Southeast-Asian J. of Sciences Vol. 3, No. 2 (2014) pp. 141-152

SAMPLE SIZE DETERMINATION FOR NON-FINITE POPULATION

Inpong Luanglath

The International College (BUIC) &
Family Enterprise Research Center (FERC)
Bangkok University
e-mail: Lecturepedia@gmail.com

Abstract

This paper reviews the conventional approach to sample size calculation with finite and non-finite population. Two alternative sample size determination methods are provided. The need and necessity of a test sample as a requisite to minimum sample size determination is explained. Standard error, margin of error and sampling errors are differentiated. This paper presents two new methods of determining minimum sample size: (i) n-hat method and (ii) multistage nonfinite population method. Under both methods, the minimum sample size is $n \approx 30$. The range under MNP method is 30-40 counts. The claim under Weisberg and Bowen that the minimum sample size for 0.05 error level $n = 400$ is wrong. The standard error equation has been wrongly applied as the sampling error and it is erroneously used as a tool for minimum sample size determination.

1 Introduction to Standard Error

The standard error is the standard deviation of the distribution of a statistic. (Evritt, 2003). The standard error is given by:

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{1}$$

where σ is the estimated standard deviation and n is the sample size. This "sample size" is not the same as the minimum sample size needed to test the

Key words: Multi-stage non-finite population method, n-hat, sample size, standard error.

condition $\bar{x} = \mu$ in the sample population inferential argument. The value for σ may be determined through the Z-equation:

$$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}} \tag{2}$$

where $\bar{x}$ = sample mean;μ = assumed population mean; S = sample standard population; and n = sample size. By solving for σ, thus

$$\sigma = \Big(\frac{\bar{x} - \mu}{Z}\Big)\sqrt{n} \tag{3}$$

The population is assumed to be normally distributed and is generally described as $N(0, \sigma^2)$: normally distributed with mean 0 and variance σ^2. This assumption is implicit. However, the standard error equation makes no explicit reference to such distribution. (Press, Flannery, Teulosky and Vetterling, 1992, p. 465). The variance of a normally distributed is $\sigma^2 = 1$ for standard normal distribution. The condition $N(0, \sigma^2)$ may be written as $N(0, 1)$. Therefore, the standard error in (1) may be written as:

$$SE = \frac{1}{\sqrt{n}} \tag{4}$$

The condition $N(0, 1)$ produces the probability density function:

$$P(x)dx = \frac{1}{\sqrt{2\pi}} exp\big[-z^2/2\big]dz \tag{5}$$

It assumes that $Z = (\bar{x} - \mu)/\sigma$; therefore $\frac{dz}{dx\sigma} = P(x)dx$. Thus, $SE = \sigma/\sqrt{n}$ becomes $SE = 1/\sqrt{n}$ in statement (4). Statement (4) has been misinterpreted and misapplied to mean "sampling error" for the purpose of calculating minimum sample size by many researchers. (Weisberg, 1971, p. 41). This misapplication starts with the assumption that the standard error is because the error level in the normal distribution curve is set at: $\alpha = \big|\big(-\frac{1}{2}\alpha + \frac{1}{2}\alpha\big)\big| = 0.025 + 0.025 = 0.05$ This leap of logic represents a misunderstanding of the standard error equation and the misuse of the normal distribution curve (5) and the normal distribution function $\phi(z)$ which gives the probability of a standard normal variate to assume a value between $[0, z]$ or:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z exp[-z^2/2]dx \tag{6}$$

which may be reduced to:

$$\phi(z) = \frac{1}{2} erf\big(\frac{z}{\sqrt{2}}\big) \tag{7}$$

where erf is the error function (Abramowitz, 1972; Spanier and Oldham, 1987). The error function for z is given by:

$$erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \tag{8}$$

The leading factor $\frac{2}{\pi}$ is sometimes omitted. (Wittiker and Watson, 1990, p. 341). The error function satisfies the identity:

$$err(z) = 1 - erfc(z) \tag{9}$$

which may be reduced to:

$$erf(z) = \frac{2z}{\sqrt{\pi}} {}_1F_1\big(\frac{1}{2}; \frac{3}{2}; -z^2\big) \tag{10}$$

where $erf(z) = erfc$ and ${}_1F_1\big(\frac{1}{2}; \frac{3}{2}; -z^2\big)$ is a confluent of hypergeometric function of the first kind:

$${}_1F_1\big(\frac{1}{2}; \frac{3}{2}; -z^2\big) = 1 + \frac{a}{b}z + \frac{a(a+1)z^2}{b(b+1)2!} + \cdots \tag{10a}$$

which may be reduced to:

$${}_1F_1\big(\frac{1}{2}; \frac{3}{2}; -z^2\big) = \sum_{k=1}^{\infty} \frac{(a)_k z^k}{(b)_k k!} \tag{10b}$$

where $(a)_k$ and $(b)_k$ are Pochhammer symbols (Humbert, 1920, pp. 490-492).

Generally, the error function has the value between 0 and 1: $erf(0) = 0$ and $erf(\infty = 1$. For the normal distribution curve, the error is fixed at 0.05 as a standard error because it is the area under the curve bounded by an interval of two units of standard deviation covering 0.95 of the area under the curve produced by the Gaussian function (equation (5)). The use of 0.05 from the Gaussian curve's precision level (random error level) as the standard error in $SE = 1/\sqrt{n}$ for the purpose of determining minimum sample size is erroneous. Defining SE as the minimum sample error is also incorrect. SE is the standard error, not a sampling error. The misapplication comes in the form of: *fix the value of SE to 0.05 and solve for n.* The following calculation illustrates this point: $0.05 = 1/\sqrt{n}; 0.005(\sqrt{n}) = 1$; therefore $n = 400$.

Under this logic, the term n is equated to mean "minimum sample size." The fallacy of this logic would lead the requirement of minimum sample size to be 400 when the error level is 0.05. The value of 0.05 comes from the 5% random error in the normal distribution curve under 0.95 confidence interval. The second error of the logic of comes from the assumption that the data is normally distributed without even testing to verify whether it is indeed normally distributed. Equation (4) and its principal component: $\sqrt{n}$ is not a distribution. The application of 0.05 as the standard value for sampling error (4) is a misuse of the equation. The logic of using 0.05 as the "error level" is directly taken from the precision level of the probability distribution function (PDF) for a normal distribution. However, equation (4) does not determine the confidence level or the error level, i.e. random error of 0.05 explained by the t-equation and Z-equation from probability distribution of the normal curve. The corruption of $n = 400$ came from Weisberg and Bowen who insinuated the standard error equation to mean sampling error.

2 Alternative Sample Size Determination n-hat

Sample size determination may be categorized into two scenarios: finite population and non-finite population. Where the population is finite, the common sample size determination method is classified as population proportion method. The population proportion method has two requisites: (i) the population total must be known, and (ii) the distribution must be normal. The paper proposes an alternative sample size determination method. The first method called n-hat appeared in the proceeding of the Silapakorn University 70^{th} Anniversary International Conference 2013. (Luanglath & Rewtrakunphaiboon, 2013, pp. 127-139). Below are the steps for n-hat calculation introduced in the proceeding of that conference.

Step-1, the population is estimated from an initial sample randomly selected from a population. The test statistic is used to estimate the population mean. The sample test statistic is given by: $t = (\bar{x} - \mu)/(S/\sqrt{n})$, solve for the population mean μ, thus: $\mu = t\big(S_x/\sqrt{n}\big) - \bar{x}$.

Step-2, use the unit normal distribution Z-equation to solve for the estimated population standard deviation. The Z-equation is given by: $Z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$, solve for the population standard deviation σ, thus, $\sigma = [(\bar{x} - \mu)/Z]\sqrt{n}$.

In the foregoing two steps, n is the initial sample size and the standard confidence interval of 0.95 is used.

Step-3, with a given initial sample n, compute the expected alpha ($\hat{E} = \hat{\alpha}$) for the sample by using the expected error equation. The expected error equation is given by:

$$\hat{E} = \frac{n - n[1 - df(\alpha)]}{n} \text{ of simply } \hat{E} = df(\alpha) \tag{11}$$

where $df = n - 1$ and α is the specified error level, i.e 0.05.

Step-4, estimate the raw sample size by using the following equation:

$$\tilde{n} = \big(\frac{\sigma^2 n}{S^2}\big)^{-\hat{E}} \tag{12}$$

The value for $\tilde{n}$ is called a raw estimate because it is calculated on the basis of the estimated population variance, actual sample variance, and the estimated error. Since the error may run from 0.01 to 0.99, the value of $\tilde{n}$ does not provide an accurate estimate for the minimum sample size. It is necessary to engage in an experiment by allowing the errors to move towards the point estimate (population mean) from the maximum error and from the minimum error. The term $\tilde{n}$ is the error migration in the experiment where the error is allowed to move from 0.01 towards the estimated population mean (μ) and $\tilde{n}$ is the maximum estimated error allowed to moved from the upper random error region towards the point estimate μ. This experiment produces the following *specified error ratio*:

$$n^* = \frac{\tilde{n}_1^{0.99}}{n_1^{0.01}} \tag{13}$$

where $\tilde{n}_1^{0.99} = \tilde{n}_1/0.99$ and $\tilde{n}_1^{0.01} = \tilde{n}_1/0.01$.

As part of the experiment, the specified error also has a *specified error range* calculated by:

$$n_r = n_1^{0.99} - n_1^{0.01} \qquad 14)$$

From this range, the *specified error median* is determined by:

$$n_r^M = n_r/2 \tag{15}$$

The minimum sample size for an unknown population N is given by taking the square root of the specified error of minimum sample median thus:

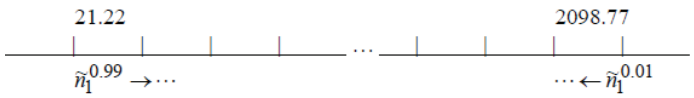
$$\hat{n} = \sqrt{n_r^M} \tag{16}$$

A numerical illustration of the calculation by the steps outlined above is in order. Assume that the following initial sample data is given as a set of series of independent events: $\{1, 1, 1, 1, 1, 11, 1, 0, 0\}$. There, eight events $n = 10$, the initial sample mean is $\bar{x} = 0.80$ and the standard deviation is $S_x = 0.4216$. The estimated population mean follows: $\mu = t(S_x/\sqrt{n}) - \bar{x} = 1.64(0.4216/3.16)$. Finally, $\mu = 0.59$.

Use the estimate error equation to calculate the estimated error among the initial sample: $\hat{E} = df(\alpha) = 9(0.05) = 0.45$.

Next, find the specified error ratio at 0.99 and 0.01 points: $n_1^{0.99} = \tilde{n}_1/0.99 = \frac{20.99}{0.99} = 21.22$, and 0.01 error: $\tilde{n}_1^{0.01} = \tilde{n}_1/0.01 = \frac{20.99}{0.01} = 2098.77$

The range of the minimum sample size estimate may be demonstrated by a line series thus:



The range of the specified error of the estimated sample size is: $n\_r = n_1^{0.99} - n_1^{0.01} = |21.22 - 2098.77| = 2077.57$.

The specified error maximum and minimum ratio is given by:

$$n^* = \tilde{n}_1^{0.99}/n_1^{0.01} = 21.22/2098.77 = 0.01.$$

The media for the range is given by: $n_r^M = \frac{n_r}{2} = 2077.57/2 = 1038.78$; the minimum sample size is given by: $n_{min} = \sqrt{n_r^M}$. Thus $\hat{n} = \sqrt{1038.78} = 32.23 \simeq 32$.

If this minimum sample is correct, it must meet the confidence interval test for the population estimate in the Z-equation $Z = \bar{x} - \mu/(\sigma/\sqrt{n})$ where n is substituted by $\hat{n} = n$. Recall that $Z_{0.95} = 1.65$. If the calculation $\hat{n} = 32.23$ is correct when substituting $\hat{n} = n$ in the Z-equation, the result must satisfy the following hypothesis statements: $H_0 : Z_{\hat{n}} < Z_{0.95}$; $H_A : Z_{\hat{n}} > Z_{0.95}$.

The test calculation follows: $Z = (\bar{x}-\mu)/(\sigma/\sqrt{\hat{n}}) = (0.80-0.59)/\sqrt{32.23}) = 21/0.07 = 2.91$. The explanatory power of $Z(2.91) = 0.9981$ or 99.81%. This means that the null hypothesis may be rejected because the decision rule for the null hypothesis is "accept the null hypothesis if $Z_{ons} > Z_{0.95}$. otherwise reject." In this case, the calculation shows that $Z_{obs} > Z_{0.95}$. The minimum sample size of $\hat{n} = 32.23$ is 99.81% accurate or the probability of error is 0.19% or 0.0019[1]

3 Multistage Nonfinite Population Method (MNP) for Minimum Sample Size Determination

In 2013, the author introduced the n-hat method. In this paper, the author proposes another alternative to minimum sample size determination called n-omega or Multi-stage Nonfinite Population (MNP) method. This new method is based on the specified alpha level. Using the random error: α level as the basis to calculate the sample size, n-omega method allows the researcher to determine minimum sample size at various level of confidence interval. A table of minimum sample size is provided for various confidence level. n-omega is an improvement over Weisberg-Bowen's heuristics approach based on $SE = 1/\sqrt{n}$

[1]Agresti, Alan and Franklin, Christine (2012). Statistics: The Art and Science Learning from Data, 3rd ed. Pearson Prentice Hall. Sect. 7.2; p. 321.

which is incorrect and not efficient. Under Weisberg-Bowen's approach, the minimum sample size for a 5% error tolerance is 400 counts whereas under MNP (n-omega) it is 33.72 or approximately 34. This number is consistent with other writers advocating for a minimum sample size of 30. Roscoe, for instance, suggested that a minimum sample size should be 30 (Roscoe, 1975, p. 163). Roscoe's rationale is that a sample of 30 ensures the benefits of the Central Limit Theorem. This minimum sample size of 30 was also echoed by Agresti and Franklin (Agresti & Franklin, 2012, p. 312). However, prior writers did not provide a precise method to arrive at the magic number 30. This writing proffers the following steps in MNP method.

Firstly, the estimated sample size called n_1 is obtained through the root of the conventional Specified Precision Estimation (SPE) method thus:

$$n_1 = \frac{Z\sigma}{E} \tag{17}$$

where $Z_{0.95} = 1.65$; $\sigma_{N(0,1)} = 1$ and $E = 0.05$ for 0.95 confidence interval. For other percentage confidence interval, the value for each parameter may be changed accordingly. For 0.95 confidence interval, the calculation follows: $n_1 = Z\sigma/E = 1.65(1)/0.05 = 1.65 = 33$.

Secondly, obtain the second estimate of the minimum sample size (n_2) by the conventional SPE method according to the following formula:

$$n_2 = \frac{Z^2\sigma^2}{E^2} \tag{18}$$

Following the above assumption for 0.95, the calculation follows: $n_2 = Z^2\sigma^2/E^2 = (1.65)^2(1)^2/(0.05)^2 = 2.723(1)/0.0025 = 1089$.

Thirdly, after knowing the raw range between 33 and 1089 of possible sample size, the square root of the median of the range between n_1 and n_2 is calculated as n_3 :

$$n_3 = \sqrt{\frac{n_1 - n_2}{2}} \tag{19}$$

The calculation for the 0.95 confidence interval example follows: $n_3 = \sqrt{(1089 - 33)/2}$ which is $n_3 = \sqrt{(1089 - 33)/2} = \sqrt{1056/2} = \sqrt{528} = 22.98$.

Fourthly, the raw estimate of 22.98 is put into a percentage range between 1% and 99% in order to codify the value of into a distribution space ω, thus:

$$\left.\begin{aligned} \omega_{max} &= n_3/0.01 \\ \omega_{min} &= n_3/0.99 \end{aligned}\right\} \longrightarrow \omega = \omega_{max} - \omega_{min}$$

The calculation for the distribution range follows: $\omega_{max} = 22.98/0.01 = 2297.83$ and $\omega_{max} = 22.98/0.99 = 23.21$. The actual range is $\omega = 2297.83 - 23.21 = 2274.61$.

Lastly, the minimum sample size is the square root of the median of the range, thus:

$$n_\omega = \sqrt{\frac{\omega}{2}} \tag{20}$$

The calculation in the example continues: $n = \sqrt{\omega/2} = \sqrt{2274.61/2} = \sqrt{1137.31} = 33.72$. The minimum sample size for 0.95 confidence interval with 0.05 error level is 33.72 or approximately 34 counts. A table for minimum sample size for variance error level and confidence interval is given in Table 2.

Table 1 Minimum Sample Table Using Multistage Nonfinite Population Method

%	$1-\alpha$	$\sigma_{N(0,1)}$	E	n_1	n_2	n_3	$\omega_{\max}$	$\omega_{\min}$	n
0.80	0.840	1.00	0.05	16.80	282.24	11.52	1,152.04	11.64	**23.88**
0.85	1.040	1.00	0.05	20.80	432.64	14.35	1,434.99	14.49	**26.65**
0.90	1.280	1.00	0.05	25.60	655.36	17.74	1,774.49	17.92	**29.64**
0.95	**1.650**	**1.00**	**0.05**	**33.00**	**1,089.00**	**22.98**	**2,297.83**	**23.21**	**33.72**
0.96	1.750	1.00	0.05	35.00	1,225.00	24.39	2,439.26	24.64	**34.75**
0.97	1.880	1.00	0.05	37.60	1,413.76	26.23	2,623.13	26.50	**36.03**
0.98	2.050	1.00	0.05	41.00	1,681.00	28.64	2,863.56	28.92	**37.65**
0.99	2.320	1.00	0.05	46.40	2,152.96	32.45	3,245.43	32.78	**40.08**
0.995	2.550	1.00	0.05	51.00	2,601.00	35.71	3,570.71	36.07	**42.04**
0.999	3.000	1.00	0.05	60.00	3,600.00	42.07	4,207.14	42.50	**45.63**
0.9999	3.291	1.00	0.05	65.82	4,332.27	46.19	4,618.69	46.65	**47.81**
1.00	5.500	1.00	0.05	110.00	12,100.00	77.43	7,742.74	78.21	**61.91**
								Mean	**38.31**
								Standard deviation	**10.34**

Note that n_ω (n-omega) adopts a different approach to minimum sample size calculation. Where n-hat requires the undertaking of a test sample, n-omega does not require an empirical test sample. The efficiency of n-omega is its reliance on the specified error level and the distribution of sample range ω-space making the link between the minimum sample and the normal distribution. Similar to the Yamane's approach, n_ω uses the alpha level as the basis. However, unlike Yamane's method, n_ω does not depend on known population size. In that aspect, n_ω is more useful than Yamane's population proportion approach. The mean for the minimum sample from 0.80 to 1.00 confidence interval is 38.31. If the confidence interval range is run from 50% to 100%, the mean minimum sample is 29.82 or approximately 30. The mathematical rationale for a minimum sample of 30 has been found.

The sample size at various confidence intervals from 0.50 to 1.0 is tested for distribution type under the Anderson-Darling test. The result of the Anderson-Darling test confirms that the sample set is not normally distributed: $A^2 = -7.90$ while the null hypothesis is $A^{*2} = -8.28$. . Under this circumstance, in order to select the range of minimum sample size, the probability of data occurrence is used as a guide. In so doing, a range of confidence interval between

Table 2 Mean Minimum Sample Size for 50% - 100% Confidence Interval is 30

%	$1-\alpha$	$\sigma_{N(0,1)}$	E	n_1	n_2	n_3	$\omega_{\max}$	$\omega_{\min}$	n
0.50	0.000	1.00	0.05	0.00	0.00	0.00	0.00	0.00	**0.00**
0.55	0.130	1.00	0.05	2.60	6.76	1.44	144.22	1.46	**8.45**
0.60	0.260	1.00	0.05	5.20	27.04	3.30	330.45	3.34	**12.79**
0.65	0.380	1.00	0.05	7.60	57.76	5.01	500.80	5.06	**15.74**
0.70	0.520	1.00	0.05	10.40	108.16	6.99	699.14	7.06	**18.60**
0.75	0.680	1.00	0.05	13.60	184.96	9.26	925.63	9.35	**21.40**
0.80	0.840	1.00	0.05	16.80	282.24	11.52	1,152.04	11.64	**23.88**
0.85	1.040	1.00	0.05	20.80	432.64	14.35	1,434.99	14.49	**26.65**
0.90	1.280	1.00	0.05	25.60	655.36	17.74	1,774.49	17.92	**29.64**
0.95	**1.650**	**1.00**	**0.05**	**33.00**	**1,089.00**	**22.98**	**2,297.83**	**23.21**	**33.72**
0.96	1.750	1.00	0.05	35.00	1,225.00	24.39	2,439.26	24.64	**34.75**
0.97	1.880	1.00	0.05	37.60	1,413.76	26.23	2,623.13	26.50	**36.03**
0.98	2.050	1.00	0.05	41.00	1,681.00	28.64	2,863.56	28.92	**37.65**
0.99	2.320	1.00	0.05	46.40	2,152.96	32.45	3,245.43	32.78	**40.08**
0.995	2.550	1.00	0.05	51.00	2,601.00	35.71	3,570.71	36.07	**42.04**
0.999	3.000	1.00	0.05	60.00	3,600.00	42.07	4,207.14	42.50	**45.63**
0.9999	3.291	1.00	0.05	65.82	4,332.27	46.19	4,618.69	46.65	**47.81**
1.00	5.500	1.00	0.05	110.00	12,100.00	77.43	7,742.74	78.21	**61.91**
								Mean	**29.82**
								Standard deviation	**15.48**

0.80 to 1.0 is recommended. The table below shows the various percentage probability of each occurrence according to confidence interval level.

The last four items: items 15, 16, 17 and 18 are used as the range of interest with corresponding sample size of 42.04, 45.63, 47.81 and 61.91 respectively. The probabilities of their occurrences are 0.785, 0.846, 0.877 and 0.981. These probabilities are subjected to the adjacency test to verify randomness. The adjacency test is given by: **???**

The purpose of the adjacency test is to verify the random nature of a data set. Assume that there is a data set: $x_1 : (i = 1.2. \cdots)$. The test statistic depends on the number of n. For $n > 25$, the test statistic is given by:

$$L_{n>25} = 1 - \left[\frac{\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2}{2\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] \tag{21}$$

For $n > 25$, equation (21) approximately follows a normal distribution with mean zero: $\bar{x} = 0$ and the variance is given by:

$$S_x^2 = \sqrt{\frac{(n-2)}{(n-1)(n+1)}} \tag{22}$$

Table 3 Percentage Probability of Occurrence of Each Minimum Sample Size

i	X_i	$2i$	$2i-1$	$\overline{X}$	$(X_i-\overline{X})$	S	$Z=(X_i-\overline{X})/S$	$F(Z)$
1	0	2	1	29.82	-29.82	15.48	-1.92636	0.0256
2	8.45	4	3	29.82	-21.37	15.48	-1.38049	0.0735
3	12.79	6	5	29.82	-17.03	15.48	-1.10013	0.1251
4	15.74	8	7	29.82	-14.08	15.48	-0.90956	0.1711
5	18.60	10	9	29.82	-11.22	15.48	-0.72481	0.2266
6	21.40	12	11	29.82	-8.42	15.48	-0.54393	0.2912
7	23.88	14	13	29.82	-5.94	15.48	-0.38372	0.3264
8	26.65	16	15	29.82	-3.17	15.48	-0.20478	0.4013
9	29.64	18	17	29.82	-0.18	15.48	-0.01163	0.4801
10	33.72	20	19	29.82	3.90	15.48	0.251938	0.599
11	34.75	22	21	29.82	4.93	15.48	0.318475	0.626
12	36.03	24	23	29.82	6.21	15.48	0.401163	0.655
13	37.65	26	25	29.82	7.83	15.48	0.505814	0.695
14	40.08	28	27	29.82	10.26	15.48	0.662791	0.745
15	**42.04**	**30**	**29**	**29.82**	**12.22**	**15.48**	**0.789406**	**0.785**
16	**45.63**	**32**	**31**	**29.82**	**15.81**	**15.48**	**1.021318**	**0.846**
17	**47.81**	**34**	**33**	**29.82**	**17.99**	**15.48**	**1.162145**	**0.877**
18	**61.91**	**36**	**35**	**29.82**	**32.09**	**15.48**	**2.072997**	**0.981**

For $n < 25$, the test statistic is given by:

$$L_{n<25} = \frac{\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{23}$$

The null hypothesis may be rejected if the test statistic lies outside of the lower and upper bound of the critical value. The hypothesis statement follows: $H_0 : L_{abs} < L_{0.05}$; $H_A : L_{obs} > L_{0.05}$. The critical value of L is given in a range of lower and upper value. If the test statistic value falls within this range, the null hypothesis cannot be rejected. Recall that the null hypothesis states that the pattern of the data is random. If the test statistic for the observation falls out of the range, i.e. lower than the lower bound or higher than the upper bound, the series is not random.

The result of the adjacency test shows that $L(obs) = 17.81$ compared to the range of the null hypothesis for random number: $0.78 < L(4) < 3.22$. The probabilities of the four minimum sample sizes are not random occurrence. The sample sizes for these four probabilities are: 42.20, 45.63, 47.81 and 61.91. The chi-square test for homogeneity follows:

$$\chi^2 \frac{(n-1)S^2}{\sigma^2} \tag{24}$$

The sample size is 4; the mean is 49.35; the sample variance is 75.80. The value for σ^2 may be estimated through the Z-equation.

Table 4 Critical Value of L at Various Significance Levels*

	Significance Level: α			
Two-sided	0.10		0.02	
One-sided	0.05		0.01	
n	*a*	*b*	*a*	*b*
4	**0.78**	**3.22**	0.63	3.37
5	0.82	3.18	0.54	3.46
6	0.89	3.11	0.56	3.44
7	0.94	3.06	0.61	3.39
8	0.98	3.02	0.66	3.34
9	1.02	2.98	0.71	3.29
10	1.06	2.94	0.75	3.25
11	1.10	2.90	0.79	3.21
12	1.13	2.87	0.83	3.17
15	1.21	2.79	0.92	3.08
20	1.30	2.70	1.04	2.98
25	1.37	2.63	1.13	2.87

*Lower bound = a and upper bound = b. Source: Hart, B.I. (1942). "Significance Level for the Mean Square Successive Difference to the Variance." *Annals of Mathematical Statistics*, **13**: 445-7.

The estimated variance is $\sigma^2 = (8.65)^2 = 74.88$. The chi-square test calculation follows:$\chi^2 = (n-1)S^2/\sigma^2 - (4-1)75.80/74.88 = 227.40/74.88 = 3.04$.

The hypothesis and decision rule follows: $H_0 : \chi^2_4 < 9.50; H_A : \chi^2_4 > 9.50$. "Reject the null hypothesis if $\chi^2_4 > 9.50$." According to the calculation, the null hypothesis cannot be rejected. It means that the data set 42.20, 45.63, 47.81 and 61.91 may be best fitted into the normal distribution curve. These minimum sample sizes are homogeneous and, therefore, the set of four minimum sizes 42.20, 45.63, 47.81 and 61.91 may used as the basis for estimating the estimated minimum sample size.

Recall that the descriptive statistics of these 4 items were: $\bar{x} = 49.35; S = 8.71$ and $n = 4$. The inferential statistics are: $\mu = 42.21$ and $\sigma = 8.65$. The range of the estimated mean is $2\sigma = 2(8.65) = 17.31$. Therefore, the upper limit is $\mu + 2\sigma = 42.21 + 17.31 = 59.52$. and the lower limit is $\mu - 2\sigma = 42.21 - 17.31 = 24.90$. The difference between the upper and lower limits is obtained by $Max - Min = 59.52 - 24.90 = 34.61$. The estimated minimum sample size is given by:

$$< n >= \sqrt{\frac{n^{0.01} - n^{-0.99}}{2}} \quad \text{or simply} \quad < n >= \sqrt{\frac{\hat{\omega}}{2}} \tag{25}$$

The calculation follows:

$$< n >= \sqrt{\frac{34.61^{-0.01} - 34.61^{-0.99}}{2}} = \sqrt{\frac{3461.42 - 34.96}{2}} = 41.39.$$

The estimated minimum sample is 41 counts or 40. The percentage confidence is as high as $\mu + 2\sigma = 0.81 + 0.16 = 0.97$ using the data set 0.785, 0.846, 0.877 and 0.981 as the basis. Recall the mean for the minimum sample size for the confidence interval from 0.50 to 1.0 was 30 counts. Therefore, a range between 30 and 40 counts would be reasonable sample size with confidence level of 0.95 to 0.97.

4 Conclusion

This paper has clarified the proper definition and use of the standard error equation. It is erroneous to use the standard error equation to determine minimum sample size. The misuse of the standard error equation as a source for minimum sample size determination is traced back to a 1971 book by Weisberg & Bowen. This paper points out that error. The paper provides two new sample size calculation methods. The first method called n-hat method was introduced in 2013. The second method called Multistage Nonfinite Population: MNP or n-omega (n_ω) method appears for the first time in this conference paper. Both methods provide efficient means for minimum sample size determination. MNP is a new contribution to the field of research methodology in social science.

References

[1] Abramowitz, M. and Stegun, I. A. (Eds.) (1972). "Error Function and Fresnel Integrals." Ch. 7 in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 297-309.

[2] Agresti, Alan and Franklin, Christine (2012), Statistics: The Art and Science Learning from Data, 3rd ed. Pearson Prentice Hall. Sect. 7.2; p. 321.

[3] Everitt, B.S. (2003), "The Cambridge Diction of Statistics", CUP. ISBN 0-521-81099-X. See entry for "standard error."

[4] Humbert, P. (1920), *Sur les function hypocylindriques,* C.R. Acd. Sci. Paris 171, 490-492.

[5] Luanglath, P. I. and Rewtrakulpaiboon (2013). *Determination of Minimum Sample Size for Film-Induced Tourism Research." Silapakorn 70th Anniversary International Conference 2013,* Towards the Next decade of Hospitality and Creative Economics: Looking Forward to 2020. December 1st - 3rd, 2013, Bangkok, Thailand, Conference Proceeding, pp. 127-139.

[6] Press, W.H.; Flannery, B.P., Teukoksky, S.A., and Vetterling, W.T. (1992), "Numerical Recipes in FORTRAN: The Art of Scientific Computing", 2nd ed. Cambridge, England: Cambridge University Press. p. 465.

[7] Spanier, J. and Oldham, K. B. (1987). *The Error Function erf(x) and Its Complement erf(x),* Ch. 40 in An Atlas of Functions, Washington, DC: Hemisphere, pp. 385-393.

[8] Watson, G. N. (1928), *Theorems Stated by Ramanujan (IV): Theorems on Approximate Integration and Summation of Series,* J. London Math. Soc. 3, 282-289.

[9] Weisberg, Jon A. & Bowen, Bruce D. (1971). "Introduction to data Analysis", p. 41.

[10] Whittaker, E. T. and Robinson, G. (1967). *The Error Function.* §92 in The Calculus of Observations: A Treatise on Numerical Mathematics, 4th ed. New York: Dover, pp. 179-182.

[11] Wittaker, E.T. & Watson, G.N. (1990), "A Course in Modern Analysis", 4th ed. Cambridge University Press, p. 341.