American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

Sampling Distribution: Impact of the Population Reliability on the Sample Size Determination

Dr. Banzi Wellars*

University of Rwanda. College of Science and Technology. Department of Mathematics.
Kigali city. PO BOX 3900. Rwanda
Email: banziwe10@yahoo.fr

Abstract

One of the spiniest problems in the elementary theory of the sampling is the determination of the size of the sample. Several authors tried to answer this question. In this paper, we present the methods of determination of the sample size in order to estimate to a given precision the population mean μ, the variance of the population being known or unknown. In the same way, we present the methods of determination of the size of the sample in order to estimate to a given precision the frequency f of a character in the population at different level of its reliability denoted here by p .the method uses the sampling error of the distribution of the specified statistic. Notice that the bigger the population reliability ($\cong 1$ or 100%), the smaller is the necessary sample size to get the best estimation of the parameter. The representativeness of the population doesn't essentially depend on the size of the sample; the size of the sample influences the precision of the measures; to achieve a highest accuracy, it is necessary to increase the sample size. The results of this research will help in determining the sample size to be drawn from different populations (finite or finite).

AMS (MSC 2000) Mathematics Subject Classification: 62D05, 62E99

Keywords: Sampling, Sample size, Estimation, Error

1. Introduction

In sampling analysis the most ticklish question is: what should be the size of the sample or how large or small should be the sample size “n”? If the sample size “n”is too small, it may not serve to achieve the objectives and if it is too large, we may incur huge cost and waste resources.

---

* Corresponding author.
E-mail address: banziwe10@yahoo.fr.

As a general rule, one can say that the sample must be of an optimum i.e., it should neither be excessively large nor too small. Technically, the sample size should be large enough to give a confidence interval of desired width and as such the size of the sample must be chosen by some logical process before sample is taken from the universe. In determining the sample size in a statistical experiment, we must know two things: How close we wish our estimate to be to the true value of the population parameter, and how certain we wish to be that our estimate will be within the selected number of units of the value of the parameter [6]. Furthermore, size of the sample should be determined by a researcher keeping in view the following points [4]:

Nature of universe: universe may be either homogeneous or heterogeneous in nature. If the items of the universe are homogeneous, a small sample can serve the purpose. But if the items are heterogeneous, a large sample would be required. Technically, this can be termed as the dispersion factor.

Number of classes proposed: If many class-groups are to be formed, a large sample would be required because a small sample might not be able to give a reasonable number of items in each class-group.

Nature of study: If items are to be intensively and continuously studied, the sample should be small. For a general survey the size of the sample should be large, but a small sample is considered appropriate in technical surveys.

Type of sampling: Sampling technique plays an important part in determining the size of the sample. A small random sample is apt to be much superior to a large bad badly selected.

Standard of accuracy and acceptable confidence level: If the standard of accuracy or the level of precision is to be kept high, we shall require relatively larger sample. For doubling the accuracy for fixed significance, the sample size has to be increased fourfold.

Availability of finance: In practice, size of the sample depends upon the amount of money available for the study purposes. This factor should be kept in view while determining the size of sample for large samples result in increasing the cost of sampling estimates.

Other considerations: nature of units, size of the population, size of questionnaire, availability of trained investigators or enumerators in case of structured interview, the conditions under which the sample is being conducted, the time available for completion of the study are few other considerations to which a researcher must pay attention while selecting the size of the sample [2,5].

There are two alternative approaches for determining the size of the sample. The first approach is "to specify the precision of estimation desired and then to determine the sample size necessary to insure it" and the second approach "uses Bayesian statistics to weight the cost of additional information against the expected value of the additional information". The first approach is capable of giving a mathematical solution, and as such is frequently used technique of determining " n ". The limitation of this technique is that it does not analyze the cost of gathering information *vis-a-vis* the expected value of information. The second approach is theoretically optimal, but it is seldom used because of the difficult involved in measuring the value of information. Hence, we shall mainly concentrate on the first approach [4].

2. Determination of Sample size trough the approach based on precision rate and confidence level

Researcher will have to specify the precision that he wants in respect of his estimates concerning the population parameters. For instance, a researcher may like to estimate the mean of the universe within “ ± 2 ” of the true mean with 95 per cent confidence. In this case we will say that the desired precision is ± 2 , i.e., if the sample mean is \$100, the true value of the mean will be not less \$98 and no more than \$102. In other words, all this means that the acceptable error denoted here by d , is equal to 2.

2.1. Sampling Distribution of Mean

2.1.1. The variance of the population known

Sample size when estimating a mean: the confidence interval for the universe mean μ is given by

$$\overline{X} \pm z_c \frac{\sigma}{\sqrt{n}} \qquad (2.1)$$

Where $\overline{X}$ is the sample mean and z_c the critical value of the standard normal variable corresponding the given confidence level. Suppose we have $\sigma = 4.8$ for our purpose. If the difference between μ and $\overline{X}$ or the acceptable error is to be kept within ± 4 of the sample mean with 95% confidence, then we can express the acceptable error d as equal to $d = z_c \times \frac{\sigma_P}{\sqrt{n}}$ or $4 = 1.96 \times \frac{4.8}{\sqrt{n}}$ hence

$$n = \frac{(1.96)^2 \times (4.8)^2}{(4)^2} = 5.531904 \approx 6$$

In a general way, if we want to estimate the mean μ in a population with standard deviation σ with an error no greater than $"d"$ by calculating a confidence interval with confidence corresponding to Z , the necessary sample size $"n"$ is obtained by [2]:

$$n = \frac{z^2 \sigma^2}{d^2} \qquad (2.2)$$

All this is applicable when the population happens to be infinite. But in case of finite population of size N , the above stated formula for determining sample size will become [4]

$$n = \frac{z^2 \times N \times \sigma^2}{(N-1)\, d^2 + z^2 \sigma^2} \qquad (2.3)$$

2.1.2. The population variance unknown

We considered the question of the determination of the number of measures previously while supposing known the standard deviation of the population. Actually this parameter is generally unknown or known pain, but often, the coefficient of variation can be considered fortunately like known or practically known. It is thus frequently for example in the biologic domain, because the errors of numerous variables are roughly proportional to the averages, so that the coefficients of variation are practically constant in conditions data.

While always supposing true the conditions of the previous section, the gotten previously relation becomes

Either $$n = \frac{z^2_{1-\frac{\alpha}{2}} \sigma^2}{d^2} \qquad (2.4)$$

or $$n = \frac{z^2_{1-\frac{\alpha}{2}} V^2}{d^2} \qquad (2.5)$$

In this last relation, V designates the coefficient of variation, expressed in percent of the average, and, d, the maximum relative mistake or the relative mistake margin, i.e., the mistake maximum corresponding d_r to one degree of confidence level $1-\alpha$ and expressed also in percent of the average.

For small values of n, the variable z doesn't follow the normal law anymore, but rather the law established by Student, the pseudonym of the English statistician W. G. Gosset that published in 1908, under this name, a survey carrying on this variable [1]. Then, the variable t following by definition a law of student with ν degrees of freedom is defined by the following probability density function

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad \text{for} \quad -\infty < t < \infty \qquad (2.6)$$

Where Γ denotes gamma function and ν the number of degrees of freedom. There exist as many probability curves densities as of values ν. Thus, for small values of n, the relation (2.5) is been able to convenient, because it must be used by successive approximation and, the formulas (2.5) become then,

$n = \frac{t^2_{1-\frac{\alpha}{2}} V^2}{d^2}$, where $t_{1-\frac{\alpha}{2}}$ dependent approximation of n, through the intermediary of number of degrees of freedom. However one can give a graphic representation of it as abacus. Inversely this abacus can also serve to determine graphically, in relative value, the half-length of one interval of confidence, knowing the coefficient of variation V and the efficient n [1, 5]

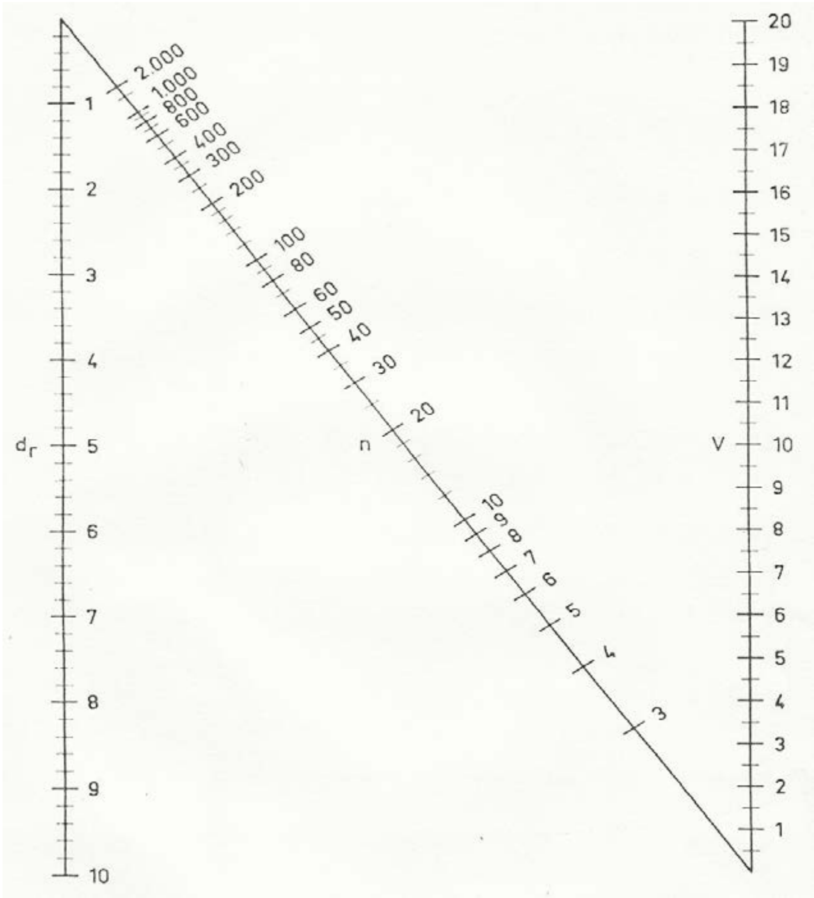

Figure1: Relation between the number of n measures, the coefficient of V variation, and the d margin mistake relative d_r , for a confidence level 95%,

When n is sufficiently big, the quantity $t_{1-\frac{\alpha}{2}}$ can be replaced by the corresponding value of the standard normal distribution. In these conditions, one has respectively for $\alpha = 0.05$ and $\alpha = 0.01$:

$$n \approx \frac{4V^2}{d_r^2} \quad \text{and} \quad n \approx \frac{7V^2}{d_r^2} \qquad (2.7)$$

Because and $z_{0.975} = 1.96 \approx 2$ and $z_{0.995} = 2.575 \approx \sqrt{7}$

Example:

The following table gives the weights of three shares of chickens, measures to three different dates: Averages, Standard deviations and Coefficient of variation estimated.

Table 2.1: Averages, standard deviation and coefficient of variation of three shares of chicken

Average	Standard deviation	Coefficient of variation
(gr)	(gr)	$\hat{V}$ (%)
1,067 1,308 1,518	134 166 164	12.5 12.7 10.8
1,022 1,242 1,421	138 162 169	13.5 13.0 11.9
876 1,094 1,310	98 136 172	11.2 12.5 13.1

Let's suppose that, by measure of prudence, that this estimated coefficient of variation is rounded to 13%, and that one wishes to estimate the average weight in the future with a maximum relative mistake of 5% (that means a maximum mistake of 50 gr for an average weight of 1,000gr, of 60 gr for a middleweight of 1,200gr, etc.). One has in these conditions:

$$n \approx 4\frac{13^2}{5^2} = 27$$

A result more exact $n = 28$ or $n = 29$ can be gotten with the help of the diagram, while joining by a right line the points $V = 13\%$ and $d_r = 5\%$, and while reading the corresponding value on the "n" axis. Inversely, one can wonder what precision one would get while weighting 50 chosen chickens at random in the considered population. One would have then, for a same degree of confidence level:

$$d_r \approx 2\frac{13}{\sqrt{50}} = 3.7\%$$

This result means that one would only commit 5% of the cases an error superior to 37 gr for an average weight of 1,000gr, to 44 gr for a middleweight of 1,200gr, etc.

2.1.3. *Approximation of the population Standard deviation under the normal law distribution*

In many cases the standard deviation of the population is not available. Since we have not yet taken the sample and are in the stage of deciding how large to make it (sample), we cannot estimate the population standard deviation. In such a situation, if we have an idea about the range (i.e., the difference between the highest and the lowest values of the population), we can use that to get a crude estimate of the standard deviation of the population for getting a working idea of the required sample size. We can get the said estimate of standard deviation as follows:

Since 99.73 per cent of the area under normal curve lies within the range ± 3 standard deviations, we may say that these limits include almost all of the distribution. Accordingly, we can say that the given range equals 6 standard deviations because of ± 3 . Thus, a rough estimate of the population standard deviation would be:

$$6\hat{\sigma} = the\ given\ range \Rightarrow \hat{\sigma} = \frac{the\ given\ range}{6}$$

If the range happens to be, say \$12, then $\hat{\sigma} = \frac{12}{6} = \$2$ and this estimate of standard deviation $\hat{\sigma}$ can be used to determine the sample size in the formulae stated above

2.1.4. *Case of Standard deviation and Coefficient of variation unknown*

When the standard deviation and the coefficient of variation are unknown or known pain, one can use a more rigorous method, presented notably by [4]. If one has an estimation $\hat{\sigma}^2$ of the variance of the population, with k_2 degrees of freedom, and if one sets an absolute error d_r , corresponding to a confidence level $1-\alpha$ and as it is only passed with a probability β , the number of measures to be done is:

$$n = t^2_{1-\frac{\alpha}{2}}\ F_{1-\beta}\ \frac{\hat{\sigma}^2}{d_r^2} \qquad (2.8)$$

The value $t_{1-\frac{\alpha}{2}}$ possessing $n-1$ degrees of freedom and the value $F_{1-\beta}$ of Snedecor possessing $k_1 = n-1$ and k_2 degrees of freedom. The most often, the introduction of the factor $F_{1-\beta}$ provokes an increase of n of the order of 25 to 75%, in relation to the case where the coefficient of variation is supposed known.

Example:

The following tables give the estimated values of “n”

Table 2.2: Approximate values of the sample size "n" needed for the error d_r =5% while the population variance is estimated to 4 with k_2 =30 degrees of freedom

d_r =5%			
α=0.05		α=0.01	
β=0.05	β=0.01	β=0.05	β=0.01
1,077,822	1,954,037	27,038,918	49,020,196
98,356	159,680	523,262	849,513
47,305	73,063	159,396	246,190
33,167	49,566	91,231	136,338
26,757	39,131	65,808	96,242
23,185	33,244	53,208	76,295
20,852	29,532	45,642	64,643
19,314	26,971	40,882	57,091
18,092	25,051	37,349	51,714
17,156	23,668	34,707	47,883
16,432	22,478	32,723	44,763
15,877	21,575	31,210	42,409
15,018	20,171	28,927	38,853
14,310	19,128	27,167	36,313
13,437	17,754	24,994	33,024
12,883	16,836	23,657	30,917
12,276	15,878	22,264	28,798
11,709	14,980	20,940	26,790
11,366	14,465	20,195	25,703
10,644	13,415	18,646	23,501
10,132	12,541	17,548	21,721
9,957	12,355	17,200	21,341

The next table gives the approximate values of the sample size "n" needed for the error d_r =10% while the population variance is estimated to 4 with k_2 =30 degrees of freedom.

We remark from the next tables that, as the error d_r increases, as the needed sample size decreases. We also remark that the sample size for all cases is relatively high; that is why this approach is not practical.

Table 2.2: Approximate values of the sample size “n” needed for the error d_r =10% while the population variance is estimated to 4 with k_2 =30 degrees of freedom.

dr=10%			
α=0.05		α=0.01	
β=0.05	β=0.01	β=0.05	β=0.01
269,456	488,509	6,759,729	12,255,049
24,589	39,920	130,815	212,378
11,826	18,266	39,849	61,548
8,292	12,392	22,808	34,084
6,689	9,783	16,452	24,060
5,796	8,311	13,302	19,074
5,213	7,383	11,410	16,161
4,828	6,743	10,220	14,273
4,523	6,263	9,337	12,929
4,289	5,917	8,677	11,971
4,108	5,620	8,181	11,191
3,969	5,394	7,802	10,602
3,754	5,043	7,232	9,713
3,578	4,782	6,792	9,078
3,359	4,438	6,249	8,256
3,221	4,209	5,914	7,729
3,069	3,970	5,566	7,200
2,927	3,745	5,235	6,697
2,841	3,616	5,049	6,426
2,661	3,354	4,662	5,875
2,533	3,135	4,387	5,430
2,489	3,089	4,300	5,335

2.2. Sampling Distribution of proportion (frequency)

The variance of the sampling distribution of proportion or frequency f is given by

$\sigma^2(f) = \frac{pq}{n}$ for infinite population or $\sigma^2(f) = \frac{pq}{n}\frac{N-n}{N-1}$ for an exhaustive sampling from a finite population of size N . Since the confidence interval for the universe proportion, $\hat{p}$ is given by

$f \pm z_c \times \sqrt{\frac{pq}{n}}$ then with the given precision rate, the acceptable error d can be expressed as follows:

$$d = z_c \sqrt{\frac{pq}{n}} \Leftrightarrow d^2 = z_c{}^2 \frac{pq}{n}$$
$$n = \frac{z_c{}^2 \times p \times q}{d^2} \qquad (2.9)$$

The formula gives the size of the sample in case of infinite population when we are to estimate the proportion in the universe. But in case of finite population the above stated formula becomes [4]

$$n = \frac{z_c{}^2 \times p \times q \times N}{(N-1) \times d^2 + \left(z_c{}^2 \times p \times q\right)} \qquad (2.10)$$

where p represents the population reliability and $q = 1 - p$. Notice that for a no reliable population we take $p = 0.5$

3. Determination of sample size trough the approach based on Bayesian statistics

This approach of determining "n" utilizes Bayesian statistics and as such is known as Bayesian approach. The procedure for finding the optimal value of "n" or the size of sample under this approach is as follows [4] expected value of the sample information (EVSI) for every possible n;

(i) Also workout reasonably approximated cost of taking a sample for every possible n;
(ii) Compare the EVSI and the cost of the sample for every possible n. In other words, workout the expected net gain (ENG) for every possible n as stated below:

For a given sample size (n):

(EVSI) – (Cost of sample) = (ENG) (2.11)

Form (iii) above the optimal sample size, that value of n which maximizes the difference between the EVSI and the cost of the sample, can be determined. The computation for every possible n and then comparing the same with the respective cost is often a very cumbersome task and is generally feasible with mechanized or computer help. Hence, this approach although being theoretically optimal is rarely used in practice.

4. Results

The following are the examples of determination of sample size knowing the error d and the population probability (reliability) p, for different population size N using the formula (2.10). the size would be rounded to the nearest positive integer.

Table 4.1: Sample size n from a population with N individuals and error d=1%

	p=0.5		p=0.6		p=0.7		p=0.8		p=0.9	
α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
100	98.98	99.41	98.94	99.38	98.79	99.30	98.41	99.08	97.22	98.37
120	118.53	119.15	118.47	119.11	118.26	118.99	117.72	118.67	116.01	117.66
150	147.71	148.67	147.61	148.61	147.28	148.42	146.45	147.93	143.80	146.36
200	195.94	197.64	195.77	197.54	195.19	197.19	193.73	196.33	189.12	193.57
250	243.68	246.31	243.43	246.16	242.51	245.62	240.27	244.29	233.20	240.02
300	290.94	294.70	290.58	294.49	289.28	293.72	286.08	291.81	276.12	285.74
400	384.04	390.63	383.41	390.25	381.15	388.90	375.62	385.56	358.61	375.02
500	475.30	485.44	474.33	484.86	470.87	482.77	462.46	477.62	436.94	461.55
700	652.51	671.78	650.67	670.66	644.18	666.66	628.52	656.89	582.28	626.86
1000	905.78	943.37	902.24	941.15	889.81	933.30	860.19	914.24	775.83	857.08
1500	1297.49	1376.05	1290.23	1371.33	1264.96	1354.72	1205.91	1314.93	1046.35	1199.79
2000	1655.43	1785.52	1643.64	1777.57	1602.84	1749.77	1509.18	1683.93	1267.29	1499.61
2500	1983.81	2173.59	1966.88	2161.83	1908.74	2120.85	1777.37	2024.88	1451.14	1764.11
3000	2286.12	2541.90	2263.68	2525.83	2187.00	2470.06	2016.24	2340.84	1606.51	1999.19
5000	3288.37	3844.96	3242.12	3808.31	3087.07	3682.91	2757.40	3402.80	2044.26	2725.61
7500	4211.54	5170.15	4135.98	5104.09	3886.92	4881.32	3378.33	4401.11	2366.72	3330.73
10000	4899.25	6246.62	4797.29	6150.43	4465.40	5829.83	3806.97	5157.69	2569.36	3746.62
20000	6488.53	9083.52	6310.89	8881.53	5748.78	8228.07	4701.80	6949.76	2947.97	4610.09
50000	8056.64	12485.74	7784.55	12107.25	6946.67	10924.51	5473.77	8780.19	3233.89	5349.87
100000	8762.53	14266.98	8441.62	13774.91	7465.19	12264.22	5790.69	9625.23	3341.93	5652.21
200000	9163.99	15362.81	8813.58	14793.76	7754.60	13065.34	5963.32	10111.82	3398.70	5816.56
500000	9423.02	16105.02	9052.93	15480.77	7939.28	13598.30	6071.93	10428.14	3433.70	5919.84
1000000	9512.65	16368.63	9135.62	15724.18	8002.81	13785.75	6109.02	10538.02	3445.53	5955.09
10000000	9594.79	16613.36	9211.35	15949.88	8060.86	13958.93	6142.78	10638.91	3456.25	5987.17
10000000	9603.08	16638.23	9218.99	15972.81	8066.71	13976.49	6146.18	10649.11	3457.32	5990.40

Table 4.2: Sample size n from a population with N individuals and error d=2 %

	p=0.5		p=0.6		p=0.7		p=0.8		p=0.9	
α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
N										
100	96.04	97.68	95.88	97.58	95.32	97.25	93.95	96.42	89.72	93.80
120	114.33	116.66	114.11	116.53	113.31	116.05	111.37	114.87	105.48	111.17
150	141.24	144.81	140.89	144.61	139.68	143.87	136.74	142.05	127.94	136.43
200	184.69	190.87	184.11	190.51	182.04	189.22	177.07	186.09	162.57	176.54
250	226.51	235.88	225.63	235.33	222.53	233.37	215.14	228.62	194.09	214.36
300	266.78	279.88	265.55	279.10	261.27	276.35	251.13	269.71	222.90	250.07
400	343.00	364.99	340.98	363.67	333.94	359.01	317.55	347.87	273.67	315.85
500	413.97	446.45	411.02	444.47	400.83	437.53	377.43	421.08	317.00	375.04
700	542.16	599.31	537.12	595.73	519.83	583.32	481.14	554.44	387.02	477.26
1000	706.18	806.37	697.64	799.91	668.75	777.68	606.02	727.17	463.87	599.87
1500	923.46	1102.69	908.91	1090.65	860.47	1049.72	759.30	959.70	548.60	749.67
2000	1091.36	1350.89	1071.09	1332.87	1004.44	1272.25	869.23	1142.35	603.74	856.63
2500	1225.00	1561.83	1199.51	1537.79	1116.54	1457.64	951.92	1289.61	642.48	936.83
3000	1333.89	1743.30	1303.72	1713.40	1206.28	1614.48	1016.38	1410.86	671.20	999.19
5000	1622.30	2271.06	1577.88	2220.57	1437.35	2057.20	1175.58	1737.61	737.09	1152.66
7500	1818.94	2676.15	1763.29	2606.30	1589.59	2384.07	1275.48	1965.17	775.13	1248.53
10000	1936.29	2938.19	1873.35	2854.20	1678.48	2589.83	1332.08	2102.87	795.67	1302.71
20000	2143.75	3444.02	2066.86	3329.19	1832.17	2974.95	1427.07	2349.85	828.59	1393.41
50000	2291.03	3840.76	2203.43	3698.49	1938.68	3266.38	1490.85	2527.99	849.69	1454.16
100000	2344.73	3994.12	2253.05	3840.50	1976.99	3376.64	1513.40	2593.53	856.96	1475.60
200000	2372.53	4075.49	2278.71	3915.67	1996.71	3434.61	1524.93	2627.59	860.64	1486.57
500000	2389.53	4125.93	2294.39	3962.20	2008.74	3470.36	1531.93	2648.46	862.87	1493.22
1000000	2395.25	4143.02	2299.66	3977.96	2012.78	3482.44	1534.28	2655.49	863.61	1495.45
10000000	2400.42	4158.52	2304.43	3992.25	2016.43	3493.39	1536.40	2661.85	864.29	1497.47

Table 4.3: Sample size n from a population with N individuals and error d=3 %

	p=0.5		p=0.6		p=0.7		p=0.8		p=0.9	
α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
N										
100	91.51	94.92	91.19	94.72	90.05	94.01	87.34	92.28	79.51	87.05
120	107.96	112.74	107.51	112.46	105.94	111.46	102.19	109.04	91.62	101.80
150	131.62	138.81	130.95	138.38	128.62	136.87	123.14	133.23	108.08	122.56
200	168.57	180.57	167.47	179.84	163.67	177.29	154.87	171.21	131.75	153.97
250	202.70	220.33	201.12	219.24	195.65	215.46	183.20	206.54	151.68	181.94
300	234.34	258.24	232.22	256.75	224.96	251.57	208.65	239.49	168.70	207.01
400	291.14	329.00	287.88	326.59	276.79	318.24	252.49	299.14	196.21	250.09
500	340.69	393.74	336.22	390.28	321.19	378.42	288.91	351.70	217.49	285.77
700	422.95	507.97	416.09	502.23	393.30	482.74	345.94	440.06	248.27	341.44
1,000	516.48	649.23	506.28	639.88	472.93	608.57	406.05	542.24	277.74	399.87
1,500	623.77	828.41	608.95	813.23	561.32	763.31	469.50	661.75	306.00	461.26
2,000	696.07	961.02	677.66	940.66	619.18	874.49	509.29	743.70	322.40	499.61
2,500	748.09	1,063.13	726.87	1,038.27	660.00	958.23	536.58	803.40	333.11	525.84
3,000	787.32	1,144.18	763.85	1,115.43	690.34	1,023.58	556.46	848.82	340.65	544.92
5,000	879.57	1,350.03	850.37	1,310.18	760.23	1,185.23	600.98	957.05	356.82	587.54
7,500	934.30	1,483.47	901.42	1,435.49	800.77	1,286.84	626.03	1,022.21	365.49	611.45
10,000	964.31	1,560.60	929.32	1,507.59	822.71	1,344.48	639.35	1,058.24	369.98	624.16
20,000	1,013.11	1,692.60	974.56	1,630.42	857.96	1,441.30	660.43	1,117.31	376.94	644.23
50,000	1,044.83	1,783.10	1,003.88	1,714.22	880.60	1,506.40	673.76	1,156.02	381.24	656.91

100,000	1,055.85	1,815.45	1,014.05	1,744.10	888.42	1,529.42	678.33	1,169.53	382.69	661.25
200,000	1,061.45	1,832.07	1,019.21	1,759.43	892.38	1,541.20	680.63	1,176.41	383.43	663.44
500,000	1,064.84	1,842.19	1,022.33	1,768.76	894.77	1,548.35	682.02	1,180.57	383.87	664.76
1,000,000	1,065.97	1,845.59	1,023.38	1,771.90	895.57	1,550.75	682.49	1,181.96	384.01	665.20
10,000,000	1,067.00	1,848.66	1,024.32	1,774.73	896.29	1,552.92	682.90	1,183.22	384.15	665.60

Table 4.4: Sample size n from a population with N individuals and error d=4 %

	p=0.5		p=0.6		p=0.7		p=0.8		p=0.9	
α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
N										
100	85.84	91.31	85.34	90.98	83.59	89.82	79.51	87.05	68.58	79.09
120	100.15	107.68	99.46	107.22	97.09	105.61	91.62	101.80	77.38	91.06
150	120.17	131.20	119.18	130.52	115.78	128.15	108.08	122.56	88.78	107.30
200	150.20	167.88	148.66	166.76	143.40	162.90	131.75	153.97	104.12	130.59
250	176.70	201.71	174.57	200.10	167.35	194.55	151.68	181.94	116.15	150.15
300	200.25	233.01	197.51	230.86	188.32	223.51	168.70	207.01	125.86	166.80
400	240.28	289.09	236.35	285.79	223.30	274.59	196.21	250.09	140.53	193.64
500	273.03	337.89	267.96	333.38	251.30	318.24	217.49	285.77	151.09	214.34
700	323.40	418.64	316.31	411.75	293.34	388.87	248.27	341.44	165.30	244.17
1,000	375.33	510.07	365.81	499.86	335.42	466.53	277.74	399.87	177.84	272.62
1,500	428.90	614.44	416.51	599.69	377.55	552.33	306.00	461.26	188.99	299.79
2,000	461.86	684.46	447.52	666.20	402.85	608.25	322.40	499.61	195.11	315.51

2,500	484.19	734.70	468.45	713.70	419.73	647.60	333.11	525.84	198.97	325.76
3,000	500.31	772.50	483.53	749.32	431.78	676.79	340.65	544.92	201.63	332.98
5,000	536.01	861.11	516.78	832.40	458.11	743.83	356.82	587.54	207.18	348.40
7,500	555.84	913.50	535.19	881.26	472.51	782.59	365.49	611.45	210.07	356.66
10,000	566.31	942.17	544.90	907.90	480.05	803.53	369.98	624.16	211.54	360.94
20,000	582.79	988.70	560.13	951.03	491.84	837.13	376.94	644.23	213.79	367.56
50,000	593.14	1,018.89	569.69	978.93	499.19	858.67	381.24	656.91	215.16	371.65
100,000	596.67	1,029.37	572.94	988.60	501.69	866.09	382.69	661.25	215.63	373.03
200,000	598.46	1,034.69	574.59	993.51	502.94	869.86	383.43	663.44	215.86	373.72
500,000	599.53	1,037.91	575.58	996.47	503.70	872.13	383.87	664.76	216.00	374.14
1,000,000	599.89	1,038.98	575.91	997.47	503.96	872.89	384.01	665.20	216.04	374.28
10,000,000	600.21	1,039.95	576.21	998.36	504.18	873.58	384.15	665.60	216.09	374.41

Table 4.5: Sample size n from a population with N individuals and error d=5%

	p=0.5		p=0.6		p=0.7		p=0.8		p=0.9	
α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
N										
100	79.51	87.05	78.84	86.59	76.52	84.96	71.29	81.14	58.28	70.76
120	91.62	101.80	90.73	101.16	87.67	98.94	80.86	93.80	64.50	80.18
150	108.08	122.56	106.84	121.64	102.62	118.44	93.40	111.13	72.21	92.49
200	131.75	153.97	129.90	152.51	123.71	147.50	110.53	136.32	82.00	109.26
250	151.68	181.94	149.24	179.90	141.11	172.97	124.21	157.78	89.27	122.60
300	168.70	207.01	165.68	204.37	155.72	195.47	135.37	176.28	94.88	133.47
400	196.21	250.09	192.13	246.24	178.85	233.43	152.51	206.55	102.96	150.09
500	217.49	285.77	212.49	280.76	196.36	264.21	165.04	230.27	108.50	162.21

700	248.27	341.44	241.77	334.31	221.09	311.09	182.15	265.07	115.62	178.71
1,000	277.74	399.87	269.63	390.12	244.15	358.85	197.50	298.95	121.60	193.46
1,500	306.00	461.26	296.17	448.32	265.71	407.51	211.36	331.95	126.70	206.74
2,000	322.40	499.61	311.51	484.47	277.98	437.14	219.04	351.35	129.41	214.09
2,500	333.11	525.84	321.50	509.09	285.90	457.09	223.93	364.11	131.10	218.75
3,000	340.65	544.92	328.52	526.95	291.44	471.43	227.31	373.15	132.25	221.97
5,000	356.82	587.54	343.52	566.70	303.19	502.99	234.38	392.63	134.60	228.71
7,500	365.49	611.45	351.55	588.92	309.42	520.41	238.09	403.16	135.81	232.24
10,000	369.98	624.16	355.71	600.69	312.64	529.58	239.99	408.64	136.42	234.05
20,000	376.94	644.23	362.13	619.26	317.59	543.96	242.89	417.15	137.35	236.80
50,000	381.24	656.91	366.10	630.96	320.63	552.96	244.66	422.42	137.92	238.49
100,000	382.69	661.25	367.44	634.96	321.66	556.03	245.26	424.21	138.11	239.06
200,000	383.43	663.44	368.12	636.98	322.18	557.58	245.56	425.11	138.20	239.34
500,000	383.87	664.76	368.52	638.20	322.49	558.51	245.74	425.65	138.26	239.52
1,000,000	384.01	665.20	368.66	638.61	322.59	558.83	245.80	425.83	138.28	239.57
10,000,000	384.15	665.60	368.78	638.97	322.68	559.11	245.86	425.99	138.30	239.62

Table 4.6: Sample size n from a population with N individuals and error d=6%

	p=0.5		p=0.6		p=0.7		p=0.8		p=0.9	
α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
N										
100	72.93	82.36	72.12	81.76	69.36	79.68	63.30	74.93	49.24	62.70
120	82.98	95.43	81.93	94.63	78.38	91.85	70.71	85.58	53.59	69.97
150	96.25	113.44	94.83	112.30	90.10	108.40	80.10	99.76	58.79	79.14
200	114.55	139.81	112.55	138.08	105.93	132.23	92.36	119.57	65.10	91.08
250	129.31	162.48	126.76	160.14	118.42	152.32	101.69	135.75	69.59	100.15
300	141.46	182.17	138.41	179.23	128.52	169.49	109.04	149.20	72.93	107.27
400	160.28	214.69	156.38	210.62	143.86	197.28	119.87	170.31	77.60	117.73
500	174.19	240.44	169.58	235.35	154.95	218.81	127.47	186.10	80.70	125.04
700	193.36	278.64	187.70	271.83	169.93	249.98	137.42	208.16	84.56	134.60

1000	210.76	316.34	204.05	307.58	183.22	279.89	145.96	228.48	87.70	142.79
1500	226.62	353.54	218.88	342.63	195.08	308.61	153.38	247.24	90.32	149.88
2000	235.48	375.62	227.13	363.33	201.60	325.30	157.38	257.83	91.68	153.70
2500	241.14	390.25	232.39	376.99	205.73	336.21	159.88	264.63	92.52	156.08
3000	245.07	400.65	236.04	386.69	208.58	343.89	161.59	269.37	93.09	157.71
5000	253.31	423.21	243.67	407.66	214.52	360.38	165.13	279.37	94.25	161.08
7500	257.65	435.47	247.68	419.02	217.62	369.22	166.96	284.65	94.84	162.82
10000	259.87	441.87	249.74	424.95	219.20	373.81	167.89	287.37	95.14	163.70
20000	263.28	451.83	252.88	434.15	221.62	380.91	169.30	291.54	95.59	165.04
50000	265.37	458.02	254.81	439.86	223.10	385.31	170.16	294.11	95.86	165.86
100000	266.07	460.13	255.45	441.80	223.59	386.79	170.45	294.97	95.95	166.14
200000	266.42	461.19	255.78	442.78	223.84	387.54	170.59	295.40	95.99	166.27
500000	266.64	461.82	255.98	443.37	223.99	387.99	170.68	295.67	96.02	166.35
1000000	266.71	462.04	256.04	443.56	224.04	388.14	170.71	295.75	96.03	166.38
10000000	266.77	462.23	256.10	443.74	224.09	388.27	170.73	295.83	96.04	166.41

Table 4.7: Sample size n from a population with N individuals and error d=10%

	p=0.5		p=0.4		p=0.3		p=0.2		p=0.1	
α	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
N										
100	49.24	62.70	48.22	61.74	44.90	58.54	38.30	51.83	25.88	37.70
120	53.59	69.97	52.39	68.77	48.48	64.82	40.87	56.67	27.02	40.18
150	58.79	79.14	57.34	77.61	52.69	72.61	43.81	62.53	28.25	43.01
200	65.10	91.08	63.32	89.06	57.69	82.52	47.20	69.72	29.60	46.28
250	69.59	100.15	67.55	97.71	61.18	89.89	49.49	74.90	30.48	48.48
300	72.93	107.27	70.70	104.47	63.74	95.57	51.16	78.79	31.09	50.07
400	77.60	117.73	75.08	114.36	67.27	103.78	53.39	84.27	31.90	52.22
500	80.70	125.04	77.98	121.25	69.59	109.41	54.83	87.95	32.40	53.59
700	84.56	134.60	81.57	130.22	72.43	116.66	56.58	92.55	32.99	55.26

1,000	87.70	142.79	84.49	137.87	74.72	122.75	57.96	96.34	33.45	56.57
1,500	90.32	149.88	86.91	144.46	76.60	127.95	59.08	99.50	33.82	57.64
2,000	91.68	153.70	88.18	148.01	77.58	130.71	59.66	101.17	34.00	58.19
2,500	92.52	156.08	88.95	150.21	78.18	132.43	60.01	102.19	34.12	58.53
3,000	93.09	157.71	89.48	151.72	78.59	133.60	60.25	102.88	34.19	58.75
5,000	94.25	161.08	90.55	154.84	79.41	136.01	60.73	104.30	34.34	59.21
7,500	94.84	162.82	91.09	156.44	79.83	137.24	60.97	105.03	34.42	59.44
10,000	95.14	163.70	91.37	157.26	80.04	137.87	61.10	105.39	34.46	59.56
20,000	95.59	165.04	91.78	158.50	80.35	138.82	61.28	105.94	34.52	59.73
50,000	95.86	165.86	92.03	159.25	80.55	139.40	61.39	106.28	34.55	59.84
100,000	95.95	166.14	92.11	159.50	80.61	139.59	61.43	106.39	34.56	59.87
200,000	95.99	166.27	92.16	159.63	80.64	139.69	61.45	106.45	34.57	59.89
500,000	96.02	166.35	92.18	159.70	80.66	139.75	61.46	106.48	34.57	59.90
1,000,000	96.03	166.38	92.19	159.73	80.67	139.77	61.46	106.49	34.57	59.90
10,000,000	96.04	166.41	92.20	159.75	80.67	139.78	61.47	106.50	34.57	59.91

5. Concluding remarks

The size of the sample affects the accuracy of the inference made about the population. The larger the sample, the closer one is to measuring the population itself. Therefore, large samples are more useful than small samples, if the large samples are unbiased, i.e., random. On the other hand, a small unbiased sample is certainly more useful than a large biased sample. Therefore, the single most important characteristic of a good sample is not its size but its representative nature, i.e., randomness. It must be realized that the time and expense involved in increasing sample sizes beyond certain limits may result in diminishing returns.

The tables above show that for the sampling distribution of the frequency, the necessary sample size to achieve the desired accuracy depends upon the reliability of the population; the bigger the population reliability, the smaller is the sample size required.

References

[1] Dagnelie P., P*rincipes d'experimentation*. Gembloux. Belgium. 2001, pp 88-92.

[2] Gurumani N., *An introduction to Biostatistics*. 2nd Edition. MJP Publishers. India, 2005, pp. 185-186.

[3] Harris H.M, Horvitz D.G., Mood A.M. On the determination of sample sizes in designing experiments. J. *Amer. Statist*. 35. USA. 1948. Pp. 391-402

[4] Kothari. C.R., *Research Methodology*. Methods and Techniques. New Age International (P) Ltd., Publishers. Dharmesh Art Process, Delhi. India, 2008, pp. 174-180

[5] Lenth R.V., Some practical guidelines for effective sample determination. *Am. Statist*. 55(3), 2001. Pp. 187-193

[6] Roger C. Pfaffenberger, James H. Patterson, *Statistical methods*. Third edition. IRWIN, Homewood. Illinois. USA, 1987, pp. 387-392