**Research Article**
Journal of Extension Education
Vol. 31 No. 1, 2019
DOI:https://doi.org/10.26725/JEE.2019.1.31.6200-6208

# Survey on Sample Sizes of Postgraduate Theses in Agricultural Education and Extension in Universities of Nigeria

**Michael Ejime Ikehi[1], Francis M. Onu[1], Florence O. Ifeanyieze[1], Patrick S. Paradang[2], Mmachukwu G. Nwakpadolu[3], Lilian U. Ekenta[1] and Clara U. Nwankwo[1]**

***ABSTRACT***

*The study focused on the representativeness of the sample sizes generated for populations under study in survey studies in Agricultural Education and Agricultural Extension. The study was carried out in Nigeria. A total of 4561 completed survey theses submitted to the Departments of Agricultural Education and Agricultural Extension in federal universities in Nigeria for the award of various postgraduate degrees between 2008 and 2018 were studied. The objectives of the study were to report the effect of sample size on the confidence level and margin of error of the generated data and to determine the frequently cited formula for generating sample sizes and their implications. The most frequently cited method for generating sample sizes was reported. Findings further revealed the problems associated with using non-statistical methods to generate sample sizes. The study recommended the use of mathematically proven formula in determining sample sizes in survey studies to generate statistically dependable, reliable and generalizable data.*

***Keywords:*** *Post graduate thesis; Confidence level; Margin of error; Population; Sampling error, Survey; Nigeria*

## INTRODUCTION

In most institutions of higher learning, research writing forms the bulk of the requirements for the award of degrees, mainly at the postgraduate levels. In most studies involving farmer education or awareness, survey research design is often adopted. In Nigeria, from observation, over 60% of postgraduate student theses majoring in Farmer Education and Extension had adopted descriptive survey research design. Descriptive survey research design studies a group of people or item by collecting and analyzing data from few individuals or items considered to be a representative of the entire group when the entire group cannot be studied for several reasons. In majority of cases, survey research design uses the sample of a population to describe, explain and document findings by collecting data based on the opinions and views of the samples studied, using questionnaire, interview, and

1. Department of Agricultural Education, University of Nigeria, Nsukka, Nigeria, 2. Federal College of Education, Pankshin, Plateau State, Nigeria, 3. Michael Okpara University of Agriculture Umudike, Imo state, Nigeria.

**Received : 24-04-2019; Accepted : 21-06-2019**

focus group discussion among others (Ali, 2006; Anaekwe, 2007; Uzoagulu, 2011). It is a method of descriptive research used for collecting primary data based on visual, verbal or written communications using a representative sample. The focus of survey research is on the ability of the sample to effectively predict the prevalent traits in the people or items studied when the entire population cannot be reached or fully studied. Therefore, the accuracy of the sample size is as important as the generated data. Calculating the right sample size is crucial to generate accurate information for generalization since different sample sizes have different statistical power. The calculation is necessary because the number of responses received based on the sample size determines the survey's confidence level and margin of error.

Confidence level describes how accurate results are, whereas the margin of error shows the range the survey results would fall between if confidence level holds true based on the sample size. From observation, several students have often expressed some level of difficulty in picking justifiable sample size in survey studies. Difficulties in clearly explaining the population and justifying the sample and sample size for some studies expectedly leads to questioning of the quality of the generated data. In most local and international standard journals, manuscripts adopting survey research design especially those originating from local institutions and authors face high possibility of rejection. The reasons adduced by some reviewers were the unreliability of the data presented due to sample size and sampling error, among numerous others. Reviewers indicated the need to collect reliable data through adequate sample size in survey studies. Some of the articles submitted for publication in reputable journals are products of postgraduate thesis modified into manuscripts.

In the rural settings of Africa, particularly in Nigeria where most population and subsequently the sample sizes for extension studies are drawn, farmers are always available and receptive. The farmers are widely available to attend to research instruments and provide quantifiable response for data presentation. However, there are different categories of farmers based on characteristics such as age, location, type and size of farm, education level among others. Therefore, a study must specify the character of the desired farmers and select adequate representatives when the entire population cannot be studied. If these and other vital conditions are not met, data collected might not be reliable. This study assessed Agricultural Education and Agricultural Extension theses that adopted survey research design to report (1) the effect of sample size on the confidence level and margin of error of data generated; and (2) determine the frequently cited method for generating the sample sizes in the completed theses studied.

## METHODOLOGY

The study was carried out in Federal Universities in Nigeria. These universities were Ahmadu Bello University (ABU), University of Abuja (UniAbuja), University of Benin (UniBen), University of Lagos (UniLag), University of Agriculture, Umudike (UniAgric) and University of Nigeria, Nsukka (UNN). In addition to being federal universities, these universities offer Agricultural Education and Extension and are respectively located in the Northern, Central, Southern, Western and Eastern Nigeria. In to have a (near) national view, these strategically located universities were chosen for this study. The study being a survey of completed studies had a population of 4,561 submitted postgraduate (M.Sc./ M.Sc (Ed. ) and Ph.D.) theses for the Departments of Agricultural Education and Department of Agricultural Extension, in the aforementioned universities.

The sample size for the study was 367 completed theses, achieved through Yamane formula. Purposive sampling technique was employed in selecting the reviewed studies. For a completed thesis to be included for review, it must; be an empirical research; adopted survey as the research design; have a population greater or equal to 1,000; be carried out between 2008-2018; have a sample derived from the population; a sample size formula.

**Table 1.**
**Universities and Theses Distribution**

| University | Region in Nigeria | Thesis Population Contribution | % Contribution | Sample Drawn | Degree Distribution | |
|---|---|---|---|---|---|---|
| | | | | | MSc/MSc (Ed) | Ph.D. |
| Ahmadu Bello University (ABU) | North | 735 | 16.1 | 59 | 42 | 17 |
| University of Abuja (UniAbuja) | Central | 378 | 8.3 | 30 | 22 | 8 |
| University of Benin (UniBen) | South | 687 | 15.1 | 55 | 39 | 16 |
| University of Lagos (UniLag) | West | 531 | 11.6 | 43 | 31 | 12 |
| University of Agriculture, Umudike (UniAgric) | East | 943 | 20.7 | 76 | 63 | 13 |
| University of Nigeria, Nsukka (UNN) | East | 1,287 | 28.2 | 104 | 87 | 17 |
| **Total** | | **4,561** | **100** | **367** | **284** | **83** |

Selected theses were grouped. To ensure homogeneity in combining the theses, the studies were grouped based on their population size. Population ranges of 1,000 – 4,999; 5,000 – 9,999; 10,000 – 14,999 and ≥ 15,000 were set thus any included study must fall within the indicated ranges. With the set population ranges, an average population for each range was used to test the sample sizes and the data that were expected to be generated using cited sample size formula. Frequency count and assumed value were used to address the research objectives.

## FINDINGS AND DISCUSSION

### Effects of Sample Size on Generated Data

**Table 2.**
**Sample size Influencing Confidence Level and Margin of Error**

| Sl. No. | Population | Sample size | Confidence level (%) | Margin of error (%) | Relative | Score | Average score | Sample Status |
|---|---|---|---|---|---|---|---|---|
| 1 | 1000 | 278 | 95 | 5 | 19 in 20 | 8.6 | 8.1-9.1 | Expected |
| 2 | 1000 | 158 | 90 | 6 | 18 in 20 | 8.6 | 8.0-9.2 | Low |

*Source: Adapted from FST, 2014.*

In surveys, a high confidence level and low margin of error are easy to achieve based on the availability and size of respondents. An example of the effects of confidence level and margin of error is shown in Table 2. Data in Table 2, revealed that when a population of 1,000 with a sample size of 278 at 95% confidence level is replicated 19 out of 20 times in a survey study, the results would be within a margin of error of 5%. The 5% margin of error permissible revealed that if the entire population was surveyed, the results can only differ with a score of ±5%. The data on Table 2 is an actual result of a study where farmers were asked to rate the effectiveness of the extension service in their locality on a scale of 0-10 and which gave a final average score of 8.6. With the 5% margin of error at 95% confidence level it was expected that if the entire population of the 1,000 farmers were surveyed the obtainable average score would be between 8.1 – 9.1 (±5%). However, a different result was obtained when the sample size moved further away from the population.

A lowered number of respondents leads to a drop in the confidence level (FST, 2014). From the data presented in Table 1, the confidence level dropped to 90%, with a margin of error of 6%. The responses of same farmers with mean value of 8.6 dropped to 18 in 20 chances (as against 19 in 20) and the results falls between average score of 8.0 – 9.2 if the total 1,000 farmers were to be surveyed. Therefore, the closer the sample size is to the population the better and more reliable the data generated becomes.

Theoretically speaking, a sample size cannot be too high. Unfortunately, it is sometimes much more expensive to incentivize or convince all target population members to take part. This could be expensive, and from a statistical perspective, ultimately frivolous (FST, 2014). For these reasons, there exists the standard confidence level of 95% with a margin of error of either 5% or 2.5%. In the end, attempting to go beyond this level of accuracy could be unrealistic and ultimately a less beneficial priority than making sure that the respondent farmers are valid for the survey and are giving reliable responses (FST, 2014).

While FST (2014) holds the view that confidence level and margin of error are better at 95% and 5% respectively, in social sciences the margin of error could be between 1-10% depending on the nature of the study and how accurate the study should be for generalization. It has been repeatedly argued that the more the best; the closer the sample is to the population the higher the power to generalise findings (Creswell, 1994; Bogdan & Biklen, 1982). Thus, going below recommended confidence level and margin of error results to sampling error which leads to lack of confidence in the generated data.

**Frequently Cited Formula for Generating Sample Size**

Data in Table 3 revealed the most frequently cited sample size formula. Out of the 367 theses studied, 329 cited percentages as "suggested" by Gall, Gall and Borg (Uzoagulu, 2011) for generating sample sizes while the remaining 38 cited Taro Yamane. The problem associated with using percentage suggestion is that it is a non-statistical approach. Uzoagulu (2011) argued that the use of percentage is not backed up with any statistics in terms of whether such percentage is a true representation of the population and whether the sampling error is reduced or not. Furthermore, the changes in the number of sample size is not systematically progressive in percentage suggestion according to population size; the higher the population size, the low the percentage value suggested to obtain as the sample size. The sample size computed statistically is more reliable than determining the size by mere approximation, except in a casual study where accuracy of data and appropriate procedure are not important. A postgraduate thesis or even a study for publication in either local or international journals is not a casual study and should not generate "casual" data for generalization.

Unarguably, increased sample size lowers sampling error, margin of error and ensures higher confidence level. However, if the procedure is not systematic, results become less reliable. Table 3, showed that citing percentage formula for a population of 3,000 (average) generated 600 (average) respondents as the sample size. The sample size is closer to the population than 353 statistically obtainable from the same population range using formula method. However, comparing populations 7,500, 12,500 and 15,000 in the percentage formula against the statistical Taro Yamane formula revealed the danger inherent in using the percentage formula. Using the percentage suggestions of 20%; 10%; 5% and 3%, a population of 3,000 generated a sample size of 600; 7,500 generated 750; 12,500 generated 625; and 15,000 generated 450, respectively. It is expected that the sample size of the 15,000 or 12,500 should

**Table 3.**
**Thesis and Sample Size Formula**

n=367

| Sl. No. | Number of Thesis | Thesis Population (Range) | Thesis Population (Average) | Cited formula | Value suggested | Sample size generated (Average) | Status |
|---|---|---|---|---|---|---|---|
| 1 | 116 | 1,000 – 4,999 | 3,000 | Percentage | 20% | 600 | Non-statistical |
| 2 | 93 | 5,000 – 9,999 | 7,500 | Percentage | 10% | 750 | Non-statistical |
| 3 | 89 | 10,000 – 15,000 | 12,500 | Percentage | 5% | 625 | Non-statistical |
| 4 | 31 | ≥ 15,000 | 15,000 | Percentage | (<5%) 3% | 450 | Non-statistical |
| 5 | 7 | 1,000 – 4,999 | 3,000 | Taro Yamane | $\frac{N}{1+N(e)^2}$ | 353 | Statistical |
| 6 | 14 | 5,000 – 9,999 | 7,500 | Taro Yamane | $\frac{N}{1+N(e)^2}$ | 380 | Statistical |
| 7 | 10 | 10,000 – 15,000 | 12,500 | Taro Yamane | $\frac{N}{1+N(e)^2}$ | 388 | Statistical |
| 8 | 7 | ≥15,000 | 15,000 | Taro Yamane | $\frac{N}{1+N(e)^2}$ | 390 | Statistical |
| Total | 367 | | | | | | |

be larger than that of 7,500 and farther from that of the 3,000 but the opposite was the case; the sample sizes decreased with increasing population which makes the progression unmathematical. Following percentage suggestions for sampling, the size began to drop as the population increased thus widening the margin of error and increasing sampling error as well as lowering the confidence level of the data generated for generalization for large populations. The formula method (Taro Yamane), respected the increase in population and increased the sample sizes steadily according to the increasing population. The statistical Taro Yamane proves to be mathematically reliable as progression where according to population size, as shown in Figure 1.

The study revealed that the number of respondents generated from a population

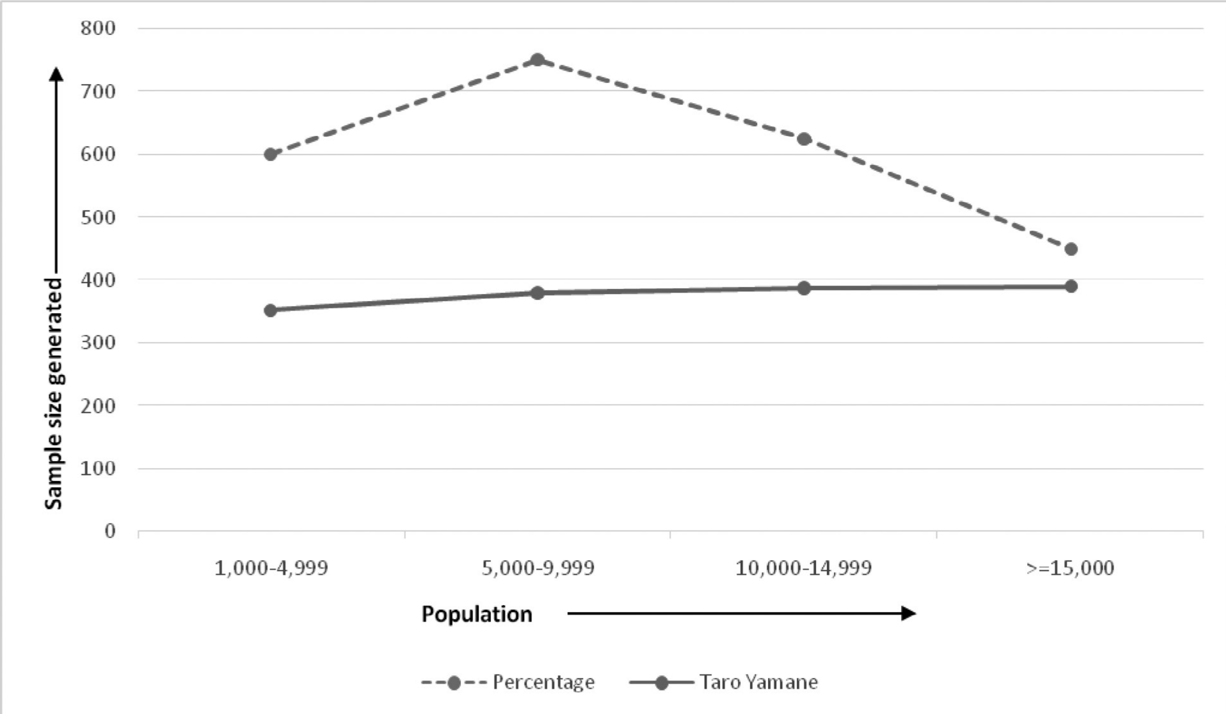


*Fig. 1. Sample Sizes for Statistical and Non-statistical Formula*

as sample, depends on the formula utilized, and the sample size generated as the representative of the population could affect the quality and generalization of the findings. The sample size is mathematically not a true representative of the population if the sample is not statistically picked from the population. Thus, the margin of error and confidence level of a sample for a study is directly influenced by the method (statistical/non-statistical) utilized in determining the sample size. The percentage suggestions appear to be the most popularly cited method among the student researcher. This is likely due to the ease in using the method for calculation. The percentage suggestions often generate a larger sample size for are small population and generate a small sample size for a large population. For a large sample size obtained through non-statistical approach for a small population, "over representation" of the population is a less beneficial priority as stated by FST (2014). When the population is over represented, chances are that the findings might skew to a direction presenting a view of the available rather than an average view generalizable on the entire population (Bogdan & Biklen, 1982; Cohen et al., 2009; Eboh, 2009). Percentage approach which is non-statistical does increase the sample size bringing it closer to the population especially for small populations but shows the opposite for larger populations. Also, in the percentage formula, the percentage chosen according to population size does not accurately address a particular population size rather based the choice of percentage on population range.

The formula requires keying-in of the "exact" population size and computing a statistically dependable sample size with even increase or decrease according to population size. Authors such as Guest et al., (2006), Francis et al., (2010), Uzoagulu (2011), Emmel (2013) and Fugard and Potts (2015) emphasized the need for a statistically generated sample size. In addition to Taro Yamane sample size generating formula, there are many available sample generating formulas for finite and infinite populations (Kothari, 2004). Also, it is possible to use software such as Microsoft Office Excel and some webpages to generate statistical sample from a given population.

Comparing the percentage suggestions and the Taro Yamane formula, obviously reveals that the percentage suggestion generates higher sample size than the later. As observed by the researchers, this large size has become a major threat to the quality and dependability of the data generated in most studies carried out by postgraduate students of agricultural education or agricultural extension in tertiary institutions. When the size is large it becomes more capital intensive and difficulty to reach all target respondents as indicated by the generated sample size. There are instances where some researchers resort to “arm-chair” data collection; a situation where the researcher fills out the instrument by him/herself or contact individuals who are not members of the population to respond to the instrument. According to Kume (1991), Onwuegbuzie and Leech (2007), FST (2014) and Mustafee (2014) no sample size is theoretically large or small as long as it is a true representative of the population of the studied. Thus, the main concern of a survey research with sampling in agriculture is to ensure that the respondent farmers are valid for the survey and are giving truthful responses. Hence a concise and statistically generated sample size presents a mathematical “true” representativeness of the entire population under study and makes it possible to generalize thedata generated statistically.

## CONCLUSION

All research studies are carried out to find a solution to an existing problem and some require obtaining data and utilizing the data to infer or generalize solutions. Therefore, no matter the number of respondents (farmers), if the approach for generating the sample is not mathematical, the sample size is statistically not a true representative of the studied population which makes the generated data unreliable for generalization. This study thus recommends the use of statistical formulas for generating sample sizes to ensure reliability of data available on postgraduate theses and subsequently journal articles. The statistical methods for generating sample size should be more reliable than non-statistical suggestion.

This study is limited in that it did not compare the sample sizes that can be generated by the various sample size generating formulas. Further study is thus suggested to compare the output sample sizes of the various documented sample size generating formulas.

## REFERENCES

Ali, A. (2006). *Conducting research in education and the social sciences*. Enugu, NG: Tian ventures.

Anaekwe, C. M. (2007). *Basic research methods and statistics in education and social sciences*. Onitsha, NG: Sofie Publicity and Printry Ltd.

Bogdan, R.C. & Biklen, S.K. (1982). *Methods of social research.* Boston; Allyn and Bacon Inc

Cohen, L., Manion, L. & Morrison, K. (2009). *Research methods in education* (6th Ed.). Abingdon, Oxon; Routledge.

Creswell, J. (1994). *Research design, qualitative and quantitative approach;* London; Sage publications

Eboh, C.C. (2009). *Social and economic research.* Enugu, Nigeria; African Institute of Applied Economics

Emmel, N. (2013). *Sampling and choosing cases in qualitative research: A realist approach*. London: Sage.

Field Surveys Team. (2014). *Calculating the right survey sample size.* Retrieved 23rd May, 2016, from http://fluidsurveys.com/university/calculating-right-survey-sample-size/

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P. & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology and Health*, 25, 1229–1245. doi:10.1080/08870440903194015

Fugard, A. J. & Potts, H. W. (2015). Supporting thinking on sample sizes for thematic analyses: A quantitative tool. *International Journal of Social Research Methodology*. doi:10.1080/13645579.2015.1005453.

Guest, G., Bunce, A. & Johnson, L. (2006). How many interviews are enough?: An experiment with data saturation and variability. *Field Methods*, 18, 59–82. doi:10.1177/1525822X05279903

Kothari, C.R. (2004) *Research Methodology Methods and Techniques.* 2nd Edition, New Age International Publishers, New Delhi

Kume, H. (1991). *Statistical methods for quality improvement*. Tokyo; 3A Corporation.

Mustafee, S. (2014). *Statistical methods for agricultural science.* New Delhi; Scientific International PVT Ltd.

Nwana, O. C. (1981). *Introductory to educational research*. Ibadan: Heinemann Educational Books Ltd.

Onwuegbuzie, A. J. & Leech, N. L. (2007). A call for qualitative power analyses. *Quality & Quantity*, *41*, 105–121. doi:10.1007/s11135-005-1098-1

Uzoagulu, A. E. (2011). *Practical guide to writing research project report in tertiary institutions*. Enugu, NG: Cheston Publishers.