Journal of Empirical Studies
2014 Vol. 1, No. 1, pp. 23-29
ISSN(e): 2312-6248
ISSN(p): 2312-623X

DEVELOPMENT OF LOGARITHMIC EQUATIONS FOR STATISTICAL SAMPLE DETERMINATION

Moayad B. Zaied[1] --- Ahmed M. El Naim[2] --- Dafalla M. Mekki[3]

[1]*Department of Agric. Eng., Faculty of Natural Resources and Environmental Studies, University of Kordofan, Elobied, Sudan*

[2]*Department of Crop Sciences, Faculty of Natural Resources and Environmental Studies, University of Kordofan, Elobied, Sudan*

[3]*Department of Animal Production, Faculty of Natural Resources and Environmental Studies, University of Kordofan, Elobied, Sudan*

ABSTRACT

Sample size estimation is a fundamental step in designing trials and studies for which the primary objective is the estimation or the comparison of parameters. In this paper, the equation of sample for proportion was used to develop two equations for sample size determination. The resultant equations are natural and normal logarithmic. The validation test was conducted for populations with different sizes from 10 to 100000 from which sample size was calculated applying the equation of sample for proportion, finite population correction for proportion equation and the developed equations at 0.05 and 0.01 levels of significance. It was found that the sample size calculated by natural logarithmic equation was larger than sample sizes calculated by proportion, finite population correction for proportion, and normal logarithmic equation. Sample size calculated by normal logarithmic equation was the smallest up to 1000 population size at 0.05 level of significance and up to 10000 population sizes at 0.01 levels.

Keywords: Sample size, Hypothesis testing, Validation tests, Population prevalence.

1. INTRODUCTION

A study that collects too much data is or insufficiently precise and lacks the power to reject a false null hypothesis is a waste of time and money. Sample size calculation for a study estimating a population prevalence has been shown in many authors (Lwanga and Lemeshow, 1991; Daniel, 1999). The basic and general concepts of statistical inference, including estimation and hypothesis testing are presented by many workers: Gardner and Altman (1986; 1988), Machin *et al.* (1997), Friedman *et al.* (1998), Altman *et al.* (2000), Sackett (2001), Armitage *et al.* (2002) and Piantadosi (2005). The objective of this study is to calculate an adequate size sample to developed of logarithmic equation for statistical sample determination. It can be determined using a simple formula.

1.1. Calculation of the Sample Size

The following equation of sample for proportion

$$n = \frac{N}{1+\alpha^2 N} \text{...}(1)$$

Where,

N = population size.

n = sample size.

α = significance level.

The following simple formula (Daniel, 1999) can be used:

$$n_0 = \frac{Z^2 P(1-P)}{d^2} \text{..}(2)$$

Where,

n_0 = the sample size without considering the finite population correction factor.

Z = level of confidence,

P = expected prevalence or proportion (in proportion of one; if 20%, P = 0.2), and

d = precision (in proportion of one; if 5%, d = 0.05).

In these studies, investigators present their results with 95% confidence intervals (CI).

Expected proportion (P): This is the proportion (prevalence) that investigators are going to estimate by the study. Sometimes, investigators feel a bit puzzled and a common response is that '*We don't know this P*. From the formula, it can be conceived that the sample size varies inversely with the square of the precision ($d2$) (Naing *et al.*, 2006).

Smaller d (good precision or smaller error of estimate)..

Again, d in the formula should be a proportion of one rather than percentage (Naing *et al.*, 2006).

The equation of finite population correction for proportion is given as follow

$$n = \frac{n_0 N}{n_0 + (N-1)} \text{...}(3)$$

2. METHODOLOGY

2.1. Assumptions

Consider the following equation of sample for proportion

$$n = \frac{N}{1+\alpha^2 N}$$

By conducting the division operation for the right side of Equation 1, the following form will result

$$\frac{1}{\alpha^2} - \frac{1}{\alpha^2} \times \frac{1}{\alpha^2 N + 1}$$

Assume that,

$$\frac{1}{1+\alpha^2 N} \text{ is low value and } 0 \le \frac{1}{1+\alpha^2 N} \le \frac{dn}{dN}$$

2.2. Boundary Conditions

n (0 – n), N (0 – N)

By differentiating the middle and right terms of the above assumed inequality,

$$\frac{dn}{dN} = \frac{1}{1+\alpha^2 N} \quad \text{...(4)}$$

2.3. Equation Derivation

The following equation will be obtained by separating the variables in Eq. (4):

$$dn = \frac{dN}{1+\alpha^2 N} \quad \text{...(5)}$$

Applying the determinate integration for the mentioned boundary conditions and proceeding in solution

$$\int_0^n dn = \int_0^N \frac{dN}{1+\alpha^2 N}$$

$$[n]_0^n = \frac{1}{\alpha^2}\left[\ln(1+\alpha^2 N)\right]_0^N$$

$$n - 0 = \frac{1}{\alpha^2}\left[\ln(1+\alpha^2 N) - \ln(1+0)\right]$$

$$n = \frac{\ln(1+\alpha^2 N)}{\alpha^2} \quad \text{...(6)}$$

Equation (6) is the new developed natural logarithmic model for large sample size, a model for smaller sample size can be obtained by replacing natural logarithm with normal logarithm as follow

$$n = \frac{\log(1+\alpha^2 N)}{\alpha^2} \quad \text{...(7)}$$

2.4. Validity Test

Sample sizes for different sizes of population were calculated using equation of sample for proportion (Eq. 1), equation of finite population correction for proportion (Eq.3) and new Equations (Eq. 6 and Eq. 7). The results of sample sizes obtained from new equations were then compared to results obtained from Equation 1 and Equation 3 in histograms.

3. RESULTS AND DISCUSSION

The sample sizes were calculated by the different equations at two levels of significance. Fig. 1 and Table 1 showed the sample sizes calculated by the different equations at 0.05 level of significance while Table 2 and Fig. 2 demonstrated the sample sizes calculated by the different equations.

In Fig. 1 It was shown that for 10 population size, the sample size calculated with proportion, finite population correction for proportion and natural logarithmic equation at 0.05 level of significance was equal to population size except in case of normal logarithmic equation, the sample size was 40 % of population size. For 20 population size, the size of sample calculated with natural logarithmic equation was equal to population size while other equations resulted in smaller sample sizes. For population sizes from 30 to 100000, the sample sizes calculated with natural logarithmic equation were larger than these calculated with other equations. The logarithmic equation resulted in smaller sample size from 10 to 750 population size as compared to proportion and finite population correction for proportion equations. In Fig. 2 It was shown that for 10 to 70 population size, the sample size calculated with proportion, finite population correction for proportion and natural logarithmic equation at 0.01 level of significance was equal to population size except in case of normal logarithmic equation, the sample sizes were the smallest for population size from 10 to 10000. For population size from 80 to 100000, the sizes of sample calculated with natural logarithmic equation were the largest while other equations resulted in smaller sample sizes. Normal logarithmic equation resulted in largest sample size for 100000population as compared to proportion and finite population correction for proportion equations. It can be concluded that the developed equations can be used for sampling.

4. CONCLUSION

The equation of sample for proportion was used to develop two equations for sample size determination. The resultant equations are natural and normal logarithmic. The validation test was conducted for populations with different sizes from which sample size was calculated applying the equation of sample for proportion, finite population correction for proportion and the developed equations at levels of significance (0.05 0.01). It was found that the sample size calculated by natural logarithmic equation was larger than sample sizes calculated by proportion, finite population correction for proportion, and normal logarithmic equation. While logarithmic equation resulted in the smallest sample sizes for most population sizes.

The developed equations were valid to be used for sampling.

REFERENCES

Altman, D., D. Machin, T. Bryant and S. Gardner, 2000. Statistics with confi dence: Confidence intervals and statistical guidelines. 2nd Edn., BMJ Books.

Armitage, P., G. Berry and J.N.S. Matthews, 2002. Statistical methods in medical research. 4th Edn., Blackwell Science.

Daniel, W.W., 1999. Biostatistics: A foundation for analysis in the health sciences. 7th Edn., New York: John Wiley & Sons.

Friedman, L., C. Furberg and D.L. DeMets, 1998. Fundamentals of clinical trials. 3rd Edn., New York: Springer-Verlag.

Gardner, M.J. and D.G. Altman, 1986. Confidence intervals rather than P values: Estimation rather than hypothesis testing, BMJ (292): 746–750.

Gardner, M.J. and D.G. Altman, 1988. Estimating with confidence, BMJ (296): 1210–1211.

Lwanga, S. and S. Lemeshow, 1991. Sample size determination in health studies: A practical manual. World Health Organization, Geneva.

Machin, D., M. Campbell, P. Fayers and A. Pinol, 1997. Sample size tables for clinical studies. 2nd Edn., London: Blackwell Science.

Naing, L., T. Winn and B.N. Rusli, 2006. Sample size calculation for prevalence studies. Archives of Orofacial Sciences, 1: 9-149.

Piantadosi, S., 2005. Clinical trials: A methodologic perspective. 2nd Edn., Hoboken, New Jersey: John Wiley and Sons.

Sackett, D.L., 2001. Randomized controlled trials . Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need. Canadian Medical Association Journal, 165: 1226–1236.

BIBLIOGRAPHY

Cochran, W.G., 1977. Sampling techniques. 3rd Edn., New York: John Wiley & Sons.

Freedman, L.S., 1982. Tables of the number of patients required in clinical trials using the logrank test. Statistics in Medicine, 1(2): 121–129.

Macfarlane, S.B., 1997. Conducting a descriptive survey: 2. Choosing a sampling strategy. Trop Doct., 27(1): 14-21.

Table-1. Sample size as calculated by proportion, finite population correction for proportion, natural logarithmic and normal logarithmic equation at $\alpha = 0.05$

N	S.P. Eq.	F. P.C.P Eq.	LN Eq.	LOG Eq.
10	10	10	10	4
20	19	19	20	8
30	28	27	29	13
40	36	35	38	17
50	44	42	47	20
60	52	48	56	24
70	60	55	65	28
80	67	61	73	32
90	73	66	81	35
100	80	71	89	39
150	109	93	127	55
200	133	111	162	70
250	154	124	194	84
300	171	135	224	97
350	187	145	251	109
400	200	153	277	120
450	212	159	302	131
500	222	165	324	141
750	261	185	422	183
1000	286	198	501	218

10000	385	240	1303	566
100000	398	245	2210	960

S.P. Eq.: Sample proportion equation, F. P.C.P Eq.: finite population correction for proportion equation, LN Eq.: natural logarithmic equation LOG Eq.: normal logarithmic equation

Table-2. Sample size as calculated by proportion, finite population correction for proportion, natural logarithmic and normal logarithmic equation at $\alpha = 0.01$

N	S.P. Eq.	F. P.C.P Eq.	LN Eq.	LOG Eq.
10	10	10	10	4
20	20	20	20	9
30	30	30	30	13
40	40	40	40	17
50	50	50	50	22
60	60	60	60	26
70	70	70	70	30
80	79	79	80	35
90	89	89	90	39
100	99	99	100	43
150	148	148	149	65
200	196	196	198	86
250	244	244	247	107
300	291	292	296	128
350	338	339	344	149
400	385	386	392	170
450	431	432	440	191
500	476	478	488	212
750	698	701	723	314
1000	909	914	953	414
10000	5000	5158	6931	3010
100000	9091	9625	23979	10414

S.P. Eq.: Sample proportion equation, F. P.C.P Eq.: finite population correction for proportion equation, LN Eq.: natural logarithmic equation LOG Eq.: normal logarithmic equation.

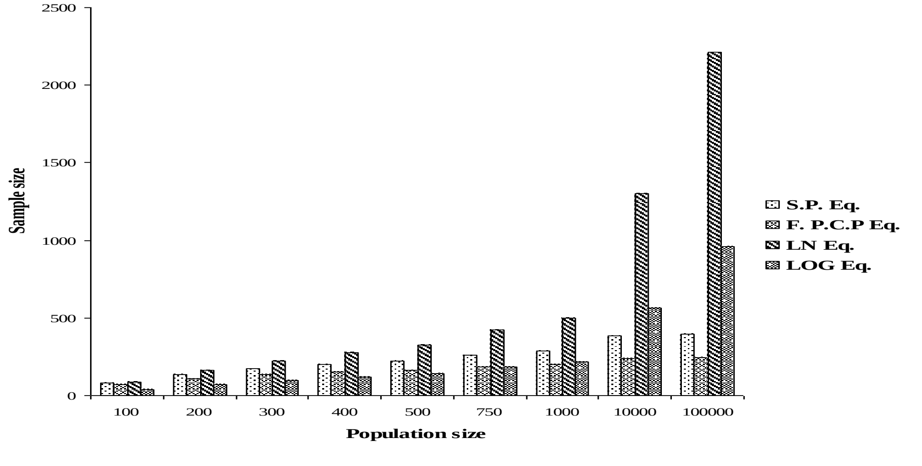

Fig. 1 Sample size by different equatioms at 0.05 level of significance

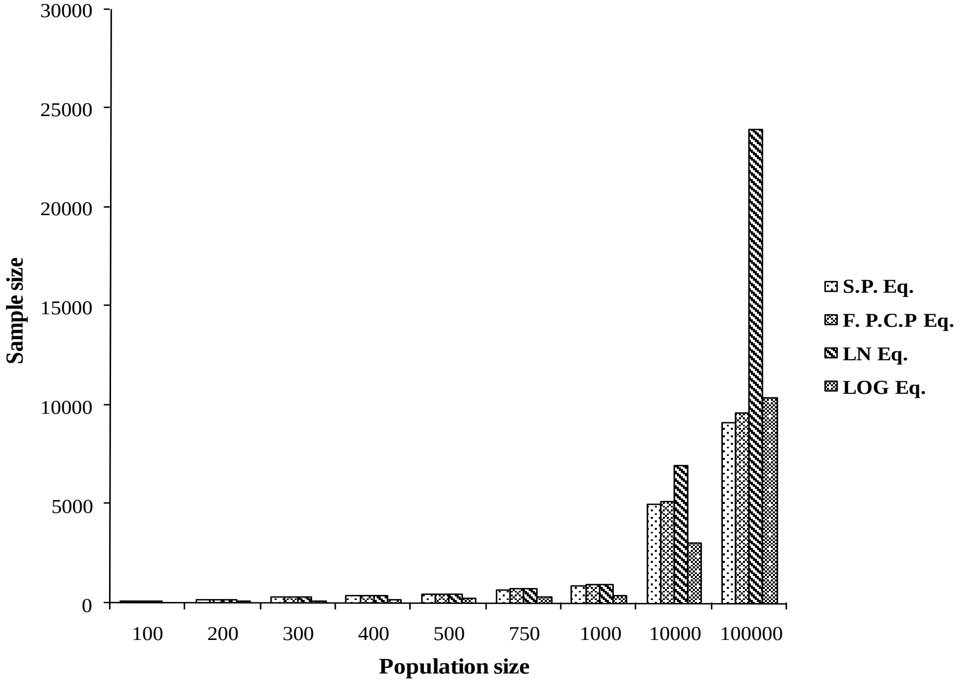

fig.2 Sample size by different equations at 0.01 level of significance