PSU - Journal of Engineering, Technology, and Computing Sciences (JETCS)

(Vol. No. 2, Issue 1 pp. 17-25, June 2020)

Sample Size Estimation Using a Preliminary Sample as Basis: A Simulation Analysis

Jan Homarl T. Alambra[1], Paulo V. Cenas, Andhee M. Jacobe, Bobby F. Roaring

College of Computing, Pangasinan State University - Urdaneta City campus
jeyhuo.alambra@gmail.com[1]

Abstract – *This research paper provided a new way in sample size estimation for continuous data using information from a preliminary or initial sample through a computer simulation analysis as its statistical modeling approach. Four simulation schemes were designed wherein each has different cases that were subjected to a simulation process. Two hundred different normal population data were generated and analyzed through the simulation process wherein it contains 30 simulation runs (repeated model building process). For every model building process, the data were fitted into four regression models namely: linear, quadratic, cubic, and power regression models. The sample size computed using the sample estimates (or the Estimated Sample size or ESS) and using the population parameters (or the true sample size) served as the independent and dependent variables for the regression analysis.*

Throughout the whole simulation analysis, thirty-two equations for each averaged regression model were established and breakdown as follows: 2 x 4 models for Scheme 1, 6 x 4 for Scheme 2, another 6 x 4 for Scheme 3, and 18 x 4 for Scheme 4. For the majority of the cases, the power regression model was considered to be the best. Nonetheless, each of the averaged models was deemed to be good and useful using the "at least 90% average R^2" rule as a criterion for simulation analysis.

Keywords: *Sample Size Estimation, Simulation Analysis, Simulation Scheme, Regression Models, Coefficient of Determination (R^2)*

INTRODUCTION

Every statistical study (surveys, experiments, clinical trials, etc.) will work better if each of the aspects in the study was carefully and meticulously planned. That is, the problem should be cautiously defined and operationalized. The experimental and observational units must be selected from the appropriate population and they must be randomized properly. The procedures and methods should be strictly followed. The instrumentation must be reliable and valid in obtaining measurements. Finally, the number of units must be adequate relative to the goals or objectives of the study.

Data had played a great role in the advancement and development of the modern world. It had served as the baseline of many decisions that had led to the invention of new things, ratification of present things, and renovation of old things. There are two ways to collect data or information, i.e. through census and sample survey.

Census, also known as complete enumeration, is the acquisition of data from every unit of the population. The key advantage of this process is that the results acquired here (assuming full compliance) are known with certainty. Its principal disadvantage is its impracticality since it involves a considerable cost (both in money and time) in the collection and compilation of the data (Sabatella & Franquesa, 2009).

On the other hand, sample surveys are the collection of data from a fraction of the population. It has served as "an efficient and economical source of providing statistical data on various subjects for both research and administrative purposes (Narasimha-Prasad, n.d.)," and it is preferred by many researchers due to its strengths, benefits and advantages (Sincero, 2013). The advantages of sample survey over the census are listed as follows (based on Cochran, 1977; Sabatella and Franquesa, 2009):

The sample size must be properly determined for a researcher to draw valid and generalized conclusions. If the sample size is too small, the study may fail to detect important aspects such as the effects or associations of a variable to other variables. Similarly, if it is too large, the study may grow into a more complicated situation and thus, producing more inaccurate results.

Sample size estimation has been a crucial part of the design of any quantitative research. The importance of the sample size is stated in the introduction of this paper. Now, according to Cochran (1977), there are six principal steps in choosing the sample size:

1. There must be a statement concerning the expected sample size. These statements may include the desired limits of error or some decision to be made.
2. Equations that connect n with the desired precision of the sample must be found. The equations vary depending on the sampling design used.
3. The unknown parameters in those equations must be estimated.
4. If the target population has certain major subdivisions that must be published and if the limits of error are set up for those subdivisions, then a separate calculation must be made for n in each subdivision. The total n is acquired by addition.
5. Some methods must be found for the reconciliation of the conflicting values of n because the sample surveys utilize many items or characteristics.
6. Finally, the chosen value of n must be consistent with the available resources for the sample survey. A decision must be faced – whether to proceed with a much smaller sample size or to abandon efforts until the resources are enough.

"A sample size generally depends on five study design parameters: minimum expected difference or also known as the effect size, estimated measurement variability, desired statistical power, significance criterion, and whether a one- or two-tailed statistical analysis is planned." (Singh & Masuku, 2014)

Aside from the population size and the objective of a study, Miaoulis & Michener (1976) listed the level of precision, the level of confidence or risk, and the degree of variability in the attributes being measured as other criteria for a good sample size.

There are many approaches, methods, or strategies in determining the sample size. One can apply formulae concerning sample size estimation, conduct a pilot study, imitate a sample size of similar studies, use published tables, or simply guess based on instincts (see Cochran, 1977; Israel, 1992; Sudman, 1976; Singh and Masuku, 2014).

For some economic reasons, adequate sample size is important. In the same paper of Lenth (2001), it was stated that:

> "An under-sized study can be a waste of resources for not having the capability to produce useful results, while an over-sized one uses more resources than are necessary. In an experiment involving human or animal subjects, the sample size is a pivotal issue for ethical reasons. An under-sized experiment exposes the subjects to potentially harmful treatments without advancing knowledge. In an oversized experiment, an unnecessary number of subjects are exposed to potentially harmful treatment, or are denied a potentially beneficial one."

With such theoretical and practical importance, sampling procedures and methods must be properly utilized to acquire the desired sample size for a study. Sample size estimation "requires care in eliciting scientific objectives and in obtaining suitable quantitative information before the study (Lenth, 2001)" and it requires a close and honest collaboration between those who are experts on sampling theory and those who have good scientific knowledge in the art and practice of the subject matter (Sathian, et al., 2010).

Generally speaking, there are three ways of obtaining a sample size. The first one is cost-based, where the convenience and practicality of the data collection are considered. The next one is variance-based, where the variance of the collected data is used as an estimate of the population parameter. The last one is the Statistical-Power Based, where sample size is assessed through the power of some statistical test (Singh & Masuku, 2014).

Many factors affect sample size. These include the purpose of the study, sampling design, accuracy of estimation, dispersion, and size of the population, level of significance, power of the test, budget, and practical constraints. Yet, there is no single rule that can be applied to determine the required sample size for all surveys (Serumaga-Zake & Arnab, 2013).

Looking at the literature, many statisticians and researchers had come with their criteria, guidelines, formulas, and tables regarding the determination of the proper sample size. In several research studies involving surveys, the so-called Slovin's Formula was often used. Yet sadly, many of these researchers used the formula inappropriately, "giving the wrong impression that it can be used in just about any sampling problem" (Tejada & Punzalan, 2012). Also, many of the statistical concepts, especially on sampling are presented incorrectly leading to the misuse of such terms.

OBJECTIVES OF THE STUDY

This research generally addressed the misconceptions by providing and citing related literature. Also, this study provided a new method in sample size estimation using some information from the preliminary sample. With these certain objectives at hand, this research sought to:

1. Determine the characteristics of the generated values of the variables used in the regression analysis?
2. Determine the best regression model predictive to the needed and actual sample size using the computer-generated data?

METHODOLOGY

Research Design

This research study utilized an exploratory and experimental quantitative design through a simulation process as a modeling approach. The design of this research substantially involved the iterative nature of the simulation process as the researcher tries to develop and generate a statistical model in sample size estimation. The simulation process itself paves a way for the exploratory part of this study as it permits the creation of a program or a system wherein parameters and variables are subjected to change depending on the perspective of the researcher about the output of the said program or system.

Simulation Process

This research study aimed to develop and generate a statistical model that could be used as a computational tool in sample size estimation that utilizes a simulation analysis as its modeling approach. The computer software 'R Studio' which utilizes the software 'R Project' and the 'R' programming language and environment for statistical computing and graphics as its inner system was used in generating the required parameters and variables in the whole simulation process. A modified source script, which was coded by the researcher, was applied in the software to create a program that will eventually run the simulation process.

Simulation Procedure

The simulation process followed a step-by-step procedure to achieve an organized system that will eventually fulfill the main objective of the study and the procedure was divided into three stages (see Figure 2, on the next page). Each step for the three stages was represented in the source code of the program (except for the last stage).

Stage 1 (Population Parameters, Sampling Aspects, and Dependent Variable)

The first stage (Stage 1) involved the two primary steps in the simulation process. The first step was the setting and generation of the population parameters and the sampling aspects.

The population parameters referred to the population mean, population variance, and population size while the Sampling aspects denoted the sampling method, sampling criterion, and sample fraction. Here, the researcher answered the first specific objective of this study.

The process generated and utilized 200 different populations from a normal distribution. The population mean and variance for each population were held fixed throughout the simulation process. Also, the values for the population size for each population were held fixed while the sampling criterion, and sample fraction were different for each population or held fixed for all the populations throughout the process depending on the simulation scheme.

Here, the sample size was computed using the formula of Cochran (1977) for the continuous data that uses the relative error as the degree of precision or accuracy. The following formula for the required sample size was utilized in the process:

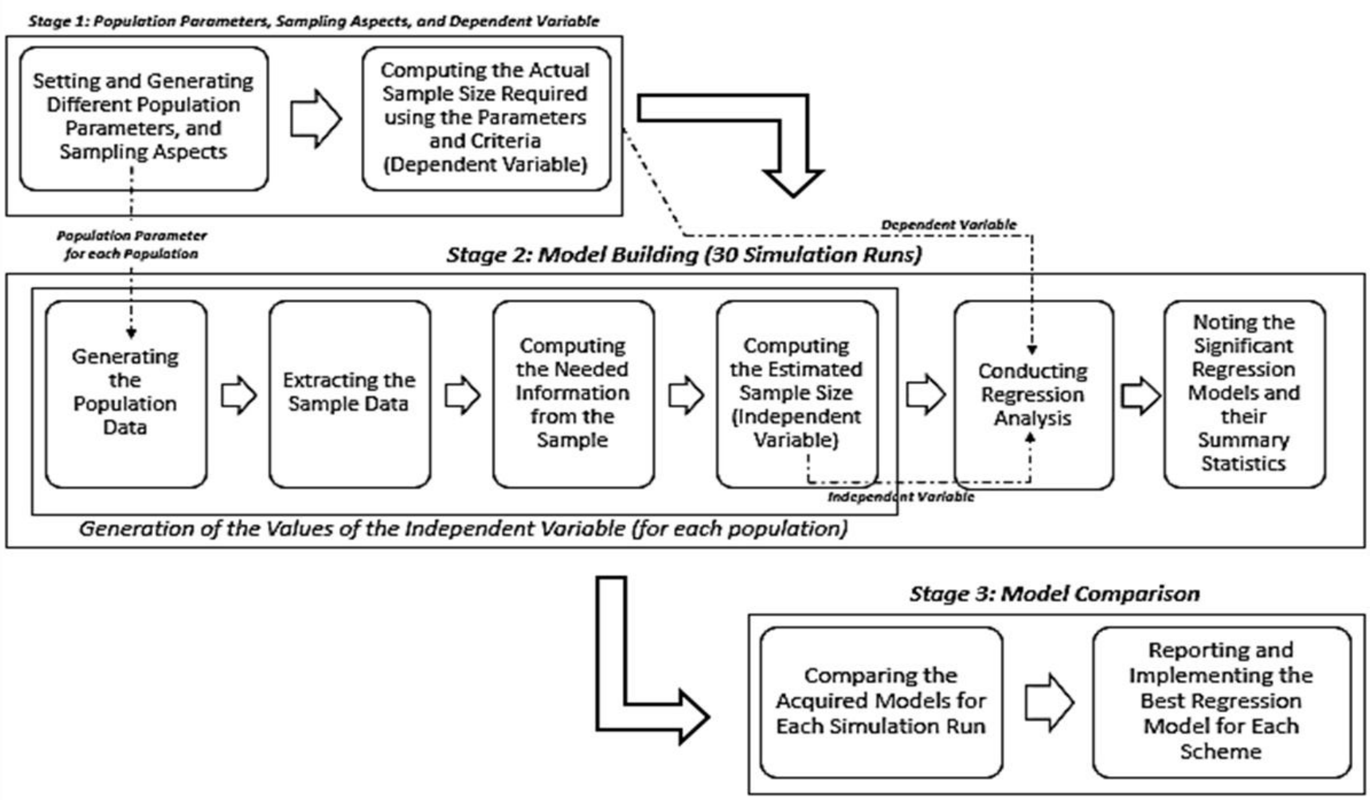

Figure 2: The Simulation Process

The sampling method that was used is the Simple Random Sampling without Replacement (SRSWOR) while the sampling criterion was the relative error.

The second step for the first stage has generated the output for this stage. This step was the computation of the values for the dependent variable of the analysis, which is the actual or true sample size for each population.

$$n = \frac{n_o}{1 + {n_o}/{N}}$$

where $n_o = \left(\frac{tS}{r\bar{Y}}\right)^2$ on which r is the relative error, t is the abscissa of the normal curve that cuts off an area of α at the tails, α is maximum probability to commit a greater error than r, $\bar{Y}$ is the population mean and S is the population standard deviation, and N is the size of the population. This formula was utilized even if the sampling ${n_o}/{N}$ is deemed to be negligible.

Stage 2 (Model Building Stage)

The second stage (Stage 2) was called the Model Building Stage. Primarily, the main objective of this stage is the development of a statistical model. An iterative process called simulation run was utilized in the statistical model building. Under each simulation run, another iterative process was employed in generating the values of the estimated sample size (ESS) for each population. These values of the ESS served as the values of the independent variable for the analysis.

The first part of this stage was the process of generating the value of the ESS. The first step in this process was the generation of the population data. The modified program generated population data from a normal distribution using the generated values of the population parameters in the first stage. The said data was held fixed throughout all the simulation runs for each simulation scheme.

The second step was the drawing of the sample from the generated population data on the previous step using a simple random sampling without replacement. The sampling procedure was automatically done by the sampling function of the modified program. Afterward, on the third step of this process, the needed information from the sample was extracted using embedded functions and computational formulas. These information were enumerated as follows: sample mean, sample variance, and coefficient of variation. After extracting those information, the fourth step was the computation of the values of the ESS using the extracted data on the previous step.

The said process was repeated for the next 199 different populations. After all the values of the independent variable – the ESS has been computed, it is now ready for the next part of this stage – the statistical analysis.

The second part of the Model Building Stage is comprised of two steps. The first step under this part is the regression analysis. Regression analysis was used as a method of analysis since it has been the most flexible and robust tool in the process of statistical modeling.

Under the first step, the data were subjected to four curve-fitting procedures to produce with the following regression models: Linear, Quadratic, Cubic, and Power Regression models. For each procedure, the data were tested using a 95% level of confidence against the null hypothesis which states that the ESS is not predictive to the actual or true sample size.

After the regression analysis, the significant statistical model(s) was noted. The beta coefficients, the value of r-squared, the mean square error of the residuals, and some other summary statistics regarding the model were also noted. This procedure was the last step of the second stage.

The two parts of this stage were followed for each of the 30 simulation runs. The output of this stage was the significant regression models and their summary statistics.

Stage 3 (Model Comparison)

After acquiring all the significant models and noting their summary statistics, the last stage (Stage 3) involved another two steps: the model comparison and the reporting of the best model. At the model comparison, all acquired model was subjected into explicit comparison using some summary statistics as criteria in determining the best regression model for the sample size estimation. The said summary statistic was the r-squared of the model. The rule of thumb for this summary statistic was that the higher the r-squared, the better the quality of the model. After determining the best model for the cases in the simulation schemes, the model was readied for deployment and implementation.

Simulation Schemes

In this research, simulation scheme refers to the systematic arrangement or combination of different conditions to create various cases that extend the range of the data analysis in statistical modeling through the created simulation process. By having different simulation schemes, the researcher has been able to determine the best statistical model for some cases in the real world. Four different simulation schemes were utilized for this study where each scheme followed the laid simulation process in the previous section.

The first simulation scheme comprised of the analysis of the 200 populations with different parameters, different sampling criteria, and different sample fraction for each population. The values of the population parameters were fixed for all simulation schemes.

Now, the second scheme involved an analysis of 200 different populations with different parameters and different sampling criteria for each population but with a constant sample fraction for all populations. Meanwhile, the analysis of the 200 populations with different parameters and different sample fractions for each population and a constant sampling criterion for all the population comprised the third simulation scheme.

The values for population parameters, sampling criterion, and sample fraction were the same as the values from the first scheme. Yet, the cases when the sampling criterion (relative error) is either 1%, 3%, or 5% were studied on the second scheme while the cases when the sample fraction is either 1%, 5%, or 10% were studied on the third scheme.

The fourth simulation comprised of the analysis of 200 different populations with different parameters for each population but with constant sampling criterion and sample fraction. This scheme involved the combinations of the cases in the second and third schemes.

Also for each simulation scheme, the population size has been divided into two cases. The first case, in which values were drawn from the range 100 to 5000 were referred to as the smaller data and the second case, in which values were drawn from the range 5000 to 10000, were referred as larger data. Hence, there were 2, 6, 6, and 18 cases analyzed for the first, second, third, and fourth simulation scheme, respectively. In total, 32 cases were analyzed.

The values of the population parameters, sampling criterion, and sample fraction were fixed on the modified computer program through the use of the random seed function. In R programming language, randomized values were held fixed through the calling of random seed by using the function set.seed( ). For the whole research simulation analysis, pi (π) which value is equivalent to 3.14159265358979 was used as the random seed. The said random seed was used for each simulation process in each simulation scheme.

The sequence 1 to 200 was used as random seeds for the generation of population data for each of the 200 population. The samples drawn from each population were indexed through the random seed formula $200*r+i$ where r corresponds to the order of the simulation run and i corresponds to the order of the population where the sample was drawn.

RESULTS AND DISCUSSIONS

Looking at its theoretical and practical importance, sample size determination is one of the most crucial and integral parts of any quantitative researches. Though some theories, principles, and rules were established on sample size estimation, one is ought to remember that there is no single rule of the thumb that could be used in all cases of sample size estimation. The goal of this research paper was to create a more efficient way in sample size estimation for continuous data through the simulation analysis as the modeling approach.

The whole simulation analysis was executed using a modified computer program through the 'R' programming language and its corresponding software. Four simulation schemes were utilized for this particular research study, which certain cases under each scheme were studied. Every case throughout the whole simulation analysis follows a certain simulation process. Two hundred different populations were generated from each of the normal distributions of smaller and larger sizes and were subjected to the 30 model building process (simulation runs) for every case in every scheme. Four regression models were considered for the model building process and were as follows: linear, quadratic, cubic, and power regression models.

In this part, some statistical descriptions for the data parameters and the sample estimates and sizes were presented. Meanwhile, the second part of the said section contains the discussion of the results of the simulation analysis for each scheme. This second part was also divided into two sub-parts: the description of the variables and the regression model findings.

For the first simulation scheme (Scheme 1), on average, the estimated sample size (ESS) produced relative errors of around 6.13% and 5.79%, for each of the cases respectively. The arrangement of the predictive power of the four regression models were different for the two cases studied here. For the first case which is the modelling under the smaller data, the best model according to its average R^2 were the cubic model, followed by quadratic, linear and lastly power model. But for the other case, the best model were the power model, followed by the cubic, quadratic, and linear models. All models, both for the smaller and larger data, have a good predictive power since the lowest value of the average coefficient of determination for this scheme was 0.9743628.

The second simulation (Scheme 2) involved three cases (5%, 3%, and 1%) of sampling criterion (relative error), both for smaller and larger data. The acquired values of the ESSs for each of the cases produced and the average percentage of error were 9.88%, 6.26%, and 1.65% for the smaller data and were 8.66%, 6.12%, and 2.03% for the larger data. All of the cases determined the power regression model as the best model according to its average R^2. It was followed by cubic, quadratic, and linear models. The lowest value of average R^2 throughout the simulation analysis for this scheme was 0.9761406.

Three cases (1%, 5%, and 10%) of sample fraction were considered and studied on the third simulation scheme (Scheme 3). An ESS deviated from its corresponding true sample size by about 12.88%, 5.56%, or 3.98% for the cases under the smaller data and by about 11.95%, 5.38%, or 3.75% for the cases under the larger data. All the cases for this scheme produced the same arrangement of the performance of the regression models based on their corresponding average R^2 except for the first case (1%) of the sample fraction under the smaller data. For the said exception of the cases, the best model was the cubic regression model and followed by the quadratic, linear, and power models, respectively. For the rest of the cases, the arrangement of the models was as follows: power, cubic, quadratic, and linear regression models. The lowest average R^2 for this scheme was 0.9551129.

The fourth simulation scheme (Scheme 4) involved the combination of the cases for Scheme 2 and Scheme 3. It was divided into three sub-schemes (referred to as Scheme 4.1, 4.2, and 4.3) on which the same cases of sample fraction were included for each value of the sampling criterion (relative error). The largest relative errors produced by the ESS concerning the true sample for each of the sub-schemes were 20.07%, 13.07%, and 4.31% while the smallest ones were 5.49%, 3.93%, and 1.08%. Except for the first case of sample fraction under the smaller data both on Scheme 4.2 and Scheme 4.3, the best model for all of the cases was also the power regression models, followed by the cubic, quadratic, and linear models. As for the exception, the best model was the cubic, followed by the quadratic, linear, and power models. The lowest regression models for each sub-scheme were as follows: 0.9128181, 0.9489014, and 0.9879447.

CONCLUSIONS AND RECOMMENDATIONS

At a 5% level of significance, the computer-generated data were analyzed using regression analysis. Based on the results of the simulation analysis for each scheme, the averaged regression models stated in the previous section (Results and Discussions) were established with the following generalizations:

1. All regression models produced throughout the whole simulation analysis were highly significant with *Pr(>F)* = 0.0000.
2. The majority of the cases acquired an averaged power model that was deemed to be the best model among the four regression models based on their corresponding average coefficient of determination.
3. Using the "at least 90% average R^2" rule as a criterion for the goodness and usefulness of the generated regression models for simulation analysis, all of the established regression models were good and useful.

The utilization of the models must be based on the fact that the data should come from a normal distribution. For a more generalized model, one must consider the models presented on the Scheme 1 where one could get an initial sample fraction of 1% to 10% and one could utilize a relative error from 1% to 5% as a sampling criterion depending on the preference of the reader for the sample size estimation using the acquired sample estimates. For a fixed value of the relative error and a sample fraction ranging from 1% to 10%, one must consider the models for the third case (1%) of Scheme 2 on which were the most accurate ones. Yet, if one would utilize a relative error of 1%, higher values of sample size would be generated. So, if there is a limit on resources, one could consider the models on the first and second cases under Scheme 2 wherein the reader can choose between 3% or 5% relative error values. If one wants wider choices of the values of the relative error, one could consider the models on the Scheme 3 wherein the initial sample size were fixed and the relative error could be from 1% to 5% continuously. And, if one is contented in using a fixed value of a sampling criterion and sample fraction, one could consider the models on the cases for Scheme 4.

Though the models could be compared using their corresponding average coefficients of determination, their individual performance in predicting real-life values of the true sample size needed is still unknown since those models were generated from hypothetical data. The researcher suggests then that one must perform a comparative analysis of the model using some real-life data to truly determine the capability of the generated models. One may also opt to utilize the simulation schemes and processes provided in establishing models for some continuous real-life data in general, or under a specific field. It is also recommended that one must consider a bigger number of population and other population parameters and extend more cases for each scheme to cater a wider range of analysis.

REFERENCES

[1] Cochran, W. (1977). *Sampling Techniques.* Canada: John Wiley and Sons.

[2] Jacobe, A. M. (2014). *Estimation of Variance: A Simulation Analysis.* Urdaneta: Pangasinan State University.

[3] Miaoulis, G., & Michener, R. D. (1976). *An Introduction to Sampling.* Dubuque, Iowa: Hunt Publishing Company.

[4] Rao, R. (1985). *Elements of Health Statistics.* Varanasi, India: R. Publication.

[5] Shannon, R. (1975). *System Simulation: The Art and Science.* Prentice-Hall.

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

[7] Yamane, T. (1967). *Statistics: An Introductory Analysis.* New York, Evanston, London: Harper International Division.

[8] Almeda, Capistrano, & Sarte. (2010). *Elementary Statistics.* Quezon: University of the Philippines Press. *Retrieved last October 13, 2016 from http://webcache.googleusercontent.com/search?q=cache:http://www.stat.upd.edu.ph/docs/course%2520notes/Chapter%25201.pdf*

[9] Bryman, A., & Cramer, D. (1996). *Quantitative Data Analysis with Minitab: A Guide for Social Scientists.* New York: Routledge. *Retrieved last November 25, 2016 from http://libgen.io/get/709B59C3250AA955A32BE74C845B63C4/Duncan%20Cramer-Quantitative%20Data%20Analysis%20with%20Minitab_%20A%20Guide%20for%20Social%20Scientists%20%281996%29.pdf*

[10] Burton, A., Altman, D., Royston, P., & Holder, R. (2006). The Design of Simulation Studies in Medical Statistics. *Statistics in Medicine 2006*, 4279-4292. *Retrieved last November 25, 2016 from https://www.researchgate.net/profile/Patrick_Royston/publication/6842399_The_design_of_simulation_studies_in_medical_statistics/links/5798daa408aed51475e87c2c/The-design-of-simulation-studies-in-medical-statistics.pdf?origin=publication_detail*

[11] Djimeu, E. W., & Houndolo, D.-G. (2016). *Power Calculation for Causal Inference in Social Science: Sample Size and Minimum Detectable Effect Determination.* International Initiative for Impact Evaluation. *Retrieved last October 13, 2016 from http://www.3ieimpact.org/media/filer_public/2016/04/06/wp26-power-calculation.pdf*

[12] Israel, G. D. (1992). Sampling the Evidence of Extension Program Impact. *Program Evaluation and Organizational Development, IFAS. Retrieved last October 13, 2016 from http://www.psycholosphere.com/Determining%20sample%20size%20by%20Glen%20Israel.pdf*

[13] Khan, A. (n.d.). *Building Statistical Models.* Australia: University of Queensland. *Retrieved last Novemeber 25, 2016 from*

http://people.umass.edu/biep691f/pdf/Linear%20Regression%20in%20Stata%20by%20Khan%2033%20slides.pdf

[14] Lenth, R. V. (2001). *Some Practical Guidelines for Effective Sample-Size Determination.* Iowa City, USA: University of Iowa. *Retrieved last October 13, 2016 from https://www.scribd.com/document_downloads/direct/249918630?extension=pdf&ft=1487234320&lt=1487237930&user_id=189349203&uahk=cMCSIJh7+s2ie26/sePirtuhAl0*

[15] Narasimha-Prasad, N. (n.d.). *Lecture Notes on Sample Surveys.* University of Alberta. *Retrieved last October 13, 2016 from http://www.stat.ualberta.ca/~prasad/stat561_lec1.pdf*

[16] Rissanen, J. (n.d.). *Lectures on Statistical Modeling Theory.* Royal Holloway, UK: University of London. *Retrieved last November 25, 2016 from http://www.lce.hut.fi/teaching/S-114.300/lectures.pdf*

[17] Robinson, S. (2004). *Simulation : The practice of Model Development and Use.* John Wiley and Sons. *Retrieved last Novemeber 25, 2016 from http://197.14.51.10:81/pmb/GENIE_DES_PROCEDES/Simulation%20The%20Practice%20of%20Model%20Development%20and%20Use.pdf*

[18] Sabatella, A., & Franquesa, R. (2009). *Manual of Fisheries Sampling Surveys: Methodologies for Estimators of Socio-Economic Indicators in the Mediterranean Sea.* Spain. *Retrieved last October 13, 2016 from ftp://ftp.fao.org/docrep/fao/006/y5228e/y5228e00.pdf*

[19] Sathian, B., Sreedharan, J., S., Baboo. N., Sharan, K., Abhilash, E. S., & Rajesh, E. (2010). *Relevance of Sample Size Determination in Medical Research* . Nepal: Nepal Journal of Epidemiology. *Retrieved last October 13, 2016 from http://xa.yimg.com/kq/groups/18751725/1475501789/name/SAMPLE+SIZE+PUBLICATION.pdf*

[20] Serumaga-Zake, P. A., & Arnab, R. (2013). *A Suggested Statistical Procedure for Estimating the Minimum Sample Size Required For a Complex Cross-Sectional Study*. South Africa. *Retrieved last October 13, 2016 from http://www.iiis.org/CDs2013/CD2013SCI/IMSCI_2013/PapersPdf/HA369WA.pdf)*

[21] Shannon, R. (1988). Introduction to the Art and Science of Simulation. In E. W. D.J. Medeiros (Ed.), *1998 Winter Simulation Conference. Retrieved last November 25, 2016 from http://cecs.wright.edu/~fciarall/ISE195/Readings/ShannonSimulationART.pdf*

[22] Sincero, S. M. (2013). *Advantages and Disadvantages of Surveys.* Explorable. *Retrieved last October 13, 2016 from https://explorable.com/advantages-and-disadvantages-of-surveys*

[23] Singh, A. S., & Masuku, M. B. (2014). Sampling Techniques & Determination Of Sample. *International Journal of Economics, Commerce and Management*. *Retrieved last October 13, 2016 from http://ijecm.co.uk/wp-content/uploads/2014/11/21131.pdf*

[24] Srivastava, V. (2014). *Lecture Notes on Sampling Theory.* Indian Institute of Technology Kanpur. *(Retrieved last October 13, 2016 from http://home.iitk.ac.in/~shalab/sampling/chapter1-sampling-introduction.pdf)*

[25] Tejada, J. J., & Punzalan, J. R. (2012). *On the Misuse of Slovin's Formula.* Manila City, Philippines: University of the Philippines. *(Retrieved last October 13, 2016 from http://www.psai.ph/sites/default/files/On%20the%20Misuse%20of%20Slovin's%20Formula.pdf)*