

## **CALCULATING THE IDEAL SAMPLE SIZE FOR TARGETING POLICY TRAINING AND CERTIFICATION**

**Nouhoun Moro**

Assistant Professor, Department of Statistics, University of Malaysia

**Boukary Soumahoro**

Assistant Professor, Department of Statistics, University of Malaysia

Article received: 28/01/2025, Article Revised: 21/02/2025, Article Accepted: 27/03/2025

**DOI:** <https://doi.org/10.55640/tpjrjsms-v02i03-01>

© 2025 Authors retain the copyright of their manuscripts, and all Open Access articles are disseminated under the terms of the [Creative Commons Attribution License 4.0 \(CC-BY\)](#), which licenses unrestricted use, distribution, and reproduction in any medium, provided that the original work is appropriately cited.

### **ABSTRACT**

Sample size calculation is a fundamental step in designing an effective training and certification policy, particularly when developing targeted programs aimed at specific groups. This article investigates the importance of calculating the appropriate sample size for training and certifying targeting policies, particularly in scenarios where policy implementation can be influenced by various factors such as skill levels, economic conditions, and demographic differences. The research outlines the steps required to determine sample sizes that provide statistical power and practical relevance for policy decisions. It introduces a methodology for sample size determination and discusses how this calculation impacts the success of training programs in meeting their objectives, ensuring valid certification, and achieving reliable outcomes. The findings suggest that an appropriate sample size can significantly enhance the accuracy of training evaluations and policy effectiveness.

### **KEYWORDS**

Sample size calculation, training programs, certification, targeting policies, statistical power, policy evaluation.

### **INTRODUCTION**

Training programs, especially those designed to implement and certify specific targeting policies, are essential components of contemporary policymaking. Whether aimed at enhancing employee skills, increasing public awareness, or certifying professionals in specific fields, training programs rely heavily on robust sampling methodologies to ensure valid, reliable, and applicable results. Effective training programs not only require proper content and methodology but also demand careful attention to the number of participants involved. This ensures that the data generated from these programs are statistically significant and generalizable to the larger population.

Sample size calculation plays a pivotal role in the design and implementation of these policies. It is a critical aspect of experimental design and statistical analysis that allows researchers and policymakers to estimate the necessary number of participants to achieve a specific level of

confidence in their results. In the context of training and certification for targeting policies, having the correct sample size ensures that the conclusions drawn from the program are robust and applicable, particularly when targeting policies are intended to address particular social, economic, or demographic groups.

In this article, we explore the methodology behind calculating sample sizes for training and certification programs. This process involves understanding factors such as expected effect sizes, statistical power, and variability within the population. We aim to provide insights into how accurate sample size estimations can influence the effectiveness of targeting policies. By discussing theoretical approaches, practical applications, and examples, this article aims to serve as a guide for policymakers and researchers involved in the design of training programs that are both efficient and impactful.

The calculation of sample size is an essential aspect of

designing training programs and certifying targeting policies. The success of any training initiative, particularly those aimed at certifying individuals for targeted roles or specific policies, hinges on the ability to measure outcomes effectively. A well-calculated sample size is a pivotal part of this process as it ensures the statistical power and reliability of the findings that shape the policy's effectiveness. Whether in social, economic, or organizational settings, targeted policies often require training programs that cater to specific groups, and calculating the right number of participants in these programs is key to ensuring their success.

In training and certification contexts, targeting policies refer to initiatives designed to focus on particular groups of individuals based on various demographic, economic, or social criteria. These programs are commonly used in areas such as workforce development, education, and professional certification, where different groups require tailored content or delivery to meet specific needs. With the increasing importance of such programs, ensuring the validity and reliability of the conclusions drawn from them becomes crucial. This is where sample size calculation plays an indispensable role.

Sample size calculation is essentially the process of determining the number of participants needed in a study or program to ensure the results are statistically significant. If too few participants are chosen, the study might fail to detect meaningful effects, while too many participants could result in wasted resources, time, and effort. An ideal sample size ensures that the conclusions derived from the training and certification process are both statistically valid and practically relevant.

The primary goal of any training or certification program is to assess whether a particular policy or intervention has a meaningful impact on participants. In targeting policies, these effects could involve improvements in skills, knowledge, or even behavioral outcomes. As such, determining the right sample size ensures that the evaluation of these policies is carried out with enough statistical power to identify whether the policy is effective or not. Without an appropriate sample size, any evaluation risks being either inconclusive or misleading.

The factors that influence sample size calculations include the effect size, variability within the population, significance level ( $\alpha$ ), and the desired statistical power ( $1-\beta$ ). The effect size refers to the magnitude of the difference the policy or training program is expected to make. The more significant the expected change, the smaller the required sample size. Conversely, when the change is subtle or small, a larger sample size is necessary to detect the effect with confidence. Variability within the population also influences the sample size—greater variability demands a larger sample to draw valid conclusions.

This article aims to explore the methods and theories behind sample size calculations in the context of training and certifying targeting policies. It will discuss the importance of a well-calculated sample size in ensuring the success of such programs, the methodology involved in determining the correct size, and the impact of these calculations on policy effectiveness. The following sections will elaborate on how various factors come into play during the sample size calculation process and why this step is crucial for obtaining reliable, actionable insights that can drive policy improvements.

By the end of this article, readers will have a better understanding of the statistical principles behind sample size determination and how to apply these principles to real-world training and certification programs. This will empower policymakers, researchers, and program designers to create more effective, data-driven policies that achieve their intended outcomes with precision.

## **METHODS**

### **Sample Size Calculation Theory**

Sample size calculation involves determining the number of observations or participants required to achieve reliable, statistically significant results in a study or program. The formula for sample size calculation often depends on several key factors:

1. **Effect Size:** The expected magnitude of the effect or difference between groups (e.g., between trained and untrained individuals). In targeting policies, this might refer to the expected difference in skill levels or certification success rates after training.
2. **Significance Level ( $\alpha$ ):** The probability of rejecting the null hypothesis when it is true (i.e., the Type I error rate). Typically, a significance level of 0.05 is used, which corresponds to a 5% chance of wrongly rejecting the null hypothesis.
3. **Statistical Power ( $1-\beta$ ):** The probability of correctly rejecting the null hypothesis when it is false. A power of 80% (0.80) is generally considered acceptable in most social science studies. Higher power means a lower likelihood of a Type II error (failing to detect a true effect).
4. **Variance or Standard Deviation:** The variability within the population being studied. Larger variability often requires a larger sample size to detect an effect reliably.

### **Practical Application of Sample Size Calculation for Targeting Policies**

To illustrate how sample size calculation works in practice for training and certifying targeting policies, we

can use a simple example of a policy aimed at training workers in a specific industry, with the goal of certifying them as qualified professionals.

1. **Determining the Effect Size:** If the objective of the training is to increase productivity or certification success rates among a group of workers, an initial estimate of the effect size is needed. Suppose historical data or pilot studies suggest that the training will increase success rates from 50% to 70%. This gives an effect size of 0.2 (Cohen's  $d$ ).

2. **Choosing a Significance Level and Power:** In many cases, a significance level of 0.05 is chosen, and a power of 0.80 is aimed for to reduce the probability of Type II errors.

3. **Assessing Variability:** If there is significant variability in the baseline success rate (e.g., some workers perform well, and others perform poorly), this must be factored into the sample size calculation. A larger sample size would be required if the variability is higher.

Using a sample size calculation tool, such as G\*Power or statistical software like R, these factors are inputted into a formula that estimates the necessary sample size.

## Example Calculation Using a Hypothetical Training Program

For this example, assume a training program designed to improve the certification success rate among a group of workers:

- Effect Size (Cohen's  $d$ ): 0.2
- Significance Level ( $\alpha$ ): 0.05
- Power ( $1-\beta$ ): 0.80
- Standard Deviation ( $\sigma$ ): 0.5 (based on variability in performance)

Using a power analysis tool, the required sample size for each group (control and treatment) is estimated at 100 participants. This suggests that to detect a statistically significant difference between trained and untrained groups, 100 participants per group are needed.

## RESULTS

The results of the sample size calculation provide several key insights for the design of training and certification policies:

1. **Appropriate Sample Size for Policy Implementation:** For the hypothetical training program, a total of 200 participants (100 in each group) would be required to detect an effect size of 0.2 with 80% power and a 5% significance level. This ensures that the policy

can be evaluated with sufficient confidence, minimizing both Type I and Type II errors.

2. **Impact of Variability:** Variability within the population plays a crucial role in determining the sample size. If the success rate of certification is highly variable, a larger sample size would be required to achieve the same level of statistical power. Therefore, understanding the expected variability in the population is key to ensuring the success of the training program.

3. **Precision and Confidence:** Calculating the appropriate sample size ensures that the results of the training program are both reliable and precise. Small sample sizes may lead to overestimating the impact of the training program or failing to detect true effects, which could result in misleading policy decisions.

4. **Considerations for Real-World Application:** In practice, policymakers may need to adjust for factors like participant attrition or non-compliance, which can influence the required sample size. A buffer is often added to the sample size to account for these real-world issues.

## DISCUSSION

The process of sample size calculation is a critical aspect in designing training and certification programs for targeting policies. A proper sample size ensures that the conclusions drawn from a study or training evaluation are reliable, statistically valid, and actionable. When dealing with targeted policies, where interventions are tailored for specific subgroups within a population, determining the correct sample size is even more crucial. In this section, we discuss the importance of calculating sample size in the context of training and certification programs, the challenges involved, and how it directly influences the effectiveness and implementation of targeted policies.

### Significance of Sample Size Calculation

Sample size calculation serves several important functions in research and policy evaluation, particularly for training and certification programs. The primary purpose is to ensure that the study or program has enough participants to detect meaningful differences or effects, should they exist. In the context of targeting policies, this is particularly important because these policies are often designed for specific groups with particular characteristics (e.g., low-income individuals, marginalized communities, or specific industry professionals).

If the sample size is too small, the program may lack statistical power, leading to an inability to detect a real effect, even if one exists. This could result in ineffective policy decisions, where policymakers might incorrectly conclude that the training program or certification policy

had no impact. On the other hand, choosing a sample size that is too large can be inefficient and costly, leading to the unnecessary allocation of resources without improving the quality of the results. Therefore, accuracy in sample size calculation ensures that resources are used effectively and that the training or certification program can be evaluated in a statistically robust manner.

### **Factors Influencing Sample Size Calculation**

Several key factors influence the calculation of sample size, and understanding these factors is essential for designing a well-powered training or certification program.

1. **Effect Size:** The effect size represents the magnitude of the difference or impact expected from the intervention. In the case of targeting policies, the effect size might refer to how much the training or certification program is expected to improve skills, knowledge, or other measurable outcomes. A larger effect size typically requires a smaller sample size, as a more substantial effect is easier to detect. However, the challenge lies in accurately estimating the effect size, especially in new or untested training programs. Often, historical data or pilot studies are used to make an informed estimate of the expected effect size. If the effect size is underestimated, the sample size might be too small, risking the failure to detect important effects.

2. **Variability (Standard Deviation):** The variability or standard deviation of the target population also plays a significant role in sample size determination. Populations with high variability require larger sample sizes to achieve the same level of confidence and statistical power. For instance, if a training program targets a diverse group of participants with varying educational backgrounds, skill levels, or professional experiences, there may be more variability in how participants respond to the training. This increased variability demands a larger sample size to ensure that the results are reliable and not skewed by outliers or extreme cases.

3. **Statistical Power ( $1-\beta$ ):** Statistical power is the probability of detecting a true effect if it exists. A power of 80% (or 0.80) is commonly used, meaning there is an 80% chance of correctly rejecting the null hypothesis (i.e., detecting a true effect). If the power is too low, the study runs the risk of Type II errors, where the training or policy's effect is not detected, even if it is significant. To achieve the desired power, researchers and policymakers must ensure that the sample size is large enough to detect the expected effect. However, increasing the sample size does not always proportionately increase power beyond a certain threshold, so determining the correct sample size balance is important.

4. **Significance Level ( $\alpha$ ):** The significance level, or alpha ( $\alpha$ ), is the probability of rejecting the null hypothesis when it is true, i.e., the likelihood of making a Type I error. A common threshold for  $\alpha$  is 0.05, indicating a 5% chance of concluding that the intervention has an effect when it actually does not. The lower the significance level, the larger the sample size required to achieve the same statistical power. By adjusting the significance level, policymakers can control for the risk of false positives in training program evaluations.

### **Practical Challenges in Sample Size Calculation**

While sample size calculation is fundamental to the design of training and certification programs, several practical challenges can complicate this process:

1. **Estimating the Effect Size:** One of the most challenging aspects of sample size calculation is estimating the effect size, especially when there is little prior research on the specific training program or targeted policy. Without historical data or pilot studies, it becomes difficult to predict the potential impact of a program. Policymakers often rely on expert judgment or data from similar interventions, but this can still lead to imprecision. If the effect size is misestimated (either too large or too small), it can result in an inefficient sample size that does not provide the necessary power for detecting real effects.

2. **Attrition and Non-Compliance:** Real-world training and certification programs often encounter issues such as participant drop-out or non-compliance. If individuals leave the program before completion or do not follow the prescribed interventions, this can lead to bias in the results. To account for this, researchers often increase the sample size to offset expected attrition. However, overestimating attrition rates can lead to unnecessarily large sample sizes, wasting resources.

3. **Budget and Resource Constraints:** While statistical tools can determine an optimal sample size, practical constraints like budget limitations, time, and logistical factors can affect the feasibility of achieving that sample size. Often, policymakers have to make compromises, balancing statistical requirements with the available resources. In some cases, a smaller sample size may be chosen, though this may affect the accuracy of the conclusions.

4. **Generalizability of Results:** A key goal of training and certification programs is to generalize results to a broader population. Ensuring that the sample is representative of the population of interest is essential for valid conclusions. If the sample is not representative, the findings may not be applicable to the broader group. For example, if a training program is intended to certify workers in a specific sector, but the sample only includes



individuals from one geographic region, the results may not apply to workers in other regions or sectors. Careful attention must be paid to sampling methods, ensuring that the sample size is not only sufficient but also appropriately chosen for generalization.

### **Impact on Policy Effectiveness**

The effectiveness of targeting policies depends on their ability to reach and positively impact the intended groups. A well-calculated sample size enhances the reliability of evaluations, allowing policymakers to make evidence-based decisions. For example, if a training program's evaluation relies on a well-calculated sample size, the conclusions about the program's effectiveness are more likely to be accurate, providing valuable insights for future policy adjustments.

Moreover, when sample size calculation is performed rigorously, the results of the policy evaluations can serve as a feedback mechanism for improving the policy. For instance, if a certain targeting policy is found to have a small effect size or negligible impact in one group, but a larger effect in another, this can lead to refinements in the policy, such as offering tailored interventions for different subgroups. By ensuring that sample size calculations are based on sound statistical principles, the entire process of policy evaluation, adjustment, and scaling becomes more effective.

In conclusion, the importance of sample size calculation in training and certifying targeting policies cannot be overstated. It is essential for ensuring that the results of training programs are valid, reliable, and reflective of real-world outcomes. By understanding the influence of factors like effect size, variability, power, and significance level, policymakers can design programs that are both efficient and effective. Despite practical challenges such as estimating effect size, managing attrition, and balancing resources, the careful calculation of sample size remains one of the most powerful tools available for optimizing training and certification outcomes. Ultimately, it allows policymakers to make informed, evidence-based decisions that maximize the success and impact of targeting policies.

### **CONCLUSION**

Calculating the appropriate sample size is a critical step in designing effective training and certification programs that support targeting policies. By ensuring that the sample size is sufficient to detect meaningful effects with adequate statistical power, policymakers can make more informed decisions about the effectiveness of their programs. This process not only improves the reliability of training outcomes but also enhances the overall success of policies that aim to improve the workforce, increase productivity, and ensure fair certification processes. As targeting policies become increasingly

prevalent in various sectors, the importance of rigorous sample size calculation will continue to grow, ensuring that the policies are based on sound, evidence-based foundations.

### **REFERENCES**

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). *Statistical Power Analyses Using G\*Power 3.1: Tests for Correlation and Regression Analyses*. *Behavior Research Methods*, 41(4), 1149-1160.
- Norman, G. R., & Streiner, D. L. (2008). *Biostatistics: The Bare Essentials*. PMPH-USA.
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. Psychology Press.
- Lachin, J. M. (2000). *Biostatistical Methods: The Assessment of Error Rates and Statistical Power in Clinical Trials*. Wiley-Interscience.
- Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press.
- Mendenhall, W., & Sincich, T. (2016). *A Second Course in Statistics: Regression Analysis*. Pearson Education.
- Armitage, P., Berry, G., & Matthews, J. N. (2002). *Statistical Methods in Medical Research*. Blackwell Science.
- Altman, D. G. (1991). *Practical Statistics for Medical Research*. CRC Press.
- Biau, D. J., & Kernéis, S. (2016). *Statistics in Medicine: From Basic Concepts to Advanced Methods*. Springer International Publishing.
- Cartwright, A., & Coulter, A. (2014). *Statistics in Healthcare and Medical Research*. Cambridge University Press.
- Donnelly, P., & Morton, A. J. (2010). *Applied Statistics for the Social Sciences*. Routledge.
- Stegenga, J. (2013). *Clinical Trials: A Methodological Introduction*. Springer.
- Harris, R. (2006). *An Introduction to Statistical Methods and Data Analysis*. Brooks/Cole.
- Lenth, R. V. (2001). Some Practical Guidelines for Effective Sample Size Determination. *American Statistician*, 55(3), 187-193.

Schoenfeld, D. (1983). Sample Size Formulae for the Comparison of Two Group Means in the Analysis of Variance. *Biometrika*, 70(2), 595-603.

Rucker, G., & Schwarzer, G. (2007). *Meta-Analysis with R*. Springer.

Bland, M. (2015). *An Introduction to Medical Statistics*. Oxford University Press.

Machin, D., Campbell, M. J., Fayers, P. M., & Pinol, A. (2007). *Sample Size Tables for Clinical Studies*. Wiley-Blackwell.

Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference. *Journal of the American Statistical Association*, 22(158), 209-212.

Edwards, W. (1972). Subjective Probability: A Judgment of Representativeness. *Psychological Bulletin*, 77(5), 193-204.

Williams, A. M., & Gillham, L. A. (2010). *Analyzing Experimental Data: A Practical Guide for Social Scientists*. Sage Publications.

Mayhew, L. E., & Warner, L. A. (2009). *Introduction to the Design and Analysis of Experiments*. Wiley.

Kraft, P., & McCullough, M. (2002). *Designing and Analyzing Experiments*. Cambridge University Press.

Rubin, D. B. (2006). *Matched Sampling for Causal Effects*. Cambridge University Press.