Choose the Right Sample Size For Your Survey Research: Six Common Principles

*by **Tran Thanh Vu**, Durham University, United Kingdom*

Email: vu.tran-thanh@durham.ac.uk

Abstract:

Determining the right sample size is crucial for valid survey research. This article explores six common principles for sizing your sample, including ratios with survey questions and variables, reference tables, online calculators, and formulae based on desired confidence levels and error margins. Suitable for both probability and non-probability sampling, these principles equip researchers to draw statistically sound conclusions from their surveys.

Determining sample size is always the most important factor in the process of conducting survey research because it not only affects the ability to generalize the results but also dominates the actual circumstances of collecting and analyzing data.

As an instructor of research methodology, I frequently encounter the question, "How big should the sample size be?" My response would always be, "The more, the better." It is, however, not always easy to obtain a good number of participants for a survey project.

I have attempted to find, in hopelessness, a definite answer to such a question. I am glad to admit that I failed, yet the knowledge I gained from that search has proved its usefulness to the many quantitative projects that I have supervised.

There are many different formulas and principles for determining the sample size for a survey research paper. I have outlined in this article some of the most common principles.

1. Sample-To-Item Ratio

This principle is often proposed for papers using exploratory factor analysis (EFA) with a sample-to-item ratio of at least 5:1 (i.e., multiplying the total number of questions in the survey by 5 to get the sample size) (Gorsuch, 1983; Hatcher, 1994; Suhr, 2006). A paper using a 30-question questionnaire would have a minimum sample size of 150. Some other researchers, such as Costello and Osborne (2005), also propose a ratio of 20:1.

2. Sample-To-Variable Ratio

Hair et al. (2018) propose a ratio of 5:1, 15:1, or 20:1 in the correlation of sample and independent variable. That is, at least 5 participants must be surveyed for a questionnaire with one independent variable. However, this principle, especially the ratio of 5:1, is very difficult to follow in practice, as a small sample size will make it difficult for inferential analyses.

3. Krejcie and Morgan's Table of Reference

This table of reference was published in 1970 and is very popular in the fields of behavioral science and social science. Krejcie and Morgan propose a number of 384 samples for populations of 1,000,000 individuals or more - and the number 384 is also considered a "magical number" in scientific research with thousands of papers using it.

However, this table of reference is actually only suitable for papers using probability sampling techniques. Papers using non-probability sampling techniques such as purposive, snowball, or quota should consider other methods.

4. Online Sample Calculators

This is one of the easiest methods for social science researchers, which I have personally employed for many different projects. The most popular tool is the Raosoft sample size calculator, which requires parameters such as: confidence level, margin of error, population proportion, and population size.

5. Roscoe's Principle

Roscoe (1975) based on the central limit theorem (CLT) proposes a number from 30 to less than 500 for behavioral science research papers - noting that a sample size greater than 500 can lead to Type II error. For comparative analysis papers (male/female; urban/rural; with/without degree...) each group should have at least 30 samples.

6. Green's Principle

Green (1991) proposed the formula $N \geq 50+8m$ (where m is the number of independent variables) to determine the sample size for the coefficient of determination. For example, a paper with 7 independent variables would need a sample size of at least $50 + 8 \times 7 = 116$ for regression analysis.

In addition, the formula $N \geq 104+m$ is also proposed for Beta independent variables. That is, a minimum of 105 samples are required for linear regression analysis and more for multiple linear regression analysis.

References

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10(7)*, 1-9.

Gorsuch, R. L. (1983). *Factor Analysis* (2nd ed.). Lawrence Erlbaum.

Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research, 26*(3), 449-510.

Hatcher, L. (1994). *A step-by-step approach to using the SAS® system for factor analysis and structural equation modeling*. SAS Institute.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate Data Analysis* (8th ed.). Cengage Learning.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*, 607-610.

Roscoe, J. T. (1975). *Fundamental research statistics for the behavioral sciences* (2nd ed.). Holt Rinehart and Winston.

Suhr, D. D. (2006). *Exploratory or Confirmatory Factor Analysis*. SAS Institute Inc.

Tran Thanh Vu is an ESRC doctoral researcher at the School of Education, Durham University, UK. Prior to his doctoral study, he worked for nine years as an EFL teacher and teacher educator in Ho Chi Minh City, Vietnam. Vu founded and administers a professional learning community called TESOL Research Collaboration Network. He also serves on the Steering Committee of the Teacher Educator Interest Section of TESOL Int'l Association.