INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH & REVIEWS

journal homepage: www.ijmrr.online/index.php/home

NAVIGATING THE COMPLEXITY OF SAMPLE SIZE DETERMINATION FOR ROBUST AND RELIABLE RESULTS

Sohil Altaf Pirani

Senior Research Fellow (SRF), Dr. Babasaheb Ambedkar Marathwada University, India.

Postgraduate Diploma, Applied Statistics, University of Mumbai, India

How to Cite the Article: ***Pirani, S. (2024).*** *Navigating the Complexity of Sample Size Determination for Robust and Reliable Results. International Journal of Multidisciplinary Research & Reviews, Vol 03, No. 02, pp. 73-86.*

Keywords

Sample Size, Determination of Sample Size, Optimum Sample Size, Sample Size Justification

Abstract

The most frequent issue that the researcher faces while conducting any research is to decide the sample size. How much sample size to take? How many respondents are enough to understand the problem being addressed in the research? It is a dilemma that everyone goes through. This research was conducted with the sole purpose of helping future researchers navigate the complexities of sample size determination for robust and reliable results. Although there are several research articles available on understanding the sample size. This paper adds to the literature by providing the principles and laws, and online calculators for sample size justification.

1.INTRODUCTION

In survey research, figuring out a sufficiently large sample size that is representative of the population is essential to guaranteeing that results fall within the range of random error (Bartlett et al., 2001). However, mishandling sample size determination tables and algorithms can result in errors. Typical missteps include failing to account for sample variance, sampling error, and nonresponse bias (Tejada & Punzalan, 2021; Wunsch, 1986).

Determining an adequate sample size that is representative of the population under study is one of the major issues faced by social science researchers doing survey research. This is done to guarantee that conclusions extrapolated from the sample and applied to the population are within the bounds of random error (Bartlett, et al. 2001).

2. Definition of Sampling

The practice of choosing a few individual participants or a subset of the population to draw statistical conclusions and estimate the characteristics of the entire population is known as sampling. It is essentially the process of gaining knowledge about the population from a sample that has been taken from it.

3. Population

The population needs to be precisely characterised for the survey to achieve its goals. In other words, the population from which the sample units are to be drawn must be specified precisely and unambiguously. Defining the population to whom study results are to be generalised is the first step in sample selection. the intended audience, etc.

4. Sample

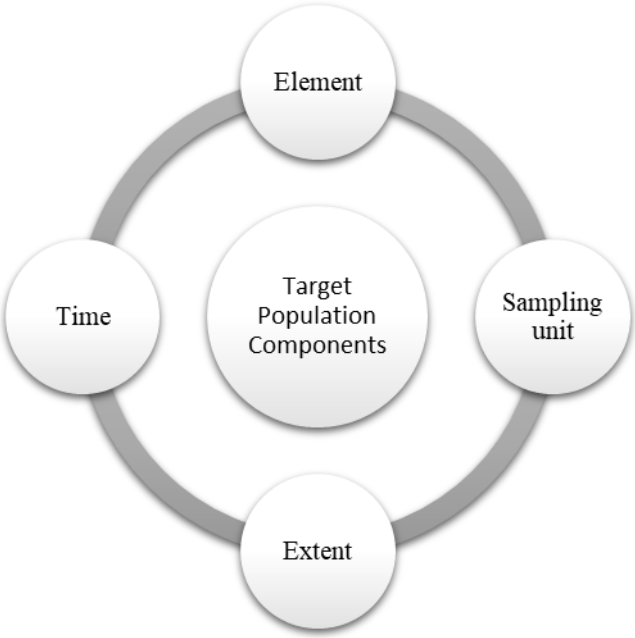
A sample is a portion of the target population that reflects the people or things being studied and is selected from a sampling frame.

5. Universal Population vs Target Population

All individuals or entities that have a common property or quality are referred to as the universal population. It stands for the larger group that an investigation or analysis is cantered on. All people or things from which a sample is taken to draw inferences or generalizations about a specific phenomenon are included in the universal population in research.

The target population definition needs to have four essential components. Primarily, the element denotes the thing or entity (respondents, products, companies) from which information is sought. Second, the smallest observation unit with the attributes under study is the sampling unit, which is the fundamental unit made up of these components. (An

individual, A firm, a livestock, etc). Third, the population's geographical bounds and the Fourth Important component is the Time component which denotes the time frame taken into consideration.

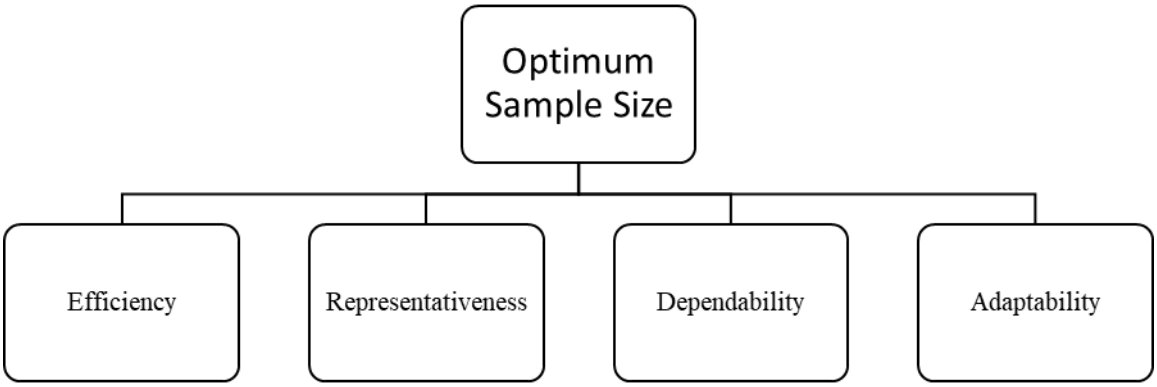


6. Meaning of sample size

One of the most important parts of research using sampling techniques is figuring out the sample size. It describes the quantity of observations used to estimate a population's characteristics. To conclude the complete reference population, a subset of the population is investigated, and this subset is represented by the sample size that is taken from the population. A suitable sample size is necessary to correctly represent the population. Furthermore, to guarantee validity and reliability in study findings, the size that is chosen must be ideal and acquired using scientific techniques.

7. Optimum sample size

A sample that satisfies the needs of efficiency, representativeness, dependability, and adaptability is considered ideal for a study.



8. Large sample theory vs small sample theory

A large sample is defined as one with a sample size of at least thirty (n>30). The sampling distribution of statistics for a large sample is normal (distribution). Large sample theory is the study of the sampling distribution of statistics for a large sample. It is possible to use the value of statistics from large samples to estimate population parameters.

A sample is considered tiny if its size is fewer than thirty (n<30). The sampling distribution for a small sample is t, F, and χ^2.

9. Factors affecting sample size

(a) **Character of the Population:** Fewer cases will be sufficient if the population is homogeneous. It can also be necessary to use a large number of examples if the population is diverse.

(b) **Resource availability:** Before choosing the sample size, one should take into account the time and financial resources that are accessible. If there is ample time and funding, a large sample can be taken.

(c) Type of Sampling Method: A small sample will suffice if restricted random sampling is used. To guarantee that every number is represented, additional numbers may be chosen in a basic random sampling. In comparison to stratified sampling, the cluster sampling approach requires a larger sample size.

(d) **Required degree of precision:** A large sample size is necessary if the maximum degree of accuracy is needed.

(e) **Nature of Analysis:** The statistical instruments and tests a researcher intended to employ for analysis may have an impact on sample size. Larger samples are necessary for complex multivariate statistics.

(f) **Population Size:** For finite, infinite, known, and unknown populations, different sample sizes are computed.

(g) Sampling Framework and Study Objective

Three important criteria to be specified to determine the appropriate sample size

(a) Margin of Error (e)

Difference between sampling estimates and population parameter is called sampling error.

Statistical expression of sampling error is called margin of error. It is expressed in terms of percentage points.

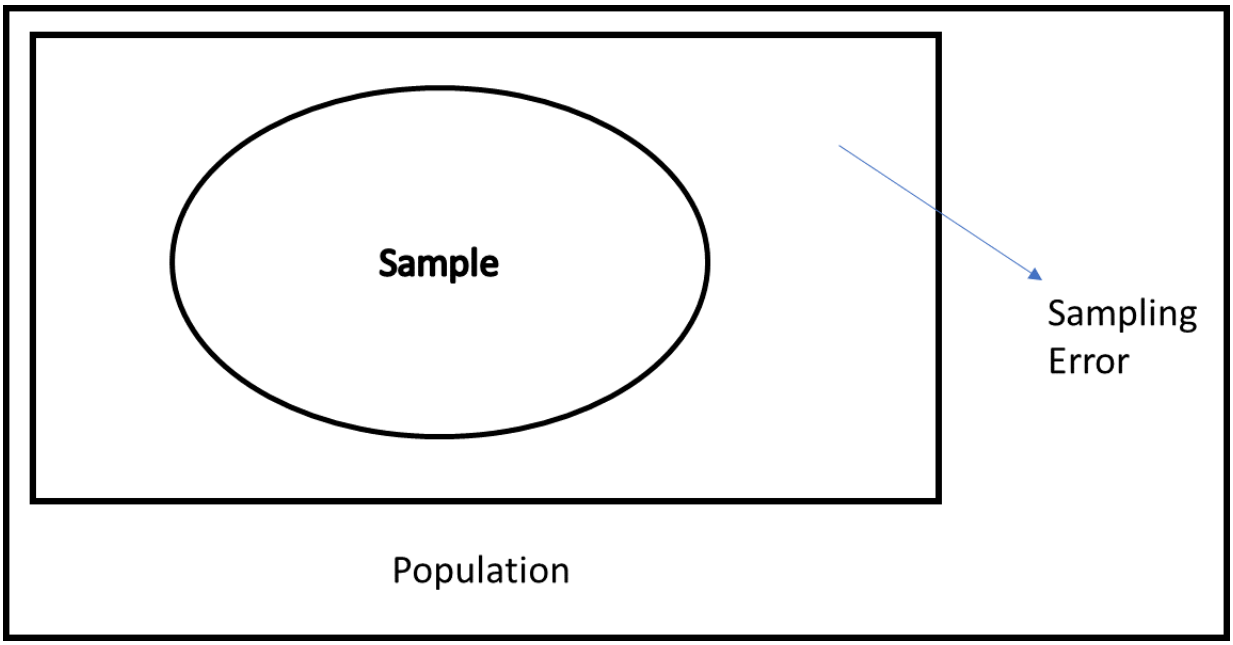


Figure 1: Sampling Error

It gives us a range in which true population values lie. It is also known as precision rate.

e = confidence interval * ½. There is inverse relation between e and sample size.

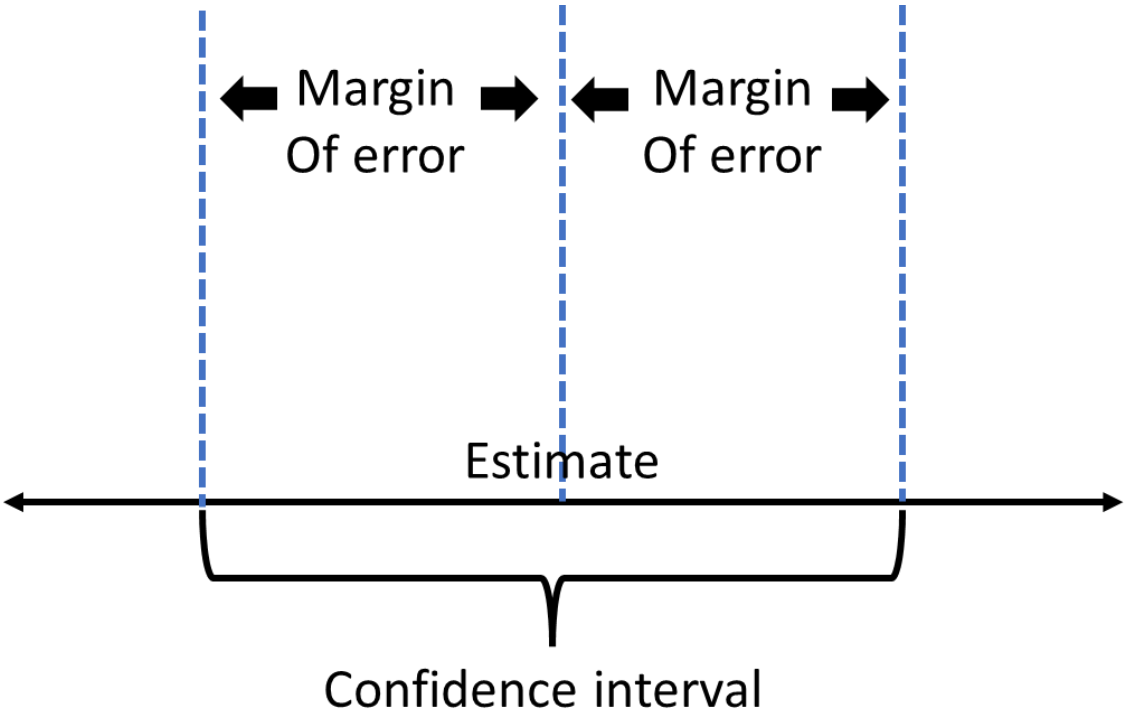


Figure 2: Margin of error

(b) ***Level of Confidence (Level of Certainty)***

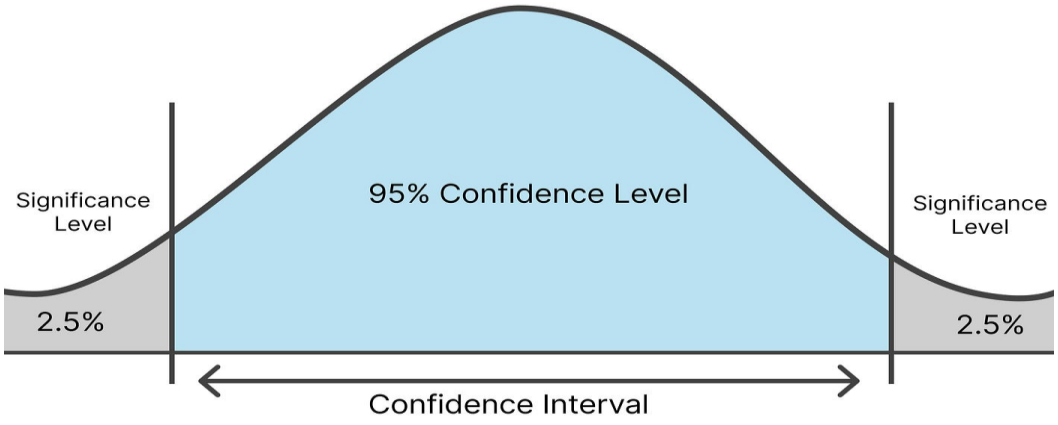


Figure 3: Level of Confidence

Probability that population estimates will lie within the given confidence interval. If 95% confidence level is selected ,95 out of 100 samples will have true population values within the range of precision and 5 % is level of significance represents that the obtained sample does not represent the true population value. There is direct relationship between level of confidence and sample size.

Confidence level	α (level of significance)	$Z_{\frac{\alpha}{2}}$
99%	1%	2.575
95%	5%	1.96
90%	10%	1.645

(c) Degree of Variability (Sampling Variability)

When we extract random sample from the given population again and again the sample statistics will vary from sample to sample. The more heterogeneous population larger sample is required and less variable (More homogeneous) population smaller sample is required.

I. Formulae method for sample size Justification

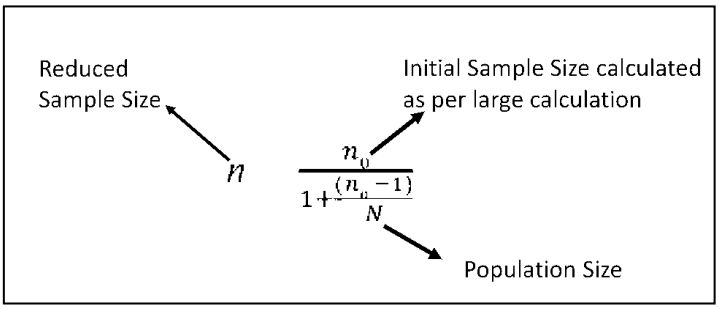
1. Cochran's infinite population (population proportion is known)

$$n = \frac{z^2 (pq)}{e^2}$$

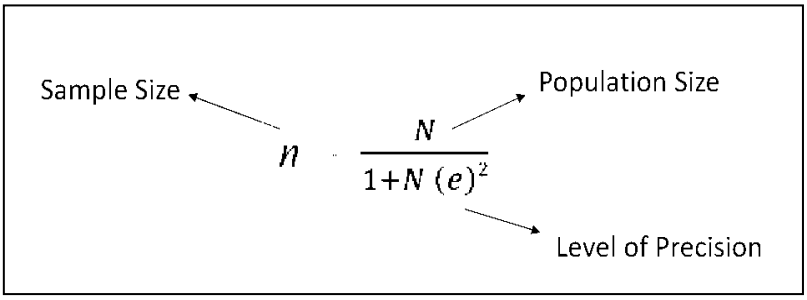
2. Cochran's infinite population (population proportion is Unknown)

$$n = \frac{1}{4} \frac{Z^2}{e^2}$$

3. Cochran's finite population



4. Yamane's finite population



II. Principles and Laws for sample size Justification

Table No. 1 Principles and Laws for sample size Justification

Principles/Laws	Explanation for justification of sample size
Nunnally principle	To build a structural equation model there must be 10 respondents for each item.
Bentler and chou Criteria	When the data is Normally distributed and if the data collection is complex from Senior managers, CEO, Proprietors minimum required sample size to build structural equation model is in the ratio 1:5 (5 cases for each item)
Kaiser Meyer Olkin (KMO)	To apply Exploratory factor analysis (EFA) if value of KMO >0.7 we can conclude that sample is adequate.
Representative sample	As per Business Research method by William Zikmund (1983) minimum required sample is as suggested by the experts.
Qualitative Study	As per Malhotra et al (2018) and Cresswell (2021) the minimum required sample size is 30 Phenomenology research (n=30) Ethnography research (n=30) Grounded theory (n=30) Narratives (n=1)

Experimental study	In case of experimental research minimum required sample is 30 in Treatment group and 30 in control group as per previous literature
SEM model using partial least square method	The minimum required sample is 10 times of construct with maximum items.
Multiple Linear Regression model	As per green (1991) minimum required sample is n=50 +8 * predictors

III. Published Tables for sample size Justification

Table No. 2 Bartletts sample Size

Required Sample Size†

Population Size	Confidence = 95% Margin of Error				Confidence = 99% Margin of Error			
	5.0%	3.5%	2.5%	1.0%	5.0%	3.5%	2.5%	1.0%
10	10	10	10	10	10	10	10	10
20	19	20	20	20	19	20	20	20
30	28	29	29	30	29	29	30	30
50	44	47	48	50	47	48	49	50
75	63	69	72	74	67	71	73	75
100	80	89	94	99	87	93	96	99
150	108	126	137	148	122	135	142	149
200	132	160	177	196	154	174	186	198
250	152	190	215	244	182	211	229	246
300	169	217	251	291	207	246	270	295
400	196	265	318	384	250	309	348	391
500	217	306	377	475	285	365	421	485
600	234	340	432	565	315	416	490	579
700	248	370	481	653	341	462	554	672
800	260	396	526	739	363	503	615	763
1,000	278	440	606	906	399	575	727	943
1,200	291	474	674	1067	427	636	827	1119
1,500	306	515	759	1297	460	712	959	1376
2,000	322	563	869	1655	498	808	1141	1785
2,500	333	597	952	1984	524	879	1288	2173
3,500	346	641	1068	2565	558	977	1510	2890
5,000	357	678	1176	3288	586	1066	1734	3842
7,500	365	710	1275	4211	610	1147	1960	5165
10,000	370	727	1332	4899	622	1193	2098	6239
25,000	378	760	1448	6939	646	1285	2399	9972
50,000	381	772	1491	8056	655	1318	2520	12455
75,000	382	776	1506	8514	658	1330	2563	13583
100,000	383	778	1513	8762	659	1336	2585	14227
250,000	384	782	1527	9248	662	1347	2626	15555
500,000	384	783	1532	9423	663	1350	2640	16055
1,000,000	384	783	1534	9512	663	1352	2647	16317
2,500,000	384	784	1536	9567	663	1353	2651	16478
10,000,000	384	784	1536	9594	663	1354	2653	16560
100,000,000	384	784	1537	9603	663	1354	2654	16584
300,000,000	384	784	1537	9603	663	1354	2654	16586

As per Research Advisor (2006) and as per Bartlett's sample size table Up to 300 million (30 crores) Target population the minimum required sample size is 384 at 95% level of confidence and 5 % precision rate.

Table No: 3 Yamane's Table

Size of Population (N)	Sample Size (n) for Precision (E) of:			
	±3%	±5%	±7%	±10%
500	A	222	145	83
600	A	240	152	86
700	A	255	158	88
800	A	267	163	89
900	A	277	166	90
1,000	A	286	169	91
2,000	714	333	185	95
3,000	811	353	191	97
4,000	870	364	194	98
5,000	909	370	196	98
6,000	938	375	197	98
7,000	959	378	198	99
8,000	976	381	199	99
9,000	989	383	200	99
10,000	1,000	385	200	99
15,000	1,034	390	201	99
20,000	1,053	392	204	100
25,000	1,064	394	204	100
50,000	1,087	397	204	100
100,000	1,099	398	204	100
>100,000	1,111	400	204	100

As per Yamane (1967) if Target population is more than 1 Lakhs the minimum required sample size is 400 at 95% level of confidence and 5 % precision rate.

IV. Online Calculator for sample size Justification

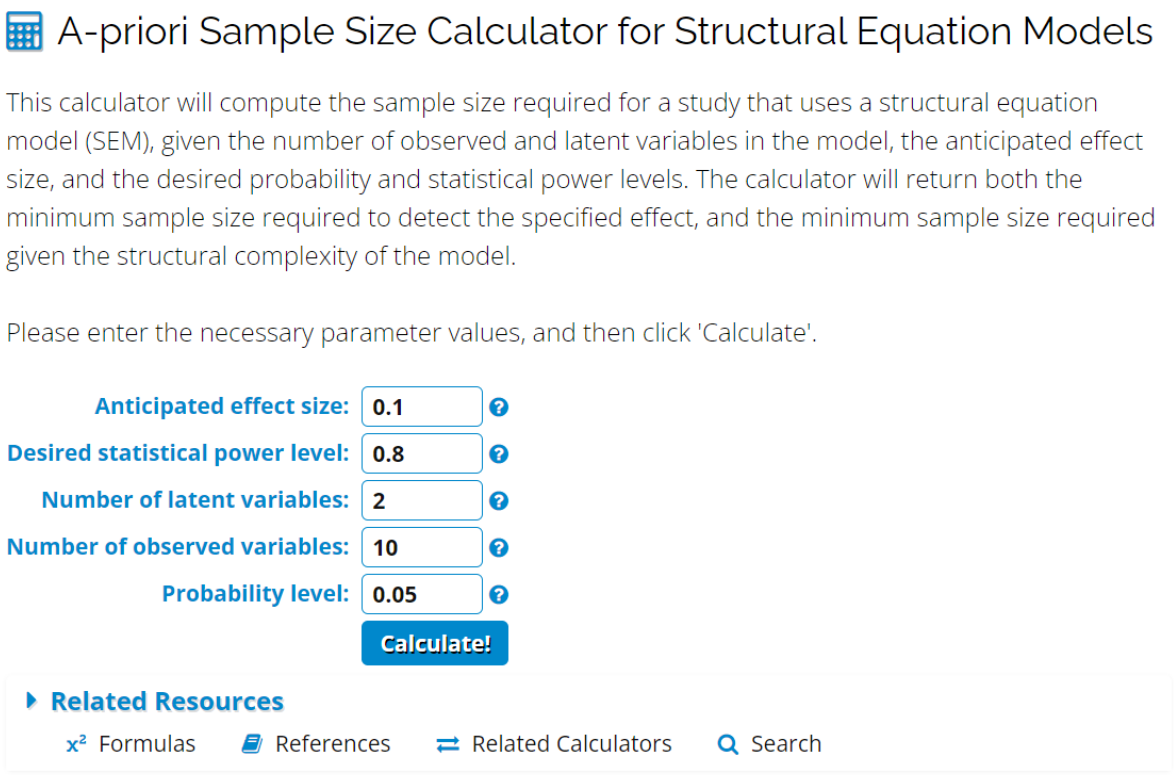


Figure 5: Daniel Soper online calculator

As per Soper, D.S. (2024) Minimum required sample size for SEM, Multiple regression, Hierarchical multiple regression & Student t test can be calculated by entering the effect size, Level of significance and Statistical power.

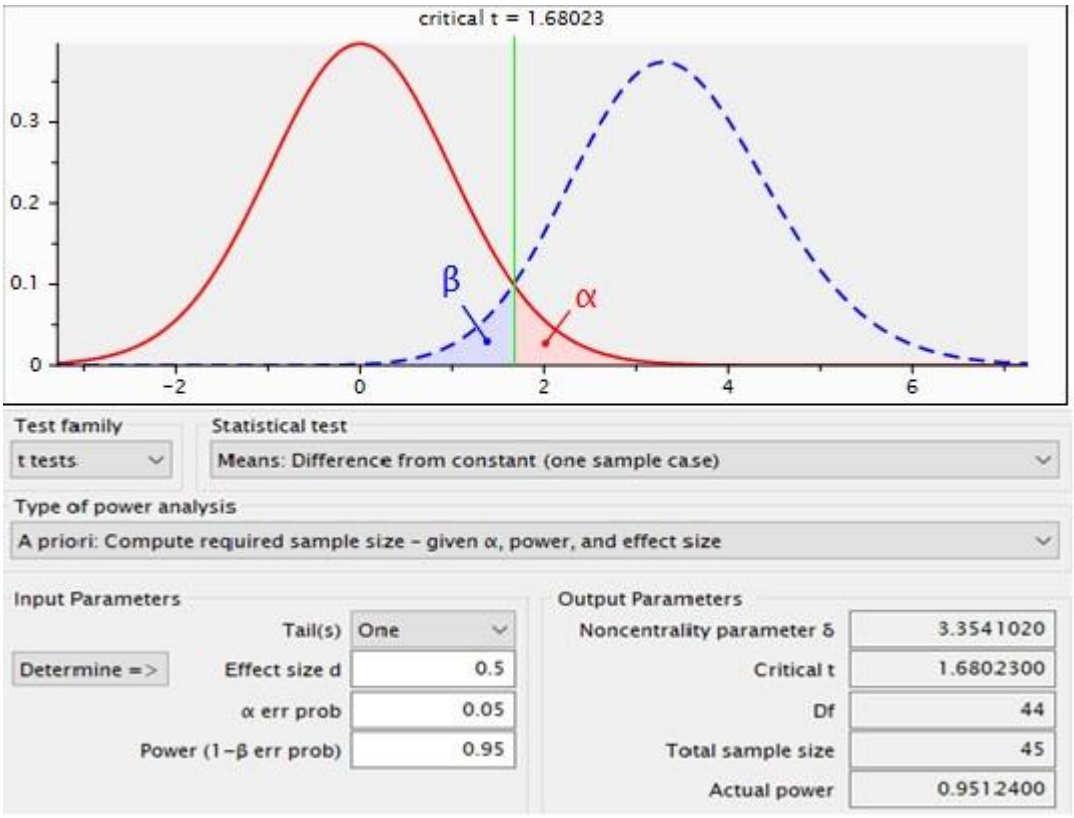


Figure 6: G-Power Online Calculator

As per Faul et al (2007) Minimum required sample size for parametric and Non-Parametric Test can be calculated by applying effect size, Alpha, and power of Test.

10. Implications of the Study

This paper provides helpful advice for researchers who are having trouble figuring out sample sizes, which is a typical problem in research. Offering guidelines for justifying sample sizes gives researchers a methodical way to guarantee accurate results. Furthermore, it contributes to the literature on research methods, boosting comprehension and raising the calibre of research outputs in all fields.

11. Authors Contribution

The writers affirm that they have no connections to, or engagement with, any group or body that provides financial or non-financial assistance for the topics or resources covered in this manuscript.

12. CONFLICT OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article**.**

13. PLAGIARISM POLICY

All authors declare that any kind of violation of plagiarism, copyright and ethical matters will taken care by all authors. Journal and editors are not liable for aforesaid matters.

14. SOURCES OF FUNDING

The authors received no financial aid to support for the research.

REFERENCES

[1] Bartlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research. Information Technology, Learning, and Performance Journal, 19(1), 43-50.

[2] Bentler, P.M. and Chou, C. (1987) Practical Issues in Structural Modeling. Sociological Methods and Research, 16, 78- 117. http://dx.doi.org/10.1177/0049124187016001004

[3] Creswell, J. W. (2021). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (5th ed.). Sage Publications, Inc.

[4] Dubey, U. K. B., Kothari, D. P. (2022). Research Methodology: Techniques and Trends. United Kingdom: CRC Press.

[5] Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191

[6] Green, S. B. (1991). How many subjects does it take to do a regression analysis? Multivariate Behavioral Research, 26(3), 499–510. https://doi.org/10.1207/s15327906mbr2603_7

[7] Handbook of Research Methodology: A Compendium for Scholars & Researchers. (n.d.). (n.p.): Educreation Publishing.

[8] Israel, G. D. (1992). Determining sample size. University of Florida Cooperative Extension Service, Institute of Food and Agriculture Sciences, EDIS.

[9] Kothari, C. R. (2004). Research Methodology: Methods and Techniques. India: New Age International (P) Limited.

[10] Kumar, R. (2018). Research Methodology: A Step By Step Guide For Beginners, 2/E. India: Pearson Education.

[11] Malhotra, N. K., Nunan, D., & Birks, D. F. (2018). Marketing Research: An Applied Orientation (7th ed.). Pearson Education Limited.

[12] Nunnally, J. C. (1997). The Study of Change Evaluation Research: Principle Conserning Measurement Experimental Design and Analysis. In E. L. Dlm. Struening, & M. Guttentag (Eds.), Handbook of Evaluation Research. Beverly Hills: Sage.

[13] PANNEERSELVAM, R. (2014). RESEARCH METHODOLOGY. India: PHI Learning.

[14] Research Methodology: A Handbook for Beginners. (2017). (n.p.): Notion Press.

[15] Soper, D.S. (2024). A-priori Sample Size Calculator for Structural Equation Models [Software]. Available from https://www.danielsoper.com/statcalc

[16] Tejada, J. J., & Punzalan, J. R. B. (2012). On the misuse of Slovin's formula. The Philippine Statistician, 61(1), 129–136.

[17] Zikmund, W. G. (1983). Business Research Methods. Fort Worth, TX: Dryden Press.