Pamukkale Üniversitesi
Sosyal Bilimler Enstitüsü Dergisi
Pamukkale University Journal of Social Sciences Institute
ISSN 1308-2922 E-ISSN 2147-6985

Article Info/Makale Bilgisi
Received/Geliş:15.05.2022 **Accepted/Kabul:**10.10.2022
DOİ:10.30794/pausbed.1117138
Research Article/Araştırma Makalesi

Şanlı, S. (2023). "Sampling Methods and Appropriate Sample Size Determination:A Concise Overview", *Pamukkale Social Sciences Institute Journal*, Issue 56, Denizli, pp. 357-375.

SAMPLING METHODS AND APPROPRIATE SAMPLE SIZE DETERMINATION: A CONCISE OVERVIEW

Sera ŞANLI*

Abstract

Sample size determination has been a controversial issue for many applied researches, such that incorrect determination of how large a sample should be may lead to erroneous inferences in the study. For this reason, in this paper it has been aimed to highlight the general criteria and formulations for detecting the required minimum sample size especially in probability sampling methods -which can be summarized as simple random sampling, systematic sampling, stratified sampling and cluster sampling- and various research designs. Besides, the place of statistical power and effect size on sample size determination and the sample size computations for regression analysis based on the study by Cohen (1988) have also been mentioned**.**

Keywords: *Sample size, Statistical power, Probabilistic sampling, Simple random sampling, Stratified sampling, Cluster sampling.*

ÖRNEKLEME YÖNTEMLERİ VE UYGUN ÖRNEKLEM BÜYÜKLÜĞÜNÜN BELİRLENMESİ: ÖZLÜ BİR BAKIŞ

Öz

Örneklem büyüklüğünün belirlenmesi çoğu uygulamalı araştırma için tartışmalı bir konu olmuştur, öyle ki bir örneklemin ne kadar büyük olması gerektiğinin yanlış belirlenmesi çalışmada hatalı çıkarımlara yol açabilmektedir. Bu nedenle bu çalışmada özellikle basit tesadüfi örnekleme, sistematik örnekleme, tabakalı örnekleme ve küme örneklemesi olarak özetlenebilecek olasılıklı örnekleme yöntemlerinde ve çeşitli araştırma tasarımlarında gerekli minimum örneklem büyüklüğünün saptanmasına yönelik ölçütlerin ve formülasyonların öne çıkarılması amaçlanmıştır. Ayrıca istatistiksel güç ve etki büyüklüğünün örneklem büyüklüğünün belirlenmesindeki yerine ve Cohen'in (1988) çalışmasına dayanarak regresyon analizi için örneklem büyüklüğü hesaplamalarına da değinilmiştir.

Anahtar kelimeler: *Örneklem büyüklüğü, İstatistiksel güç, Olasılıklı örnekleme, Basit tesadüfi örnekleme, Tabakalı örnekleme, Küme örneklemesi.*

*Res. Assist. Dr., Çukurova University, Faculty of Economics and Administrative Sciences, Department of Econometrics, ADANA. e-mail: sanlis@cu.edu.tr (https://orcid.org/0000-0002-4827-1032)

1. INTRODUCTION

Sampling is known as the process of drawing a sample from a population of interest. Undoubtedly, conducting this process is a hard task bringing some technical issues with it. Especially for quantitative researches, since it is the primary desirable aim for statisticians and practitioners to work with the samples which reflect a fair and perfect match of the overall population under the investigation and thus to feel themselves confident in making sound inferences, it would be useless to use the samples in case they do not satisfy the representative feature for the entire reference population. More precisely, one should understand from the *representative* term that distinctive traits of any sample such as age or gender are marked by resemblance in many respects as possible to the distinctive traits of a population (Paternoster and Bachman, 2017). An accurate representation of population implied by the representativeness notion is a *generalizability* affair -that is regarded as a fundamental standard in order to assess the research quality and associated with external validity which accounts for the scope of whether inferences drawn from a sample in interest are feasible for a larger population (Chasalow and Levy, 2021: 80; Findley et al., 2021: 366; Polit and Beck, 2010: 1452). Generalizability of the sample findings to the target population is influenced due to some margin of sampling error which shows the divergence between the sample statistic and the population parameter and occurs when the relevant sample may not depict an accurate behavior of the wider population from which the sample is drawn (Altman and Bland, 2014). Thus, it would be reasonable to say that the usage of an imperfect method concerning sample choice leads to sampling bias which also contributes to the sampling error that will continue to exist although the sample size is large. Even there is no sampling bias, sampling errors are likely to occur due to the chance factor. Most prevalently, the standard error of a sample statistic is mentioned as a measure of sampling error. Correspondingly, the sampling error will depend on the sample size and be minimized by using an appropriate sampling method. The larger the size of the sample, the smaller the sampling error, and sampling error may pose serious problems in inferences made especially for small sample sizes. In this respect, the correct determination of the sample size is of great importance in order to evaluate the sampling error accurately. Even though the sampling error is known to have an inverse relationship with the square root of the sample size in general, this is not the case for non-sampling errors such that these types of errors are inclined to be related in the same direction as the sample size (Medhi, 1992: 372-373).

Considering the importance of the sample size, in this study it has been aimed to present a comprehensive review for sample size calculations by presenting some formulas for different research and sampling designs as a guide for researchers. Two basic sampling designs often adopted by the researchers could be mentioned as probability and non-probability samplings. This classification also reflects the random or non-random withdrawal of the sample units to some extent and will be mentioned in Section 2. The rest of this paper has been organized as follows: Section 3 presents the formulations for sample size determination, Section 4 refers to some further considerations, finally the last section states a brief evaluation of the study.

2. AN OVERVIEW OF SAMPLING METHODS

2.1. Probability Sampling Techniques

Probability sampling methods are used to formulate samples that guarantee the *representativeness* -one of the crucial criteria to assess the sample quality- of the population to a precise manner (Farmer and Farmer, 2020) and the randomization principle constructs the dominant part of these techniques. Each unit/element in the population has certain probabilities of being incorporated into the sample. Most common sampling methods can be summarized as *simple random sampling*, *stratified random sampling*, *cluster sampling* and *systematic sampling*. The choice of the suitable sampling type is detected according to the structure of the target population such that when worked with the homogeneous populations, the usage of simple random sampling is ideal; while the choice of stratified sampling -which displays the representativeness characteristic more compared with the simple random sampling for a given sample size (thus, can provide a greater accuracy)- would be appropriate in the contrary (heterogeneous) case (Verma, 2019: 312).

Simple Random Sampling: In this sampling method, the probability that each member in the population is assigned for being included in the sample is known and equal for all units. When the sampling frame includes 1000 members and the sample to be drawn consists of 100 members, the probability of each member being included in the sample will be equal and 1/10 (Greenstein and Davis, 2013: 51). Random selection can be performed in three ways, namely lottery, random numbers table or software for random number generation. For instance, Microsoft Excel could produce the random numbers through the function given as RANDBETWEEN

(Farmer and Farmer, 2020). Unfortunately, for most real-world applications, creating such a random sample is almost not likely, and that is why pseudorandom numbers are recoursed. On the other hand, there is a limitation coming with this type of sample that it requires to make a complete list of all elements in the population under investigation, because of this the method would not be an appropriate one for infinite populations. Besides, when the members of interest are people (often so in the social sciences), there is no way to compel everyone for taking part in the study because of possible refusals. Therefore, simple random samples are of an uncommon usage when considering other easy-to-apply sampling techniques (Hedrih and Hedrih, 2022: 37).

Stratified Random Sampling: In this sampling method, population is partitioned into two or more non-overlapping subgroups (i.e., strata) that seem relevant (homogeneous) and are based on some shared attributes like gender and then a simple random sample is drawn from each stratum. This method is advantageous in terms of being cost-efficient. However, the hardship of stratifying each element of the population lies among its advantages (Greenstein and Davis, 2013: 51-52). Additionally, the investigator should make a decision about whether the size for the sample concerning each subgroup will be proportional or disproportional in its representation to the population. If the answer is proportional, then a proportional stratified sampling technique will be implemented. One advantage of it is that sampling errors can be reduced in the case of large variability between subgroups (Farmer and Farmer, 2020). The main consideration for stratified sampling is that having a population with different characteristics (different subgroups) pretty much will ensure that these subgroups are sufficiently represented in the sample (Hedrih and Hedrih, 2022: 38). However, the necessity for having a complete knowledge on sampling frame forms a disadvantage for the method (Verma, 2019: 307).

Cluster Sampling: Cluster sampling (also called as area sampling) is the process of selecting a cluster of participants from the population and every element of a population should be included in only one cluster. These clusters can be exemplified as cities, school districts, classes, area codes or shopping malls etc. Cluster sampling method can be conducted as one stage and two stage (or multistage sampling). In the one-stage cluster sampling, the sample is obtained through the choice of the cluster. In two-stage cluster sampling, at the first stage clusters are selected and then sample is constructed from each pre-selected clusters (i.e., units that contain elements) by using random sampling. Apart from this, the concepts of stratified and cluster samplings are often confused. The difference between cluster sampling and stratified sampling is that in stratified sampling samples are constructed from the elements belonging to each stratum while the elements are drawn only from the sampled clusters in cluster sampling. One crucial requirement for this type of sampling is that clusters must be homogenous between each other and the content of them must be heterogenous. Among the major advantages, it lies to draw probability samples even in cases in which the researcher does not have the knowledge about the complete list of the target population. Cluster sampling is also more preferable when the researcher wants to examine cluster-level characteristics, not the individual-level; and it offers a cost-efficiency relative to the stratified random sampling. On the other hand, this method is disadvantageous in terms of that it may not be able to maximize the representativeness of the population in the constructed sample as much as the random sampling method, given the equal size of the sample. Also in parallel with this inference, it produces a lower precision degree in estimations when compared with the simple random and stratified samplings and tends to have larger sampling errors than stratified random samples. This negativity can be compensated by an increase in the sample size which in turn may lead to higher costs (Farmer and Farmer, 2020; Greenstein and Davis, 2013: 52-53; Hedrih and Hedrih, 2022: 41; Verma, 2019: 308-309).

Systematic Sampling: In this type of sampling, first a complete list of population members being in an arranged sequence is formed and by using the systematic sampling method, it is aimed to create a sample of n units from the population. Assuming the population size as N and using the number to be obtained when N is divided by n (let it be denoted as k), primarily a random number is chosen out of these first k units. Then, every kth unit is selected from the list until the list is exhausted and the sampling of n units is completed. For instance, consider that a sample of 25 is planned to be constructed from the list of 1000 students. Thus, k would be equal to 40. First, a student is drawn from 40 students (or choose a random number between 1 and 40), let it be the 15th one. Subsequent to incorporating this 15th student into the sample, and then every 40th student is chosen from the list. So, the sample will include every 40th student from the population in that manner until it is completed to the 25 students: 15th, 55th, 95th, 135th etc. In concise, in the systematic sampling, a random selection is realized among the first k numbers at first, and then the sampling process is started. Systematic sampling also displays the characteristics of simple random and stratified samplings depending on the randomness of the first selection and on the view of as if the entire population is partitioned into k strata. Since

only the first unit is obtained at random, a systematic bias may occur in the sample. On the other hand, the usage of this method is an advantage when the study is subject to budgetary constraints (Hedrih and Hedrih, 2022: 42; Verma, 2019: 307-308).

2.2. Non-probability Sampling Techniques

In non-probability sampling techniques, contrary to the probability sampling techniques each member in the population does not have equal chance for the inclusion in the sample (Verma, 2019: 310); thus the likelihood of the inclusion is not known. These sampling techniques do not exert any control for possible investigator bias in the choice of members for inclusion in the sample and are preferable in case it is not feasible to use the probability sampling (e.g., it is not desirable to apply probability sampling for extremely small sample sizes). Generally, the usage of four types of nonprobability sampling procedures is prevalent: convenience sampling, quota sampling, purposive sampling and snowball sampling (Greenstein and Davis, 2013: 53). These sampling methods are subject to subjective judgements, theoretical knowledge and experiences of the researcher in general for choosing the sample (or its certain traits). Samples generated from these techniques do not often reflect the true picture of the population -so, as being not truly representative, the principle of generalization to the population does not operate (Verma, 2019: 310).

3. SAMPLE SIZE DETERMINATION

Determination of optimal sample size is a controversial issue for many research areas. As a matter of fact, the wrong selection of the sample size poses a danger in that it may lead to erroneous inferences intended for the research. It is likely to avert from an intrinsic part of the measurement process, namely random error, and improve the precision by expanding the sample size. As expressed by Cook and Campbell (1979), sample size is also one of the basic concerns which has an influence on the statistical conclusion validity apart from research design, the level of the significance, the variance of the population, effect size (ES), the type of the hypothesis and the method of the statistical test (Dattalo, 2008: 12).

In the general sense, there are some fundamental components that have to be certainly known about sample size calculations. Power analysis is one of those key components of study designs in detecting the adequate sample size, and Rothman (1986) defines the statistical "power" term as "the probability of detecting (as 'statistically significant') a postulated level of effect" (Loue, 2002: 43). Power, the probability of correctly rejecting the null hypothesis (true-positive rate), is computed as $1-\beta$ where β represents the chance of committing a Type II error (i.e., false negative). It tends to increase with the size of the sample in the general sense, and is contingent on the chosen sampling design and the way the elements are assigned (Hirji, 2006: 197). To have a better understanding about Type II error, one should also consider Type I error in the hypothesis testing framework. A Type I error rate (indicated by α) shows the significance level that is set by the researcher and represents the possibility of falsely rejecting the valid null hypothesis given that it is actually true (false-positive outcome) in the entire population. α and β parameter values that have a common usage are given respectively as $\alpha = 0.05$ and 0.01 (i.e., 5% and 1% significance levels respectively) and $\beta = 0.20$ indicate 0.10 (being equivalent to 80% and 90% power, respectively - i.e., the probability that the test will have meaningful results). Thus, it would be reasonable to say that the tests under the given research study should have a high statistical power for acquiring desirable well-grounded findings. Undoubtedly, there will be situations where those arbitrary limits for Type I and Type II parameters might take distinct values according to the scope of the study. The justification for these choices of error rates is stated as because they are traditionally used in this way, rather than being grounded upon a scientific validity in essence. Since the required sample size is easily affected by the selection of those parameters, it is of great importance to take the trade-off between two types of errors into account, such that a decrease in the rate of Type I error tends to increase the rate of Type II error, and vice versa. So the researcher should counterbalance between two types of errors depending on this inverse proportionality.

Some softwares being feasible for sample size calculations and/or statistical power analyses could be referred to as *G*Power* (popular among the frequently used softwares and freely available), *PS* (freely available), *PASS*, *Piface* (freely available), *nQuery*, *R packages* (pwr, TrialSize, PowerUpR, powerSurvEpi -all freely available), *SAS* (PROC POWER), *SPSS* (SamplePower), *STATA* (power), *Microsoft Excel* (PowerUp); as for the websites, some of them are *Power and Sample Size* (http://powerandsamplesize.com/Calculators/), *Sample Size Calculators* (http://www.sample-size.net/), *Genetic Power Calculator* (http://zzz.bwh.harvard.edu/gpc/) (Hickey et al., 2018; Serdar et al., 2021).

Another input for the sample size determination is the confidence level. How large the sample size should be is among the central questions that survey studies also focus on. Generally speaking, the researcher mostly tends to determine the desirable level of confidence as 95% while selecting the appropriate sample size. The "confidence level" term could be stated as "the percentage of time when the result is expected to be accurate not due to chance" (Chiu and Tavella, 2008: 111) and calculated as one minus the significance level. To sum up, one would say that the researcher wants to be 95% confident to obtain accurate results of the entire population. If the confidence level is set to 99%, this means the requirement of a larger sample depending on the likelihood that the sample will not reflect the population to a great extent will only be 1% (Johnson, 2002: 113-114).

Sample size also emerges as a function of the population size such that Krejcie and Morgan (1970) presents a short-cut guide for an appropriate sample size choice based on the given population size. Thus, a researcher could utilize from the predefined proposed values without the need for additional calculations considering it as a feasible benchmark for the analyses. Some of the sample size values given by Krejcie and Morgan (1970) have been presented in Table 1.

Besides the statistical power, significance level, confidence level and population size; the other sample size criteria can be considered as the precision level and the degree of variability as well (Miaoulis and Michener, 1976). The choice of the appropriate sample size is also influenced by the sampling design to be applied in the research to a large extent. Taking the size of the samples as equal, a simple random sample will generate a higher level of standard error than a stratified random sample, but a lower value of the standard error for the clustered samples. This association between the standard error of a complex design and that of a simple random sample raises an important concept "*design effect*" which evaluates the impact of the sample design on the accuracy degree of the population estimates (Adams et al., 2007: 91).

Table 1: Appropriate sample size selection based on a given size of the population

N	n	N	n
10	10	600	234
20	19	700	248
30	28	850	265
40	36	1000	278
50	44	1200	291
75	63	1500	306
100	80	2000	322
130	97	3000	341
160	113	4500	354
200	132	6000	361
240	148	8000	367
300	169	10000	370
360	186	20000	377
400	196	40000	380
460	210	50000	381
500	217	75000	382

Note. N and n represent the population size and sample size respectively. Results are given for 95% confidence level.
(Source: Krejcie and Morgan, 1970: 608).

3.1. General Formulas for Sample Size Determination

There are numerous widely accepted formulas to perform sample size calculations that may differ based on the structure and type of the research. For different study designs, we can summarize some of the calculations as follows:

3.1.1. Sample Size Determination on the Basis of Accuracy Factor (For a Single Population Mean Estimation)

Cost and the degree of accuracy are the crucial factors that influence the sample size of the research. If the main target in the study is to obtain accurate estimates, then the sample size calculation should be grounded upon the accuracy factor rather than the cost consideration. Therefore, such studies are usually implemented by the government agencies or private sectors where no concern is about the cost. Consider that the investigator plans to develop $(1-\alpha)\%$ confidence interval of the population mean on the basis of the normally distributed

sample in which the variance is definite. Taking the estimation range for the population mean as within *d* scores of the true parameter value (i.e., *d* indicates the precision range -regarding the closeness to the actual population mean), the confidence interval would be constructed as

$$\left(\overline{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) - \left(\overline{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 2d$$

or

$$2Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 2d \;\Rightarrow\; n = \frac{Z_{\alpha/2}^2\sigma^2}{d^2} \tag{1}$$

where n gives us the required sample size for the population mean estimation, σ is the population standard deviation (obtained from the previous published works or pilot studies), $\sigma/\sqrt{n}$ is the standard error of the mean and $Z_{\alpha/2}$ represents the standardized normal score (1.96 for $\alpha = 5\%$ significance level). In case the result is not an exact number, the optimal sample size is found as the value being rounded to the nearest exact number (Adams et al., 2007: 93; Verma, 2019: 314-315).

3.1.2. Sample Size Determination for Estimating the Population Proportion

The formula for computing the adequate sample size in the estimation of population proportion is given as

$$n = \frac{Z_{\alpha/2}^2 p(1-p)}{r^2} \tag{2}$$

where n is the required sample size, $Z_{\alpha/2}$ represents the standardized normal score (1.96 for $\alpha = 5\%$ showing the significance level), p shows the estimated proportion and r shows the precision range (i.e., maximum estimation error).

For simple random samples, if the researcher realizes that the initial sample size is proportionally very small when compared with the population size, in this case sample size should be adjusted according to a finite population correction for the computation of the sufficient final sample size as $\tilde{n} = Nn/(N+n)$. To adjust the initial sample size for other sampling methods, the size of the sample for simple random sample should be multiplied by the design effect (Adams et al., 2007: 94-95).

3.1.3. Sample Size Determination Regarding the Difference Between Two Population Means and Two Proportions for a Continuous Outcome

Continuous response measures have a prevalent usage in social sciences sphere where the principal variables for measurement such as job productivity, marital satisfaction or reading ability are generally continuous from a conceptual standpoint (Koricheva et al., 2013: 413). Under the assumption that the null hypothesis states the equality of population means $H_0 : \mu_1 = \mu_2$ while the alternative is $H_1 : \mu_1 \neq \mu_2$, where μ_1 is the true mean in the control group and μ_2 is the mean for the experimental group. Utilizing from the 2-tailed independent samples t-test, sample size to estimate the difference between two means pertaining to a continuous response can be computed using the formula given below when the researcher wants to get the same sample sizes (n) for both groups:

$$n = \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{(\mu_1 - \mu_2)^2} \tag{3}$$

where σ^2 represents the population variance for the variables under investigation, and $z_{1-\alpha/2}$ & $z_{1-\beta}$ represent the $100(1-\alpha/2)$ and $100(1-\beta)$ percentiles of the standard normal distribution. On the other hand, a simple formula for estimating the difference regarding the two proportions p_1 and p_2 (between group I and group II) based on a normal approximation can be expressed as

$$n = \frac{\left\{ z_{1-\alpha/2}\sqrt{2\overline{p}(1-\overline{p})} + z_{1-\beta}\sqrt{p_1(1-p_1)+p_2(1-p_2)} \right\}^2}{(p_1-p_2)^2} \quad (4)$$

where $\overline{p}$ is the arithmetic mean of p_1 and p_2 (Fleiss et al., 2004; Hickey et al., 2018: 5).

3.1.4. Sample Size Determination for Correlational Research Designs

According to Fraenkel and Wallen (2009), the minimum acceptable sample size for conducting correlational studies is proposed as should not be less than 30. A correlational design which explores the relationships among two or more variables has a practical usage in many research areas, and among its pros can be counted as not requiring any manipulation of variables and any rigid control over them when compared to the experimental method. It can be also implemented for archival data and enables the investigator to examine the associations among the variables in a single study. However, as a disadvantage, it is not feasible to detect cause-and-effect relationships by using these studies. As a matter of fact, it is notable to recall that correlation does not imply causation (Riggio, 2016: 29-30). Despite the prevalent usage of the correlational studies, not a vast majority of the literature studies allows for the sample size estimations in this research design. Moving from here, Bolarinwa (2020: 73) presents the sample size estimation formulas for correlational studies as

1) $n = \{\frac{Z_\alpha Z_\beta}{C_r}\}^2 + 3$ where $C_r = \frac{1}{2}\ln\left\{\frac{1+r}{1-r}\right\}$ (One-Sample Case) (5)

2) $n = \{\frac{Z_\alpha Z_\beta}{C_{r1}-C_{r2}}\}^2 + 3$ where $C_{r1} = \frac{1}{2}\ln\left\{\frac{1+r_1}{1-r_1}\right\}$ and $C_{r2} = \frac{1}{2}\ln\left\{\frac{1+r_2}{1-r_2}\right\}$ (Two-Sample Case) (6)

Note: C_r denotes the Fisher's transformation of the correlation coefficient and r denotes the correlation coefficient, Z_α denotes the standard normal deviate for Type I error and Z_β denotes the standard normal deviate for the statistical power (the complement of Type II error rate).

3.1.5. Sample Size Determination in Survey Researches

The main target in survey researches is to gather the data as the representative of the population as a whole. Concerning the sample size calculation *for the proportion* in survey researches, the relevant formula could be represented as follows:

$$n = \frac{(Z_{\alpha/2})^2 P(1-P)D}{M^2} \quad (7)$$

where *n* is the required sample size, $Z_{\alpha/2}$ denotes the standard normal deviate which is conventional (taken as 1.96 for 5% significance level), P denotes the prevalence (ratio) of the occurrence under investigation, *M* denotes the precision level (or the acceptable margin of error, conventionally taken as 5% for social sciences) and *D* denotes the design effect that reflects the performed sampling design. For simple random sampling, its value is considered as 1 and as for the other sampling designs like stratified, systematic or cluster etc., *D* tends to take higher values (generally 1 to 2). For cluster random sampling, *D* is considered as 1.5 to 2. The value of *D* will exceed 10 in the case of purposive or convenience samplings. When systematic, stratified, cluster sampling etc. sampling methods are performed rather than the simple random sampling which is not covered as a candidate sampling method for the real-life documented surveys; the larger the value of design effect, the greater the required sample size to conduct the research (Suresh and Chandrashekara, 2012).

Another formula for calculating the appropriate sample size *for questionnaire (survey)* study can be stated as:

$$n = \frac{\frac{p(1-p)Z^2}{e^2}}{1+\frac{p(1-p)Z^2}{Ne^2}} \quad \text{or} \quad n = \frac{\ln(1-\text{power})}{\ln(1-p)} \tag{8}$$

where *n* is the sample size, N is the size of the population, *power* is the usual statistical power, p is the proportion of population, e is the margin of error (expressed as percentage in the decimal form) and Z is the standardized normal score (Serdar et al., 2021: 6).

3.2. Sample Size Calculations in Sampling Strategies

3.2.1. Probabilistic Sampling Strategies

This section focuses on the sample size calculation strategies for different probabilistic samples based on the valuable work by Favero and Belfiore (2019: 178-185) presenting a detailed study on the sample size determination.

3.2.1.1. Sample Size Determination for a Simple Random Sample

Optimal sample size for a simple random sample can be determined based on the usual formulas given in subsections 3.1.1. and 3.1.2. which cover the calculations for estimating the population mean and proportion respectively. But these formulas are valid for infinite populations. For this reason, some formulas for finite populations will be given in the following:

Mean Estimation of a Finite Population

For the quantitative variables, sample size for estimating the mean for a finite population is computed according to the following formula:

$$n = \frac{N\sigma^2}{(N-1)\frac{d^2}{Z_\alpha^{\,2}}+\sigma^2} \tag{9}$$

where N is the population size, σ^2 is the population variance, d is the precision range (estimation error) based on $P(|\overline{X}-\mu| \le d) = 1-\alpha$ (μ: population mean, $\overline{X}$: sample mean) and Z_α is the coordinate of the standard normal distribution at the significance level α.

Proportion Estimation of a Finite Population

If the variable in interest is a binary variable, given that $P(|\hat{p}-p| \le d) = 1-\alpha$, the required sample size can be computed as in the following:

$$n = \frac{Np(1-p)}{(N-1)\frac{d^2}{Z_\alpha^{\,2}}+p(1-p)} \tag{10}$$

where N is the population size, p is the population proportion including the desired feature, d is the precision range (estimation error) and Z_α is the coordinate of the standard normal distribution at the significance level α. When the estimate of p ($\hat{p}$) is not available, in order to obtain a conservative sample size that will produce accurate results, $\hat{p}$ is usually taken as 0.50.

3.2.1.2. Sample Size Determination for a Systematic Sample

For systematic samples, the same general formulas mentioned for the simple random samples can be applied to get the required sample size according to the structure of the population (infinite or finite).

3.2.1.3. Sample Size Determination for a Stratified Sample

Definitions of the notations that have been used in the following formulas are given as:

n : required sample size

d : precision range (estimation error) (Note that for the *mean* estimations, *d* indicates the maximum difference that is acceptable by the investigator between μ and $\overline{X}$; for proportion estimations, *d i*ndicates the maximum difference that is acceptable by the investigator between proportion of the population and that of the sample)

k : number of strata

$w_i = N_i / N$ (weight or proportion of ith stratum, given that $\sum_{i=1}^{k} w_i = 1$)

N_i : size of ith stratum ($i = 1,2,....,k$)

N : size of the population $(N_1 + N_2 + ... + N_k)$

$\sigma_i^{\,2}$: variance of the population for stratum i

p_i : proportion of elements showing the desired feature in stratum i

Z_α : coordinate of the standard normal distribution at the significance level α

Sample Size Determination for Estimating the Mean (Infinite Population)

For the quantitative variables, sample size for estimating the mean for an infinite population is computed according to the following formula:

$$n = \frac{\sum_{i=1}^{k} w_i \sigma_i^{\,2}}{d^2 / Z_\alpha^{\,2}} \tag{11}$$

Sample Size Determination for Estimating the Mean (Finite Population)

For the quantitative variables, sample size for estimating the mean for a finite population is computed according to the following formula:

$$n = \frac{\sum_{i=1}^{k} N_i^{\,2} \sigma_i^{\,2} / w_i}{N^2 \frac{d^2}{Z_\alpha^{\,2}} + \sum_{i=1}^{k} N_i \sigma_i^{\,2}} \tag{12}$$

Sample Size Determination for Estimating the Proportion (Infinite Population)

For the binary variables, sample size for estimating the proportion for an infinite population is computed according to the following formula:

$$n = \frac{\sum_{i=1}^{k} w_i p_i (1 - p_i)}{d^2 / Z_\alpha^{\,2}} \tag{13}$$

Sample Size Determination for Estimating the Proportion (Finite Population)

For the binary variables, sample size for estimating the proportion for a finite population is computed according to the following formula:

$$n = \frac{\sum_{i=1}^{k} N_i^2 p_i(1-p_i)/w_i}{N^2 \frac{d^2}{Z_\alpha^2} + \sum_{i=1}^{k} N_i p_i(1-p_i)} \tag{14}$$

3.2.1.4. Sample Size Determination for a Cluster Sample

In this section, sample size selection for the cluster sampling has been presented for one-stage cluster sample and two-stage cluster sample separately. General notations utilized in the following formulas are given as:

N: size of the population

$\mathbf{C}$: number of clusters into which the population is separated

N_i: size of i[th] cluster ($i = 1,2,....,\mathbf{C}$)

n: required sample size

c: number of drawn clusters ($c < \mathbf{C}$)

n_i: size of i[th] cluster from the sample drawn in the 1[st] stage ($i = 1,2,....,c$) where $n_i = N_i$

e_i: size of i[th] cluster from the sample drawn in the 2[nd] stage ($i = 1,2,....,c$) where $e_i < n_i$

$\overline{N} = N / \mathbf{C}$ (average size of the clusters for population)

$\overline{n} = n / c$ (average size of the clusters for sample)

X_{ij}: j^{th} observation value for i[th] cluster

σ_{wc}^2: variance of the population in the clusters

σ_{bc}^2: variance of the population between clusters

σ_i^2: variance of the population for i[th] cluster

μ_i: mean of the population for i[th] cluster

$\sigma_c^2 = \sigma_{wc}^2 + \sigma_{bc}^2$ (total population variance)

d: precision range (estimation error)

p_i: proportion of elements showing the desired feature in cluster i

Z_α: coordinate of the standard normal distribution at the significance level α

Bolfarine and Bussab (2005) have presented the information about how to calculate σ_{wc}^2 and σ_{bc}^2 in their study. Considering the equal-sized clusters, population variances within and between clusters are calculated according to those formulations:

$$\sigma_{wc}^2 = \frac{1}{\mathbf{C}} \sum_{i=1}^{\mathbf{C}} \sigma_i^2 \tag{15}$$

and

$$\sigma_{bc}^2 = \frac{1}{\mathbf{C}} \sum_{i=1}^{\mathbf{C}} (\mu_i - \mu)^2 \tag{16}$$

Size of a One-Stage Cluster Sample

Sample Size Determination for Estimating the Mean (Infinite Population)

For the quantitative variables, the number of clusters drawn in the 1st stage (c) is computed as

$$c = \frac{\sigma_c^{\,2}}{d^2 / Z_\alpha^{\,2}} \tag{17}$$

For the equal-sized clusters, Bolfarine and Bussab (2005) state that total population variance $\sigma_c^{\,2}$ is generally not known and should be estimated based on the previous surveys.

Sample Size Determination for Estimating the Mean (Finite Population)

For the quantitative variables, the number of clusters drawn in the 1st stage (c) is computed for finite populations as

$$c = \frac{\mathbf{C}\sigma_c^{\,2}}{\mathbf{C}\frac{d^2\overline{N}^2}{Z_\alpha^{\,2}} + \sigma_c^{\,2}} \tag{18}$$

Sample Size Determination for Estimating the Proportion (Infinite Population)

For the binary variables, the number of clusters drawn in the 1st stage (c) is computed as

$$c = \frac{1/\mathbf{C}\sum_{i=1}^{\mathbf{C}} \frac{N_i}{\overline{N}} p_i(1-p_i)}{d^2 / Z_\alpha^{\,2}} \tag{19}$$

Sample Size Determination for Estimating the Proportion (Finite Population)

For the binary variables, the number of clusters drawn in the 1st stage (c) is computed as

$$c = \frac{\sum_{i=1}^{\mathbf{C}} \frac{N_i}{\overline{N}} p_i(1-p_i)}{\mathbf{C}\frac{d^2\overline{N}^2}{Z_\alpha^{\,2}} + 1/\mathbf{C}\sum_{i=1}^{\mathbf{C}} \frac{N_i}{\overline{N}} p_i(1-p_i)} \tag{20}$$

Size of a Two-Stage Cluster Sample

Initially, all clusters are regarded as having the equal sizes. For two-stage cluster sample size choice, Bolfarine and Bussab (2005) present a linear cost function which is defined as

$$R = r_1.n + r_2.e \tag{21}$$

where

n: sample size in the 1st stage

e: sample size in the 2nd stage

r_1: observation cost of one unit from the 1st stage

r_2: observation cost of one unit from the 2nd stage

Optimal sample size for the second stage (e) is given in a way to minimize the R function as follows:

$$e^* = \frac{\sigma_{wc}}{\sigma_{bc}}\sqrt{\frac{r_1}{r_2}} \tag{22}$$

3.2.2. Non-Probabilistic Sampling Strategies

For the non-random sampling designs, the sample size determination is mostly realized under the constraint of a specific budget level; or in the choice of sample size it is utilized from the suitable sizes that are already pre-calculated in the previous published works having the same attributes with the current study (Favero and Belfiore, 2019: 178). As for the ideal size choice, non-probability samples may be more efficient in working with small sample sizes compared to the large samples especially when the focus point is the quality of the gathered information rather than the quantity of it (Moule, 2020).

3.3. Statistical Power and ES in Sample Size Determination

The required sample size (n) can be determined as a function of significance level (α), population ES and statistical power. Likewise, power is a function of sample size. When there is an increase in the desired power level, the sample size *n* will also increase. On the other hand, declines in ES and α lead to sample size to increase. The degree to which H_0 is false is indexed by the discrepancy between the null and alternative hypotheses which is called population ES. In power analysis, ES also represents an index that indicates a scale-free measure and is continuous, ranging upward from zero; and this index differs for various statistical tests performed by the researcher. The notion of ES is crucial in terms of giving an insight about the practical significance of the research.

Cohen (1988: 20) proposes an ES index *d* for the t-test of the difference between two independent means as $d = \frac{m_A - m_B}{\sigma}$ for one-tailed case, where m_A and m_B are population means in raw units for Group A and Group B respectively. It is clear that *d* states the difference expressed in units of the within-population standard deviation (σ) as assumed equal for two populations. For this test, the null hypothesis is $d = 0$ and the small, medium and large ES values (i.e., alternative hypotheses) are represented as $d = 0.20,\ 0.50$ and 0.80 respectively. When $|m_A - m_B|$ takes place in the nominator, the formula for d becomes valid for two-tailed case.

A preliminary work by Neyman and Pearson (1928) puts forth the notion of statistical power, and additionally Fisher's (1925) approach handles analogous matters in his debates. Undoubtedly, a pioneering work on statistical power analyses in behavioral researches has been manifested by Cohen's (1962) (Dattalo, 2009: 229-230). In the general sense, the desirable power level is considered as 0.80. The interpretation of research findings becomes too tough when the power of study is low, such that statistically non-significant results can be attributed to the low-powered studies or the lack of a true relationship in the population. Cohen (1992) puts forward some statistical tests in computing the sample size needed to attain an ideal power level of 0.80 [thusly, indicating the presence of 80% probability of finding a significant relationship (Suter, 2012: 405)] for the specified significance criterion α and ES [thus, the complement of power -probability of a Type 2 error- is $\beta = 0.20$] as *t-test for the difference between independent means*, *t-test for the significance of a product-moment correlation coefficient* r , *test for the difference between independent* r values, *the sign test*, *the difference between independent proportions*, *chi-square tests for goodness of fit and contingency tables*, *one-way analysis of variance* and *the significance of a multiple or multiple partial correlation*. Examples of the required sample size under the power level of 0.80 for those eight situations are presented in Cohen (1992). For the power of 0.80 and $\alpha = 0.05$, a medium difference between two independent sample means ($d = 0.50$) is determined through $n = 64$ for each group. On the other hand, for the large ES, in testing the significance of r , the necessary n is 28. For conducting t-test based on $\alpha = 0.05$, the power level of the test becomes 0.80 in the case of $n = 85$ when a researcher assumes a medium size for the population r (ES= r =0.30). If t is detected to be insignificant for the given sample size, it is concluded that two situations can occur: 1) r is below 0.30 or 2) the researcher suffers from the exposure to Type 2 error risk with $\beta = 0.20$ (Cohen, 1992: 156-158).

Power analyses are possible to be conducted both before (a priori) or after gathering the research data (post hoc). A "priori" analysis mainly aims to estimate the suitable sample size to attain a desired power level. As to "post hoc" power analysis which is performed subsequent to that the study is completed, the target is to determine the power of the study (i.e., the possibilities of having correctly rejected a null hypothesis that is false)

utilizing from the information on the obtained sample size, a certain determination coefficient R^2 (as a measure of ES) and a desired significance criterion α to reject the null hypothesis (Tienken and Mullen, 2016: 88). G*Power facilitates the computations of sample size needed and power analyses for various statistical tests covering t, F, chi-square, Z apart from some exact distributions. A screenshot of G*Power software (version 3.1.9.2) on testing the difference between two independent means is displayed in Figure 1. More information on G*Power can be found in specified sources as Erdfelder et al. (1996), Erdfelder et al. (2005), Faul et al. (2007) and Faul et al. (2009). In their study, Verma and Verma (2020) provide the information on G*Power software (for version 3.1.9.2.) and its applications for sample size computations. Here, the sample size determination for testing the significance of correlation coefficient (r) is covered. Let us assume that a researcher investigates the significance of the relationship (r) between resting respiratory rate and fat%. Considering a power level of 0.9 and the significance criterion $\alpha = 0.05$ for a two-tailed test; what is the required sample size in order to test whether this relationship is significantly different from zero in detecting the determination coefficient as being at least 45%? For this aim, the first step to be carried out in G*Power is to choose "Exact" tests for correlation coefficient, "Correlation: Bivariate normal model" from statistical test drop-down menu and then designate the option "A priori: Compute required sample size – given a, power, and effect size" as the type of power analysis. Subsequently, push the "determine" button to calculate the correlation value. In the opened window, the desired coefficient of determination is typed as 0.45, and correlation value under H_1 appears automatically as 0.6708204. A further step requires to click on "calculate and transfer to main window" for ES computations and transferring the obtained value to the main window of G*Power. After this procedure, "two" is chosen for "Tail(s)" menu in the "input parameters" field; other inputs are entered as $\alpha = 0.05$, power= 0.9, correlation under H_0=0 and it is clicked on the "calculate" button. Herewith, total sample size is computed as 19 in the "output parameters" field to reject the null hypothesis by establishing 45% coefficient of determination with an actual power level of 0.9163 in a two-tailed test.

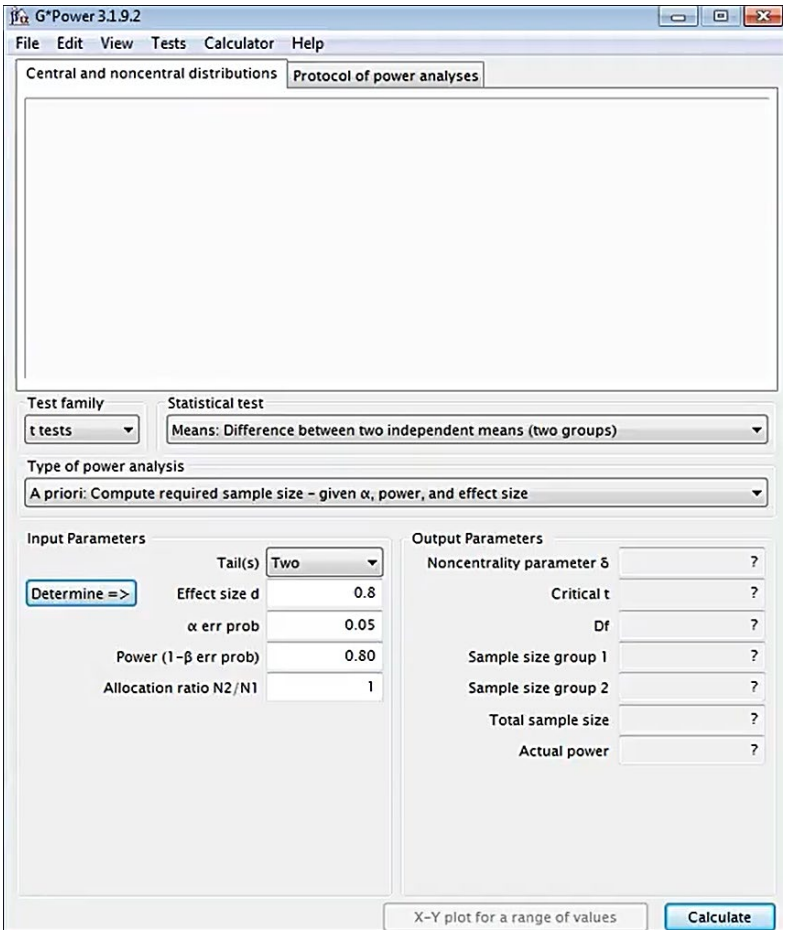

Figure 1: A display of G*Power software for testing the difference between two independent means

Recently, Kang (2021) presents the stages of G*Power computations for sample size & power analyses and five practical examples. According to this, the first step should be to set up the research aim, the null and alternative hypotheses. Secondly, from either the "test family" or "statistical test" drop-down menus, researchers should designate a suitable statistical technique. Thirdly, one of the five types of power analysis should be chosen. Subsequently, to conduct the analysis, required parameters are saved in the sphere of "input parameters". In case there is a priori information that can be obtained from previous researches for the ES computations, that information is typed in a separate calculation window through clicking on "determine" button. When the researcher deals with the two-sample or independent t-test; the alternative hypothesis states that the difference between the means of two samples differs from zero while the null hypothesis states that this divergence is zero.

Kang (2021) gives an example about comparing the analgesic efficacy of two types of drugs using G*Power software (version 3.1.9.7). To this end, a pain score will be gauged at six hours after the medical operation. Assuming that both groups include the same number of patients, the investigators aim to detect the required sample size as a priori of power analysis to obtain a significant difference under a two-tailed test with $\alpha = 0.05$ and a power level of 0.80. If the information on ES is known, the sample size computations become straightforward. Amongst the common ESs provided by G*Power, that example presents the sample size computation for a medium ES (*d*) of 0.5 in the "input parameters" area. Subsequent to opening the main window of the software, choose "test > means > two independent groups", set the "type of power analysis" as "A priori: compute required sample size-given α, power, and effect size". Besides the information given above, enter the allocation ratio as 1 in the "input parameters" field. Finally, pressing on the "calculate" button generates the same sample size (64) for each group, computing the total sample size as 128 in the "output parameters" field. Contrary to the given situation, in case the ES is not determined, it is possible to assume ES by utilizing from pilot or previous studies. Since no previous study is available for Kang's (2021) example, the means and standard deviations (SDs) for pain scores of both drugs are postulated as 7 ± 3 and 5 ± 2, respectively based on a pilot study implemented with 10 patients for both groups. Hereby, input values are 7 for "mean group 1", 3 for "SD group 1", 5 for "mean group 2" and 2 for "SD group 2". To enter these inputs in an ES calculator window, researcher should press on the "determine" button, provide the required information in " $n_1 = n_2$ " field and then push the "calculate and transfer to main window" button. At the end of all these steps, ES "*d*" computed value will appear as "0.7844645" on the "input parameters" field. After this computation, the option "two" is chosen from "Tail(s)" tab for two-sided testing and other inputs ($\alpha = 0.05$, power=0.80, allocation ratio=1) are entered in the same field. In conclusion, pushing the "calculate" button will generate the total sample size as 54 (27 for each sample) (Kang, 2021).

The concept of power relies on the average t-value for some certain alternative hypothesis distribution, which is represented here by the letter delta (δ). Assuming equal-sized groups for a two-sample t test and homoscedasticity for both groups, the general formula for δ can be stated as $\delta = \sqrt{\frac{n}{2}} \frac{(\mu_1 - \mu_2)}{\sigma}$ where n denotes the number of subjects in each group, μ_i is the population mean for group i, and the second term $(\mu_1 - \mu_2)/\sigma$ denotes the ES. Re-expressing the formula in that manner $\delta = \sqrt{\frac{n}{2}} d$ where d is substituted for $(\mu_1 - \mu_2)/\sigma$ and rearranging it as $d = \sqrt{\frac{2}{n}} \delta$ will facilitate the computation of smallest ES that yields a desired power level. The value of δ can be determined through the table provided in Cohen (2001) (see Table A.4 in Appendix A, p. 694). For example, for the power level of 0.80 and two-tailed test ($\alpha = 0.05$), the value for δ is found as 2.80. Considering the acceptable power level that depends on researchers' own judgements and the significance level will yield the δ . Using an estimate of d and δ , the required sample size is computed as $n = 2\left(\frac{\delta}{d}\right)^2$. If the result is not an exact number, it is rounded off to the nearest exact number. For instance, in case n is computed as 8.82, the inference becomes in the way that to achieve a desirable power, both groups should involve nine subjects. However, in case the sample sizes are not equal for each group, the value of n is

computed through the harmonic mean of n_1 and n_2 in order to study with the most accurate power formulas: $n = \frac{2n_1n_2}{n_1+n_2}$ where n_1 and n_2 represent the initial number of subjects for Group 1 and Group 2 respectively (Cohen, 2001: 227-229, 232).

3.3.1. Sample Size Computation for Regression Analysis

In a fixed model multiple regression and correlation analysis (MRC) system where Y denotes the dependent variable, given the various research factors in interest, each one of them (A, B etc.) is composed of one or more independent variables. Small, medium and large ESs for F-test (MRC) are f^2 = 0.02, 0.15 and 0.35 respectively (f^2 as a standardized ES measure). Consider λ as the noncentrality parameter of the noncentral F distribution. It can be expressed as a straightforward function of the ES index (f^2) and degrees of freedoms (d.f.) as follows:

$$\lambda = f^2(u+v+1) \tag{23}$$

where u indicates d.f. for numerator (i.e., the number of independent variables) and v is d.f. for denominator (i.e., d.f. of error variance). In fixed model MRC analyses, F-test which involves λ indirectly can be stated in the general sense as:

$$F = \frac{PV_S / u}{PV_E / v} \tag{24}$$

where PV_S represents the proportion of Y (i.e., a quantitative dependent variable) variance accounted for by some source S in the sample and PV_E represents the proportion of error variance. Consider a set B made up of u variables. When the source of Y variance of interest is defined as set B, for example, PV_S and PV_E are defined as in the following:

$$\begin{aligned} PV_S &= R^2_{Y.B} \\ PV_E &= 1 - R^2_{Y.B} \end{aligned} \tag{25}$$

PV (proportion of variance) terms can also be considered as the functions of squared multiple correlations (R^2). When the formula (24) is reorganized as follows:

$$F = \frac{PV_S}{PV_E} \times \frac{v}{u} \tag{26}$$

and described in population terms, it is translated into $f^2 = \frac{PV_S}{PV_E} = \frac{R^2_{Y.B}}{1-R^2_{Y.B}}$, giving a measure of ES. In order to obtain the required sample size *n*, modifying Equation (23) as $u+v+1 = \frac{\lambda}{f^2}$ will generate v as $v = \frac{\lambda}{f^2} - u - 1$ and in the simplest case that only u variables are used to describe the error variance, v would be equal to $v = n - u - 1$ with $n = \frac{\lambda}{f^2} = \frac{\lambda(1-R^2_{Y.B})}{R^2_{Y.B}}$. To determine *n*, we should determine the value of λ which also depends on the d.f. of the denominator of the F ratio (v). In the study by Cohen (1988), Table 9.4.1. (pp. 448-451) and Table 9.4.2. (pp. 452-455) present the unknown λ values of the F test that are required to determine the sample size *n* as a function of the given power, u and v values under the significance criteria 0.01 and 0.05 respectively.

Power is also obtained as a function of *n*. λ and power are also related, and this connection depends on v - which is a function of *n* being what we are trying to determine. Iteration is a good way to reveal such a connection. To obtain a given power level for the F test, suitable λ values are determined based on the trial values of v , and then *n* is calculated. If the computed sample size indicates a v value with a large discrepancy

from the trial value, a recalculated v value is exploited. For a specific value of v, interpolation between λ values is linear in the reciprocals of the v's, and interpolated λ value is given by the following formula

$$\lambda = \lambda_L - \frac{\frac{1}{v_L} - \frac{1}{v}}{\frac{1}{v_L} - \frac{1}{v_U}}\left(\lambda_L - \lambda_U\right) \tag{27}$$

where v takes place between lower and upper tabled values (v_L, v_U). λ_L and λ_U are λ's corresponding to v_L and v_U.

An Example: If an investigator works with 5 independent variables (u) that accounted for 0.10 ($R^2_{Y.B}$) of the criterion variance in the population. Using the $\alpha = 0.05$, what must n be for the power of the test to be 0.80?

Consider the lower and upper tabled v values as $v_L = 60$ and $v_U = 120$. In Table 9.4.2 (Cohen, 1988) if we take the trial value as $v = 120$, based on the information on $u = 5$, $\alpha = 0.05$ and power=0.80, the value of λ is found as 13.3. Under the given information, n and v are calculated as $n = \frac{13.3(1-0.10)}{0.10} \cong 120$ and $v = 120 - 5 - 1 = 114$. Having the results reviewed, it is apparent that there exists a quite minor disparity between $v = 120$ and the implied value of 114 (i.e., less than 10%). Therefore, this inexact calculated $n = 120$ seems sufficient in the general sense. To reach an exact size of the experiment (n), one would proceed with reiteration by interpolating for λ between $v_L = 60$ and $v_U = 120$. Based on $\lambda_L = 14$ (for $v_L = 60$) and $\lambda_U = 13.3$ (for $v_U = 120$) in Table 9.4.2., the interpolated value of λ for the given v value of 114 will be computed as 13.3 with a round-off to one decimal place:

$$\lambda = 14 - \frac{\frac{1}{60} - \frac{1}{114}}{\frac{1}{60} - \frac{1}{120}}\left(14 - 13.3\right) \cong 13.3 \tag{28}$$

Thusly, it is seen that there is no difference between the interpolated λ value for $v = 114$ and the tabled λ value for $v = 120$, and the original value of n calculated as 120 remains valid. The content of Table 9.4.2. uncovers an important point that λ values for a given power level do not differ so much according to v values of 20, 60, 120 and ∞. It is stated that the value of n in a way to have an adequate accuracy will be obtained as grounded upon the trial value of λ that corresponds to $v = 120$ for most practical studies.

An Example*:* For a medium ES of f^2 = 0.15, $\alpha = 0.01$ and $u = 3$; what is the sufficient value of n to achieve a power level of 0.80?

Table 9.4.1 (Cohen, 1988) presents the λ values corresponding to the lower and upper tabled v values as $\lambda_L = 16.9$ (for $v_L = 60$) and $\lambda_U = 16.1$ (for $v_U = 120$). In that case, using the trial value of lambda that corresponds to $v = 120$, the value of n would be approximately equal to $n = \lambda / f^2 = 16.1 / 0.15 \cong 107$ implying $v = 107 - 3 - 1 = 103$. On the other hand, in order to get a more accurate value for n, the interpolated value of λ for this given value of v is obtained as 16.2 when the result is rounded to one decimal place: $\lambda = 16.9 - \frac{\frac{1}{60} - \frac{1}{103}}{\frac{1}{60} - \frac{1}{120}}\left(16.9 - 16.1\right) = 16.2$. For this λ value, n would be equal to the value of 108 ($n = \lambda / f^2 = 16.2 / 0.15 = 108$) which shows a very minor difference relative to the previous value of *n* (=107). In brief, performing another iteration beyond it would be redundant (Cohen, 1988).

4. FURTHER CONSIDERATIONS

In some cases, it may be necessary to adjust the sample size at the onset of the analysis, since for some investigations, comparison groups may not contain an equal number of subjects. In such a situation, sample size can be converted by using the actual ratio between the uneven branches of the study (m) as $\tilde{n} = n(1+m)^2 / 4m$ where $\tilde{n}$ is the adjusted sample size and n is the available size of the sample at the initial calculation. Apart from this, it is quite possible to come across the issues like incomplete responses, consent refusals, withdrawals or dropouts during the research process. While these factors give rise to the changes in the eventual sample size of the study, they will also affect the inferences drawn from the research to an unfavourable manner, especially in long follow-ups like clinical trials and therefore the validity of the research as well. Assuming the required sample size in the final study as denoted by n , n should be adjusted as $\tilde{n} = n/(1-q)$ depending on the attrition proportion (q) which reflects the knowledge of the likelihood about any type of losses defined above in order to guarantee the final desirable sample size. For instance, if $n = 135$ and maximum 20% of recruited subjects are anticipated to drop out during the research process, the recruitment target sample size $\tilde{n}$ should be approximately as 169 subjects in such a way to ensure 135 subjects to carry through the study till the end (Bolarinwa, 2020: 74; Hazra and Gogtay, 2016).

As in many theoretical sources, up to this section it has been intensely stated that hypothetically large samples produce results that are more precise and more representative of the population. However, of course, a larger sample may not always mean the better sample. Krejcie and Morgan (1970) states that an important contribution about the sample size determination has been addressed by the research division staff of the National Education Association (NEA). The formula published by the NEA Research Bulletin (1960) highlights the importance of small-sample techniques in constructing the contours of doing more in less time in the future as a pioneer work for sample size determination; and this study which aims to acquire a sound representative picture of the opinions of 1.3 million classroom teachers (based on the findings of the period 1958-1959 that at least 26.4% of them are men) and therefore tries to detect the number of men teachers to be incorporated into the sample presents the following formula as the required sample size computations method:

$$n = \frac{\chi^2 N\pi(1-\pi)}{d^2(N-1) + \chi^2\pi(1-\pi)} \tag{29}$$

where n = the required sample size, χ^2 = chi-square table value for 1 degree of freedom and desired confidence level, N = the size of the population, π = the population proportion (assumed as 0.5 to obtain the maximum sample size), d = the degree of accuracy stated as a proportion (0.05). As a summary, this study has examined how representative a small sample could be as depending on two factors: the degree of accuracy and desired confidence level and revealed that for approximately $N = 341{,}000$ men teachers (the "men" category was investigated for constructing the total sample size because of being the smallest subgroup among the classroom teachers), the adequate minimum sample size as the representative of separate opinions the relevant population has been found as 271 ($\chi^2 = 2.706$) (NEA Research Bulletin, 1960: 99).

Additionally, more detailed information about sample size determination could be found in the recent studies by Verma and Verma (2020), Bolarinwa (2020) and Abebe (2022). For the interested readers, more specifically Hsieh et al. (1998) focus on the sample-size determination in the regression context for linear and logistic regression; Bonett & Wright (2000) handle the sample size determination for the confidence interval estimation regarding the commonly used Pearson, Kendall tau- α and Spearman correlations in behavioral researches.

5. GENERAL ASSESSMENT

The correct determination of the required sample size at the onset of any study is of crucial importance in terms of obtaining solid inferences at the end of the study. This study has been prepared with the aim of being a concise guide that brings various formulations concerning the determination of the required sample size together for researchers by giving both the criteria affecting the sample size and significant aspects of the sampling designs. Among the sampling designs, probability sampling methods stand out with their representativeness characteristic and thus regarded as a gold standard of the sampling designs, which also means obtaining a greater level of accuracy for inferences compared with the non-probability sampling designs. Considering the importance of probabilistic sampling methods in this regard, besides the formulations for sample

size determination in these methods, the present paper is an attempt to review the sample-size calculation formulations in survey research types and regression analyses as well.

REFERENCES

Abebe, H. T. (2022). "Determination of Sample Size and Errors", *Promoting Statistical Practice and Collaboration in Developing Countries,* (Ed: O. O. Awe, K. Love and E. A. Vance), Chapman and Hall/CRC.

Adams, J., Khan, H. T., Raeside, R. and White, D. I. (2007). *Research Methods for Graduate Business and Social Science Students*, Response Books (SAGE publications), New Delhi.

Altman, D. G. and Bland, J. M. (2014). "Uncertainty and Sampling Error", *BMJ*, 349.

Bolarinwa, O. A. (2020). "Sample Size Estimation for Health and Social Science Researchers: The Principles and Considerations for Different Study Designs", *Nigerian Postgraduate Medical Journal*, 27/2, 67.

Bolfarine, H. and Bussab, W.O. (2005). *Elementos de Amostragem*, Edgard Blücher, São Paulo.

Bonett, D. G. and Wright, T. A. (2000). "Sample Size Requirements for Estimating Pearson, Kendall and Spearman Correlations", *Psychometrika*, 65/1, 23-28.

Chasalow, K. and Levy, K. (2021, March). "Representativeness in Statistics, Politics, and Machine Learning", *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 77-89.

Chiu, S. and Tavella, D. (2008). *Data Mining and Market Intelligence for Optimal Marketing Returns*, Butterworth-Heinemann.

Cohen, J. (1962). "The Statistical Power of Abnormal-Social Psychological Research: A Review", *The Journal of Abnormal and Social Psychology*, 65(3), 145-153.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, Second Edition, Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ.

Cohen, J. (1992). "A Power Primer", *Psychological Bulletin*, 112(1), 155-159.

Cohen, B. H. (2001). *Explaining Psychological Statistics*, Second Edition, John Wiley & Sons.

Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Houghton Mifflin, Boston.

Dattalo, P. (2008). *Determining Sample Size: Balancing Power, Precision, and Practicality*, Oxford University Press.

Dattalo, P. (2009). "A Review of Software for Sample Size Determination", *Evaluation & the Health Professions*, 32(3), 229-248.

Erdfelder, E., Faul, F. and Buchner, A. (1996). "GPOWER: A General Power Analysis Program", *Behavior Research Methods, Instruments, & Computers*, 28(1), 1-11.

Erdfelder, E., Faul, F. and Buchner, A. (2005). "Power Analysis for Categorical Methods", *Encyclopedia of Statistics in Behavioral Science*, (Eds: B. S. Everitt and D. C. Howell), John Wiley & Sons, Chichester, UK.

Farmer, A. Y. and Farmer, G. L. (2020). *Research Methods for Social Work: A Problem-Based Approach*, SAGE Publications.

Faul, F., Erdfelder, E., Lang, A. G. and 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.

Faul, F., Erdfelder, E., Buchner, A. and 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.

Favero, L. and Belfiore, P. (2019). *Data science for Business and Decision Making*, Academic Press.

Findley, M. G., Kikuta, K. and Denly, M. (2021). "External Validity", *Annual Review of Political Science*, 24, 365-393.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh, UK.

Fleiss, J. L., Levin, B. and Paik, M. C. (2004). *Statistical Methods for Rates and Proportions*, 3rd Edition, Wiley-Interscience, Hoboken, NJ.

Fraenkel, J. R. and Wallen, N. E. (2009). *How to Design and Evaluate Research in Education*, McGraw-Hill, New York.

Greenstein, T. N. and Davis, S. N. (2013). *Methods of Family Research*, 3rd Edition, Sage Publications.

Hazra, A. and Gogtay, N. (2016). "Biostatistics Series Module 5: Determining Sample Size", *Indian journal of Dermatology*, 61(5), 496.

Hedrih, V. and Hedrih, A. (2022). *Interpreting Statistics for Beginners: A Guide for Behavioural and Social Scientists*, Routledge, New York.

Hickey, G. L., Grant, S. W., Dunning, J. And Siepe, M. (2018). "Statistical Primer: Sample Size and Power Calculations—Why, When and How?", *European Journal of Cardio-Thoracic Surgery*, 54(1), 4-9.

Hirji, K. F. (2006). *Exact Analysis of Discrete Data*, Chapman and Hall/CRC.

Hsieh, F. Y., Bloch, D. A. and Larsen, M. D. (1998). "A Simple Method of Sample Size Calculation for Linear and Logistic Regression", *Statistics in Medicine*, 17(14), 1623-1634.

Johnson, G. (2002). *Research Methods for Public Administrators*, Quorum Books, Westport CT.

Kang, H. (2021). "Sample Size Determination and Power Analysis Using the G*Power Software", *Journal of Educational Evaluation for Health Professions*, 18.

Koricheva, J., Gurevitch, J. and Mengersen, K. (Eds.). (2013). *Handbook of Meta-Analysis in Ecology and Evolution*, Princeton University Press.

Krejcie, R. V. And Morgan, D. W. (1970). "Determining Sample Size for Research Activities", *Educational and Psychological Measurement*, 30(3), 607-610.

Loue, S. (2002). *Gender, Ethnicity, and Health Research*, Kluwer Academic Publishers, New York.

Medhi, J. (1992). *Statistical Methods: An Introductory Text*, New Age International Limited Publishers, New Delhi, India.

Miaoulis, G. and Michener, R. D. (1976). *An Introduction to Sampling*, Kendall/Hunt Publishing Company, Dubuque, Iowa.

Moule, P. (2020). *Making Sense of Research in Nursing, Health and Social Care*, Sage.

Neyman, J. and Pearson, E. S. (1928). "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I", *Biometrika*, 175-240.

Paternoster, R. and Bachman, R. D. (2017). *Essentials of Statistics for Criminology and Criminal Justice*, Sage Publications.

Polit, D. F. and Beck, C. T. (2010). "Generalization in Quantitative and Qualitative Research: Myths and Strategies", *International Journal of Nursing Studies*, 47(11), 1451-1458.

Riggio, R. (2016). *Introduction to Industrial and Organizational Psychology*, Sixth Edition, Routledge, New York.

Rothman, K. J. (1986). *Modern Epidemiology*, Little, Brown and Company, Boston.

Serdar, C. C., Cihan, M., Yücel, D. and Serdar, M. A. (2021). "Sample Size, Power and Effect Size Revisited: Simplified and Practical Approaches in Pre-Clinical, Clinical and Laboratory Studies", *Biochemia Medica*, 31(1), 27-53.

Small-Sample Techniques (1960). *The NEA Research Bulletin*, Vol. 38, p. 99.

Suresh, K. P. and Chandrashekara, S. (2012). "Sample Size Estimation and Power Analysis for Clinical Research Studies", *Journal of Human Reproductive Sciences*, *5*(1), 7-13.

Suter, W. N. (2012). *Introduction to Educational Research: A Critical Thinking Approach*, Second Edition, SAGE Publications.

Tienken, C. H. and Mullen, C. A. (Eds.). (2016). *Education Policy Perils: Tackling the Tough Issues*, Routledge & Kappa Delta Pi, New York

Verma, J. P. (2019). *Statistics and Research Methods in Psychology with Excel*, Springer Singapore.

Verma, J. P. and Verma, P. (2020). *Determining Sample Size and Power in Research Studies*, Springer Singapore.

Beyan ve Açıklamalar (Disclosure Statements)

1. Bu çalışmanın yazarları, araştırma ve yayın etiği ilkelerine uyduklarını kabul etmektedirler (The authors of this article confirm that their work complies with the principles of research and publication ethics).

2. Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir (No potential conflict of interest was reported by the authors).

3. Bu çalışma, intihal tarama programı kullanılarak intihal taramasından geçirilmiştir (This article was screened for potential plagiarism using a plagiarism screening program).