ISSN: 2645-8489
Original Article

Statistical Approach: Science and Application for Determining Optimal Sample Size in empirical study

Nabaraj Poudel[1], Murari Karki[2], Kapil Shah[3]

[1] National Animal Science Research Institute
Email: npoudel43@yahoo.com
[2] Saraswati Multiple Campus
Tribhuvan University
Email: murari.karki@smc.tu.edu.np
https://orcid.org/0009-0007-0451-0808
[3] Nepal Commerce Campus
Tribhuvan University
Email: kapil.shah@ncc.tu.edu.np
https://orcid.org/0009-0003-8320-3312

Corresponding Author
Kapil Shah
Email: kapil.shah@ncc.tu.edu.np

Funding: This research received no specific grant from any funding agency in the Public, commercial, or not-for-profit sectors.

Abstract

To ensure optimal sample size selection, researchers can use a variety of formulae and procedures created by statisticians to find a suitable sample size. The objects selected have, on average, the same characteristics as the population units if a sufficiently large sample of the population is chosen at random. This is referred to as the statistical regularity principle. According to the statistical regularity principle, larger sample sizes generally produce more accurate and dependable results, assuming that all other variables remain constant.

An ideal sample minimizes the cost for a predefined mean squared error value or lowers the cost for a constant mean squared error. Descriptive cross-sectional secondary sources using readily available statistical tools such as Stata, Survey Monkey, Rao Soft, and G Power provided estimates of the sample size.

Keywords: *Mean Square Error, Sample Size, Statistical Regularity, User-friendly Statistical software*

JEL Classification: *C12, C13, C18, C51, C90, I21*

Introduction

The selection of a representative subset of the population is influenced by sample size. The fact that, on average, a sizable number of samples drawn from the population have the characteristics of the population is also emphasized by the laws of statistical regularity and big number inertia. A number of formulas and methods developed by statisticians can be used by researchers to determine an appropriate sample size, guaranteeing that they obtain the optimal sample size. By satisfying the needs of the study's goals, the optimal sample size gathers more pertinent data. The statistical regularity principle guarantees that the items chosen have, on average, the characteristics of the population units, and a sizable sample of the population is randomly selected.
According to the statistical regularity principle, if all other variables stay the same, a bigger sample size typically yields more reliable and accurate results.

How to cite this article (APA): Poudel, N., Karki, M., & Shah, K. (2024). Statistical Approach: Science and Application for Determining Optimal Sample Size in empirical study. *DEPAN, 6*(1), 108-117.

The sample size for a given purpose is determined by a variety of factors. The following are the main factors that determine the size of the sample for a particular study: the goal and scope of the study; the population and sampling unit characteristics; the sampling technique and estimation procedure to be used; the variability structure of the population; the time and cost component structure; and the population size, among other things. Either the cost for a fixed mean squared error or the cost for a fixed mean squared error value is minimized by an efficient and optimal sample.

Thus, sample size determination is the process of choosing how many observations or duplicates to include in a statistical sample. The sample size is an important consideration in any empirical study that uses a sample to draw conclusions about the population. Descriptive cross-sectional secondary sources, including easily navigable statistical tools such as Rao soft, G power, Epi info, Stata, and Survey Monkey, were used to estimate the sample size. Some of them were free, while others were trial versions. The secondary sources used for the literature citation for sample size computation included papers by various authors and formulas developed by various statisticians.

The size of the field experiment plots is used in crop science research to calculate the sample size. The sample size can be the number of plants per plot used to measure plant height, the number of plots per plant used to measure plant height, the number of leaves per plot used to measure leaf area, or the number of hills per plot used to count tillers. In a field trial experiment, each plot acts as a population. (Gomez and Gomez, 1984). Sample sizes can now be determined using a wide range of computer-based statistical programming and software. The statistical software is the true ally of any researcher from the beginning to the end.

Cochran's formula is still the most basic method for determining sample sizes, even if there are several methods available. The first step in every quantitative investigation is to determine the sample size. The sample size calculation is essential for establishing the study design in a quantitative investigation. In terms of the false positive rate, there is a trade-off between type - I mistake or false positive and type - II error or false negative. Data science theory places a high priority on reducing type - I error (Smith, 2020).

Objective

To investigate the various statistical techniques used in determining sample size in the fields of animal health and agricultural research. The objective is to investigate different statistical techniques and technologies for optimal sample size determination.

Methods

The current work was essentially prepared using the fundamental formula created by Cochran's formula (Cochran, 1977). The secondary sources of literature used in a variety of research studies were from reputable, published sources. In addition to statistical software such as Raosoft, G Power, Stata, Epi info, and R Studio, there are more than six different statistical formulas. To guarantee the validity and consistency of an optimal sample size in respect to statistical formulas, sample size calculation software was employed. G power examined the characteristics of sampling and non - sampling error in relation to test power, effect size, and sample size. Descriptive cross - sectional secondary sources were used to get the estimated sample size. Numerous sample size formulas and easily navigable statistical tools, such as Rao soft, G power, Epi info, Stata, Survey Monkey, and others, were included in these sources. A combination of trial and free versions was available. The literature's references for calculating sample size were secondary sources, such as publications written by different authors and formulas created by different statisticians.

Results

Calculating the sample size is an essential step in both quantitative and qualitative research. In essence, sample size was calculated using both a manual method and statistical software. Comprehending statistical formulas, tables, and calculators is necessary for calculating sample size manually. Correct use of formulas and table interpretation, given the fixed amount of type I error, statistical test power, and effect size are necessary for correctness.

Sample size with varying values of margin of error, population size and confidence interval

Table 1 provides an illustration of sample size with a variety of error values, population size, and confidence level. Three key factors to determine the sample size are population size, confidence interval, and the margin of error. Each has a specific purpose in figuring out how big of a sample is appropriate for a statistical analysis.

Table 1

Sample size with varying values of margin of error, population size and confidence interval

Population Size (N)	Margin of Error %	Confidence level	Response Rate	Sample Size(n)
20000	5	99	50	643
20000	5	95	50	373
20000	1	99	50	9068

Table 2 reveals the calculations are performed for various levels of Confidence (80%, 90%, 95%, 97%, 99%, 99.9%, and 99.99%). The results show how the required Sample Size and Total Sample Size increase with higher confidence levels.

Table 2

Distribution of sample size by using stat calculator in EPI Info

Population size	Expected Frequency	Margin of Error	Confidence Level	Cluster Size	Sample size	Total sample size
2000	50%	5%	80%	2	76	152
2000	50%	5%	90%	2	119	238
2000	50%	5%	95%	2	161	322
2000	50%	5%	99%	2	249	498
2000	50%	5%	99.90%	2	351	702
2000	50%	5%	97%	2	191	382
2000	50%	5%	99.99%	2	431	862

Cochran (1977), stated preliminary formula to estimate sample size for study about population proportion when population size is finite and known as

$$n = \frac{\frac{Z^2 P(1-P)}{e^2}}{1+\frac{Z^2 P(1-P)}{Ne^2}} \quad (1)$$

Where z is tabulated value of standard normal variate at α level of significance, e is margin of error which commonly varies 1% to 5%, P being the prevalence rate.

The equation (1) is used to calculate the sample size when prevalence rate is known and population size being finite. Krejcie & Morgan (1970) reduces equation (1) in the form as equation (2) for estimating sample size when population is infinite or unknown.

$$n = \frac{Z^2 P(1-P)}{e^2} \quad (2)$$

In order to deal with qualitative data such as possessing attributes of dichotomous variables these formulas are used in research.

When estimating a population mean with a given margin of error, this formula (equation 3) shows how to calculate the sample size for a finite population. When the population is not infinitely large, it takes into consideration of population size N, which lowers the sample size. The formula is based on the general principles of sample size estimation for finite populations (Cochran, 1977). This formula is helpful for finite population.

$$n = \frac{\frac{Z^2\sigma^2}{e^2}}{1+\frac{Z^2\sigma^2}{Ne^2}} \quad (3)$$

There are some circumstances in which population size is unknown, in such situation, we should assume $N \to \infty$. So, equation (3) can be reduced as

$$n = \frac{Z^2\sigma^2}{e^2} \quad (4)$$

Gupta and Kapoor (2020) discussed the relationship between range and standard deviation in their book Fundamentals of Mathematical Statistics. Sometimes standard deviation can be approximated through thumb rule, i.e. Relation between standard deviation and range is given by

$$\sigma = \frac{X_{max}-X_{min}}{4} \quad (5)$$

Sahu (2016), stated another formula for calculating sample size with reference to chi square value at 5% level of significance is given by

$$n = \frac{\chi^2 Pq/e^2}{1+\chi^2 Pq/Ne^2} \quad (6)$$

Modified version of equation (6) is stated later as

$$n = \frac{\chi^2\ Npq}{d^2(N-1)+\chi^2\ pq} \quad (7)$$

Yanni and solvi (1968), stated another useful formula to calculate sample size based on standard deviation and two probabilities of committing error, which is given as

$$n = \frac{\left(Z_{1-\frac{\alpha}{2}}+Z_{1-\beta}\right)^2 ({\sigma_1}^2+{\sigma_2}^2)}{e^2} \quad (8)$$

Cohen (1988), elaborated sample size computing methods for one group sample and two group samples with a dichotomous outcome variable as

The sample size for one group samples

$$n = \frac{(Z_\alpha+Z_{1-\beta})^2}{(ES)^2} \quad (9)$$

The sample size for two group samples,

$$n = \frac{2(Z_\alpha+Z_{1-\beta})^2}{(ES)^2} \quad (10)$$

where Z_α is tabulated value of standard normal Variate at α level of significance and $Z_{(1-\beta)}$ is the tabulated value of standard normal Variate at power level maintaining at least 80% power, ES is effect size which is about 0.2 for small size effect, 0.5 as medium effect size and 0.8 as large size effect.

To study about the correlation between two variables, the sample size calculated by using following equation (G power, 2023).

$$n = \frac{(Z_{\frac{\alpha}{2}}+Z_\beta)^2}{\frac{1}{4}[log_e\left(\frac{1+r}{1-r}\right)]} + 3 \quad (11)$$

Where r is correlation coefficient $Z_{(\alpha/2)}$ and Z_β are the tabulated value of Z at α/2 and β level of significance.
We can compute sample size based on success rate in Hypothesis as in following equation (Machin et al., 2018)

$$n = \frac{Z_\alpha\sqrt{p_0(1-p_0)}+Z_\beta\sqrt{p_1(1-p_1)}}{(p_0-p_1)^2} \quad (12)$$

where p_0 is success rate under null hypothesis and p_1 is success rate under alternative hypothesis.
Z_α and Z_β are tabulated value of standard normal Variate at α and β level of significance.

We use specific formula to determine sample size in crop science research which is based upon the plot size of the field experiments. In a field trial experiments, each plot is a population, sample size could be the number of plants per plot used for measuring the plant height , the number of plots per plant used for measuring the plant height , the number of leaves per plot used for measuring the leaf area, or the number of hills per plot used for counting tillers (Gomez and Gomez,1984).

$$n = \frac{Z_\alpha{}^2 v_s}{d^2\bar{X}^2} \quad (13a)$$

$$n = \frac{Z_\alpha{}^2 v_s}{r(d^2)\bar{X}^2 - Z_\alpha{}^2 v_p} \quad (13b)$$

Where n is required sample size, Z_α is the standardized normal variate corresponding to the level of significance, v_s is sampling variance, v_p is the variance between the plots of the same treatment, $\bar{X}$ is mean value and d is margin of error.

Illustration 1:
A researcher may wish to measure the number of panicles per hill in the transplanted rice plots with the single hill as the sampling unit from a previous experiment. He estimates the variance in the panicle number between individual hills within the same plots to be 5.0429 or a covariance of 28.4% based on the average number of panicles per hill of 17.8.He prescribes that the estimate of the plot mean should be within 8% of the true value(Gomez & Wiley, 1984). The sample size at 5% level of significance is

$$n = \frac{(1.96)^2(5.0429)}{(0.08)^2(17.8)^2}$$

$$= \frac{19.3647}{2.0277}$$

$$= 9.55$$

$$\approx 10 \text{ hills/plot}$$

Now a days, there are so many computers based statistical programming and software designed and created for the sample size estimations. The statistical software are true friends are every researcher from initial point to terminating point of investigations. In order to carry out sample size calculation, following software and APPs are useful to researchers and academicians.

Raosoft Sample size calculator
G Power
Medical Calculator
STATA
Epinfo

Table 3

Sample size with varying values of margin of error, population size and confidence interval

Population Size (N)	Margin of Error %	Confidence level	Response Rate	Sample Size(n)
20000	5	99	50	643
20000	5	95	50	373
20000	1	99	50	9068

Table 3 reveals the calculations are performed for various levels of Confidence (80%, 90%, 95%, 97%, 99%, 99.9%, and 99.99%). The results show how the required Sample Size and Total Sample Size increase with higher confidence levels.

Table 4

Distribution of sample size by using stat calculator in EPI Info

Population size	Expected Frequency	Acceptable Margin of Error	Confidence Level	Cluster Size	Sample size	Total sample size
2000	50%	5%	80%	2	76	152
2000	50%	5%	90%	2	119	238
2000	50%	5%	95%	2	161	322
2000	50%	5%	97%	2	191	382
2000	50%	5%	99%	2	249	498
2000	50%	5%	99.90%	2	351	702
2000	50%	5%	99.99%	2	431	862

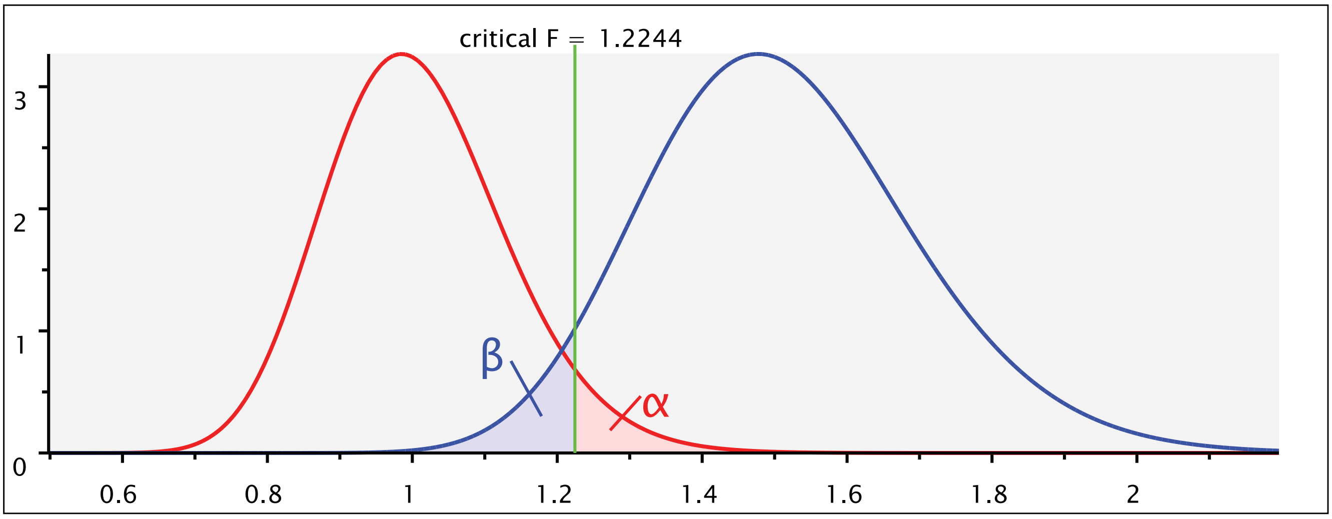

Figure 1: Illustration of type I Error and Type II Error in F - Distribution

The graph illustrates the trade - off between Type I and Type II mistakes in hypothesis testing. While testing for hypotheses often aims to minimize both α and β, lowering one often raises the other. Statistical power (1 − β) can also be inferred from the portion of the alternative hypothesis curve that continues past the critical F value (G * Power 3.1, 2023).

The total sample size goes up as the amount of confidence goes up. Because more data is needed to lower uncertainty at a higher level of trust. There are between 152 and 862 samples, with 152 giving 80% confidence and 862 giving 99.99% confidence.

Table 3

Distribution of sample size with varying confidence level for finite known population

Population Size	Confidence level	Expected frequency	Cluster size	Total sample size
2000	80	50%	152	152
2000	90%	50%	238	238
2000	95%	50%	322	322
2000	97%	50%	381	381
2000	99%	50%	498	498
2000	99.90%	50%	702	702
2000	99.99%	50%	862	862

Mead's Resource Equation

Mead's resource equation is a useful tool for estimating sample size, especially when precise parameters like expected standard deviations or predicted differences between groups are unavailable (Mead, 1988). It is expressed as:

$$E = N - B - T + 1 \quad (14)$$

Where:

E: Degrees of freedom for error (desired to be between 10 and 20 for reliable results).
N: Total number of experimental units (animals, samples, etc.).
B: Number of blocks or groups.
T: Number of treatments being compared.

Key Features:

Flexibility:
Works when precise statistical data (e.g., effect size) is unavailable.
Adaptable to various experimental designs.

Applications:
Biology & Agriculture: For experiments involving plants, animals, or microbial studies.
Clinical Research: Estimating initial group sizes in pilot studies.
Laboratory Investigations: Used when standard sample size calculations are challenging.

Practical Use:

It ensures that experiments have sufficient replication without wasting resources or using unnecessarily large sample sizes, aligning with ethical considerations (e.g., reducing animal use in research).

Mead's equation provides a simple and efficient starting point for researchers across diverse disciplines while promoting resource-efficient and ethical research practices.

For instance, if a study involving laboratory animals is designed with four treatment groups (T = 3), comprising eight animals per group, resulting in a total of 32 animals (N = 31), without additional stratification (B = 0), then E would equal 28, exceeding the threshold of 20, suggesting that the sample size may be excessive, and six animals per group could be more suitable.

Discussion

Statistical formulas and statistical software play a crucial role in determining the sample size. There is a close connection between sample size, sampling error, and non-sampling error. When the sample size increases, sampling error decreases and vice versa. On the other hand, non-sampling error increases with increasing sample size.

Statistical software is very useful to calculate sample size without handy and tedious calculation. Statistical software is open source or proprietary. The researcher's experience, the target variance, the target for the statistical test's power, and the confidence level all influence the sample size choice. As the sample size increased, the precision of the unknown population parameter increased. A power analysis calculates. It gives the chance of finding a result that is statistically significant at a certain level of significance (±) for a population effect size (Cohen, 1988). A researcher will often use a scientific sample design to take into account how the characteristics of the population change over time and to pick a sample that will make it less likely that the population as a whole will have a biased view. The researcher can get rid of a sample using a simple random sampling method or a more complex multistage sampling method that includes stratification, clustering, and an uneven probability of selection. The number required for the data analysis is another factor to take into account when determining sample size. Descriptive statistics (such as mean, SD, frequencies, etc.) will include the estimated parameter. Social sciences frequently use percentages like 90%, 95%, or 99% to convey this level of precision and certainty. If we choose a 95% confidence level, the real population parameter will be within that range in 95 out of 100 samples.

Effect size is the variance between two or more distinct groups that may produce various outcomes. Therefore, it's crucial to choose a sample size that results in the least amount of variation in results and yields outcomes that are more closely related to the larger population size. The standard deviation aids in highlighting how far out from the mean the data are. This happens when we analyze continuous data and multiple datasets to determine the statistical dispersion of the mean. The postulate effect concerns the cause-and-effect connection. As a result, the selection of the sample size must satisfy the necessity to demonstrate the rationale for certain outcomes. The significance level is represented in data analysis by the letter alpha (α). It describes the likelihood of risk involved in drawing the incorrect conclusion. To put it another way, it establishes the proportion of a sample size that is under the null hypothesis, making it simpler to select one with the lowest likelihood of drawing the incorrect conclusion. Any significant errors in the sample size estimate will impact the study's power and value. Common sample size errors include failing to complete any computations, making irrational assumptions, failing to take into account probable losses throughout the investigation, and failing to evaluate sample size over a variety of assumptions. Failure to complete sample size calculations, choosing a sample size based on convenience, insufficient financing for the research, and inefficient use of the funding that is available are all causes of studies with insufficient sample sizes that do not reach statistical significance. Because a large number of quantitative data in livestock, fisheries, and plant science can be generated in agricultural research, sample size determination is a primary concern. Agricultural research widely uses the experimental design and sample survey. However, agricultural research faces some challenges in terms of obtaining an adequate number of sampling units within the time, money, and energy constraints. In general, a sufficient amount of sample size is lacking during an experimental design in livestock and fisheries research. Sample size influences the efficiency of both the sample design and the experimental design. In agricultural research, the allocation of treatment, replication, plot size, and so on is also concerned with sample size in experiment design. Plot size and shape play an important role in coping with soil heterogeneity. The contribution of soil heterogeneity to experimental error stems from the differences within the block. The smaller is the experimental error. The choice of the suitable plot size and shape, therefore, should reduce the differences in soil productivity from plot to plot within a block and consequently reduce the experimental error. Cochran's formula is a renowned equation for calculating sample size, from which other modified formulas have been derived. Most statistical software designed for sample size calculation is based on known or unknown prevalence rates and population sizes. It calculates a sample size for extensive populations, considering simple random sampling. It provides an acceptable sample size for surveys regardless of whether the population size is known or unknown. This formula is used in case of estimating proportion of characteristics in the population. Yamni and Solvi formula was introduced to estimate a sample size when population size is known, however margin of error is preassigned by the researcher 1 to 5 % but this formula is less precise to maintain the power of the statistical test. The Power of the statistical test requires minimum 80% during sample size calculation (Cohen,1988). The power analysis formula calculates sample size according to the desired power and effect size. It ensures sufficient power to identify a significant difference or impact, used frequently in clinical trials of animal health research and hypothesis testing. Finite population adjust the sample size for small population. Now a days, there are so many modified formulae developed and utilized for sample size calculation whose basis is Cochran's formula. Many statistical software offers a sample size calculation menu to provide rapidity and easiness in sample allocation. It is necessary to apply manual approach as well as software-based approach to calculate sample size to study to accuracy of sample size.

Conclusion

Sample size calculation is a critical aspect of research design, ensuring the reliability, validity, and generalizability of study results. Selecting an appropriate sample size strikes a balance between statistical power and resource constraints, reducing the likelihood of Type I and Type II errors. Various methods and formulas, such as Cochran's formula, Slovin's formula, and power analysis, cater to different research needs, whether estimating proportions, means, or group differences. The choice of a suitable method depends on the study's objectives, demographic characteristics, and data type. Proper sample size planning enhances the precision of findings and upholds research ethics by optimizing resources and minimizing the burden on participants.

References

Altman, D. G. (1991). *Practical statistics for medical research*. Chapman and Hall.

Agresti, A. (1990). *Categorical data analysis*. John Wiley & Sons.

Campbell, M. J., Julious, S. A., & Altman, D. G. (1995). Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two-group comparisons. *BMJ, 311*(7013), 1143–1148. https://doi.org/10.1136/bmj.311.7013.1143

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). Wiley.

Desu, M., & Raghavarao, D. (1990). *Sample size methodology*. Academic Press.

Donner, A. (1984). Approaches to sample size estimation in the design of clinical trials: A review. *Statistics in Medicine, 3*(3), 199–214. https://doi.org/10.1002/sim.4780030304

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). Wiley.

Gomez, K. A., & Gomez, A. A. (1984). *Statistical procedures for agricultural research* (2nd ed.). Wiley.

Hintze, J. L. (2008). *Power analysis and sample size system (PASS) for Windows: User's Guide I.* NCSS.

James, B. E., Joe, W. K., & Chadwick, C. H. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal, 19*(1), 43–50.

Johnson, P. O. (1959). Development of the sample survey as a scientific methodology. *Journal of Experimental Education, 27,* 167–176.

Jones, E. (2018). *Introduction to statistical theory and applications*. Springer.

Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials, 2*(2), 93–113. https://doi.org/10.1016/0197-2456(81)90001-5

Lwanga, S. K., & Lemeshow, S. (1991). *Sample size determination in health studies: A practical manual* (pp. 1–3). World Health Organization.

Mohar, D., & Dulbarg, C. S. (1994). Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA, 272*(2), 122–124. https://doi.org/10.1001/jama.1994.03520020074014

Shuster, J. J. (1990). *Handbook of sample size guidelines for clinical trials*. CRC Press.

Smith, J. (2020). The balance between Type I and Type II errors in statistical analysis. *Journal of Data Science and Statistics, 12*(3), 45–60.

Thomas, L. (1997). Retrospective power analysis. *Conservation Biology, 11*(2), 276–280.

Thomas, L., & Juanes, F. (1996). The importance of statistical power analysis: An example from animal behavior. *Animal Behaviour, 52*(4), 856–859. https://doi.org/10.1006/anbe.1996.0236

Whitley, E., & Ball, J. (2002). Statistics review 4: Sample size calculations. *Critical Care, 6*(4), 335–341. https://doi.org/10.1186/cc1521

Wittes, J. (2002). Sample size calculations for randomized controlled trials. *Epidemiologic Reviews, 24*(1), 39–53.

https://doi.org/10.1093/epirev/mxf002
Wunsch, D. (1986). Survey research: Determining sample size and representative response. *Business Education Forum, 40,* 31–34.
Cohen, J. (1988). *Statistical power analyses for the behavioral sciences* (2nd ed.). Erlbaum.
Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155
Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. Wiley.
Freedman, L. S. (1982). Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine, 1*(2), 121–129.
Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1998). *Fundamentals of clinical trials* (3rd ed.). Springer-Verlag.
Greenland, S. (1985). Power, sample size, and smallest detectable effect determination for multivariate studies. *Statistics in Medicine, 4*(1), 117–127.
Ludbrook, J. (2003). Interim analyses of data as they accumulate in laboratory experimentation. *BMC Medical Research Methodology, 3,* 15. https://doi.org/10.1186/1471-2288-3-15
Pocock, S. J. (1983). Clinical trials: A practical approach. Wiley.
Russell, W. L. S., & Burch, R. L. (1992). *The principles of humane experimental technique* (Special ed.). Methuen & Co Ltd.
Schipper, M., & Van der Tweel, I. (2002). Efficiënte analyse van accumulerende biologische en medische data. *Intern Rapport, 2.* Centrum voor Biostatistiek, Universiteit Utrecht.
Todd, S., et al. (2001). Interim analyses and sequential designs in phase III studies. *British Journal of Clinical Pharmacology, 51*(5), 394–399.
Van der Tweel, I., & Schipper, M. (2002). Sequentiële analyses in klinisch en epidemiologisch onderzoek. *Nederlands Tijdschrift voor Geneeskunde, 146*(52), 2348.
Efron, S., & Ravid, R. (2019). *Writing a literature review: A practical guide* (1st ed.). Guilford Press.
Fox, N., Hunn, A., & Mathers, N. (2009). *Sampling and sample size calculation*. University of Sheffield.
Sekaran, U., & Bougie, R. (2014). *Research methods for business: A skill-building approach*. Wiley.