EUROPEAN ACADEMIC RESEARCH
Vol. VI, Issue 10/ January 2019

Impact Factor: 3.4546 (UIF)
DRJI Value: 5.9 (B+)

ISSN 2286-4822
www.euacademic.org

The effect of variance and tails of test on sample size determination

MOHAMMED OMER ADAM[1]
Department of Statistics
Sudan University of Science and Technology
ADIL MOUSA YOUNIS
Qassim University
AFRAA HASHIM
Department of Statistics
Sudan University of Science and Technology

Abstract

When different techniques of sampling are used, there are different formulas for determination of appropriate sample size.

The aim of this study was to employ formula for determining representative sample size, when simple random sampling technique is used test from one tail and two tails. at the level of significant $\propto = 0.05$, *depending on that, the value of test from two tails* $Z_{\frac{\propto}{2}} = 1.96$ *and test from one tail* $Z_{\propto} = 1.64$ *with variance ranging (0.1 to 2) and Standard Error* $e = 0.01$, *was used to obtain different sizes of samples.*

Excel office and SPSS package were used for sample size generation and analysis. To obtained results showed that, the sample sizes increases, as the variance increases. The result also showed significant Correlation between variance value (s^2) *and sample size (n) in the two cases. Significant differences were observed between sample sizes estimator when Z test from one tail or two tails. The obtained samples sizes showed significant goodness of fitting with Poisson distribution, in the two cases.*

[1] The corresponding author email: mamoh99@gmail.com

Key words: sample size, variance effect, precision, random sampling, confidence interval

1. INTRODUCTION

Sample size estimation is usually conducted through a pre-study power analysis. To select a sample size for achieving a desired power of correct detection of a meaningful difference at a given level of significance [(1)]

If we desire to estimate, with a confidence interval, the mean of a population (μ), one of the first question arise is how large the sample should be? This question must be given serious consideration, because it is a waste of time and resources to take a larger sample than is needed for the required results. On the other hand, if the number of samples is too small it may lead to results of no practical value. The key questions are how do you want our estimate to be close to the true value? Or, how wide would we like to make the confidence interval? The second question is, how much confidence do we want to place in our interval? That is, what confidence coefficient do we wish to employ? [(2)].

To answer these questions may need to use the confidence interval

$$\bar{y} \pm z_{\propto/2} * \frac{\sigma}{\sqrt{n}} \qquad \rightarrow 1$$

Where $z_{\propto/2} * \frac{\sigma}{\sqrt{n}}$ is equal one half of the confidence interval, accordingly question one can be answered, by the set up of the following equation:

$$e = z_{\propto/2} * \frac{\sigma}{\sqrt{n}} \qquad \rightarrow 2$$

Where e indicates how close the true mean we want our estimate to be. From equations 1 and 2 sample size equation can be written as

$$n = \frac{{z_{\propto/2}}^2 * \sigma^2}{e^2} \qquad \rightarrow 3$$

Sample size estimation is referred to the calculation of required number of samples for achieving some desired statistical confidence of accuracy and reliability (2).

In many researches, sample size calculation may be performed based on precision analysis, variance of population, probability assessment or other statistical inferences) (1).

This study was focusing on the estimation of sample size depending on the effect of variance of population. In surveying studies, once data are collected, the most important objective of a statistical analysis is to draw inferences about the population using sample information, e.g. how big a sample is required?. as one of the most frequently asked questions by the investigators. If the sample size is not taken properly, conclusions drawn from the investigation may not reflect the real situation for the whole population (5).

The computer programming or computation may be effectively used to determine an appropriate sample size, as one of the most important steps in statistical studies. Sample size computation should be done appropriately because if the sample size is not appropriate for a particular study then the inference drawn from the sample may lead to some wrong conclusions (5).

According to Gwowen Shieh (2013), the a priori determination of a proper sample size is necessary to achieve some specified power as an important problem frequently encountered in practical studies. To establish the needed sample size for a two-sample t test, researchers may conduct the power analysis by specifying scientifically important values as underlying population means while using a variance estimate obtained from related research or pilot study. In order to take account of the variability of sample variance, two approaches to sample size determinations should be considered. One provides the sample size required to given assurance probability, that, the actual power exceeds the planned power. The other provides the necessary sample size , such that the expected power attains the designated power level. The suggested paradigm of adjusted sample variance combines the existing procedures into one unified framework. Numerical results may be presented to illustrate the usefulness and advantages of the proposed approaches that accommodate the stochastic nature of the sample variance. More importantly, supplementary computer programs are developed to aid the implementation of the suggested techniques. The exposition may helps to clarify discrepancy and to extend the development of sample size methodology [(2)]. Simple regression and correlation analysis of some data collected, to test, hypotheses and make conclusions on the factors affecting the sample size for audit purposes of the internal auditors in the public sector in Kenya, was found that , materiality of audit issue, type of information available, source of information, degree of risk of misstatement, auditor skills and independence are some of the factors influencing the sample size determination for the purposes of internal audit evidence collection in public sector in Kenya [(3)].

2. METHODOLOGY

Simple random sampling technique was used to determine the representative sample size from one and two tails. Z value for various levels were employed (table1), where level of significance is $\propto = 0.05$, variance values ranging from (0.1 to 2), Standard Error $e = 0.01$, and total population size was assumed to be ($N = 100{,}000$). This was used to find the main sample size. Microsoft excel office was used to generate sample size values and ratios (tables 2 and 3). SPSS program was used for a test of significances (tables 4, 5) and (figs 1, 2, 3, 4).

Table 1: Z test values

	$\propto = 0.05$	$\propto = 0.01$
Two tails	1.96	2.58
One tail	1.645	2.33

3. RESULTS AND DISCUSSION

Z-test value from two tails, $z_{\propto/2} = 1.96$ with level of significance $\propto = 0.05$, variance values ranging from (0.1 to 2) and Standard Error $e = 0.01$ were used to obtain different sizes of samples (table 2).

Table 2: size of samples and ratio

$z_{\propto/2}$	s^2	e	n_0	n	Ratio
1.96	0.1	0.01	3842	3700	1
1.96	0.2	0.01	7683	7135	2
1.96	0.3	0.01	11525	10334	3
1.96	0.4	0.01	15366	13319	4
1.96	0.5	0.01	19208	16113	5
1.96	0.6	0.01	23050	18732	6
1.96	0.7	0.01	26891	21192	7
1.96	0.8	0.01	30733	23508	8
1.96	0.9	0.01	34574	25691	9
1.96	1	0.01	38416	27754	10
1.96	1.1	0.01	42258	29705	11
1.96	1.2	0.01	46099	31553	12
1.96	1.3	0.01	49941	33307	13
1.96	1.4	0.01	53782	34972	14
1.96	1.5	0.01	57624	36558	15
1.96	1.6	0.01	61466	38067	16
1.96	1.7	0.01	65307	39506	17
1.96	1.8	0.01	69149	40880	18
1.96	1.9	0.01	72990	42193	19
1.96	2	0.01	76832	43449	20

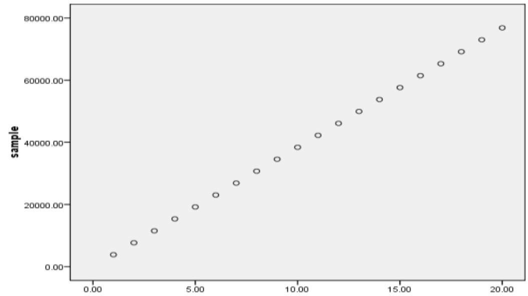

Fig 1: scatter diagram for sample sizes(n_0)

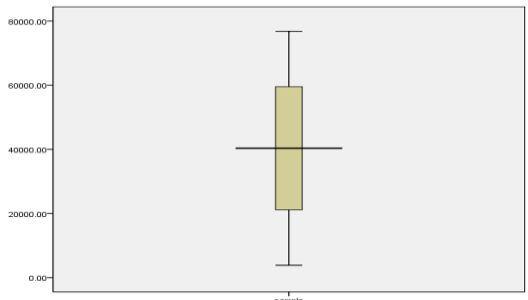

Fig 2: box plot for sample sizes(n_0)

According to the obtained results in table 2 and figures (1, 2) the sample size increases, as the variance increases.

Table 3: size of samples and ratio

$z_\propto$	s^2	E	n_0	n	Ratio
1.64	0.1	0.01	2690	2619	1
1.64	0.2	0.01	5379	5104	2
1.64	0.3	0.01	8069	74667	3
1.64	0.4	0.01	10758	9713	4
1.64	0.5	0.01	13448	11854	5
1.64	0.6	0.01	16138	13896	6
1.64	0.7	0.01	18827	15844	7
1.64	0.8	0.01	21517	17707	8
1.64	0.9	0.01	24206	19489	9
1.64	1	0.01	26896	21195	10
1.64	1.1	0.01	29587	22832	11
1.64	1.2	0.01	32275	24400	12
1.64	1.3	0.01	34965	25907	13
1.64	1.4	0.01	37654	27354	14
1.64	1.5	0.01	40344	28747	15

1.64	1.6	0.01	43034	30087	16
1.64	1.7	0.01	45723	31377	17
1.64	1.8	0.01	48413	32620	18
1.64	1.9	0.01	51102	33820	19
1.64	2	0.01	53792	34977	20

By using z-test value from one tail $z_\propto = 1.64$, with level of significance $\propto = 0.05$, variance values (s^2) ranging from $(0.1\ to\ 2)$ and standard error $e = 0.01$, different sizes of samples were obtained (table 3).

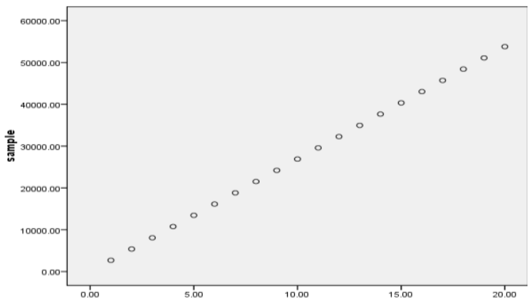

Fig 3: Scatter diagram for sample sizes(n_0)

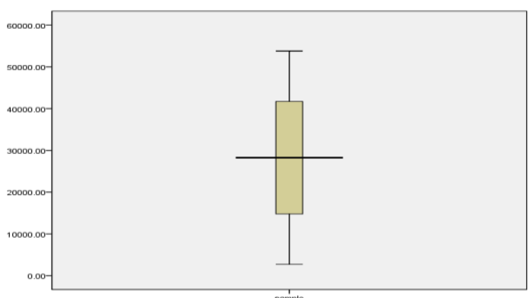

Fig 4: box plot for sample sizes(n)

From the table 3 and fig (3, 4) we can observe that, the sample size increases, as the variance increases without out layer fit and normal ratio.

Table (4): Independent sample T test

	t-test for Equality of Means	
	t	Sig
Equal variances assumed	**2.519**	0.017
Equal variances not assumed	**2.615**	0.013

Table (4) showed that the value of t-test is (**2.519**) with (p-value= 0.017<0.05). According to tables (1 and 2), there is significant difference between sample sizes estimators of Z test from one tail and two tails, at the significant level 0.05.

Table 5: One-Sample Kolmogorov-Smirnov Test

		$z_{\propto/2}$	$z_{\propto}$
Poisson Parameter	Mean	4.0337E4	2.8241E4
Kolmogorov-Smirnov Z		2.236	2.236
Sig		0.000	0.000

SPSS analysis showed that, the value of chi square test is (2.236) with (p-value= 0.000< 0.05). From table (5), and depending on tables (2 and 3), we may concluded that, there is a significant goodness of fit, in cases of Z test from one tail and two tails, at the significant level 0.05.

Table 6: Correlations

		n	
		$z_{\propto/2}$	$z_{\propto}$
s^2	Pearson Correlation	0.939	0.938
	Sig	0.000	0.000

From table (6), the Pearson Correlation value was found to be (1.000) with (p-value= 0.000<0.05). Depending on that, there is significant Correlation between the variance values (s^2) and sample size (n) in the case of one tail and two tails, at the level of significant 0.05.

4. CONCLUSION

According to the obtained results we may conclude that, the sample size was affected by the increases or decreases of variance. To select appropriate sample size for any study, we need to know the hypothesis, where the test from one tail or two tails. The value of correlation between estimated sample sizes and variance when the test from two tails is equal 0.938, and equal 0.939, when, the test from one tail, may indicate the existence of the strong correlation between sample sizes and variance. In a sample surveys, it may be use test Z from two tails to giving larger sample size.

REFERENCES

1. Shein-Chung Chow, Jun Shao, Hansheng Wang, (2017), Sample Size Calculations in Clinical Research- print edition Marcel Dekker, Inc. New York.
2. Gwowen Shieh, (2013), On using a pilot sample variance for sample size determination in the detection of differences between two means, *Psicológica*, **34**: 125-143.
3. Charles Guandaru, Kariuki, Samuel Nduati, (2012), Factors Influencing Sample Size for Internal Audit Evidence Collection in the Public Sector in Kenya Kamau. *International Journal of Advances in Management and Economics*, **1**: 42-49.
4. H.K. Sarmah, B. Bora Hazrkia, G. Choudhury, (2013) An investigation on effect of bias on determination of sample size on the basis of data related to the students of schools of Guwahati, *International Journal of Applied Mathematics and Statistical Sciences* **2** (1):
5. William G. Cochran – John Willey & Sons, (1977), Sampling techniques-

6. Sharon L. lohr, (2010), Sampling design and analysis. Second edition-Brooks/Cole- USA
7. Paula Lagares Barreiro - Justo Puerto Albandoz – MaMaEuSch, (2001), Population and sample. Sampling techniques, *International Journal of Applied Mathematics and Statistical Sciences*. **2** (1).
8. Mohd Asraf and James K. Brewer, (2004), Conducting test of hypotheses the need for an adequate Sample size, *The Australian educational researcher,* **31**(1).