SAMPLE-SIZE DETERMINATION AND INFERENTIAL STATISTICAL TECHNIQUE

Vernon W. Stone and Russell W. Irvine, Georgia State University

INTRODUCTION

More frequently it is the rule, rather than the exception, for a researcher to report the use of a given sample size without indicating the basis for the determination thereof. Just as we expect to be apprised of the research design and the statistical design, we should expect sufficient information with respect to the sampling design, of which the determination of sample size is a fundamental dimension.

This presentation may be considered an initial taxonomical effort of sample-size formulae. However, the formulae presented, of course, constitute nothing more than a sample of research situations.

Part I, "Selected Sample-Size Computational Approaches," is predicated upon a priori data, affording, for the most part, the direct sample-size determinations. Included therein are formulae for the estimation of the population mean, the estimation of the population proportion, determination of sample size for t-test, determination of sample size for X^2, and determination of sample size for the F-test. Some direct and indirect duplication of formulae are to be observed. Numerical values may be substituted for the purpose of demonstrating the means of attaining the desired applicatory results.

Part II, "Selected Sample-Size Tabular Approaches," is predicated upon a posteriori data, affording, for the most part, the indirect sample-size determinations. Included therein are formulae for the estimation of the population mean, the estimation of the population proportion, the determination of sample size for t-test, the determination of sample size for X^2, the determination of sample size for the F-test, and the determination of sample size for r, including correlation coefficients based upon the continuous interval scale.

It must be recognized that the Type I and II Errors may be either one-tailed or two-tailed, that is, one side or two sides. The Type I Error refers to alpha (α), the probability of erroneously rejecting the null hypothesis. Accordingly, one minus alpha $(1-\alpha)$ represents the probability of not rejecting the null hypothesis, that is, not making a Type I Error. The Type II Error refers to beta (β), the probability of erroneously accepting the null hypothesis. Accordingly, one minus beta $(1-\beta)$ represents the probability of not accepting the null hypothesis, that is, not making a Type II Error. The standard normal deviates ($z\alpha$ and $z\beta$, as well as $t\alpha$ and $t\beta$), are employed, respectively, for indicating the probabilities of the two errors. The use of t values, however, usually requires iterative stabilization.

It should be noted that more and more emphasis is being placed upon the use of the power function $(1-\beta)$ in hypothesis testing. Traditionally, there has been a focus on the significance criterion (α), while ignoring power $(1-\beta)$. Some statisticians, of course, are of the convincing opinion that the consideration of power results in an abundance of liberality, for significance is more likely to be uncovered on the basis of the amount of effort put forth rather than on the basis of an empirically pragmatic meaningfulness. On the other hand, reputable statisticians maintain that the traditional approach--focusing on only alpha--results in a desirable degree of conservatism. While one may tend to align himself or herself definitively on the side of one of the positions, it must be remembered that sampling is served by the inclusion of the power consideration, for such does result in the determination and use of a larger sample size.

The finite population correction (fpc) is basically employed when n represents 5 percent or more of the population. When the percent is less than 5 percent, the effect on the sample size is negligible. In this regard, however, a defendable position is that fpc can profitably be applied in connection with all finite populations, thereby elevating the finite population to an infinite population.

A sample size is, at best, a tentative, operative estimate. Accordingly, it must be recognized and understood that various and sundry approaches to sample size are, can be, and should be employed. The basic consideration in this regard is the rationale for the use, justification, and defense thereof. The formulae relating, for example, to the estimation of the population mean and the population proportion are cases in point. The former focuses on measurement (the amount on a continuous basis), and the latter focuses on counting (the number on a discontinuous, or discrete, basis). To the extent that the population estimations can be justified--whether or not such estimations are actually effected--one of the foregoing estimation formulae may be in order, assuming, of course, random selection from a finite population. Hence, the basic consideration should not be a parametric estimation in fact; it should be a parametric estimation in theory.

Finally, it must be emphasized that a researcher is not, ordinarily, cognizant of the specific statistical technique(s) which will ultimately be employed. Since n must be known in order to collect the desired data, n must, usually, be known prior to the decision with respect to statistical techniques in the data analysis. That is to say, a given set of data can and will lend itself to the use of optional statistical techniques. Therefore, a means of determining initial sample size--prior to the decision to use the t-test, for example--must be available.

I. SELECTED SAMPLE-SIZE COMPUTATIONAL APPROACHES

A. ESTIMATION OF THE POPULATION MEAN

1. Type I Error without fpc

$$n=\left(\frac{z\sigma}{E}\right)^2$$

2. Type I Error with fpc

$$n=\frac{N\,(z\sigma)^2}{(NE^2)+(z\sigma)^2}$$

3. Type I and II Errors without fpc

$$n=\left(\frac{[z_\alpha+z_\beta]\,\sigma}{E}\right)^2$$

4. Type I and II Errors with fpc

$$n=\frac{N\left([z_\alpha+z_\beta]\,\sigma\right)^2}{(NE^2)+\left([z_\alpha+z_\beta]\right)^2}$$

5. Type I Error without fpc (requiring iteration for stability)

$$n=\left(\frac{t\sigma}{E}\right)^2$$

B. ESTIMATION OF THE POPULATION PROPORTION

1. Type I Error without fpc

$$n=\left(\frac{z^2}{E^2}\,pq\right)$$

2. Type I Error with fpc

$$n=\frac{N\,(z^2\,pq)}{(NE^2)+(z^2\,pq)}$$

3. Type I and II Errors without fpc

$$n=\left(\frac{[z_\alpha+z_\beta]^2}{E^2}\,pq\right)$$

4. Type I and II Errors with fpc

$$n=\frac{N\left([z_\alpha+z_\beta]^2\,pq\right)}{(NE^2)+\left([z_\alpha+z_\beta]^2\,pq\right)}$$

5. Sokal and Rohlf (15, 17); cf. I-D-5

$$n=\frac{\text{Tabular Value (Table 1)}}{\text{(Angular Transformation Squared)}}=\frac{\alpha/1-\beta}{\delta^2}$$

C. DETERMINATION OF SAMPLE SIZE FOR t-TEST

1. Lacey (8) -- derived from t-test formula

$$n=\frac{2\,(s^2t^2)}{D^2}$$

2. Walker and Lev (19) -- $\underline{d}$ represents the departure from hypothesis which it is desired to detect

$$n=\left(\frac{\sigma}{d}(z_\alpha+z_\beta)\right)^2=\frac{\sigma^2}{d^2}(z_\alpha+z_\beta)^2$$

3. Walker and Lev (19) -- using equal-size samples and determining $\underline{n}$ for each sample

$$n_1=\frac{\sigma_1^2+\sigma_1\sigma_2}{d^2}(z_\alpha+z_\beta)^2$$

$$n_2=\frac{\sigma_2^2+\sigma_1\sigma_2}{d^2}(z_\alpha+z_\beta)^2$$

4. Hadley (5) -- Type I and II Errors -- $\underline{n}$ is sample size for one sample; $\underline{2n}$ is the sample size for both

$$n=2\sigma^2(z_\alpha+z_\beta)^2/\mu_1$$

5. Dixon and Massey (2) -- derived from $\underline{z}$ formula -- Type I Error

$$n=\frac{2\,(sz)^2}{D^2}$$

6. Dixon and Massey (2) -- derived from $\underline{z}$ formula -- Type I and II Errors

$$n=\frac{2\left([z_\alpha+z_\beta]\,s\right)^2}{D^2}$$

7. Marascuilo (10) -- Type I Error

$$n=\frac{2\,(st)^2}{D^2}$$

8. Marascuilo (10) -- Type I and II Errors

$$n=\frac{(t_\alpha+t_\beta)^2\,(2s^2)}{D^2}$$

9. Sokal and Rohlf (15, 17) -- Type I and II Errors without fpc (cf. I-D-5)

$$n=\frac{\text{Tabular Value (Table 1)}}{\text{(Angular Transformation Squared)}}=\frac{\alpha/1-\beta}{\delta^2}$$

D. DETERMINATION OF SAMPLE SIZE FOR χ^2

1. Lacey (8) -- anticipated observed percentages vs. theoretical percentages

$$\chi^2=\frac{(.8n-.5n)^2}{.5n}+\frac{(.2n-.5n)^2}{.5n}=.36n$$

$$n=6.635\ (1df,\ 1\%\ \text{level})\ /\ .36n$$

$$=18.43\ \ (19)$$

2. Lacey (8) -- anticipated observed percentages vs. theoretical percentages for 2 X 2 Control and Experimental Groups

	Failed	Passed	Totals
Control	.159n	.841n	n
	(.113n)	(.887n)	
Exp'l	.067n	.933n	n
	(.113n)	(.887n)	
	.226n	1.774n	2n

$$x^2 = \frac{(.046n)^2}{.113n} + \frac{(.046n)^2}{.113n} + \frac{(.046n)^2}{.887n} + \frac{(.046n)^2}{.887n} = .042222n$$

$n = 3.841$ (1 df, 5% level) / $.042222n = 90.97$ (91)

3. Lacey (8) -- assuming results of dichotomous survey with 10% maximum fiducial range

Owners	Non-owners	Total
.7n	.3n	n

$$x^2 = \frac{(.05n)^2}{.7n} + \frac{(.05n)^2}{.3n} = .0119n$$

$n = 6.635$ (1 df, 1% level) / $.0119n = 557.56$ (558)

4. Krejcie, et al. (7) and NEA (16) -- Cf. I-B-2

$$n = \chi^2 Npq \div d^2(N-1) + \chi^2 pq$$

5. Sokal and Rohlf (15, 17) -- Although the following is employed for detecting a true difference between two given percentages, the approach is applicable to natural and artificial dichotomies or rows and/or columns. The rationale is predicated upon the noncentrality parameter discerned from Table 2.

Assume that alpha is 5% and power is .80; moreover, that p_1 is 0.65 and p_2 is 0.55. By means of Rohlf and Sokal's Table K for angular transformation, the two proportions are converted to arcsines, angles, in degrees, whose sines correspond to the values given.

The value from Table 2 is 12,884.8; and delta square is $(53.73 - 47.87)^2 = 5.86^2 = 34.3396$

$$n = \frac{12{,}884.8}{34.3396} = 375.21 \quad (376)$$

2n (for two samples) = 752

N.B.: (1) When one of the percentages is theoretical, divide by two delta square ($2\delta^2$).
(2) When this overall approach yields a sample size of $n < 20$, the estimated n should be increased by the value of one (1).

E. DETERMINATION OF SAMPLE SIZE FOR F-TEST

1. Sokal (17) -- the basic formula uses t values for taking Type I and II Errors into account.

In studying four populations by means of ANOVA, the number of items from each population can be determined. The appropriate formula is

$$n \geq 2\left(\frac{\sigma}{\delta}\right)^2 \left\{t_{\alpha[\nu]} + t_{2(1-P)[\nu]}\right\}^2$$

where n = number of replications

σ = True standard deviation

δ = the smallest true difference which it is desired to detect (N.B.: It is necessary to know only the ratio of σ to δ, not their actual values)

ν = degrees of freedom of the sample standard deviation ($\sqrt{MS_{within}}$) with a groups and n replications per group

α = significance level (such as 0.05)

P = desired probability that a difference will be found to be significant (if it is as small as delta (δ))

$t_{\alpha[\nu]}, t_{2(1-P)[\nu]}$ values from a two-tailed t-table with degrees of freedom and corresponding to probabilities of α and $2(1 - P)$, respectively

Iterative Solution: Iterate to stability when necessary.

The ratio is given as 6/5. The initial n is 20. Then, ν is $(4(20 - 1)) = 4 \times 19 = 76$.

Substituted values on the basis of an n of 20 are:

$$n \geq 2\left(\frac{6}{5}\right)^2 \left\{t_{.01[76]} + t_{2(1-0.80)[76]}\right\}^2$$
$$= 2\left(\frac{6}{5}\right)^2 [2.64 + 0.847]^2 = 2(1.44)\ 12.16 = 35.0$$

Next, try an n of 35. Substituted values are:

$$n \geq 2(1.44)[2.61 + 0.845]^2 = 2.88\ (11.94) = 34.4\ (35)$$

Hence, 35 replications per population (a total of 140) are required for the four populations.

II. SELECTED SAMPLE-SIZE TABULAR APPROACHES

A. ESTIMATION OF THE POPULATION MEAN

1. Welkowitz, et al. (20) -- Gamma (γ), the effect size of the population, is determined by $\mu_1 - \mu_0 / \sigma$. Delta (δ), a function of n, $\gamma\sqrt{n}$, is read from Table 1.

$$n = \left(\frac{\delta}{\gamma}\right)^2$$

B. ESTIMATION OF THE POPULATION PROPORTION

1. Welkowitz, et al. (20) -- Gamma (γ), the population effect size, is determined by $p_1 - p_0 / \sqrt{p_0(1-p_0)}$. Delta ($\delta$), a function of n, $\gamma\sqrt{n}$, is read from Table 1.

$$n = \left(\frac{\delta}{\gamma}\right)^2$$

C. DETERMINATION OF SAMPLE SIZE FOR t-TEST

1. Dixon and Massey (2) -- for one-sample case -- d is read from the authors' table.

$$d = \frac{\mu - \mu_0}{\sigma / \sqrt{n}}$$

2. Dixon and Massey (2) -- for two-sample case -- d is read from authors' table.

$$d = \frac{\mu_1 - \mu_2}{\sigma\sqrt{(1/n_1) + (1/n_2)}}$$

3. Welkowitz, et al. (20) -- Gamma (γ), the population effect size, is determined by $\mu_1 - \mu_2 / \sigma$. Delta (δ), a function of n, $\gamma\sqrt{\frac{n}{2}}$, is read from Table 1. The resultant n is for each sample size; 2n (equal n's) is required for the computation.

$$n = \left(\frac{\delta}{\gamma}\right)^2 \qquad 2n = \left(\frac{\delta}{\gamma}\right)^2$$

4. Welkowitz, et al. (20) -- when the two sample sizes have unequal n's

$$n = \frac{2\, n_1 n_2}{n_1 + n_2}$$

5. Dixon and Massey (2) -- for collected and analyzed paired data ($\sigma_1 = \sigma_2$) -- n provides the number of pairs of observations. Read d from the authors' table.

$$d = (\mu_1 - \mu_2) / \sqrt{(\sigma_1^2 + \sigma_2^2)/n}$$

6. Dixon and Massey (2) -- for deviation of μ from μ_0 -- one population mean -- one sample size -- d is read from the authors' table.

$$d = \frac{\mu - \mu_0}{\sigma / \sqrt{n}}$$

7. Dixon and Massey (2) -- differences in two population means -- two-sample cases -- d is read from the authors' table.

$$d = \frac{\mu_1 - \mu_2}{\sigma\sqrt{(1/n_1) + (1/n_2)}}$$

8. Guenther (3) -- Delta (δ) read from Owen's table (13).

$$\delta = (\mu_1 - \mu_2)/\sigma \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}}$$

D. DETERMINATION OF SAMPLE SIZE FOR X^2

1. Sokal and Rohlf (15, 17) -- Cf. I-D-5, which is also applicable.

$$n \geq 2\left(\frac{\sigma}{\delta}\right)^2 \left\{t_{\alpha[\nu]} + t_{2(1-P)[\nu]}\right\}^2$$

E. DETERMINATION OF SAMPLE SIZE FOR F-TEST

1. Daniel (1) -- Phi (ϕ), a noncentrality parameter is read, converted, and interpreted on the basis of, for the most part, the Pearson and Hartley charts (14).

$$\phi' = \frac{\sqrt{\sum_{j=1}^{k} \alpha_j^2 / k}}{\sigma_e / \sqrt{n_j}}$$

2. Guenther (3) -- One-way ANOVA.

$$\phi = \left[\frac{n}{r}\sum_{j=1}^{r} (\mu_j - \mu)^2\right]^{\frac{1}{2}} \Big/ \sigma$$

3. Guenther (3) -- Formula for an indirect determination of n.

$$n = \left[\sigma^2 \Big/ \sum_{j=1}^{r} (\mu_j - \mu)^2\right] r\phi^2$$

4. Guenther (3) -- Randomized Complete Blocks.

$$\phi = \left[\frac{n}{r}\sum_{j=1}^{r} (\mu_j - \mu)^2\right]^{\frac{1}{2}} \Big/ \sigma$$

5. Kirk (6) -- Basic formula.

$$\phi = \frac{\sqrt{\sum_{1}^{k}(\mu_j - \mu)^2 / k}}{\sigma_\varepsilon / \sqrt{n}}$$

6. Dixon and Massey (2) -- Basic formula.

$$\phi = \sqrt{\frac{\sum_{i=1}^{k}(\mu_i - \bar{\mu})^2 / k}{\sigma^2 / n}}$$

7. Winer (21) -- The noncentrality parameter is read, converted, and interpreted on the basis of the Tiku tables (18).

$$\phi = \sqrt{\frac{n \sum \tau_j^2}{k\, \sigma_\epsilon^2}}$$

F. DETERMINATION OF SAMPLE SIZE FOR r

1. Welkowitz, et al. (20) -- Gamma (γ), the population effect size, $\gamma = \rho_1$, is determined by p_1, the correlation coefficient. Delta (δ), $\gamma\sqrt{n-1} = \rho_1\sqrt{n-1}$, is read from Table 1.

$$n = \left(\frac{\delta}{\gamma}\right)^2 + 1 = \left(\frac{\delta}{\rho_1}\right)^2 + 1$$

Table 1
A Function of Significance Criterion (α) and Power $(1-\beta)$

	One-tailed test (α)			
	.05	.025	.01	.005
	Two-tailed test (α)			
Power $(1-\beta)$	.10	.05	.02	.01
.25	0.97	1.29	1.65	1.90
.50	1.64	1.96	2.33	2.58
.60	1.90	2.21	2.58	2.83
.67	2.08	2.39	2.76	3.01
.70	2.17	2.48	2.85	3.10
.75	2.32	2.63	3.00	3.25
.80	2.49	2.80	3.17	3.42
.85	2.68	3.00	3.36	3.61
.90	2.93	3.24	3.61	3.86
.95	3.29	3.60	3.97	4.22
.99	3.97	4.29	4.65	4.90
.999	4.37	5.05	5.42	5.67

Table 2
Alpha and Power (Sokal and Rohlf (15, 17))

Power $(1-\beta)$	α .1	.05	.01	.001
.50	4,442.2	6,306.4	8,883.7	10,891.5
.80	10,150.2	12,884.8	16,474.3	19,171.6
.90	14,059.3	17,249.8	21,368.5	24,426.2
.99	25,890.0	30,161.4	35,536.7	39,450.1

REFERENCES

1. Daniel, Wayne W. INTRODUCTORY STATISTICS WITH APPLICATIONS. Boston: Houghton Mifflin Company, 1977.

2. Dixon, Wilfrid J. and Frank J. Massey, Jr. INTRODUCTION TO STATISTICAL ANALYSIS, 3/e. N.Y.: McGraw-Hill Book Company, 1969.

3. Guenther, William C., "Power and Sample Size Determination When the Alternative Hypotheses Are Given in Quantiles," THE AMERICAN STATISTICIAN, Vol. 31, No. 3, August, 1977, pp. 117-18.

4. Guilford, J.P. and Benjamin Fruchter. FUNDAMENTAL STATISTICS IN PSYCHOLOGY AND EDUCATION, 5/e. N.Y.: McGraw-Hill Book Company, 1973.

5. Hadley, G. INTRODUCTION TO PROBABILITY AND STATISTICAL DECISION THEORY. San Francisco: Holden-Day, Inc., 1967.

6. Kirk, R.E. EXPERIMENTAL DESIGN: PROCEDURES FOR THE BEHAVIORAL SCIENCES. Belmont, Calif.: Brooks/Cole Publishing Company, 1968.

7. Krejcie, Robert V. and Daryle W. Morgan, "Determining Sample Size for Research Activities," EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1970, 30, pp. 607-10.

8. Lacey, Oliver L. STATISTICAL METHODS IN EXPERIMENTATION: AN INTRODUCTION. N.Y.: The Macmillan Company, 1953.

9. Lehmer, E., "Inverse Table of Probabilities of Errors of the Second Kind," ANNALS OF MATHEMATICAL STATISTICS, 15, 1944, pp. 388-98.

10. Marascuilo, Leonard A. STATISTICAL METHODS FOR BEHAVIORAL SCIENCE RESEARCH. N.Y.: McGraw-Hill Book Company, 1971.

11. National Bureau of Standards. TABLES OF POWER POINTS OF ANALYSIS OF VARIANCE TESTS (Non-Central F Tables), pre-publication copy. Washington, D.C.: National Bureau of Standards, undated.

12. Odeh, Robert E. and Martin Fox. SAMPLE SIZE CHOICE. N.Y.: Marcel Dekker, Inc., 1975.

13. Owen, D.B., "The Power of Student's t-Test," JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, 60, 1965, pp. 320-33.

14. Pearson, E.S. and H.O. Hartley, "Charts of the Power Function of Analysis of Variance Tests, Derived from the Non-Central F-Distribution," BIOMETRIKA, 38, 1951, pp. 112-30.

15. Rohlf, F. James and Robert R. Sokal. STATISTICAL TABLES. San Francisco: W.H. Freeman and Company, 1969.

16. "Small Sample Techniques," THE NEA RESEARCH BULLETIN, 28, December, 1960, pp. 99-104.

17. Sokal, Robert F. and F. James Rohlf. BIOMETRY: THE PRINCIPLES AND PRACTICE OF STATISTICS IN BIOLOGICAL RESEARCH. San Francisco: W.H. Freeman and Company, 1969.

18. Tiku, M.L., "Noncentrality F Distribution" (Abbreviated from "Tests of the Power of the F Test," JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, 1967, 62, pp. 525-39.

19. Walker, Helen M. and Joseph Lev. STATISTICAL INFERENCE. N.Y.: Henry HOlt and Company, Inc., 1953.

20. WElkowitz, Joan, Robert B. Ewen, and Jacob Cohen. INTRODUCTORY STATISTICS FOR THE BEHAVIORAL SCIENCES. N.Y.: Academic Press, 1971.

21. Winer, B.J. STATISTICAL PRINCIPLES IN EXPERIMENTAL DESIGN, 2/e. N.Y.: MCGraw-Hill Book Company, 1971.