International Journal of Business and Management Invention (IJBMI)
ISSN (Online): 2319 – 8028, ISSN (Print): 2319 – 801X
www.ijbmi.org || Volume 13 Issue 7 || July, 2024 || PP: 152-167

Selecting the Right Sample Size: Methods and Considerations for Social Science Researchers

Sathyanarayana S	T. Mohanasundaram,	Pushpa B V	Hema Harsha
Professor	*Associate Professor,*	*Associate Professor,*	*Associate Professor,*
MPBIM	*MSRIT*	*MPBIM*	*MPBIM*
Bengaluru.	*Bengaluru.*	*Bengaluru.*	*Bengaluru.*

ABSTRACT

The aim of this study is to provide a comprehensive framework of various sampling techniques utilized in social science research. Sampling is a critical step in research design, influencing the accuracy and reliability of study findings. This article covers essential methodologies including estimating a population proportion (single proportion), estimating a population mean, estimating the difference between two population means, and estimating the difference between two population proportions. Additionally, it delves into sample size determination methods, highlighting Cochran's formula for survey research, Nunnally's formula for scale development, Yamane's formula, and Krejcie and Morgan's table. The concept of confidence intervals and confidence levels is thoroughly explored, elucidating their significance in inferential statistics. By examining how confidence intervals work, the study emphasizes the importance of precision and reliability in research estimates. The article also addresses critical sampling considerations that researchers must account for to ensure robust and valid results. The findings provide a detailed comparison of these methodologies, offering insights into their applicability and limitations in various research scenarios. This guide serves as a valuable resource for researchers, aiding them in selecting appropriate sampling techniques for their studies, thereby enhancing the quality and credibility of their research outcomes.

JEL Classification: C83, C10, C13, C18

Keywords: Sampling Techniques, Population Proportion, Population Mean, Confidence Intervals, Sample Size Determination

---

Date of Submission: 15-07-2024 Date of acceptance: 31-07-2024

I. INTRODUCTION

Sampling is a fundamental component of social science research, serving as a cornerstone for drawing reliable conclusions about populations. By selecting representative samples, researchers can generalize findings to broader contexts (Trochim and Donnelly (2008)). This process is essential for ensuring the generalizability of research findings and enhancing the external validity of studies. Moreover, sampling enables researchers to optimize resource allocation, minimizing costs and time requirements while still obtaining meaningful results (Babbie (2020)). Through well-designed sampling techniques, such as probability sampling, researchers can achieve accuracy and precision in estimating population parameters (Fowler (2013)). This approach helps mitigate bias and ensures that each member of the population has an equal chance of being included in the sample. Additionally, sampling addresses ethical considerations by minimizing the burden on participants and protecting their rights and welfare (Bryman (2016)). By selecting representative subsets and anonymizing data, researchers can uphold confidentiality and privacy standards. Therefore, sampling plays a vital role in social science research, facilitating generalizability, resource efficiency, accuracy, precision, and ethical considerations, thus contributing to the integrity and validity of research findings.

The sample size in research significantly influences the power, precision, and generalizability of the study's findings. Understanding these implications is crucial for designing robust studies and interpreting results accurately. Power refers to the probability of detecting a true effect when it exists. A larger sample size increases the statistical power of a study, making it more likely to identify significant effects. Smaller sample sizes can result in underpowered studies, increasing the risk of Type II errors (Cohen (1992); Maxwell, Kelley, and Rausch (2008)). Precision refers to the degree to which repeated measurements under unchanged conditions show the same results. Larger sample sizes reduce the standard error of the estimate, leading to narrower confidence intervals and more precise estimates of population parameters (Kish (1965); Hair et al. (2010)). Generalizability is the extent to which the findings of a study can be applied to the broader population. A larger and more representative sample size improves the external validity of the study, ensuring that the results are applicable to a wider population (Babbie (2016); Henrich, Heine, and Norenzayan (2010)).

Sampling in the context of social sciences is a fundamental methodological approach used to gather data from a subset of a larger population to draw conclusions about the whole population (Handwerker (2005)). In social sciences, the population often consists of people or groups with diverse characteristics, behaviours, and opinions, making it impractical or impossible to study everyone. Sampling constitutes a critical phase in the research journey as it directly impacts the validity of conclusions drawn from gathered data. Whether conducting quantitative or qualitative inquiries, researchers face the pivotal tasks of determining appropriate sample sizes and devising effective sampling methodologies to ensure the robustness of their findings. Thus, sampling allows researchers to study a representative sample that reflects the characteristics of the entire population. Sampling involves selecting a subset of individuals or units from the population using various techniques such as random sampling, stratified sampling, cluster sampling, or convenience sampling. Drawing from the seminal works of Patton (1990) and Miles and Huberman (1994), Onwuegbuzie and Leech (2007) identified 24 sampling strategies applicable to both qualitative and quantitative research. These strategies are categorized into two main classes: random (probabilistic) sampling and non-random (non- probabilistic) sampling. Additionally, minimum sample sizes for various prevalent research designs are summarised. Whether in the social sciences, natural sciences, or business, sampling is essential for several reasons, ranging from practicality to statistical validity. Sampling is essential in social sciences and research in general for several reasons: (i) Practicality: In many cases, it is simply not feasible to study an entire population due to constraints such as time, resources, or accessibility. Sampling allows researchers to study a subset of the population, making research more manageable and cost-effective. (ii) By studying a representative sample, researchers can make inferences about the larger population with a certain level of confidence; (iii) Sampling provides a way to estimate population parameters (such as means, proportions, or correlations) with a degree of accuracy. Through statistical techniques, researchers can quantify the uncertainty associated with their estimates and assess the reliability of their findings; (iv) Sampling allows researchers to gather data more efficiently by focusing resources on the most relevant segments of the population, and (v) In many cases, it may not be ethically appropriate or feasible to study an entire population, especially if it involves sensitive topics or vulnerable populations. Sampling allows researchers to gather data in a way that respects ethical guidelines and protects the rights and well-being of participants. Therefore, sampling is essential because it enables researchers to study populations effectively, make valid inferences about them, and produce meaningful insights that contribute to our understanding of social phenomena and human behaviour.

Recent developments in sampling methodologies have seen a significant shift towards innovative and more efficient techniques, driven by advancements in technology and changes in research paradigms. One notable trend is the increasing utilization of digital platforms and online resources for sampling purposes, allowing researchers to reach broader and more diverse populations while minimizing logistical challenges. Additionally, there has been a growing emphasis on the integration of multiple sampling methods within a single study, such as combining probability-based sampling with non-probability approaches to improve representativeness and generalizability. Moreover, developments in machine learning and big data analytics have facilitated the exploration of novel sampling strategies, including adaptive and dynamic sampling methods that can adjust in real-time based on incoming data streams. These advancements not only offer opportunities for enhancing the rigor and efficiency of sampling in social science research but also present new avenues for addressing longstanding methodological challenges and advancing our understanding of complex phenomena in diverse populations.

The main purpose of this paper is to provide a comprehensive exploration of sampling methodologies and sample size determination in research. Through this paper, the researchers aim to explain various sampling techniques, including both probabilistic and non-probabilistic approaches, as well as different formulas employed for computing sample sizes across diverse research contexts.

II. CLASSIFICATION OF SAMPLING

Sampling can be classified into two main categories: random (probabilistic) sampling and non-random (non-probabilistic) sampling.

RANDOM (PROBABILISTIC) SAMPLING

Random sampling involves the selection of sample members from a population in such a way that each member has an equal chance of being chosen. This method relies on chance or probability, ensuring that the sample is representative of the population and minimizing bias. The following are some random or probabilistic sampling techniques: (i) Simple Random Sampling: Every member of the population has an equal chance of being selected. (ii) Systematic Sampling: Selecting every nth member from the population after a random starting point; (iii) Stratified Sampling: Dividing the population into homogeneous subgroups (strata) and then randomly selecting samples from each subgroup; (iv) Cluster Sampling: Dividing the population into clusters and then randomly selecting some clusters for inclusion in the sample; (v) Multi-stage Sampling: Combining two or more sampling techniques in sequence (e.g., stratified followed by cluster sampling); (vi) Probability Proportional to Size (PPS) Sampling: Probability of selection for each unit is proportional to its size or some measure of importance; (viii) Random Digit Dialing (RDD): Used in telephone surveys where respondents are selected randomly using random numbers generated by a computer, and (ix) Random Walk Sampling: A method used in spatial sampling where samples are taken at random locations within a defined area by moving randomly from a starting point. These techniques aim to ensure randomness and reduce bias in the selection of samples from a population.

NON-RANDOM (NON-PROBABILISTIC) SAMPLING

Non-random sampling, also known as non-probabilistic or purposive sampling, does not rely on random selection. Instead, sample members are chosen based on specific criteria or judgment of the researcher, aiming to include individuals who are most relevant to the research objectives. While non-random sampling methods may not ensure the same level of representativeness as random sampling, they are often practical when studying hard-to-reach populations or in qualitative research where depth of understanding is prioritized over statistical generalizability. Non-random (non-probabilistic) sampling techniques include: (i) Convenience Sampling: Selection of individuals who are easily accessible or readily available to the researcher; (ii) Purposive Sampling: Selection of individuals based on specific characteristics or criteria determined by the researcher's judgment or purpose of the study; (iii) Quota Sampling: Selection of individuals based on pre-defined quotas for certain characteristics such as age, gender, or occupation, ensuring representation of various groups in the sample; (iv) Snowball Sampling: Initial participants in the study refer or nominate additional participants, creating a chain-like sample; (v) Volunteer Sampling: Participants self-select into the study by responding to a call for volunteers; (vi) Judgmental Sampling: Selection of individuals based on the researcher's judgment or expertise in identifying relevant cases; (vii) Expert Sampling: Selection of individuals who are considered experts in the field under study, and (viii) Accidental Sampling: Selection of individuals who happen to be present at a particular place and time, without any deliberate effort to make the sample representative. These techniques are often used when it is impractical or impossible to use random sampling methods, but they may introduce bias into the sample and limit the generalizability of the results.

III. METHODS FOR DETERMINING SAMPLE SIZE

A. RULE OF THUMB METHODS: SIMPLE GUIDELINES FOR QUICK ESTIMATIONS

1. 10% RULE: The 10% rule suggests using 10% of the total population as a sample size. This rule provides a quick estimation method, particularly useful when detailed statistical calculations are impractical.

Example: If a researcher is studying a small town with a population of 5,000 people, the 10% rule would suggest a sample size of 500 individuals. This ensures that the sample is large enough to provide a good representation of the town's population. In a larger setting, such as a university with 30,000 students, applying the 10% rule would result in a sample size of 3,000 students. This helps to capture a diverse range of opinions and characteristics within the student body.

While the 10% rule is convenient, it is important to note that for very large populations, this rule might result in unnecessarily large samples. In such cases, more sophisticated methods, or guidelines, such as those provided by Cochran (1977), Krejcie and Morgan (1970), and Israel (1992), might be more appropriate to balance precision and practicality.

2. MINIMUM SAMPLE SIZES: GENERAL RECOMMENDATIONS

Surveys: For general surveys, a commonly recommended minimum sample size ranges from 100 to 200 respondents to ensure sufficient statistical power and representativeness (Krejcie and Morgan (1970); Bartlett, Kotrlik, and Higgins (2001); Fowler (2013)).

Example: A national survey studying consumer preferences for a new product might aim for a minimum of 200 respondents to capture diverse opinions and provide reliable estimates of preferences across different demographic groups.

3. EXPERIMENTAL STUDIES

In experimental research, minimum sample sizes can vary, but a common guideline is to have at least 30 participants per group to ensure the robustness of statistical comparisons (Roscoe (1975); Cohen (1992)).

Example: In a clinical trial comparing the efficacy of two medications, having at least 30 participants in each treatment group (totaling a minimum of 60 participants) helps to detect significant differences between the groups.

4. QUALITATIVE RESEARCH

For qualitative studies, smaller sample sizes are often acceptable, with recommendations ranging from 5 to 30 participants, depending on the research method and depth of inquiry (Guest, Bunce, and Johnson (2006); Creswell (2013)).

Example: A qualitative study exploring patient experiences with a new healthcare intervention might involve in-depth interviews with 15-20 participants to gather detailed and rich data.

5. STRUCTURAL EQUATION MODELLING (SEM)

For SEM, a minimum sample size of 200 is frequently recommended to ensure the stability of parameter estimates and the overall fit of the model (Jackson (2003); Wolf, Harrington, Clark, and Miller (2013); Kline (2015)).

Example: A study investigating the relationships between organizational culture, employee engagement, and job performance using SEM would aim for at least 200 participants to ensure reliable and valid model estimation.

B. STATISTICAL METHODS: MORE PRECISE METHODS BASED ON STATISTICAL CALCULATIONS

When determining sample size, more precise methods rely on statistical calculations that consider various factors such as effect size, power, significance level, and population variability. These methods ensure that the sample size is adequate to achieve reliable and valid results.

1. POWER ANALYSIS

Power analysis is a critical component in the design of scientific studies, especially when determining the necessary sample size. The primary goal is to ensure that the study is capable of detecting a true effect if it exists. This involves balancing the risks of Type I and Type II errors, alongside practical constraints like available resources. (Cohen, J. (1988); Maxwell, S. E., & Delaney, H. D. (2004); Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009)).

KEY CONCEPTS IN POWER ANALYSIS

Power (1 - β): The probability of correctly rejecting the null hypothesis when it is false. A power of 0.80 (80%) is commonly used, meaning there is an 80% chance of detecting an effect if it exists. The probability of rejecting the null hypothesis when it is true (Type I error). A common α level is 0.05. The magnitude of the difference or relationship the study aims to detect. Larger effect sizes generally require smaller sample sizes, and vice versa. The number of participants or observations required to achieve the desired power level. The standard deviation (σ) of the population. Higher variability necessitates larger sample sizes to detect the same effect size.

STEPS IN CONDUCTING POWER ANALYSIS

Specify the Hypotheses:

H0: Assumes no effect or difference.

H1: Assumes there is an effect or difference.

Choose the Significance Level (α): Typically set at 0.05, but can vary depending on the field and specific study requirements.

Estimate the Effect Size (d): Based on previous research, pilot studies, or theoretical expectations. Cohen's d is a common measure, where 0.2 is small, 0.5 is medium, and 0.8 is large.

Determine the Desired Power (1 - β): Commonly set at 0.80 or 0.90.

Select the Statistical Test: The choice depends on the study design (e.g., t-test, ANOVA, regression).

Calculate or Use Software for Sample Size:

Example: A researcher planning a study to detect a medium effect size (Cohen's d = 0.5) with 80% power at a 5% significance level would use power analysis to determine the necessary sample size. Software like G*Power can facilitate these calculations.

2. CONFIDENCE INTERVALS

are a key concept in statistics, providing a range of values that likely contain the true population parameter. They offer a measure of the uncertainty or precision of an estimate, such as a mean or proportion, and are widely used in various fields of research (Cochran, W. G. (1977); Israel, G. D. (1992)). The percentage of all possible samples that can be expected to include the true population parameter. Common confidence levels are 90%, 95%, and 99%.

Formula to compute confidence interval

$$CI = \bar{x} \pm (z \text{ x } \frac{\sigma}{\sqrt{n}})$$

where: $\bar{x}$ is the sample mean, Z is the Z-score corresponding to the desired confidence level, σ is the population standard deviation (or sample standard deviation, n is the sample size. (Gardner, M. J., & Altman, D. G. (1986); Cumming, G., & Finch, S. (2005); Moore, D. S., McCabe, G. P., & Craig, B. A. (2017)

STEPS TO CALCULATE A CONFIDENCE INTERVAL

Select the Confidence Level: Determine the level of confidence desired (e.g., 95%).

Find the Z-score or t-score: Depending on the sample size and whether the population standard deviation is known, use the Z-score for large samples ($n > 30$) or the t-score for smaller samples ($n \leq 30$).

Calculate the Margin of Error: Multiply the Z-score or t-score by the standard error of the mean (SEM):

$$\text{Margin of Error} = \text{Z x } \frac{\text{s}}{\sqrt{\text{n}}}$$

Determine the Confidence Interval: Add and subtract the margin of error from the sample mean:

$$\text{CI} = \bar{\text{x}} \ \pm \text{Margin of Error}$$

Scenario: You conducted a survey of 100 individuals to estimate the average height of adults in a city. The sample mean height is 170 cm, and the sample standard deviation is 10 cm. You want to calculate a 95% confidence interval.

Given: Confidence Level: 95%, Z-score for 95% confidence level: 1.96 (from Z-tables), Sample Mean ($\bar{x}$) = 170 cm, Sample Standard Deviation (s): 10 cm, Sample Size (n): 100.

Calculate the standard error of the mean (SEM):

$$SEM = \frac{s}{\sqrt{\text{n}}} = \frac{10}{\sqrt{10}} = 1$$

Calculate the margin of error:

Margin of Error=1.96×1=1.96

Determine the confidence interval:

CI = 170 $\pm$ 1.96 ⇒ (168.04,171.96)

So, the 95% confidence interval for the average height is 168.04 cm to 171.96 cm.

A 95% confidence interval means that if we were to take 100 different samples and compute a CI for each sample, we would expect about 95 of those intervals to contain the true population mean.

3. EFFECT SIZE CALCULATIONS

Effect size is a measure that quantifies the magnitude of a relationship or difference between groups in a study. It is essential in the context of sampling because it helps determine the practical significance of findings, guiding the design of studies, particularly in determining sample size and power analysis. Effect size is a standardized measure that indicates the size of an effect, independent of the sample size. It allows for comparison across different studies and is critical in meta-analysis. (Maxwell, S. E., & Delaney, H. D. (2004); Ellis, P. D. (2010); Sullivan, G. M., & Feinn, R. (2012); Funder, D. C., & Ozer, D. J. (2019); Ellis, P. D. (2019); Lakens, D. (2021)).

TYPES OF EFFECT SIZES

(i) Cohen's d: Used for the difference between two means. (ii) Pearson's r: Used for the correlation between two variables. (iiii) Odds Ratio: Used for the association between binary variables. (iv) Eta Squared (η^2): Used for the proportion of variance explained in ANOVA.

Example: In a psychological study measuring the impact of a new therapy, researchers may estimate a small effect size (Cohen's d = 0.2) and calculate the required sample size to ensure the study is adequately powered to detect this effect.

STEPS TO CALCULATE EFFECT SIZE

Identify the Type of Effect Size Needed: Decide based on the study design and the nature of the data (e.g., difference between means, correlation, etc.).

Calculate the Effect Size: Cohen's d

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_p}$$

where $\bar{X}_1$and $\bar{X}_2$ are the sample means, and s_p is the pooled standard deviation.

Pearson's r: $r = \frac{\Sigma\ (\text{x}-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(\text{x}-\bar{\text{x}})^2}\Sigma\sqrt{(\text{y}-\bar{\text{y}})^2}}$

Odds Ratio (OR): $OR = \frac{(a/c)}{(b/d)}$

where a, b, c, and d are the counts of the outcomes in a 2x2 table.

Eta Squared (η^2) $\eta^2 = \frac{SS_{effect}}{SS_{total}}$

where SS_{effect} is the sum of squares for the effect and SS_{total} is the total sum of squares.

Interpret the Effect Size

Cohen's d: Small (0.2), Medium (0.5), Large (0.8)

Pearson's r: Small (0.1), Medium (0.3), Large (0.5)

Odds Ratio: OR > 1 indicates a positive association, OR < 1 indicates a negative association.

Eta Squared: Small (0.01), Medium (0.06), Large (0.14)

Example Calculation: Cohen's d

Scenario: Comparing test scores between two groups of students using a new teaching method vs. a traditional method.

Group 1 Mean ($\bar{X}_1$): 75

Group 2 Mean ($\bar{X}_2$): 70

Pooled Standard Deviation (s_p): 8

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_p}$$

$$d = \frac{75 - 70}{8} = 0.625$$

Interpretation: Cohen's d of 0.625 indicates a medium to large effect size, suggesting a moderate practical significance of the new teaching method over the traditional one.

FORMULA TO ESTIMATE THE SAMPLE SIZE

Determining the appropriate sample size is a crucial step in the design of any social science research study. A carefully chosen sample size ensures that the study has adequate statistical power to detect meaningful effects, while also balancing practical considerations such as time, resources, and feasibility. In social science research, where the goal is often to draw conclusions about populations based on data collected from a subset of individuals, standardized formulas are commonly employed to estimate the required sample size. These formulas take into account various factors such as the desired level of precision, the expected variability within the population, the significance level, and the desired power of the study. There are several standardized formulas commonly used to determine sample size for social science research. The choice of formula depends on the specific research design, the type of data being collected, and the statistical analysis planned for the study. The following are some commonly used formulae for different types of studies in social science research:

1. ESTIMATING A POPULATION PROPORTION (SINGLE PROPORTION)

Estimating a population proportion, also known as a single proportion, is a common task in research when we want to infer the proportion of a specific characteristic or attribute within a population based on data collected from a sample. This estimation is crucial in various fields, including social sciences, public health, market research, and quality control. The formula used to estimate a population proportion (p) from a sample proportion ($\hat{P}$) is:

$$n = \frac{z^2.\hat{P}.(1 - \hat{P})}{E^2}$$

Where:

n is the required sample size.

Z is the critical value from the standard normal distribution corresponding to the desired confidence level (e.g., 1.96 for a 95% confidence level).

$\hat{P}$= is the estimated proportion from a pilot study or previous research.

E is the desired margin of error or precision.

This formula provides an estimate of the sample size needed to estimate the population proportion with a specified level of confidence and precision. A larger sample size (n) generally leads to a smaller margin of error (E), providing more precise estimates (Kothari, C.R. (2004)).

Example 1: A researcher is conducting a survey to estimate the proportion of adults in a city who support a proposed environmental policy. He wants to estimate this proportion with a 95% confidence level and a margin of error of 3%. Based on previous similar surveys or pilot studies, the researcher estimates that approximately 60% of adults in the city support the policy ($\hat{P}$=0.60). Using the formula for estimating sample size for a population proportion:

$$n = \frac{z^2.\hat{P}.(1 - \hat{P})}{E^2}$$

where Z = 1.96 (corresponding to the 95% confidence level); $\hat{P}$=0.60; E= 0.03 (3% margin of error)

$$n = \frac{1.96^2.0.60.(1 - 0.60)}{0.03^2}$$

$$n \approx 1024.48$$

Rounding up to the nearest whole number (since one can't have a fraction of a person in the sample), researcher would need a sample size of approximately 1025 adults from the city to estimate the population proportion of support for the policy with a 95% confidence level and a margin of error of 3%.

Example 2: A researcher is conducting a study to estimate the proportion of smartphone users who prefer a particular brand in a large metropolitan area. He wants to estimate this proportion with a 99% confidence level and a margin of error of 2%. Unfortunately, the researcher does not have any prior information or pilot study results to estimate the proportion of brand preference ($\hat{P}$). In such cases, researchers often use a conservative estimate of 50% ($\hat{P}$=0.50) to maximize the required sample size, assuming that this proportion will result in the largest sample size needed.

Using the formula for estimating sample size for a population proportion:

$$n = \frac{z^2 . \hat{P} . (1 - \hat{P})}{E^2}$$

where Z = 2.576 (corresponding to the 99% confidence level); $\hat{P}$=0.50; E= 0.02 (2% margin of error)

$$n = \frac{2.576^2 . 0.50 . (1 - 0.50)}{0.02^2}$$

$$n \approx 4147.36$$

Rounding up to the nearest whole number, the researcher would need a sample size of approximately 4148 smartphone users from the metropolitan area to estimate the population proportion of brand preference with a 99% confidence level and a margin of error of 2%.

Kothari, C. R. (2004). Research Methodology: Methods and Techniques. New Age International.

2. Estimating a population mean

It is also known as a single mean, is a common objective in research when we want to infer the average value of a continuous variable within a population based on data collected from a sample. To compute the ideal sample size for estimating a population mean with a specified level of confidence and precision, we typically use the formula for sample size estimation. The formula used to estimate the required sample size for estimating a population mean (μ) is based on the standard normal distribution and is given by:

$$n = \frac{z^2 . \sigma^2}{E^2}$$

Where:

n is the required sample size.

Z is the critical value from the standard normal distribution corresponding to the desired confidence level. For example, for a 95% confidence level,

Z is approximately 1.96.

σ is the population standard deviation (if known).

E is the desired margin of error or precision.

This formula provides an estimate of the sample size needed to estimate the population mean with a specified level of confidence and precision. A larger sample size generally leads to a smaller margin of error, providing more precise estimates of the population mean. It is important to note that if the population standard deviation (σ) is unknown, researchers often use the sample standard deviation (s) as an estimate. In such cases, the formula for sample size estimation becomes:

$$n = \frac{z^2 . s^2}{E^2}$$

Where:

s is the sample standard deviation.

This adjusted formula allows researchers to estimate the required sample size based on the variability observed in the sample when the population standard deviation is unknown.

Let us consider a scenario where a researcher is conducting a study to estimate the average weekly income of employees in a certain industry. The researcher wants to estimate this population mean with a 95% confidence level and a margin of error of $50. Suppose the researcher has some prior information or knowledge suggesting that the population standard deviation of weekly income in this industry is $400.

Using the formula for estimating sample size for a population mean:

$$n = \frac{z^2 . \sigma^2}{E^2}$$

where Z = 1.96 (corresponding to the 95% confidence level); σ= $400 (population standard deviation); E= $50 (margin of error)

$$n = \frac{z1.96^2 . 400^2}{50^2}$$

$$n = 245.8624$$

Rounding up to the nearest whole number, the researcher would need a sample size of approximately 246 employees to estimate the average weekly income of employees in the industry with a 95% confidence level and a margin of error of $50.

Example 2: Let us consider a more complex example where a researcher is conducting a study to estimate the average blood pressure of adults in a city. He wants to estimate this population mean with a 99% confidence level and a margin of error of 5 mmHg. Compute the sample size for the study.
Solution: Since the researcher does not have any prior information or knowledge about the population standard deviation (σ). In such cases, researchers often use a conservative estimate or conduct a pilot study to estimate σ. Let us assume the researcher has conducted a pilot study and found that the sample standard deviation (s) of blood pressure measurements was 15 mmHg. Using the formula for estimating sample size for a population mean when the population standard deviation is unknown:

$$n = \frac{z^2 . s^2}{E^2}$$

where Z = 2.576 (corresponding to the 99% confidence level); s =15(sample standard deviation); E= 5 (margin of error)

$$n = \frac{z2.576^2 . 15^2}{5^2}$$

$$n = 59.8019$$

Rounding up to the nearest whole number, the researcher would need a sample size of approximately 60 adults to estimate the average blood pressure of adults in the city with a 99% confidence level and a margin of error of 5 mmHg, assuming a sample standard deviation of 15 mmHg.

ESTIMATING THE DIFFERENCE BETWEEN TWO POPULATION MEANS

Estimating the difference between two population means is a statistical procedure used when researchers want to compare the means of a continuous variable in two different populations or groups. This method is commonly employed in research studies to investigate whether there is a significant difference in the average values of certain characteristics, outcomes, or measures between two populations or groups. The equation used to estimate the required sample size for comparing two population means is based on the difference between the means *(μ_1−μ_2)* and is given by:

$$n = \frac{2(Z_{\frac{\alpha}{2}} + Z_\beta)^2 . \sigma^2}{(\mu_1 - \mu_2)^2}$$

Where:
n is the required sample size for each group.
$Z_{\frac{\alpha}{2}}$ is the critical value from the standard normal distribution corresponding to the desired significance level (α/2).
Z_β is the critical value from the standard normal distribution corresponding to the desired statistical power (1−β).
σ is the common standard deviation of the populations.
$\mu_1 - \mu_2$ is the difference in the means of the populations or groups.
This formula provides an estimate of the sample size needed in each group to detect a difference between the means with a specified level of significance and statistical power. A larger sample size generally leads to a smaller margin of error, providing more precise estimates of the difference between the population means. It is important to note that researchers typically use prior information, pilot studies, or literature reviews to estimate the common standard deviation (σ) and select appropriate values for the significance level (α), statistical power (1−β), and desired difference in means ($\mu_1 - \mu_2$) based on the research objectives and practical considerations.
Example 1: Let us consider an example where a researcher wants to compare the effectiveness of two different teaching methods, Method A and Method B, in improving students' test scores. He wants to estimate whether there is a significant difference in the average test scores between the two methods with a significance level of 0.05 and a statistical power of 0.80. Suppose you conducted a pilot study or reviewed previous research and found that the common standard deviation of test scores for both methods is 10 points (σ=10). He wants to detect a difference of at least 5 points ($\mu_1 - \mu_2 = 5$) between the average test scores of the two methods.
Solution
Using the formula for estimating sample size for comparing two population means:

$$n = \frac{2(Z_{\frac{\alpha}{2}} + Z_\beta)^2 . \sigma^2}{(\mu_1 - \mu_2)^2}$$

where:
$Z_{\frac{\alpha}{2}}$ =1.96 (corresponding to a significance level of 0.05)
Z_β=0.84 (corresponding to a statistical power of 0.80)
σ =10 (common standard deviation)
($\mu_1 - \mu_2$)=5 (difference in means)

$$n = \frac{2(1.96 + 0.84)^2 . 10^2}{(5)^2}$$
$$n = 62.72$$

Rounding up to the nearest whole number, the researcher would need a sample size of approximately 63 students for each teaching method group to compare the effectiveness of Method A and Method B in improving test scores with a significance level of 0.05 and a statistical power of 0.80, assuming a common standard deviation of 10 points and a difference in means of 5 points.

Example 2: Let us consider a more complex example involving the comparison of the effectiveness of two different medications, Medication A and Medication B, in reducing blood pressure for patients with hypertension. A researcher wants to estimate whether there is a significant difference in the average reduction in blood pressure between the two medications with a significance level of 0.01 and a statistical power of 0.90. Suppose he has conducted a pilot study or reviewed previous research and found that the common standard deviation of blood pressure reduction for both medications is 12 mmHg (σ=12). He wants to detect a difference of at least 8 mmHg ($\mu_1 - \mu_2 = 8$) in the average reduction in blood pressure between the two medications.

Solution: Using the formula for estimating sample size for comparing two population means:

$$n = \frac{2(Z_{\frac{\alpha}{2}} + Z_\beta)^2 . \sigma^2}{(\mu_1 - \mu_2)^2}$$

where:

$Z_{\frac{\alpha}{2}}$ =2.576 (corresponding to a significance level of 0.01)

Z_β=1.282 (corresponding to a statistical power of 0.90)

σ =12 (common standard deviation)

$(\mu_1 - \mu_2)$=8 (difference in means)

$$n = \frac{2(2.576 + 1.282)^2 . 12^2}{(8)^2}$$
$$n \approx 46.5288$$

Rounding up to the nearest whole number, the researcher would need a sample size of approximately 47 patients for each medication group to compare the effectiveness of Medication A and Medication B in reducing blood pressure with a significance level of 0.01 and a statistical power of 0.90, assuming a common standard deviation of 12 mmHg and a difference in means of 8 mmHg.

ESTIMATING THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS

Estimating the difference between two population proportions is a statistical procedure used when researchers want to compare the proportions of two categorical variables in different populations or groups. This method is commonly employed in research studies to investigate whether there is a significant difference in the proportions of certain characteristics, behaviours, or preferences between two populations or groups.

The equation used to estimate the required sample size for comparing two population proportions is based on the difference between the proportions ($P_1 - P_2$) and is given by:

$$n = \frac{z^2 . (p_1 . (1 - p_1) + p_{2.} (1 - p_2)}{E^2}$$

Where: n is the required sample size for each group.

Z is the critical value from the standard normal distribution corresponding to the desired confidence level.

P_1 and P_2 are the estimated proportions in each population or group.

E is the desired margin of error or precision.

This formula provides an estimate of the sample size needed in each group to detect a difference between the proportions with a specified level of confidence and precision. A larger sample size generally leads to a smaller margin of error, providing more precise estimates of the difference between the population proportions. It is important to note that researchers typically use prior information, pilot studies, or literature reviews to estimate the proportions (*P_1 and P_2*) and select an appropriate margin of error (E) based on the research objectives and practical considerations.

Example 1: Let us consider an example where a researcher wants to compare the proportions of adults who own smartphones in two different cities, City A and City B. He wants to estimate whether there is a significant difference in smartphone ownership between the two cities with a 95% confidence level and a margin of error of 3%. Suppose the researcher has conducted a pilot study or reviewed previous research and found that approximately 60% of adults in City A own smartphones (p_1 =0.60) and 55% of adults in City B own smartphones (p_2=0.55).

Solution: Using the formula for estimating sample size for comparing two population proportions:

$$n = \frac{z^2.(p_1.(1-p_1)+p_{2.}(1-p_2)}{E^2}$$

Where, Z = 1.96 (corresponding to the 95% confidence level); *p₁* = 0.60(proportion in City A); *p₂* = 0.55 (proportion in City B); E= 0.03 (3%margin of error)

$$n = \frac{1.96^2.(0.60.(1-0.60)+0.55(1-0.55)}{0.03^2}$$

$$n = 2079.31$$

Rounding up to the nearest whole number, the researcher would need a sample size of approximately 2080 adults in each city to compare the proportions of smartphone ownership between City A and City B with a 95% confidence level and a margin of error of 3%.

Example: Let us consider a more complex example where the researcher wants to compare the proportions of patients who respond positively to two different treatments for a particular medical condition. The researcher wants to estimate whether there is a significant difference in treatment response rates between the two treatments with a 99% confidence level and a margin of error of 2%. Suppose you conducted a pilot study or reviewed previous research and found that approximately 70% of patients respond positively to Treatment A (p_1=0.70) and 65% of patients respond positively to Treatment B (p_2=0.65).

Solution

Using the formula for estimating sample size for comparing two population proportions:

$$n = \frac{z^2.(p_1.(1-p_1)+p_{2.}(1-p_2)}{E^2}$$

Where, Z = 2.576 (corresponding to the 99% confidence level); p_1 = 0.70(proportion for treatment A); p_2 = 0.65 (proportion of treatment B); E= 0.02 (2%margin of error)

$$n = \frac{2.576^2.(0.70.(1-0.70)+0.65.(1-0.65)}{0.02^2}$$

$$n = 7248.1025$$

Rounding up to the nearest whole number, the researcher would need a sample size of approximately 7249 patients for each treatment group to compare the proportions of positive treatment responses between Treatment A and Treatment B with a 99% confidence level and a margin of error of 2%.

FOR SAMPLE SIZE IN SURVEY RESEARCH

COCHRAN'S FORMULA

Cochran's formula is a widely used method to determine the appropriate sample size needed in survey research to ensure the results are statistically reliable and representative of the population being studied. This formula is particularly useful when the population size is large and the researcher aims to achieve a desired level of precision in estimating population proportions. The formula for Cochran's sample size (n) is given by:

$$n_0 = \frac{z^2.p.(1-p)}{E^2}$$

Where: n is the required sample size, Z is the critical value from the standard normal distribution corresponding to the desired level of confidence, p is the estimated proportion of the population with the attribute of interest (or the maximum expected proportion if no prior estimate is available), E is the desired margin of error or precision. Cochran's formula allows researchers to determine the minimum sample size required to estimate a population proportion with a specified level of confidence and precision. A larger sample size leads to a smaller margin of error, providing more precise estimates of the population proportion. It is important to note that Cochran's formula assumes simple random sampling from an infinite population or a large finite population with replacement. If the population size is relatively small and sampling without replacement is conducted, adjustments to the formula may be necessary using finite population correction factors.

Example 1: A researcher is conducting a survey to estimate the proportion of adults in a city who supports a proposed environmental policy. He wants to estimate this proportion with a 95% confidence level and a margin of error of 3%. Suppose he does not have any prior information about the proportion of adults who support the policy in the city. In such cases, he can use a conservative estimate of 0.50 (50%) for p, assuming that the support for the policy is equally likely as not.

Using Cochran's formula:

$$n_0 = \frac{z^2.p.(1-p)}{E^2}$$

Where: Z=1.96 (corresponding to the 95% confidence level); p=0.50 (conservative estimate) E=0.03 (margin of error of 3%)

$$n_0 = \frac{1.96^2.0.5.(1-0.5)}{0.03^2}$$

$$n \approx 1067.11$$

Rounding up to the nearest whole number, the researcher would need a sample size of approximately 1068 adults in the city to estimate the proportion of support for the environmental policy with a 95% confidence level and a margin of error of 3%.

Example 2: Let us consider a more complex example where a researcher is conducting a survey to estimate the proportion of customers who are satisfied with a new product launched by a company. He wants to estimate this proportion with a 99% confidence level and a margin of error of 2%. Suppose the company has a large customer base, and he estimates that approximately 70% of customers are satisfied with the new product (p=0.70). However, due to the importance of this survey and the desire for high precision, he wants to ensure a very small margin of error.

$$n_0 = \frac{z^2.p.(1-p)}{E^2}$$

Where: Z=2.576 (corresponding to the 99% confidence level); p=0.70 (conservative estimate) E=0.02 (margin of error of 2%)

$$n_0 = \frac{2.576^2.0.7.(1-0.7)}{0.02^2}$$

$$n \approx 5163.45$$

Rounding up to the nearest whole number, the researcher would need a sample size of approximately 5164 customers to estimate the proportion of satisfaction with the new product with a 99% confidence level and a margin of error of 2%.

When the population size is unknown and you are estimating the population mean, Cochran's formula can be adapted for this scenario. The formula is:

$$n_0 = \frac{z^2.\sigma^2}{e^2}$$

where: n_0 = the initial sample size, z = z-score (the number of standard deviations from the mean corresponding to the desired confidence level, e.g., 1.96 for 95% confidence), σ = estimated standard deviation of the population, e = desired margin of error (expressed as a decimal)

Example: Suppose you want to determine the sample size for a survey where the population mean is estimated, with an estimated standard deviation σ of 10, a margin of error e of 2, and a confidence level of 95% (z = 1.96).

Calculate the initial sample size

$$n_0 = \frac{z^2.\sigma^2}{e^2}$$

$$n_0 = \frac{1.96^2.10^2}{2^2}$$

$$n_0 = 96.04$$

So, the required sample size would be approximately 96.

THE SAMUEL B. GREEN FORMULA

The Samuel B. Green formula (Green, 1991) is a heuristic used in multiple regression analysis to estimate the minimum sample size needed. Green proposed two different rules of thumb depending on whether you are interested in testing individual predictors or the overall model.

For Testing Individual Predictors:

$$n \geq 104 + k$$

where n is the sample size and k is the number of predictors.

For Testing the Overall Model:

$$\mathrm{n} \geq 50 + 8\mathrm{k}$$

where n is the sample size and k is the number of predictors.

Example: Suppose you have 5 predictors:

1. **For Testing Individual Predictors:** n ≥ 104 + 5 = 109
2. **For Testing the Overall Model:** n ≥ 50 + 8 × 5 = 50 + 40 = 90

In this case, the researcher would choose the larger of the two estimates to ensure you have enough power to test both individual predictors and the overall model, so you would need a sample size of at least 109.

NUNNALLY'S FORMULA (FOR SAMPLE SIZE IN SCALE DEVELOPMENT)

Nunnally's Formula is used in scale development to determine the sample size needed for conducting psychometric analyses, such as factor analysis, to assess the reliability and validity of a measurement instrument. It helps researchers ensure they have a sufficiently large sample to yield reliable results. The formula is based on the desired level of reliability (usually expressed as Cronbach's alpha) and the number of items in the scale. It assumes a normal distribution of responses and provides an estimate of the minimum sample size required to achieve a specified level of reliability.

$$n = \frac{k}{1 - r}$$

Where: n = the required sample size, k = the number of items in the scale, r = the desired reliability coefficient (typically Cronbach's alpha). In this formula, r represents the expected reliability of the scale. Researchers often set a threshold for the acceptable level of reliability, such as 0.70 or 0.80.

For example, if a researcher has a scale with 20 items (k = 20) and you want to achieve a reliability of at least 0.80 (r = 0.80), you can use the formula to calculate the required sample size:

$$n = \frac{20}{1 - 0.80}$$

$$n = 100$$

So, according to Nunnally's Formula, the researcher would need a minimum sample size of 100 participants to achieve a reliability of at least 0.80 for the scale with 20 items.

Example 2: A researcher is developing a new scale to measure job satisfaction among employees in a large multinational company. His scale consists of 30 items designed to capture various aspects of job satisfaction. He aims to achieve a reliability coefficient (Cronbach's alpha) of at least 0.85 for the scale. Using Nunnally's Formula, determine the minimum sample size required for the study.

Solution: Number of items in the scale (k) = 30; Desired reliability coefficient (r) = 0.85

$$n = \frac{30}{1 - 0.85}$$

$$n = 200$$

So, according to Nunnally's Formula, the researcher would need a minimum sample size of 200 participants to achieve a reliability of at least 0.85 for the job satisfaction scale with 30 items.

YAMANE'S FORMULA

Yamane's formula is a straightforward method used for determining sample sizes in survey research, particularly in social science studies where the population size is known or easily determinable. It is commonly used in situations where researchers want to obtain a representative sample from a large population. The formula was proposed by S. Yamane in his book "Statistics: An Introductory Analysis" (1967).

$$n = \frac{N}{1 + Ne^2}$$

Where: n= sample size, N= population size, e = margin of error (expressed as a proportion, usually between 0 and 1)

Yamane's formula assumes a simple random sampling method and is based on the population size and the desired margin of error for the sample estimate. The margin of error represents the acceptable amount of variability between the sample estimate and the true population parameter.

Example: Suppose a researcher wants to conduct a survey on a university campus with a total student population of 10,000 students. They aim to achieve a margin of error of 0.05 (5%).

$$n = \frac{N}{1 + Ne^2}$$

Using Yamane's formula:

$$n = \frac{10{,}000}{1 + 10{,}000 \text{ x } 0.05^2}$$

$$n \approx 385$$

So, according to Yamane's formula, the researcher would need a sample size of approximately 385 students to achieve a margin of error of 5% in their survey.

Example 2: A nonprofit organization is conducting a survey to assess the satisfaction levels of citizens regarding the quality of healthcare services in a large city. The city has a total population of 500,000 residents. The organization wants to ensure a representative sample with a margin of error of no more than 2%. They plan to use Yamane's formula to determine the required sample size for their survey.

Solution

Population size (N) = 500,000; Margin of error (e) = 0.02 (2%):

Using Yamane's formula:

$$n = \frac{N}{1 + Ne^2}$$

$$n = \frac{5{,}00{,}000}{1 + 5{,}00{,}000 \text{ x } 0.02^2}$$

$$n = \frac{N10{,}000}{1 + 10{,}000 \text{ x } 0.05^2}$$
$$n \approx 2{,}487.56$$

So, according to Yamane’s formula, the nonprofit organization would need a sample size of approximately 2,488 residents to achieve a margin of error of 2% in their survey. However, since the calculated sample size is not a whole number, the organization must decide whether to round up or down. Rounding up to the nearest whole number ensures a slightly larger sample size, providing additional assurance in the survey’s representativeness and reliability. Therefore, the organization might decide to round the sample size up to 2,488 participants.

The adjusted Yamane formula incorporating the population variance (π) and the z-score (z) for a given significance level (α) is designed to increase accuracy, especially for dichotomous variables. The modified formula is:

$$n = \frac{N.z^2.\pi(1-\pi)}{N.e^2 + z^2.\pi(1-\pi)}$$

where: n = sample size, N = population size, z = z-score corresponding to the desired confidence level (e.g., 2 for α=0.05, 3 for α=0.01), π = population variance (for a dichotomous variable, typically π=0.50, e = margin of error (expressed as a decimal)

Example Calculation: Suppose you have a population size N of 10,000, a margin of error e of 5% (0.05), a population variance π of 0.50, and you want to use a z-score of 2 (for α=0.05).

$$n = \frac{10000.\,2^2.\,0.50(1-0.50)}{10000.\,2^2 + 2^2.\,0.50(1-0.50)}$$
$$n \approx 384.56$$

So, the required sample size would be approximately 385.

KREJCIE AND MORGAN’S TABLE

Krejcie and Morgan’s Table is a widely used statistical tool for determining the appropriate sample size for a survey or research study based on a given population size. It was developed by Robert V. Krejcie and Daryle W. Morgan in their 1970 paper, "Determining Sample Size for Research Activities," published in the journal Educational and Psychological Measurement. The table provides a quick reference for researchers to ensure that their sample is large enough to yield statistically significant results.

KEY CONCEPTS OF KREJCIE AND MORGAN’S TABLE

Population Size (N): This refers to the total number of people or units in the group being studied.

Sample Size (S): This is the number of people or units selected from the population to participate in the study.

Confidence Level: Typically, a confidence level of 95% is used, which means that if the same population were sampled multiple times, the sample mean would fall within the confidence interval 95% of the time.

Margin of Error (e): This is the range within which the true population parameter is expected to fall. It is often set at 5%, indicating a 95% confidence level.

HOW THE TABLE WORKS

Population Size (N): The table lists population sizes in one column.

Sample Size (S): The corresponding recommended sample sizes are listed in an adjacent column.

The sample sizes provided in the table are calculated based on the formula for sample size determination for a given level of precision, confidence, and variability. The formula used is:

$$S = \frac{X^2.N.P.(1-P)}{d^2.(N-1) + (X^2.P.(1-p)}$$

Where: S=is the required sample size, X^2 is the chi-square value for the desired confidence level (e.g., 1.96 for 95% confidence), N is the population size, P is the population proportion (assumed to be 0.5 for maximum sample size), d is the degree of accuracy (the margin of error, e.g., 0.05 for ±5%).

PRACTICAL APPLICATION

Ease of Use: Researchers can simply look up the population size in the table and find the corresponding sample size without performing complex calculations.

Accuracy: The table ensures that the sample size is sufficient to make reliable inferences about the population.

Standardization: It provides a standardized method to determine sample size, promoting consistency across different studies.

Example: Imagine you are conducting a survey to understand the job satisfaction levels of employees at a large corporation. The total number of employees at the corporation (population size, N) is 1,200. Using Krejcie and Morgan’s Table:

Population Size (N): Find the row in the table that corresponds to a population size of 1,200.

Sample Size (S): Look at the recommended sample size for a population of 1,200.

According to Krejcie and Morgan's Table, for a population size of 1,200, the recommended sample size (S) is approximately 291. This is the number of employees you need to survey to obtain statistically significant results with a 95% confidence level and a margin of error of ±5%.

$$S = \frac{X^2.N.P.(1-P)}{d^2.(N-1)+(X^2.P.(1-p)}$$

Parameters: N= 1,200; X^2 ≈3.841 (for 95% confidence level); P=0.5 (population proportion, for maximum variability); *d*=0.05 (degree of accuracy, or margin of error)

$$S = \frac{3.841^2.1200.0.5.(1-0.5)}{0.05^2.(1200-1)+(3.841.0.5\,.(1-0.5)}$$

$$S = 291$$

A CONFIDENCE INTERVAL AND CONFIDENCE LEVEL

A confidence interval is a range of values, derived from sample statistics, that is likely to contain the value of an unknown population parameter. For example, if we are trying to estimate the average height of people in a city, we might take a sample of 100 individuals and calculate the average height from that sample. The confidence interval gives us a range of heights within which we are reasonably confident the true average height of the entire city population lies.

Confidence Level: The confidence level is the probability that the confidence interval does contain the population parameter. It is often expressed as a percentage, like 95% or 99%. For instance, a 95% confidence level means that if we were to take 100 different samples and compute a confidence interval for each sample, then approximately 95 of the 100 confidence intervals would contain the true population parameter. A z-score, also known as a standard score, is a statistical measure that indicates how many standard deviations a data point is from the mean of a dataset. It is calculated by subtracting the mean of the dataset from the data point and then dividing by the standard deviation. A z-score of 0 indicates that the data point is exactly at the mean of the dataset. Positive z-scores indicate data points above the mean, while negative z-scores indicate data points below the mean. The further the z-score is from 0, the more unusual or extreme the data point is relative to the rest of the dataset. In the context of confidence intervals, z-scores are used to determine the critical values for constructing confidence intervals when the population standard deviation is known. The z-score associated with a given confidence level corresponds to the number of standard deviations away from the mean that captures a certain percentage of the data in a normal distribution. The following is the formula to compute the Z score

$$z = \frac{x-\mu}{\sigma}$$

These z-scores are used to calculate the margin of error and construct confidence intervals around sample statistics. The higher the confidence level, the wider the confidence interval will be, because we want to be more certain that it contains the true population parameter.

Confidence Level	z-score (±)
0.7	1.04
0.75	1.15
0.8	1.28
0.85	1.44
0.92	1.75
0.95	1.96
0.96	2.05
0.98	2.33
0.99	2.58
0.999	3.29
0.9999	3.89
0.99999	4.42

Source: https://www.calculator.net/sample-size-calculator.html

IV. DISCUSSION AND CONCLUSION

Sampling is a cornerstone of research in the social sciences, enabling researchers to make inferences about populations based on data collected from a subset. The primary goal of sampling is to ensure that the sample accurately represents the larger population, which is essential for the validity and reliability of the research findings. The choice of sampling method significantly impacts the quality and generalizability of the research outcomes. Random sampling, which includes methods like simple random sampling and systematic sampling, ensures that each member of the population has an equal chance of being included in the sample. This randomness helps minimize selection bias, making the sample more representative of the population (Groves et al., 2018). However, in practical settings, achieving perfect randomness can be challenging, especially in large and diverse populations. Stratified sampling enhances the representativeness of the sample by dividing the population into distinct subgroups and sampling from each subgroup (Kozak et al., 2021). This approach is beneficial when researchers aim to ensure that specific subgroups within the population are adequately represented. For instance, stratified sampling can improve the precision of estimates for minority groups that might otherwise be underrepresented in a simple random sample. Cluster sampling, on the other hand, is particularly useful for geographically dispersed populations (Valliant et al., 2020). By dividing the population into clusters and sampling entire clusters, researchers can reduce the logistical and financial burden of data collection. While this method can be more cost-effective, it may introduce additional variability if clusters are not homogeneous, potentially affecting the precision of the estimates. Non-probability sampling methods, such as convenience sampling and purposive sampling, are often employed when random sampling is impractical or when the research aims to explore specific phenomena in-depth (Etikan et al., 2016). These methods can provide valuable insights, particularly in exploratory research, but they may lack the generalizability of probability-based methods. Recent trends in sampling reflect advancements in technology and data collection techniques. Big data and computational methods have introduced new possibilities for sampling from large and complex datasets. Machine learning algorithms and data mining techniques allow researchers to analyze vast amounts of data and refine their sampling strategies, enhancing both efficiency and accuracy (Kang et al., 2020). Adaptive sampling methods are gaining traction, especially in dynamic research contexts where conditions change during the study. These methods adjust sampling strategies based on interim findings, improving the focus and relevance of the research (Thompson & Seber, 1996). For example, adaptive cluster sampling can concentrate resources on areas with higher variability or importance, optimizing the research process (Morstatter et al., 2013). The rise of online platforms and social media has also transformed sampling practices. Researchers are increasingly using web-based tools and social media analytics to reach diverse and large populations. While these methods offer new opportunities for data collection, they also pose challenges related to representativeness and data quality. Addressing these challenges requires careful consideration of the sample's composition and potential biases. Integrative sampling approaches, which combine multiple sampling methods, reflect a sophisticated trend in contemporary research. By leveraging the strengths of different methods, researchers can address various aspects of sampling challenges and enhance the robustness of their findings. For instance, combining stratified and cluster sampling can balance precision and cost-efficiency, providing a more comprehensive view of the population.

Sampling methods are fundamental to the practice of social science research, influencing the accuracy and generalizability of study findings. Traditional methods, such as random, stratified, and cluster sampling, provide robust frameworks for ensuring representative samples and minimizing bias. However, emerging trends in technology and data collection are expanding the possibilities for sampling, offering new tools and techniques to handle large and complex datasets. The integration of computational methods, adaptive sampling, and online data collection represents significant advancements in the field. These innovations enhance the ability to capture diverse and dynamic populations, addressing some of the limitations of traditional methods. Researchers must stay abreast of these developments and select sampling strategies that align with their research objectives and contexts. Ultimately, the choice of sampling method should be guided by the research goals, the nature of the population, and the available resources. By applying both established and innovative sampling techniques, researchers can improve the validity and relevance of their findings, contributing valuable insights to the field of social sciences.

REFERENCES

[1]. Babbie, E. (2016). The Practice of Social Research (14th ed.). Belmont, CA: Wadsworth.

[2]. Babbie, E. R. (2020). The practice of social research. Cengage AU.

[3]. 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.

[4]. Brewer, J., & Hunter, A. (1989). Multimethod research: A synthesis of style. Newbury Park, CA: Sage.

[5]. Bryman, A. (2016). Social research methods. Oxford university press.

[6]. Cochran, W. G. (1977). Sampling Techniques (3rd ed.). New York: John Wiley & Sons.

[7]. Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

[8]. Cohen, J. (1992). Quantitative methods in psychology: A power primer. Psychological Bulletin, 112(1), 155-159.

[9]. Cohen, J. (1992). Statistical power analysis. Current Directions in Psychological Science, 1(3), 98-101.

[10]. Creswell, J. W. (2002). Educational research: Planning, conducting, and evaluating quantitative and qualitative research. Upper Saddle River, NJ: Pearson Education.

[11]. Creswell, J. W. (2013). Qualitative Inquiry and Research Design: Choosing Among Five Approaches (3rd ed.). Thousand Oaks, CA: SAGE Publications.

[12]. Creswell, J. W., & Clark, V. L. P. (2017). Designing and conducting mixed methods research. Sage publications.

[13]. Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. American Psychologist, 60(2), 170-180. DOI: 10.1037/0003-066X.60.2.170

[14]. Ellis, P. D. (2010). The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results. Cambridge University Press.

[15]. Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. American Journal of Theoretical and Applied Statistics, 5(1), 1-4. DOI: 10.11648/j.ajtas.20160501.11

[16]. Faul, F., Erdfelder, E., Buchner, A., & 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.

[17]. Fowler, F. J. (2013). Survey Research Methods (5th ed.). Thousand Oaks, CA: SAGE Publications.

[18]. Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. Advances in Methods and Practices in Psychological Science, 2(2), 156-168. DOI: 10.1177/2515245919847202

[19]. Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than P values: Estimation rather than hypothesis testing. British Medical Journal (Clinical Research Ed.), 292(6522), 746-750. DOI: 10.1136/bmj.292.6522.746

[20]. 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

[21]. Groves, R. M., Presser, S., & Dipko, S. (2018). The role of sampling in survey research. Annual Review of Sociology, 44, 55-82. DOI: 10.1146/annurev-soc-073117-041157

[22]. Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. Field methods, 18(1), 59-82.

[23]. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). Multivariate Data Analysis (7th ed.). Upper Saddle River, NJ: Pearson Education.

[24]. Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33(2-3), 61-83.

[25]. Israel, G. D. (1992). Determining sample size. Program Evaluation and Organizational Development, IFAS, University of Florida. PEOD-6.

[26]. Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates in structural equation modeling: A Monte Carlo study. Structural Equation Modeling: A Multidisciplinary Journal, 10(1), 128-141.

[27]. Kang, J., Carlin, B. P., & Chow, S.-L. (2020). A comparison of machine learning algorithms for large-scale surveys. Journal of Computational and Graphical Statistics, 29(4), 741-758. DOI: 10.1080/10618600.2020.1797878

[28]. Kish, L. (1965). Survey Sampling. New York: John Wiley & Sons.

[29]. Kline, R. B. (2015). Principles and Practice of Structural Equation Modeling (4th ed.). New York: Guilford Press.

[30]. Kozak, M., Parsons, M., & Brown, R. (2021). Advances in stratified sampling techniques for improving precision. Journal of Statistical Planning and Inference, 207, 161-171. DOI: 10.1016/j.jspi.2020.10.008

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

[32]. Lakens, D. (2021). Sample size justification. Collabra: Psychology, 7(1), 13.

[33]. Lee, S., Raju, P. S., & Hsu, J. (2017). Integrative sampling techniques for complex population structures. Journal of Statistical Planning and Inference, 181, 51-61. DOI: 10.1016/j.jspi.2016.10.007

[34]. Li, S., Marquart, J. M., & Zercher, C. (2000). Conceptual issues and analytical strategies in mixed-method studies of preschool inclusion. Journal of Early Intervention, 23, 116- 132.

[35]. Maxwell, S. E., & Delaney, H. D. (2004). Designing Experiments and Analyzing Data: A Model Comparison Perspective (2nd ed.). Lawrence Erlbaum Associates.

[36]. Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. Annual Review of Psychology, 59, 537-563.

[37]. Miles, M. and Huberman, A.M. (1994) Qualitative Data Analysis: An Expanded Sourcebook (2nd edn). Thousand Oaks, CA: Sage

[38]. Moore, D. S., McCabe, G. P., & Craig, B. A. (2017). Introduction to the Practice of Statistics (9th ed.). W.H. Freeman.

[39]. Morgan, D. L. (1998). Practical strategies for combining qualitative and quantitative methods: Applications to health research. Qualitative Health Research, 3, 362-376.

[40]. Morstatter, F., Pfeffer, J., & Liu, H. (2013). When is a liability not a liability? The role of data quality in big data. ACM Transactions on Intelligent Systems and Technology, 4(4), 60. DOI: 10.1145/2460276.2460283

[41]. Nunnally, J. C. (1967). Psychometric Theory. New York: McGraw-Hill.

[42]. Onwuegbuzie A.J. (2007). Mixed methods research in sociology and beyond. In: Ritzer, G. (eds) The Blackwell Encyclopedia of Sociology, vol. VI, pp. 2978–2981. Blackwell Publishers Ltd, Oxford.

[43]. Onwuegbuzie, A. J., & Leech, N. L. (2004). Enhancing the interpretation of “significant” findings: The role of mixed methods research. The Qualitative Report, 9(4), 770-792.

[44]. Onwuegbuzie, A. J., Jiao, Q. G., & Bostick, S. L. (2004). Library anxiety: Theory, research, and applications. Lanham, MD: Scarecrow Press.

[45]. Patton, M. Q. (2002). Qualitative research & evaluation methods. sage.

[46]. Patton, M.Q. (1990) Qualitative Research and Evaluation Methods (2nd edn). Newbury Park, CA: Sage.

[47]. Roscoe, J. T. (1975). Fundamental Research Statistics for the Behavioral Sciences (2nd ed.). New York: Holt, Rinehart and Winston.

[48]. Sandelowski, M. (2001). Real qualitative researchers do not count: The use of numbers in qualitative research. Research in Nursing & Health, 24, 230-240.

[49]. Sudman, S. (1976). Applied Sampling. New York: Academic Press.

[50]. Sullivan, G. M., & Feinn, R. (2012). Using Effect Size or Why the P Value Is Not Enough. Journal of Graduate Medical Education, 4(3), 279-282. DOI: 10.4300/JGME-D-12-00156.1

[51]. Tashakkori, A., & Teddlie, C. (Eds.). (2003a). Handbook of mixed methods in social and behavioral research. Thousand Oaks, CA: Sage.

[52]. Thompson, S. K., & Seber, G. A. F. (1996). Adaptive Sampling. Wiley.

[53]. Trochim, W.M.K. and Donnelly, J.P. (2008) The Research Methods Knowledge Base. 3rd Edition, Atomic Dog, Mason, 56-65.

[54]. Valliant, R., Dorfman, A., & Royall, R. M. (2020). Finite Population Sampling and Inference: A Prediction Approach. Wiley.

[55]. Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. Educational and Psychological Measurement, 73(6), 913-934.