Sample size analysis for conducting research

Mariia Pokushko[1,2,3*], *Margarita* Karaseva[1,2], *Roman* Pokushko[2], *Ilias* Aslanidis[1], *Julia* Weitzel[1], and *Artur* Nikiforov[1]

[1]Siberian Federal University, Department of Digital Management Technologies, 79, Svobodny av., Krasnoyarsk, 660041, Russian Federation
[2]Reshetnev Siberian State University of Science and Technology, Department of System Analysis and Operations Research, 31, Krasnoyarsky Rabochy av., Krasnoyarsk, 660037, Russian Federation
[3]University of Cadiz, Department of Computer Engineering, 10, street University of Cadiz, Puerto Real, 11519, Spain

Abstract. The paper describes a method for analyzing a sample size for conducting research. The main stages of sample size analysis with different error rates are considered. The stages of the work in the software for calculating the error rate and sample size are described. Experiments were carried out concerning the application of the described method and program in practice when studying the operating efficiency of electric motors for pumping stations.

1 Introduction

Current scientific trends lead to the development of new methods and tools for sample analysis. The ability to determine the sample size before starting experiments is an important task in conducting research. The application of Statistical Power Analysis tools makes it possible to solve problems of this kind. Researchers from different countries are working on this problem [1-4]. Statistical Power Analysis tools were applied to analyze a sample used in different fields of activity, such as medicine, economics, engineering, etc. [4-8]. In their publications, scientists come to conclusions about the efficiency of applying these tools for conducting research. The authors of the paper will apply Statistical Power Analysis method and the G*Power software to analyze and determine a sample size when conducting research on the efficiency of electric motors of pumping stations.

2 Problem statement

The statement of the sample analysis problem in technology is a rather complex process. The quality and results of the research itself depend on the quality and quantity of the analyzed data. The application of Statistical Power Analysis methods and tools makes it possible to minimize this problem and calculate data errors and the amount of data for sample analysis correctly. Statistical Power Analysis helps determine the required amount of data for conducting research before data collection. It makes it possible to minimize the time of data

* Corresponding author: mvp1984@mail.ru

collection and experiments. Therefore, the authors believe that this kind of the problem is quite relevant for conducting research in technology and technical systems analysis.

3 Research questions

The authors will describe Statistical Power Analysis and how it can be used to determine the required sample size in the following sections. The features of applying the G*Power software and its capabilities for analyzing a sample will be considered while the research. The authors will calculate a sample size for analyzing the operating efficiency of electric motors in pumping stations at different error probabilities. The paper will analyze the obtained sample size data at different probabilities of an error and make conclusions about the possibility of applying this analysis when studying the operating efficiency of electric motors of pumping stations.

4 Purpose of the study

The aim of the study is to analyze a sample size for conducting research on the efficiency of operating electric motors of pumping stations. In accordance with this goal, the following tasks will be solved:

- method and tools of Statistical Power Analysis will be described;
- G*Power method will be considered;
- application of this method and tools will be described while conducting research of the operating efficiency of electric motors of pumping stations;
- sample analysis will be carried out when conducting research of the operating efficiency of electric motors of pumping stations;
- sample volume will be calculated while conducting research of the operating efficiency of electric motors of pumping stations at different error rates;
- conclusions will be made about the possibility of applying this analysis while studying the operating efficiency of electric motors of pumping stations.

5 Research methods

The paper will consider the main methods and tools of sample analysis. The authors will look at 2 basic concepts when studying and preparing a sample for research. Then, the authors will look at the statistical power and sample size calculation. These concepts should be considered when preparing data research before data collection process starts.

The statistical efficiency of the experiment is the probability of the significance of the research result given a correctly defined research hypothesis [9].

The paper considers types errors while analyzing statistical power. Errors include those errors that may be made during statistical power analysis. Such errors are Type I and Type II errors [10]. Type I error is called alpha(α). Alpha (α) is a level of significance, usually it sets to 0.05 or 0.01. α is the probability that the null hypothesis is true even though it is false [9]. Type II error is called beta (β). It consists of the probability of holding the null hypothesis, which is actually false. An extreme level of significance (0.001) implies less efficiency of the research, since it increases the likelihood of committing any of the specified errors [10]. A less extreme level of significance (within the mathematical limits of 0.05 or 0.01) means greater statistical power because it reduces the likelihood of committing any of the errors considered.

Thus, power is determined by the probability of not making a Type II error. A statistical power indicator is estimated as [9]:

 https://doi.org/10.1051/itmconf/20245903006

$$M = 1 - \beta \tag{1}$$

The presence of the right sample is critical for the research to adequately demonstrate its stated objectives. Sampling influences, on the one hand, representativeness, and on the other, statistical power.

After that, in the experiments, the authors will calculate the statistical power and the required sample size for conducting specific research, namely the research of the operating efficiency of electric motors of pumping stations. The calculations will be carried out with the G*Power software [11].

5.1 Experiments

Experiment 1.

The experiment assumes that there are no restrictions on obtaining a certain required amount of data. We believe that it is possible to collect any amount of data required to conduct the study with a minimum error rate of α is 0.05 while the research. Calculate a sample size required for a minimum Type I and Type II error rate, i.e., the optimal number of observations for conducting the research.

Figure 1 presents calculation results.

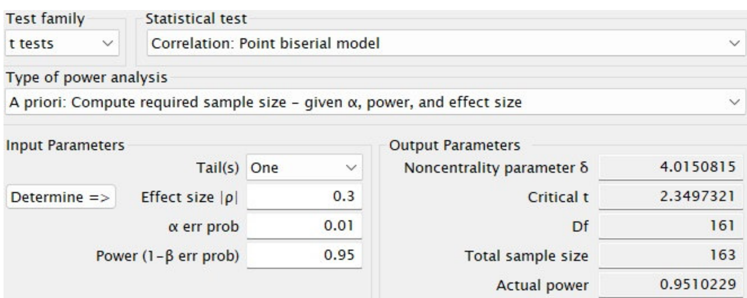

Fig. 1. Results of sample analysis for calculations in the G*Power software.

These calculations prove that the required number of observations is 163. That means that the minimum sample size when minimizing Type I and Type II errors is 163. Therefore, the statistical power indicator will be quite high and amount to 0.9510229.

Figure 2 presents the graphical distribution of Type I and Type II errors.

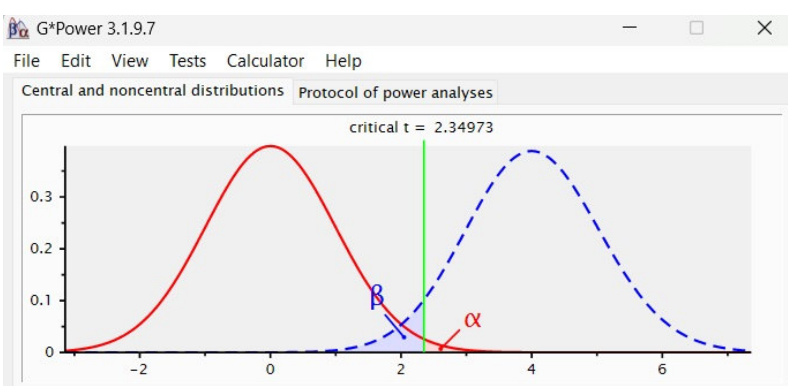

Fig. 2. 1. Graphical representation of the α and β distribution in calculations for experiment 1.

The figure presents graphical distribution of Type I and Type II errors. According to the given figure it is clear that critical t is 2.34973.

If the research encounters difficulties in obtaining sample data, the margin of error in sample size calculations can be increased. In this case, it is necessary to describe what magnitude of error the given experimental results were obtained during the research when analyzing the results of an experiment. Let's conduct experiment 2. Here we increase the error rate α; it will be 0.05.

Experiment 2.

The authors will assume that there are limitations in obtaining a certain required amount of data in this experiment. Therefore, when calculating a sample size, the authors will increase the error rate α; it will be 0.05.

Figure 2 present the calculation results.

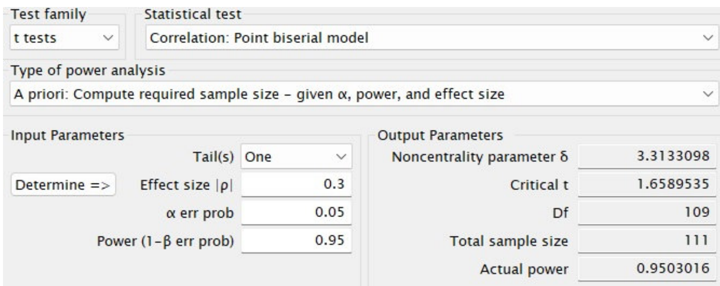

Fig. 3. Results of sample analysis for calculations in the G*Power software.

These calculations show that the required number of observations is 111. It means that the minimum sample size when minimizing Type I and Type II errors will be 111. And the statistical power indicator will be lower than in Experiment 1; it will be 0.9503016.

Figure 4 present the graphical distribution of Type I and Type II errors in the experiment.

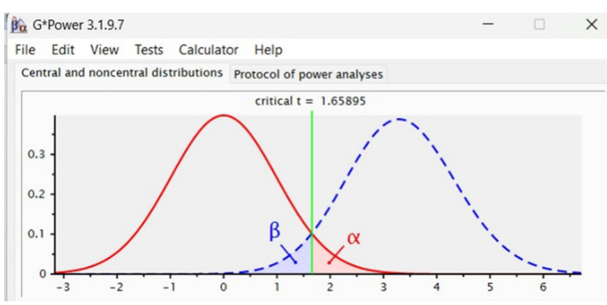

Fig. 4. Graphical representation of the α and β distribution in calculations for experiment 2.

The figure presents the graphical distribution of Type I and Type II errors. According to this figure it is clear that critical t is 1.65895.

Thus, when conducting experiments, we can come to the conclusion that the minimum amount of required data sample for analysis increases by minimizing the magnitude of Type I and Type II errors. As the magnitude of Type I and Type II errors increases, the minimum required data sample for analysis decreases. Accordingly, under different research conditions and the possibility of obtaining the data required for analysis, according to the research results, scientists provide data on the magnitude of type I and type II errors when conducting research to indicate the statistical power of a given research.

6 Conclusion

The paper solves the problem of analyzing the sample volume for conducting research on the operating efficiency of electric motors of pumping stations. The method and tools of Statistical Power Analysis were described. The G*Power method was considered. The application of this method and tools while conducting the research of the operating efficiency of electric motors of pumping stations was described. A sample analysis was carried out while conducting the research of the operating efficiency of electric motors of pumping stations. The sample size was calculated while conducting the research of the operating efficiency of electric motors of pumping stations at different error rates. The conclusions were made about the possibility of applying this analysis when studying the operating efficiency of electric motors of pumping stations. It was concluded that the minimum amount of required data sample for analysis increases by minimizing the magnitude of Type I and Type II errors. As the magnitude of Type I and Type II errors increases, the minimum required data sample for analysis decreases. Therefore, according to the research results scientists provide data on the magnitude of type I and type II errors under different research conditions and possibility of obtaining the data required for analysis while conducting research to denote the statistical power.

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant № 075-15-2022-1121).

References

1. M. Doll, I. Klein, Communications in Statistics - Theory and Methods 52(24) (2023). https://www.doi.org/10.1080/03610926.2022.2068029
2. D.M. Goldenholz, H. Sun, W. Ganglberger, M.B. Westover, Biomedicines 11(3), 685 (2023). https://www.doi.org/10.3390/biomedicines11030685
3. H. Kang, Journal of Educational Evaluation for Health Professions 18 (2021). https://www.doi.org/10.3352/JEEHP.2021.18.17
4. K.K. Harrall et al., BMC Medical Research Methodology 23(12) (2023). https://www.doi.org/10.1186/s12874-022-01819-y
5. M. van Iterson et al., BMC Genomics 10(439) (2009). https://www.doi.org/10.1186/1471-2164-10-439
6. G. Rogers, M. Szomszor, J. Adams, Scientometrics 125, 777–794 (2020). https://www.doi.org/10.1007/s11192-020-03647-7
7. J. M. Cárdenas Castro, Salud & Sociedad 5(2), 210-244 (2014). https://www.doi.org/10.22199/s07187475.2014.0002.00006
8. M.A. Constantin, N.K. Schuurman, J.K. Vermunt, Psychological Methods (2023). https://www.doi.org/10.1037/met0000555
9. T.A. Kyriazos, Psychology **9(8)**, 2207-2230 (2018). https://www.doi.org/10.4236/psych.2018.98126
10. E. Erdfelder, F. FAul, A. Buchner, A.G. Lang, Behavior Research Methods **41**, 1149-1160 (2009). https://www.doi.org/10.3758/BRM.41.4.1149
11. G*Power. Statistical Power Analyses for Mac and Windows (2023). https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower