adjustment to the sample mean and found no evidence for the consistent effectiveness of this adjustment; the nature of the adjustment used in this study is but one way for a sample mean to be adjusted in the pursuit of a value that more closely represents the population mean. I made a limited exploration of one other adjustment procedure. Other procedures exist, and they may offer better results than what are reported herein.

Conclusions

Researchers conducting studies with low levels of sample eligibility are playing with fire as even extremely weak correlations between sample eligibility and the focal variable lead to disastrous results. The resultant bias in the sample mean is not adequately addressed by confidence interval equations, cannot be ameliorated with large sample sizes, and will only be successfully corrected with a post hoc adjustment by chance.

References

Brand, A., Bradley, M. T., Best, L. A., & Stoica, G. (2008). Accuracy of effect size estimates from published psychological research. *Perceptual and Motor Skills*, *106*(2), 645–649. https://doi.org/10.2466/pms.106.2.645-649Carlson, S. M., Shoda, Y., Ayduk, O., Aber, L., Schaefer, C., Sethi, A., Wilson, N., Peake, P. K., & Mischel, W. (2018). Cohort effects in children's delay of gratification. *Developmental Psychology, 54*(8), 1395–1407. https://doi.org/10.1037/dev0000533

Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). John Wiley & Sons.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Fisher, R. A. (1966). *The Design of Experiments* (8th ed.). Edinburgh: Hafner.

Frick, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods*, *1*(4), 379–390. https://doi.org/10.1037/1082-989X.1.4.379

Harrison, R. L. (2010). Introduction to Monte Carlo simulation. In: AIP Conference Proceedings, *1204*, 17–21. https://doi.org/10.1063/1.3295638

Hays, W. L. (1988). *Statistics* (4th ed.). Holt, Rinehart, and Winston.

Hershcovis, M. S., Cameron, A.-F., Gervais, L., & Bozeman, J. (2018). The effects of confrontation and avoidance coping in response to workplace incivility. *Journal of Occupational Health Psychology*, *23*(2), 163–174. https://doi.org/10.1037/ocp0000078

Ilies, R., Hauserman, N., Schwochau, S., & Stibal, J. (2003). Reported incidence rates of work-related sexual harassment in the United States: Using meta-analysis to explain reported rate disparities. *Personnel Psychology, 56*(3), 607–631. https://doi.org/10.1111/j.1744-6570.2003.tb00752.x

Kalton, G., & Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics*, *19*(2), 81–97.

Klenke, A. (2020). *Probability Theory: A Comprehensive Course* (3rd ed.). Springer Science & Business Media.

Lee, S. (2006). An evaluation of nonresponse and coverage errors in a prerecruited probability Web panel survey. *Social Science Computer Review*, *24*, 460–475.

Lohr, S. L. (1999). *Sampling: Design and Analysis*. Duxbury Press.

Mooney, C. Z. (1997). *Monte Carlo Simulation*. SAGE Publications, Inc., https://dx.doi.org/10.4135/9781412985116

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation, structural equation modeling, *A Multidisciplinary Journal*, *8*(2), 287-312, https://doi.org/10.1207/S15328007SEM0802_7

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rivers, D., & Bailey, D. (2009). Inference from matched samples in the 2008 U.S. national elections. *Proceedings of the Joint Statistical Meetings*, *1*, 627–639.

Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmüller, C., Aziz, S., & Knight, W. E. (2003). Profiling Active and Passive Nonrespondents to an Organizational Survey. *Journal of Applied Psychology, 88*(6), 1104–1114. https://doi.org/10.1037/0021-9010.88.6.1104

Rubinstein, R. Y., & Kroese, D. P. (2016). *Simulation and the Monte Carlo Method* (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/9781118631980

Tourangeau, R., Conrad, F. G., & Couper, M. P. (2013). *The Science of Web Surveys*. Oxford University Press.

Wimbush, J. C., & Dalton, D. R. (1997). Base rate for employee theft: Convergence of multiple methods. *Journal of Applied Psychology, 82*(5), 756–763. https://doi.org/10.1037/0021-9010.82.5.756

Table 1

Sampling Model

Observation	*X*	Likelihood
1	*-0.128253*	*0.95395466*
2	*-0.8434122*	*0.93762798*
3	*0.31487059*	*0.72391977*
4	*0.5649171*	*0.56395474*
5	*0.81337363*	*0.51698005*
6	*-0.6620843*	*0.50839876*
7	-1.0430537	0.48307724
8	0.91170886	0.4542518
9	1.05047946	0.44881648
10	0.62219059	0.33055836
11	0.17958144	0.28119662
12	0.53588559	0.27977928
13	-0.4753217	0.27373595
14	1.01882272	0.20931332
15	-1.2491926	0.12762911
16	-1.7228676	0.07990412

Note. Observations eligible for selection (Likelihood > .5; i.e., eligibility = .5 in this example) are in italics. Study sample would be a random sample of those eligible. Correlation between score on *X* and likelihood variable is .10.

Table 2

Homogeneous Population Results

	Sample likelihood correlation						
	0	.01	.02	.03	.04	.05	.1
			Sample eligibility = .5				
$E(\bar{X})$	-.0010	.0070	.0152	.0232	.0307	.0403	.0817
SD	.100	.100	.100	.101	.100	.099	.099
95% CI	.949	.949	.951	.940	.939	.935	.876
			Sample eligibility = .1				
$E(\bar{X})$	.0008	.0171	0.0350	.0544	.0696	.0901	.1761
SD	.100	.101	.100	.100	.100	.100	.100
95% CI	.950	.941	.933	.917	.890	.853	.582
			Sample eligibility = .01				
$E(\bar{X})$	.0001	.0271	.0524	.0816	.1062	.1337	.2666
SD	.099	.100	.099	.100	.100	.099	.099
95% CI	.951	.939	.922	.873	.817	.736	.246
			Sample eligibility = .001				
$E(\bar{X})$	.0002	.033	.0675	.1006	.1340	.1666	.3365
SD	.100	.100	.101	.099	.099	.100	.098
95% CI	.948	.937	.897	.833	.741	.619	.081

Note. $N = 100$ for each sample. $E(\bar{X})$ and *SD* are the mean and standard deviation of the 10,000 sample means. 95% CI indicates the proportion of the 10,000 replications for which the population mean was contained in the confidence interval.

Table 3

Heterogeneous Population: Constant Sample Eligibility and Constant Likelihood Correlation

	Sample eligibility		
	.05	.1	.5
$E(\bar{X})$	-.0002	.0007	.0006
SD	.103	.103	.102
95% CI	.947	.944	.946
Adjusted *N*	85.507	85.466	85.512
Adjusted $E(\bar{X})$	.0005	.0007	.0005
Adjusted *SD*	.110	.110	.109
Adjusted 95% CI	.950	.951	.953

Note. Sample likelihood correlation = 0 for all groups. *N* = 100 for each sample. $E(\bar{X})$ and *SD* are the mean and standard deviation of the 10,000 sample means. 95% CI indicates the proportion of the 10,000 replications for which the population mean was contained in the confidence interval.

Table 4

Heterogeneous Population: Varied Sample Eligibility and Constant Likelihood Correlation

Sample eligibility: Group 1	.5	.5	.5	.5
Sample eligibility: Group 2	.3	.3	.3	.3
Sample eligibility: Group 3	.1	.1	.1	.1
Likelihood correlation: Group 1	0	.03	.05	.1
Likelihood correlation: Group 2	0	.03	.05	.1
Likelihood correlation: Group 3	0	.03	.05	.1
$E(\bar{X})$	-.1114	-.0819	-.0603	-.0093
SD	.101	.101	.102	.101
95% CI	.812	.874	.910	.951
Adjusted *N*	33.220	33.389	33.460	33.368
Adjusted $E(\bar{X})$	.0005	.0341	.0620	.122
Adjusted *SD*	.182	.181	.179	.182
Adjusted 95% CI	.957	.955	.944	.901

Note. $N = 100$ for each sample. $E(\bar{X})$ and *SD* are the mean and standard deviation of the 10,000 sample means. 95% CI indicates the proportion of the 10,000 replications for which the population mean was contained in the confidence interval.

Table 5

Heterogeneous Population: Constant Sample Eligibility and Varied Likelihood Correlation

Sample eligibility: Group 1	.05	.05	.1
Sample eligibility: Group 2	.05	.05	.1
Sample eligibility: Group 3	.05	.05	.1
Likelihood correlation: Group 1	.05	.1	.1
Likelihood correlation: Group 2	.03	.05	.05
Likelihood correlation: Group 3	0	0	0
$E(\bar{X})$	.05438	.1012	.0878
SD	.102	.099	.103
95% CI	.918	.835	.857
Adjusted *N*	85.466	85.443	85.417
Adjusted $E(\bar{X})$	.055	.101	.087
Adjusted *SD*	.109	.506	.110
Adjusted 95% CI	.923	.853	.875

Note. $N = 100$ for each sample. $E(\bar{X})$ and *SD* are the mean and standard deviation of the 10,000 sample means. 95% CI indicates the proportion of the 10,000 replications for which the population mean was contained in the confidence interval.

Table 6

Heterogeneous Population: Varied Sample Eligibility and Varied Likelihood Correlation

Sample eligibility: Group 1	.1	.5	.5
Sample eligibility: Group 2	.05	.3	.3
Sample eligibility: Group 3	.05	.1	.1
Likelihood correlation: Group 1	.1	.05	.1
Likelihood correlation: Group 2	.05	.03	.05
Likelihood correlation: Group 3	0	0	0
$E(\bar{X})$	.05035	-.0781	-.0492
SD	.101	.101	.101
95% CI	.920	.880	.9258
Adjusted *N*	66.617	33.398	33.410
Adjusted $E(\bar{X})$	.091	.023	.045
Adjusted *SD*	.123	.181	.181
Adjusted 95% CI	.887	.952	.946

Note. N = 100 for each sample. $E(\bar{X})$ and *SD* are the mean and standard deviation of the 10,000 sample means. 95% CI indicates the proportion of the 10,000 replications for which the population mean was contained in the confidence interval.

Copyright Permission

Name: Hong, Julia

Email (to receive future readership statistics): juliahong39@gmail.com

Type of document: ['Thesis']

Title: A Monte Carlo Analysis of Nonprobability Sampling and Post Hoc Corrections

Keywords (3-5 keywords not included in the title that uniquely describe content): nonprobability sampling, selection bias, data interpretation, monte carlo method

Committee Chair: Reagan Brown

Additional Committee Members: Xiaowen Chen, Katrina Burch

Select 3-5 TopSCHOLAR® disciplines for indexing your research topic in TopSCHOLAR®: Industrial and Organizational Psychology, Statistical Methodology, Applied Statistics