PalArch's Journal of Archaeology of Egypt / Egyptology

ON THE INFLUENCE OF SAMPLE SIZE TO PRECISION OF PERCENTAGES AND MEANS IN A SURVEY RESEARCH

Wilfredo P. Mariño

University of Rizal System Antipolo Campus Antipolo City, Philippines

wilfredo.marino@urs.edu.ph

Wilfredo P. Mariño. On the Influence of Sample Size to Precision of Percentages and Means in a Survey Research. – PalArch's Journal of Archaeology of Egypt/Egyptology 17(2), 539-546. ISSN 1567-214X

Keywords: Sample Size, Influence of sample size, Precision of means and percentage.

ABSTRACT

One of the most frequently asked questions in statistics is how to determine the sample size. For many educational research pieces, inappropriate, inadequate, or excessive sample sizes continue to influence research findings' quality and accuracy. The study demonstrates the influence of sample size on the precision of measurements like percentage and means in survey research. In the study, the researcher utilized a sample size formula based on the margin of error and confidence level to determine the different sample sizes (100, 150, 300, and 1200) based on the 1761 total population of a state university study was conducted. A census on the population was conducted to determine the students' profile concerning sex, age, year level, and course and seven benchmark statements on the students' attitude towards the university using a validated instrument. Accordingly, a number was assigned to each of the respondents during the encoding in statistical software. After this, a random number generator was utilized to determine the respondents in different sample sizes from the census's encoded data. Based on the statistical analyses, the result shows the precision of percentages in the four profile variables. The means of the seven benchmark statements do not vary significantly in terms of sample size. On inferential statistics, only the sample sizes 300 and 1200 had the same decision with the population. To conclude, the proper calculation can determine the sample size necessary to have a specific percent chance that the result will not vary from the population.

INTRODUCTION

In many educational research types, the determination of sample size and sampling techniques to be used is a common undertaking in the first stage of research planning. Inappropriate, insufficient, or disproportionate sample sizes influence the quality and validity of research findings. Still, appropriate sample sizes are usually overlooked since it is not required for the most commonly used statistics test. However, some researchers estimate the effect of sample size on their analysis but rarely done in survey and market research.

The study tried to show the influence of sample size determined through a sample size formula to the precision of percentages and means and demonstrate the effect empirically in standard inferential statistics to emphasize the importance of sample size in any research type.

Specifically, the following problems were addressed by the study:

1. What are the different sample sizes utilized in the study based on a population using the confidence level and error margin?
2. What is the precision of the different percentages of respondents' profiles (sex, age, year level, and course) in varying sample sizes from the population percentages?
3. What is the precision of the different means of respondents' responses (Students' attitudes towards the university) in varying sample sizes from the population means?
4. How does the comparison of the students' attitude towards the university in various sample sizes in terms of sex, age, year level, and course differ with the population analysis of variance?

Review of literature

According to Bartlett et al. (2001), inappropriate, inadequate, or excessive sample sizes continue to influence research quality and accuracy. In the paper of Naing et al. (2006), they suggested utilizing software to check the normal approximation assumption and incorporate finite correction in the sample size calculation to address practical issues that may arise in sampling. As stated by Diong, 2016, "the standard error is dependent on sample size: larger sample sizes produce smaller standard errors, which estimate population parameters with higher precision." Also, Cohen (1988) describes a sample statistic's statistical precision as "the closeness with which it can be expected to approximate the relevant population value." Usually, precision is estimated using a standard error or the amount of chance that the estimate is different from the population's parameter.

According to Brown (2006), a population is "the entire group of people that a particular study is interested in." Moreover, it is identified that numerous factors can affect the precision of a parameter estimate, and one of them is the sample size.

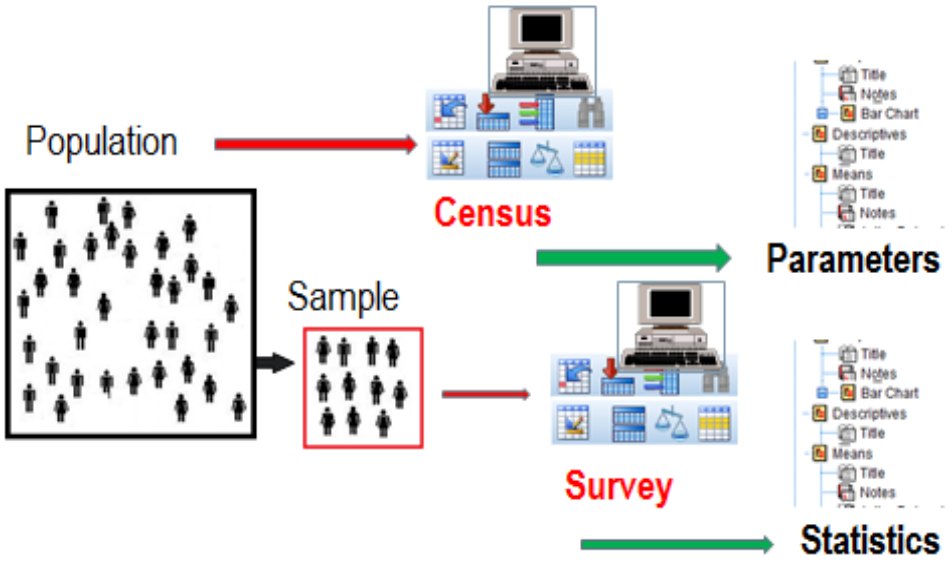
As mentioned by Cohen (1988, p. 6), "depending upon the statistics in question, and the specific statistical model on which the test is based, reliability [i.e., precision] may or may not be directly dependent upon the unit of measurement, the population value, and the shape of the population distribution. However, it is always dependent upon the size of the sample."

Also, in the study of Hassel et al. (2017), it was shown that the sample size needed for time-use research based on a time-sampling technique depends on the design and aim of the study.

With these readings, the researcher is interested in confirming whether the profile and perception of all students' entire population would vary with the estimate in different sample sizes.

The study was based primarily on the definition of statistics as used in research.

Figure 1: A paradigm of the interplay of the variables in the study



The model vividly shows the definition of statistics: the mathematics of collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling. In the study, the researcher compared the result of the statistics in different sample sizes to the parameters and showed its precision for percentage and means.

METHODS

The research used data mining. It is a process that, using statistical tools, transforms raw data into useful information. To outline the processed data, descriptive methods were employed. In addition, in the study, survey analysis was used. "In survey research, information was gathered from a group of individuals to describe some aspects or characteristics such as the attitude of the population," according to Fraenkel and Wallen (2007).

A population of 1761 respondents from a university was used to assess the sample size's effect on the accuracy of percentages and precision of means. Before its use, experts validated a study-made questionnaire in the field of research.

For the empirical analysis, a population census was performed. Each respondent was questioned about their profile and their reaction. Data collected was encoded and labeled correctly. A random number generator from the population of variables was used to perform random sampling based on the desired sample sizes.

Statistical software was used to achieve the basic goals, and a set of tables containing percentages and means of the varying sample sizes were provided for comparison with the parameter.

RESULTS AND DISCUSSION

The following were the results of the study.

VARIOUS SAMPLE SIZES BASED ON A POPULATION

The different sample sizes were determined based on a population using the desired confidence level and margin of error. Cochran's sample size formula below was utilized

$$\mathbf{n} = \frac{(\mathbf{z})^2(\mathbf{p})(\mathbf{q})}{\mathbf{e}^2}$$

n = sample size
z = value from the normal table at 95% confidence level (**z=1.96**)
p = % that an event would occur
q = % that event would not occur
e = margin of error (MOE)
Note: (p)(q) is at a maximum at p=0.5 and q=0.5

The table below shows the sample sizes obtained using the sample size formula above based on the error margin.

Table 1: Margin of error and the corresponding number of sample sizes based on Cochran's sample size formula

Margin of Error	Sample Size
$\pm$ 10.0%	100
$\pm$ 8.0%	150
$\pm$ 5.6%	300
$\pm$ 2.3%	1200

As obtained, with 100, 150, 300, and 1200 samples, the result of the survey research will fall between the intervals within the bounds of their corresponding margin of error ($\pm\ MOE$) in 95% of the survey.

Table 2: Goodness of fit test of the percentages of respondents' profile in varying sample sizes to the population percentages

Profile	Sample Size	Chi-square value	p-value	Decision
Sex (Male, Female)	100	0.149	0.699	Fail to reject
	150	0.050	0.823	Fail to reject
	300	0.027	0.871	Fail to reject
	1200	0.015	0.903	Fail to reject
Age (18 & below, 19-20, 21 & above)	100	2.685	0.259	Fail to reject
	150	2.716	0.873	Fail to reject
	300	0.294	0.985	Fail to reject
	1200	0.021	0.989	Fail to reject
Year Level (First Year, Second Year, Third Year, Fourth Year, Fifth Year)	100	5.311	0.150	Fail to reject
	150	0.366	0.947	Fail to reject
	300	0.341	0.952	Fail to reject
	1200	0.098	0.992	Fail to reject
Course (Education, Hotel and Restaurant Management, Tourism, Engineering, Business, Information, and Technology)	100	9.953	0.354	Fail to reject
	150	6.573	0.682	Fail to reject
	300	2.242	0.987	Fail to reject
	1200	0.711	0.999	Fail to reject

Table 2 presents the profile, sample sizes, Chi-square value, probability value, and decision on the influence of different sample sizes to the precision of percentages of respondents' profile.

The result shows no sufficient evidence at a 0.05 level of significance to show that the percentages of male and female respondents in different sample sizes vary from that of the population.

Table 3: Comparative analysis of the different means of student-respondents' attitudes towards the university in various sample sizes and population means

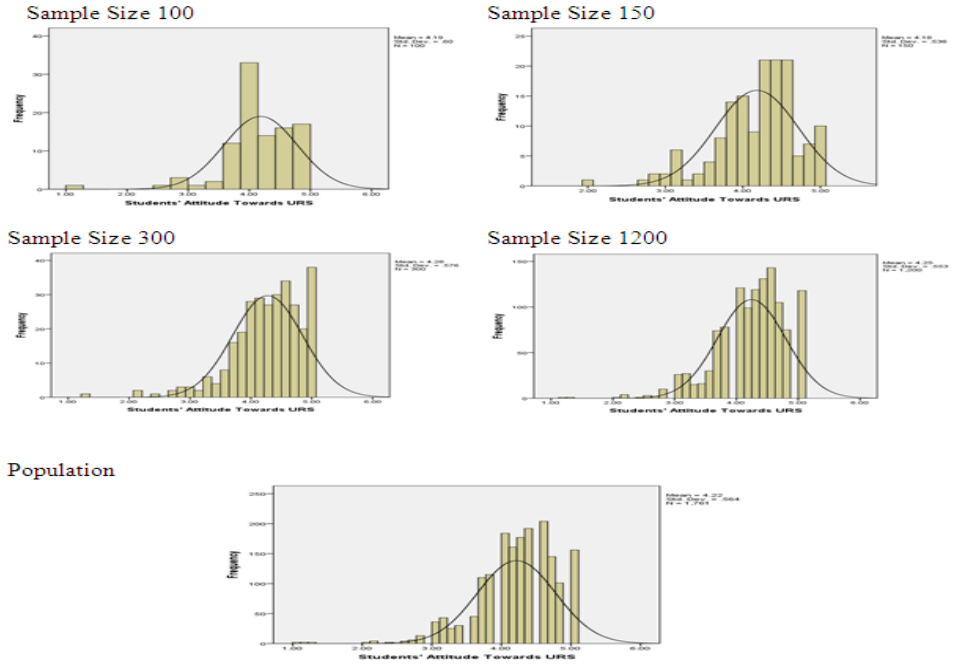
Sample Size	Degree of Freedom	t- value	p-value	Decision
100	1859	0.550	0.582	Fail to reject
150	1909	0.753	0.451	Fail to reject
300	2059	1.613	0.107	Fail to reject
1200	2959	1.289	0.198	Fail to reject

Table 3 presents the sample sizes, degree of freedom, t-value, probability value, and decision on the influence of sample size on the precision of different means of student-respondents' attitude towards the university.

As revealed, based on the result of t-test analysis between the means in various sample sizes and the population mean, it shows that the computed t-values and its corresponding probability values, which are all greater than the 0.05 level of significance, are not enough to warrant rejection of the null hypothesis.

As shown, a sample of size 100, 150, 300, and 1200 closely matches the means of its population. The result implies that the sample mean very close to the population mean.

Figure 4: Histograms of means based on the different sample size



As shown in the figures above, the different histograms closely match the population, revealing that the sample size increases. The shape of the distribution remains skewed to the right, regardless of sample size.

Table 4: Analysis of students' attitude towards the university in various sample size in terms of student-respondents' profile

Profile	**Sample Size**	**Degree of Freedom**	**t-value**	**p-value**	**Decision**
Sex	100	98	0.304	0.762	Fail to reject
	150	148	0.747	0.456	Fail to reject
	300	298	0.166	0.869	Fail to reject
	1200	1198	3.189	0.001	Reject
	(Population)1761	1759	2.438	0.015	Reject
Profile	**Sample Size**	**Degree of Freedom**	**F value**	**p-value**	**Decision**
Age	100	2/97	0.645	0.527	Fail to reject
	150	2/147	0.132	0.877	Fail to reject
	300	2/297	1.632	0.197	Fail to reject
	1200	2/1197	2.897	0.048	Reject
	(Population)1761	2/1758	2.978	0.042	Reject
Profile	**Sample Size**	**Degree of Freedom**	**F-value**	**p-value**	**Decision**
Year Level	100	3/96	1.830	0.147	Fail to reject
	150	3/146	1.177	0.321	Fail to reject
	300	2/296	2.396	0.068	Fail to reject
	1200	3/1196	7.010	0.000	Reject
	(Population)1761	3/1757	8.350	0.000	Reject
Profile	**Sample Size**	**Degree of Freedom**	**F-value**	**p-value**	**Decision**
Course	100	9/90	1.340	0.228	Fail to Reject
	150	9/140	1.566	0.131	Fail to Reject
	300	9/290	2.411	0.011	Reject
	1200	9/1190	4.654	0.000	Reject
	(Population)1761	9/1751	5.398	0.000	Reject

Table 4 presents the sample size, degree of freedom, t-value, probability value, and decision to analyze the students' attitude towards the university regarding student-respondents' profile in various sample sizes.

As revealed in the table, the result shows no sufficient evidence at the 0.05 level of significance to show a significant difference between the perception of male and female respondents in sample size 100, 150, and 300. On the other hand, with a 1200 and a 1761 population sample size, the study shows that male response differs significantly from female responses.

On age, with a sample size of 100,150 and 300, the null hypothesis was not rejected, which shows that age is not a variable to the respondents' attitude towards the university. On the other hand, the 1200 sample size and the 1761 population show that there is s sufficient evidence to claim that age has something to do with the respondents' attitude towards the university.

The variance analysis in terms of year level shows that with a sample of 100, 150, and 300, a significant difference in respondents' attitude towards the university cannot be confirmed. However, with a 1200 sample size, it demonstrates the same result with the population. The results show that there is sufficient evidence at the 0.05 level of significance that students attitude towards the university differs significantly in terms of their year level.

The result revealed that there is no sufficient evidence to show that respondents attitude towards the university differs significantly in terms of course, in the sample size 100 and 150. Nevertheless, on the sample size 300 and 1200 and in the population, the null hypothesis was rejected, proving that course has significantly affected the respondents' attitude.

As observed, the result of statistical analysis change with different sample sizes. The t-test and ANOVA result was presented as part of the presentation of the effect of varying sample sizes.

CONCLUSION AND RECOMMENDATION

With the aid of the data gathered and analyzed, the following summary and conclusions are now formulated.

1. The researcher can utilize a sample size formula to have a result that falls between intervals within the bounds of the margin of error.
2. In the tables, means and percentages' precision does not vary significantly in terms of sample size.
3. Only 1200 samples have the same decision with the parameters on rejecting the hypothesis that students' attitude differs significantly in terms of sex, age, year level, and course.

In light of the preceding conclusions, the following recommendations are hereby forwarded.

1. The result implies that as the sample size increased, the more likely to conclude that there was a statistical difference between the two groups. The result can be attributed to the precision with which the estimate of the population's percentage increases with increasing sample size.
2. A calculation can determine the sample size necessary to have a specific percent chance that a significant difference will be found between sample means when performing a statistical test at a particular level of significance. The researcher can evaluate whether the sample is large enough for the study.
3. Researchers may consider the proper sampling technique to identify the respondents in survey research regardless of size.

ACKNOWLEDGMENT

The researcher would like to extend wholehearted gratitude to the University of Rizal System Administrations, who made the study came into reality, Dr. Danilo M. Pascual for giving the researcher a permit to conduct the study, the different Deans and Program Heads of Antipolo Campus for the assistance in the data gathering; and the students and student-leaders for their cooperation in answering the questionnaire checklist.

References

Barlett, James E; Kotrlik, Joe W; Higgins, Chadwick C (2001). Information Technology, Learning, and Performance Journal; Morehead 19.1 (Spring 2001): 43-50.

Brown, J. D. (2006). Statistics Corner. Questions and answers about language testing statistics: Generalizability from second language research samples. Shiken: JALT Testing & Evaluation SIG Newsletter, 10(2), 24-27. Retrieved from the World Wide Web at http://jalt.org/test/bro_24.htm

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Diong, Joanna (2016). How Does Sample Size Affect the Precision of Estimates? Reproducible Research in the Digital Age. Scientifically Sound. 2016

Fraenkel, Jack R, Wallen Norman E (2007). How to Design and Evaluate Research in Education. McGraw Hill Co. Inc. New York

Hassel, DV, Bakker DD, Velde LVD, and Hoek LVD (2017). Assessing the Precision of Time-Sampling Based Study among GP's: Balancing Sample Size and Measurement Frequency

Naing, L; Winn T; Rusli BN (2006). Practical Issues in Calculating the Sample Size for Prevalence Studies. Archives of Orofacial Sciences 2006; 1:9-14