A CONSULTANT'S GUIDE TO SAMPLE SIZE

Karen A. Summers

May 1990

Stat 570

1.0 Introduction

The determination of sample size is such an important step in the experimental process that many books and journal articles in many different fields are devoted to the subject. So many, in fact, that the task of tracking down the best method for some particular problem becomes a cumbersome chore. The sample size problem is a universal one. No matter what the field of study, there comes a time when the researcher must ask "How large should my sample be?" And anyone who has done consulting knows that (too) many times this is when the researcher first decides to contact a statistician. If the statistician is lucky, the researcher will have some understanding of power and distribution considerations, otherwise these need to be explained and determined. Then, when researching sample size methodology, it becomes apparent early on that many authors have their own method of notation.

The purpose of this paper is two-fold: (i) to summarize some of the more common, most used sample size determination methods, and (ii) to describe these methods using uniform notation whenever possible. In this paper I have provided test statistics, sample size formulas, power formulas, examples, and/or important references for the following procedures: (i) one-sample inference about the mean which includes methods for the z-test, normal approximation of the binomial test on a proportion, t-test, and achievement of a specified confidence interval width; (ii) the significance of a product moment correlation coefficient, r; (iii) nonparametric one-sample inference about location which includes methods for the sign test and the Wilcoxon one-sample test; (iv) two-sample inference about the mean which includes methods for the z-test for both equal and unequal variances, the normal approximation to the binomial test for equivalence of two proportions, and the t-test; and (v) the Wilcoxon two-sample test.

1.1 Sample Size and Power

Ideally, whenever statistical tests are done, the investigator would like to be assured that the null hypothesis is rejected whenever it is false and "accepted" whenever it is true. While absolute assurance is not possible without sampling all (or nearly all) of the population, it is possible to find the probability of rejecting the null hypothesis (Guenther, 1965). Rejecting the null hypothesis when it is true is called a Type I error. We will require that the probability of a Type I error is no larger than a specified value denoted by α, often called the significance level. Not rejecting a false null hypothesis is called a Type II error and its probability is denoted by β. The power of the test, then, is defined to be the probability of rejecting a false null hypothesis. It is equal to $1 - \beta$ and will be denoted by Π. Power depends on the sample size, the level of significance α, and the chosen alternative parameter value.

When sample size is to be determined, power, along with the level of significance, must be predetermined by the investigator. These values are then used in the sample size formulas.

2.0 One-Sample Inference About the Mean

Four methods are presented for inference about the population mean μ based on one-sample data: (i) normal test when the population variance σ^2 is known, (ii) t-test when σ^2 is unknown, (iii) sample size needed to achieve a specified confidence interval width when σ^2 is known, and (iv) sample size to achieve a specified confidence interval width when σ^2 is unknown.

2.1.a Normal Distribution z-Test — σ^2 Known

For a variety of experimental and observational studies, it is of interest to test $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$ based on a random sample from a $N(\mu, \sigma^2)$ population (Kupper and Hafner, 1989). In such situations, one may wish to specify a $100\alpha\%$ significance level and the power at a chosen alternative value of μ in order to determine the necessary sample size.

Suppose that $x_1, \ldots, x_n$ is an observed random sample from $N(\mu, \sigma^2)$ where σ^2 is known, and let

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \quad .$$

A size α test of $H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_0$ is to reject H_0 if $z_0 \geq z_{1-\alpha}$, where $z_{1-\alpha}$ is the $100(1-\alpha)$ percentile of the standard normal distribution. The power function for this test is

$$\Pi(\mu) = 1 - \Phi\left(z_{1-\alpha} + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) = 1 - \beta \quad ,$$

where Φ is the standard normal cumulative distribution function. The sample size required to achieve a size α test with power $1-\beta$ for an alternative value $\mu^* > \mu_0$, for a one-tailed test, is given by

$$n = \left[\frac{(z_{1-\alpha} + z_{1-\beta})\sigma}{\mu_0 - \mu^*}\right]^2 \quad ,$$

and for a two-tailed test ($H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$) is

$$n = \left[\frac{(z_{1-\alpha/2} + z_{1-\beta})\sigma}{\mu_0 - \mu^*}\right]^2$$

(Bain and Engelhardt, 1987).

Example 2.1

A paint company sells a type of house paint that has a drying time of 75 minutes with a variance of 9 minutes. One of the research scientists has developed a paint additive which he believes will reduce drying time to 72.5 minutes. It is known that σ^2 is the same for the distribution of paint drying time whether it has the additive or not. He decides to test the following hypotheses with a significance level of $\alpha = .05$ and power $1-\beta = .90$: $H_0 : \mu = 75$ minutes versus $H_a : \mu < 75$ minutes, where $\mu^* = 72.5$ minutes. He then uses the following formula to determine the sample size he needs to accomplish the significance and power he desires:

$$n = \left[\frac{(z_{.95} + z_{.90})\sigma}{\mu_0 - \mu^*}\right]^2 = \left[\frac{(1.65+1.28)3}{75-72.5}\right]^2 \approx 13 \quad .$$

He needs to randomly select 13 cans of paint, add the additive, and measure their drying times.

Example 2.2

Suppose the additive used in Example 2.1 has a peculiar property where if it did not decrease drying time, it could increase it. With α, $1-\beta$, and μ^* being the same as in Example 2.1, he would now want to test $H_0 : \mu = 75$ versus $H_a : \mu \neq 75$ and would use the following formula to determine his sample size:

$$n \doteq \left[\frac{(z_{.975} + z_{.90})\sigma}{\mu_0 - \mu^*}\right]^2 = \left[\frac{(1.96+1.28)3}{75-72.5}\right]^2 \approx 16 \quad .$$

2.1.b Normal Approximation of Binomial

Often, in experimental situations, we need to determine sample size when the problem leads to a test of hypothesis about a proportion. Suppose the experiment consists of n trials which have the following properties: (i) the result of each trial will be classified into one of two categories (i.e., success and failure), (ii) the probability p of a success will be the same for each trial, and (iii) each experiment will be independent of all others. Then the number of successes follows a binomial distribution. Binomial tables can be used quite easily for most such experiments, but if they are not available or the range of entries is exceeded, then the following normal approximation may be used.

Define the test statistic

$$z_0 = \frac{x - np}{\sqrt{np(1-p)}} \quad .$$

A size α test of $H_0 : p = p_0$ versus $H_a : p > p_0$ is to reject H_0 if $z_0 > z_{1-\alpha}$, where z_α is the $100(1-\alpha)$ percentile of the standard normal distribution. The power function for this test is

$$\Pi(p^*) \doteq p\left(z > \frac{z_{1-\alpha}\sqrt{p_0(1-p_0)} + \sqrt{n}\,|p_0 - p^*|}{\sqrt{p^*(1-p^*)}}\right) = 1-\beta$$

where z is a standard normal random variable and p^* is the alternative value where $p^* > p_0$. The sample size required to achieve a size α test with power $1-\beta$ for an alternative value p^* for a one-tailed test is

$$n \geq \left(\frac{z_{1-\alpha}\sqrt{p_0(1-p_0)} + z_{1-\beta}\sqrt{p^*(1-p^*)}}{p_0 - p^*} \right)^2$$

(Guenther, 1965). For the two-sided test, replace α with $\alpha/2$.

Example 2.3

A certain manufacturing process produces 10% defective parts per lot on the average. An industrial engineer develops a new process which he believes will reduce the percentage of defectives to 9% per lot. It is important to know if this difference is significant since start-up costs for the new process are high. The engineer decides to draw a random sample of parts to inspect for testing the hypothesis $H_0 : p = .10$ versus $H_a : p < .10$ at $p^* = .09$ at a significance level of $\alpha = .01$. He requires power of $1 - \beta = .95$. The sample size he then needs is

$$n \geq \left[\frac{z_{.99}\sqrt{(.10)(.90)} + z_{.95}\sqrt{(.09)(.91)}}{.10 - .09} \right]^2 \approx 13,700 \quad .$$

2.2 t-Test — σ^2 Unknown

When σ^2 is unknown, one can estimate it using the sample variance, S^2, where

$$S^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \quad .$$

Let $x_1, \ldots, x_n$ be an observed random sample from $N(\mu, \sigma^2)$, where σ^2 is unknown, and let

$$t_0 = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \quad .$$

For a one-tailed size α test of $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$, the power to reject H_0 in favor of H_a when $\mu > \mu_0$ is

$$\Pi(\mu) = P\left[t'^{\sqrt{n}\theta}_{n-1} > t^{1-\alpha}_{n-1} \mid \mu > \mu_0 \right]$$

where $t^{1-\alpha}_{n-1}$ is the $100(1-\alpha)$ percentile of the central t_{n-1} distribution and where $t'^{\sqrt{n}\theta}_{n-1}$ has a noncentral t_{n-1} distribution with $n-1$ df and noncentrality parameter $\sqrt{n}\theta = \sqrt{n}(\mu - \mu_0)/\sigma$ (Kupper and Hafner, 1989).

Sample size determination can be accomplished using tables for the noncentral t such as the table on page 535 in Bain and Engelhardt (1987) which is reproduced in Table 1 of the Appendix. This tables gives the sample size required to achieve $1-\beta$ for specified $d = |\mu - \mu_0|/\sigma$ and α for a one-tailed test. The table also gives the approximate n for two-tailed significance at 2α.

Example 2.4

Returning to the paint drying time problem in Example 2.1, suppose the scientist is not willing to assume that σ^2 is the same after the additive is added, and therefore, he decides to use a t-test. In order to calculate d, however, he needs to specify a value for σ. He chooses $\sigma = 3$. Then

$$d = (\mu_0 - \mu^*)/\sigma = (75 - 72.5)/3 \approx .8 \quad .$$

From Table 1 in the Appendix, one finds a sample size of approximately 15 cans of paint is required for the $\alpha = .05$ one-tailed test to ensure a power of .90. Similarly, if a two-tailed test was to be used in Example 2.2, he would need a sample size of 18 cans of paint.

2.3 Determination of σ^2 When It Is Not Known

At this point, one might become confused since the t-test for σ^2 unknown requires a value for σ to obtain the noncentrality parameter used to determine sample size and since S^2 cannot be calculated until after the sample is drawn. There are three commonly used methods for determining σ^2 when it is unknown: (i) using historical or published results as in Example 2.4, (ii) using knowledge about the mechanisms of action in the material to derive a mathematical model for σ^2, and (iii) using the sample variance S^2 from a pilot study. In most cases, the calculated n is an estimate, not an exact answer.

2.4 Sample Size to Achieve Specified Confidence Interval Width — σ^2 Known

Because the length of a confidence interval can be decreased or increased by the size of the sample, it is reasonable to know how large the sample size must be in order to make the length of the interval no greater than some specified length L. To meet this requirement we must have

$$2z_{1-\alpha/2}\sigma/\sqrt{n} \leq L$$

which gives the following sample size formula

$$n \geq \left(\frac{2\sigma z_{1-\alpha/2}}{L} \right)^2$$

(Guenther, 1965).

Example 2.5

Returning once again to the paint drying time problem in Example 2.1 where $\alpha = .05$ and $\sigma = 3$, suppose the scientist wants to determine a 95% confidence interval for the average drying time for paint containing the additive so that the interval is no greater than 2 minutes. The minimum sample size required is

$$n \geq \left(\frac{2\sigma z_{.975}}{L} \right)^2 = \left(\frac{2(3)(1.96)}{2} \right) \approx 35 \quad .$$

2.5 Sample Size to Achieve Specified Confidence Interval Width — σ^2 Unknown

There is an excellent article by S.L. Beal (1989) which contains formulas and tables for computing n such that

$$1 - \alpha = P[\text{width of the } CI \leq L \text{ \underline{and} the } CI \text{ is correct}] \quad .$$

3.0 The Significance of a Product Moment Correlation Coefficient, r

Many times, especially in the behavioral sciences, the linear relationship between two variables is of interest to the researcher. When this is the case, a frequently used statistic is the Pearson product-moment correlation coefficient, r.

The t-test for the significance of r is given by

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1 - r_s^2}}$$

where r_s = the sample r, n is the sample size, and t follows a t-distribution having $n - 2$ degrees of freedom. Solving for r_s gives

$$r_s = \sqrt{\frac{t^2}{t^2 + (n-2)}} \quad .$$

Cohen provides tables for the sample size required to test both one-tailed and two-tailed tests at specified α and $1-\beta$. To test $H_0 : r = 0$ versus $H_a : r > 0$ use α for the tabled a_1, and to test $H_0 : r = 0$ versus $H_a : r \neq 0$ use α for the tabled a_2. These tables are reproduced in the Appendix as Table 2 (Cohen, 1977, pp. 101–2).

Example 3.1

An educational psychologist is consulted by the dean responsible for admissions at a large university with regard to the desirability of supplementing their criterion for admission by using a personality questionnaire. The plan is to administer the test to a random sample of entering freshman and determine whether scores on this test (X) correlate with the freshman year grade point average (Y). The decision is made that if $r = .10$, then it is worth adding to the selection procedure. It is also decided that power of $1-\beta = .90$ and a significance level of $\alpha = .05$ are desired. Then using Table 2 with $\alpha = a_1 = .05$, $r = .10$, and power $= .90$, the required sample size is 864.

For a good, recent review of the sample size determination problem in simple regression and correlation, see the paper by Gatsonis and Sampson (1989).

4.0 Nonparametric One-Sample Inference About Location

"Methods based on ranks form a substantial body of statistical techniques that provide alternatives to the classical parametric methods" (Lehmann, 1975). The Sign Test and the Wilcoxon one-sample test are presented here.

4.1 Sign Test

An application of the test that a proportion is .50 is the nonparametric sign test. Let $x_1, \ldots, x_n$ constitute a random sample from a population with median η. We want to find the sample size necessary to test the hypothesis $H_0 : \eta = \eta_0$ versus $H_a : \eta > \eta_0$ at a specified significance level α and power $1-\beta$. Suppose $P\{X = \eta_0\} = 0$.

As our test statistic, we use the quantity $S =$ the number of observations greater than η_0. Then S is a binomial random variable with $P = P\{X > \eta_0\}$. Under H_0, $P = 1/2$. For specified alternative $\eta^* > \eta_0$, let $p^* = P\{X > \eta_0 \mid \eta = \eta^*\} > 1/2$. (One can specify p^* directly.) Applying the results of Section 2.1.b with $p = 1/2$

$$n = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{2(p^* - 1/2)} \right)^2$$

(Noether, 1987).

Example 4.1

Steel rods produced by a certain company have a median length of 10 meters when the process is operating properly. It is suspected that the process is not operating properly and further that the rods are twice as likely to be greater than 10 meters as they are likely to be less. The hypothesis $H_0 : \eta = 10$ versus $H_a : \eta > 10$ is to be tested at $\alpha = .05$ and $1 - \beta = .90$ where $P(X > 10) = 2/3$. The sample size needed then is

$$n = \left(\frac{z_{.95} + z_{.90}}{2(2/3 - 1/2)} \right)^2 = \left(\frac{1.65 + 1.28}{1/3} \right)^2 \approx 78 \quad .$$

4.2 Wilcoxon One-Sample Test

Given here is Noether's (1987) method of sample size determination for the Wilcoxon one-sample test. Another approach can be found in Lehmann (1975).

This procedure is a test of symmetry about a hypothesized population median. We will test H_0 : the population is symmetric about 0 versus H_a : the population is shifted to the right (Gibbons, 1985).

Let $X_1, \ldots, X_n$ be a random sample with W = number of (i, j) pairs for which $\{X_i + X_j\}$ is positive where $1 \le i \le j \le n$. Then W is the Wilcoxon test statistic for testing $H_0 : p' = 1/2$ versus $H_a : p' > 1/2$, where $p' = P\{X + X' > 0\}$, where X and X' are two independent observations. Power is defined as

$$\Pi(p') = \frac{[n(p - 1/2) + 1/2n(n-1)(p' - 1/2)]^2}{n(n+1)(2n+1)/24} \doteq 3n(p' - 1/2)^2$$

and the necessary sample size for detecting p' under the alternative hypothesis is given by

$$n = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\sqrt{3}\,(p' - 1/2)} \right)^2$$

(Noether, 1987).

Example 4.2

Assume that the company in Example 4.1 is having more problems with their process. Then under the same conditions and significance criterion so that $p' = 2/3$ in this problem also, the sample size required is

$$n = \left(\frac{z_{.95} + z_{.90}}{\sqrt{3}\,(2/3 - 1/2)} \right)^2 = \left(\frac{1.65 + 1.28}{.29} \right)^2 \approx 100 \quad .$$

5.0 Two-Sample Inference About the Mean

Up to this point we have only considered the one-sample problem. We will now look at the two-sample problem.

5.1 Normal Distribution

Let $x_1, \ldots, x_{n_1}$ and $y_1, \ldots, y_{n_2}$ be two independent random samples of size $n_1 = n_2 = n$ with distributions $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$ respectively.

5.1.a z-Test — Variances Known and Equal — $\sigma_x^2 = \sigma_y^2 = \sigma^2$

A size α test of $H_0 : \mu_x - \mu_y = 0$ versus $H_a : \mu_x - \mu_y > 0$ is to reject if

$$z_0 = \frac{\sqrt{n}\,(\overline{X} - \overline{Y})}{\sqrt{2}\,\sigma}$$

is greater than $z_{1-\alpha}$. The power of this test at alternative $\mu_x - \mu_y = d > 0$ is

$$\Pi(d) = 1 - \Phi\left(z_{1-\alpha} + \frac{d}{\sqrt{2}\,\sigma}\right)$$

and the sample size necessary to attain a specified significance level α and power $1-\beta$ is

$$n = \left[\frac{(z_{1-\alpha} + z_{1-\beta})\sqrt{2}\,\sigma}{d}\right]^2$$

for a one-tailed test. For a two-tailed test of $H_0 : \mu_x - \mu_y = 0$ versus $H_a : \mu_x - \mu_y \neq 0$, replace α with $\alpha/2$.

5.1.b z-Test — Variances Known but Unequal — $\sigma_x^2 \neq \sigma_y^2$

A size α test of $H_0 : \mu_x - \mu_y = 0$ versus $H_a : \mu_x - \mu_y > 0$ is to reject if

$$z_0 = \frac{\sqrt{n}\,(\overline{X} - \overline{Y})}{\sqrt{\sigma_x^2 + \sigma_y^2}}$$

is greater than $z_{1-\alpha}$. The power of this test at alternative $\mu_x - \mu_y = d > 0$ is

$$\Pi(d) = 1 - \Phi\left(z_{1-\alpha} + \frac{d}{\sqrt{\sigma_x^2 + \sigma_y^2}\,/\sqrt{n}}\right)$$

and

$$n = \left[\frac{(z_{1-\alpha} + z_{1-\beta})\sqrt{\sigma_x^2 + \sigma_y^2}}{d} \right]^2$$

is the sample size necessary for a specified significance level α and power $1-\beta$ for a one-tailed test. Replace α with $\alpha/2$ to determine the sample size for a two-tailed test (Snedecor and Cochran, 1980).

Example 5.1

Look at the paint additive problem again in Example 2.1. The lab has developed another additive, which is much more expensive to use, but is believed to be able to reduce drying time to 72. This is not a big difference so they decide to set power $(1-\beta)$ at .95 and the significance level (α) at .05 to test $H_0 : \mu_x - \mu_y = 0$ versus $H_a : \mu_x - \mu_y \neq 0$ to see if there is a real difference. The sample size they need at $d = 0.5$ is

$$n = \left[\frac{(z_{.975} + z_{.95})\sqrt{2}\,3}{72.5 - 72} \right]^2 = \left[\frac{(1.96 + 1.65)\sqrt{2}\,3}{.5} \right]^2 \approx 938 \quad .$$

They will need to test 938 cans of paint with each additive to detect a .5 difference between the averages with power equal to .95.

5.1.c Graphical z-Test Method

The nomogram, Figure 1 in the Appendix, is appropriate for approximating the sample size for a two-sample comparison of a continuous measurement with the same number of subjects in each group. It makes use of the standardized difference which is equal to the hypothesized true difference (usually the smallest relevant difference) divided by the estimated standard deviation. The only restriction is the common requirement that the variable that is being measured is roughly normally distributed.

The nomogram gives the relation between the standardized difference $\left(\frac{\mu_x - \mu_y}{\sigma}\right)$, the total study size $(2n)$, the power $(1-\beta)$, and the level of significance $(\alpha/2)$ for a two-tailed test. Given the two-tailed significance level (5% or 1%), by joining with a straight line the specific values for two of the variables, the required value for the other variable can easily be read off the third scale. By using the nomogram it is both simple and quick to assess the effect on the power of varying the sample size, the effect on the required sample size of changing the difference of importance, and so on.

An estimate of the standard deviation should usually be available, either from previous studies or from a pilot study. Note the nomogram is not strictly appropriate for retrospective calculations. Although it will be reasonably close for samples larger than 100, for smaller samples it will tend to over-estimate the power (Altman, 1982).

Example 5.2

Return again to the paint additive problem in Example 5.1. For a two-tailed test, they want $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 \neq 0$. The standardized difference is $(72.5 - 72)/3 = .17$. They will test at $\alpha/2 = .05$ and $1 - \beta = .95$. Then using the nomogram $2n \approx 1800$ cans of paint so they will need to add each of the additives to 900 cans.

5.1.d Normal Approximation of Binomial

Often, especially in the field of epidemiology, the investigator will want to test the proportion of successes of two binomial distributions. For example, if a new drug is being tested, the success rate, p_1, of the control group will be tested against the success rate, p_2, of the treatment group. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two independent samples distributed as $\text{BIN}(n_1, p_1)$ and $\text{BIN}(n_2, p_2)$ respectively where n_1 does not necessarily equal n_2. Let

$$z_0 = \frac{|p_1 - p_2| - 1/2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} .$$

Then an approximate size α test is to reject $H_0 : p_1 = p_2$ versus $H_a : p_1 > p_2$ is to reject H_0 if $z_0 > z_{1-\alpha}$ with power

$$\Pi(p_1 - p_2) = \Phi\left(\frac{z_{1-\alpha}\sqrt{p_1(1-p_1) + p_2(1-p_2)} - (p_2 - p_1)\sqrt{n}}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}}\right)$$

(Fleiss, 1981).

The minimum sample size required for each sample to achieve a size α test with power $1 - \beta$ for a one-tailed test is

$$n' = \left[\frac{z_{1-\alpha} + z_{1-\beta}}{p_2 - p_1}\right]^2 [p_1(1-p_1) + p_2(1-p_2)]$$

(Snedecor and Cochran, 1980).

In more recent research it has been found that a better approximation for sample size can be found by incorporating the continuity correction in the test statistic. After n' is computed, n is found using

$$n = \frac{n'}{4}\left[1 + \sqrt{1 + \frac{4}{n'|p_2 - p_1|}}\right]^2$$

(Fleiss, 1981, pp. 39–42).

Example 5.3

A new drug has been developed for some common dreadful disease. It is expensive and causes very unpleasant side effects. The existing drug has a success rate of .60, is less expensive and does not have any side effects. Since failure means the patient dies, it is decided that if the new drug causes even a 5% increase in the success rate, it will be highly valuable in the treatment of the disease regardless of the cost. To test the hypothesis of $H_0 : p_1 = p_2$ versus $H_a : p_1 < p_2$ with $p_1 = .60$, $p_2 = .65$, $\alpha = .01$, and $1 - \beta = .99$, the minimum sample size for the control group (those who get the existing drug) and the treatment group (those who get the new drug) is

$$n' = \left[\frac{z_{.99} + z_{.99}}{.65 - .60}\right]^2 [(.60(.40) + .65(.35)] = \left[\frac{2.33 + 2.33}{.05}\right]^2 (.4675) \approx 4060$$

then employing the continuity correction

$$n = \frac{4060}{4}\left[1 + \sqrt{1 + \frac{4}{4060|.65 - .60|}}\right]^2 \approx 4100 \quad .$$

5.2 t-Test — σ^2 Unknown and Equal

Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be random samples distributed as $N(\mu_x, \sigma^2)$ and $N(\mu_y, \sigma^2)$ respectively with $n_1 = n_2 = n$. Define

$$S_x^2 = (n-1)^{-1} \sum_{i=1}^{n_1} (X_i - \overline{X})^2$$

and

$$S_y^2 = (n-1)^{-1} \sum_{i=1}^{n_2} (Y_i - \overline{Y})^2 \quad .$$

Then the pooled variance is defined as

$$S_p^2 = (S_x^2 + S_y^2)/2 \quad .$$

For a one-tailed size α test of $H_0 : \mu_y = \mu_x$ versus $H_a : \mu_y > \mu_x$, the power to reject H_0 in favor of H_a when $\mu_y - \mu_x$ has a specified positive value $\sigma\theta$ is

$$\pi(\mu) = P\left\{t'^{(n/2)^{1/2}\theta}_{2(n-1)} > t^{1-\alpha}_{2(n-1)} \mid (\mu_y - \mu_x) = \sigma\theta > 0\right\} \quad ,$$

where

$$t'^{(n/2)^{1/2}\theta}_{2(n-1)} = (\overline{Y} - \overline{X})/S_p(2/n)^{1/2}$$

has a noncentral t distribution with $2(n-1)$ df and noncentrality parameter $(n/2)^{1/2}\theta$.

Table 1 in the Appendix can now be used to find the sample size required to achieve $1-\beta$ for specified $d' = |\mu_x - \mu_y|/\sigma$ (Kupper and Hafner, 1989).

Example 5.3

Look again at the paint drying time problem in Example 2.4. A second additive has now been developed which the scientist believes will reduce drying time by another two and a half minutes. He wants to test it against the first additive. He chooses $\sigma = 3$ and lets $d = (72.5 - 70)/3 = .8$, $2\alpha = .05$, and $1 - \beta = .95$. From Table 1 he finds he needs a sample size of 38 for each additive. The nomogram, introduced in section 5.1.c, may also be used and gives $2n \doteq 80$.

6.0 Wilcoxon Two-Sample Test

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent random samples. Define the test statistic as $U =$ the number of (i,j) pairs such that $\{Y_j > X_i\}$, $i = 1, \ldots, m$; $j = 1, \ldots, n$. We want to test H_0 : the 2 samples come from the same population versus H_a : the Y-observations tend to be larger than the X-observations. Let $p' = P\{Y > X\}$ and $N = m + n$. The alternative can now be stated as $H_a : p' > 1/2$.

Setting $m = CN$, we find

$$\Pi(p') = \frac{12(1-C)N^2(p'-1/2)}{N+1}$$

and, approximately,

$$N = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{12C(1-C)(p'-1/2)^2} \quad .$$

The combined sample size m and n may now be found using the fact that $m = CN$ (Noether, 1987).

Collings and Hamilton (1988) provide a bootstrap method for using observed data (or pilot data) to approximate power for this test.

7.0 Closing Remarks

This paper has attempted to provide its readers with a quick reference to a variety of methods for determining sample size. Many of these methods have been cited directly from their authors with minor notational changes.

Each sample size method presented requires that stated independence and/or distributional properties be met in order to attain reliable and accurate sample sizes. For example, one should not use the t-test formula to find the necessary sample size for exponential data. However, methods, such as the bootstrap methods described by Collings and Hamilton (1988) that do not require distributional assumptions, do exist.

Specialized software packages are available for power and/or sample size calculation. Goldstein (1989) reviews several of these packages.

The reader may also wish to consult various issues of the *Current Index to Statistics* (Burdick) for which the 1988 issue includes a list of 48 additional articles on sample size.

References

- Altman, Douglas G. (1982), "How Large a Sample," *Statistics in Practice,* England: The Devonshire Press.
- Bain, Lee J. and Engelhardt, Max (1987), *Introduction to Probability and Mathematical Statistics,* Boston: Duxbury Press.
- Beal, S.L. (1989), "Sample Size Determination for Confidence intervals on the Population Mean and on the Difference Between Two Population Means," *Biometrics,* 45, 969–977.
- Burdick, Richard K. (ed.) (1988), *Current Index to Statistics,* Vol. 14.
- Cohen, Jacob (1977), *Statistical Power Analysis for the Behavioral Sciences,* New York: Academic Press.
- Collings, Bruce J. and Hamilton, Martin A. (1988), "Estimating the Power of the Two-Sample Wilcoxon Test for Location Shift," *Biometrics,* 44, 847–860.
- Fleiss, Joseph L. (1981), *Statistical Methods for Rates and Proportions,* New York: John Wiley & Sons.
- Gatsonis, Constantine and Sampson, Allan R. (1989), "Multiple Correlation: Exact Power and Sample Size Calculations," *Psychological Bulletin,* 106, 516–524.
- Gibbons, Jean Dickinson (1985), *Nonparametric Methods for Quantitative Analysis,* Ohio: American Sciences Press, Inc.
- Goldstein, Richard (1989), "Power and Sample Size via MS/PC-DOS Computers," *The American Statistician,* 43, 4, 253–260.
- Graybill, Franklin A. (1961), *An Introduction to Linear Statistical Models,* Vol. 1, New York: McGraw-Hill.
- Guenther, William C. (1965), *Concepts of Statistical Inference,* New York: McGraw-Hill.
- Kupper, Lawrence L. and Hafner, Kerry B. (1989), "How Appropriate Are Popular Sample Size Formulas?" *The American Statistician,* 43, 2, 101–105.

- Lehmann, E.L. (1975), *Nonparametrics: Statistical Methods Based on Ranks,* San Francisco: Holden-Day, Inc.

- Noether, Gottfried E. (1987), "Sample Size Determination for Some Common Nonparametric Tests," *The American Statistician,* 82, 398, 645–647.

- Snedecor, George W. and Cochran, William G. (1980), *Statistical Methods,* Iowa: The Iowa State University Press.

APPENDIX

Sample Size for *t* Test

Sample size n to achieve power $1-\beta$ for $d = |\mu - \mu_0|/\sigma$ in one-sample case and $n = n_1 = n_2$ for $d' = |\mu_1 - \mu_2|/\sigma$ in two-sample case, for one-sided test at significance level α. These are approximate n for two-sided test at significance level 2α.

α (2α)	d	One-sample test							Two-sample test							d'
		$1-\beta$														
		0.50	0.60	0.70	0.80	0.90	0.95	0.99	0.50	0.60	0.70	0.80	0.90	0.95	0.99	
0.005 (0.01)	0.1	669	805	966	1173	1493	1785	2403	1327	1602	1922	2337	2977	3567	4806	0.1
	0.2	169	204	244	296	377	450	605	333	403	484	588	749	894	1206	0.2
	0?4	42	53	64	77	97	115	154	87	101	124	150	189	226	304	0.4
	0.6	22	26	31	36	45	53	71	37	44	55	65	85	100	138	0.6
	0.8	14	16	19	22	27	32	41	23	27	32	39	49	56	85	0.8
	1.0	10	12	13	14	19	22	28	15	18	21	26	32	38	49	1.0
	1.2	8	9	10	12	14	16	21	11	13	16	18	23	27	36	1.2
	1.4	7	8	9	10	12	13	16	9	10	12	14	18	20	27	1.4
	1.6	6	7	8	8	10	11	13	7	9	10	11	14	16	21	1.6
	1.8	6	6	7	8	9	9	11	6	7	8	9	11	13	17	1.8
	2.0	4	6	6	7	8	8	10	6	6	7	8	10	11	14	2.0
	3.0	3	4	5	5	6	6	7	4	4	5	5	6	6	8	3.0
0.0125 (0.025)	0.1	506	625	768	954	1245	1514	2090	1004	1244	1529	1901	2482	3020	4180	0.1
	0.2	127	158	194	240	312	380	524	255	316	388	482	629	766	1057	0.2
	0.4	34	42	52	63	81	98	135	64	81	97	120	157	191	265	0.4
	0.6	17	19	24	30	37	45	61	30	37	44	54	71	84	117	0.6
	0.8	11	13	15	18	23	27	36	18	21	26	32	40	49	69	0.8
	1.0	8	9	11	13	16	18	24	12	14	17	21	27	32	44	1.0
	1.2	7	7	8	10	12	14	18	9	11	13	15	19	23	31	1.2
	1.4	6	6	7	8	10	11	14	7	8	10	12	15	17	23	1.4
	1.6	5	6	6	7	8	9	11	6	7	8	9	12	14	18	1.6
	1.8		5	6	6	7	8	10	5	6	7	8	10	11	15	1.8
	2.0			5	6	6	7	9	4	5	6	7	8	9	12	2.0
	3.0							6		3	4	4	5	5	7	3.0
0.025 (0.05)	0.1	386	492	619	788	1054	1302	1840	771	982	1237	1574	2106	2603	3680	0.1
	0.2	98	124	156	201	265	327	459	193	245	310	395	527	650	922	0.2
	0.4	26	33	41	52	68	85	117	49	62	80	100	133	164	231	0.4
	0.6	13	16	20	24	32	39	53	23	29	36	45	60	77	104	0.6
	0.8	9	10	12	14	18	23	31	14	17	21	26	34	42	59	0.8
	1.0	6	7	9	10	13	16	21	9	11	14	17	22	28	38	1.0
	1.2	6	6	7	8	10	12	15	7	8	10	12	16	19	27	1.2
	1.4	5	5	6	7	8	9	12	5	7	8	10	12	15	20	1.4
	1.6			5	6	7	8	10	5	5	6	8	10	12	16	1.6
	1.8				5	6	7	8	4	4	5	6	8	10	13	1.8
	2.0					5	6	7	4	4	5	6	7	8	11	2.0
	3.0											4	4	5	6	3.0
0.05 (0.10)	0.1	272	362	473	620	858	1084	1580	543	722	943	1235	1715	2166	3160	0.1
	0.2	69	92	119	156	215	272	396	138	182	237	312	430	543	793	0.2
	0.4	19	24	31	41	55	70	101	35	46	59	79	109	137	199	0.4
	0.6	9	12	15	19	26	32	46	16	21	28	35	48	62	89	0.6
	0.8	6	8	9	12	15	19	27	10	12	16	20	28	35	50	0.8
	1.0	5	6	7	8	11	13	18	7	8	11	13	18	23	33	1.0
	1.2		5	5	6	8	10	13	5	6	8	10	13	16	23	1.2
	1.4				5	6	8	10	4	5	6	7	10	12	17	1.4
	1.6					6	6	8	4	4	5	6	8	10	14	1.6
	1.8					5	6	7		4	4	5	7	8	11	1.8
	2.0						5	6			4	4	6	7	9	2.0
	3.0												3	4	5	3.0

Table 2

	a_1 = .01 (a_2 = .02)								
	r								
Power	.10	.20	.30	.40	.50	.60	.70	.80	.90
.25	273	68	31	18	12	9	7	5	4
.50	540	134	59	31	20	14	10	7	5
.60	663	164	72	39	24	16	11	8	6
2/3	757	187	81	44	28	18	13	9	6
.70	809	200	87	48	29	19	13	9	6
.75	897	221	96	53	32	21	14	10	7
.80	998	246	107	58	36	23	16	11	7
.85	1126	277	120	65	40	26	17	12	8
.90	1296	319	138	75	45	29	20	13	8
.95	1585	389	168	91	55	35	23	16	10
.99	2154	529	228	123	74	47	31	20	13

	a_1 = .05 (a_2 = .10)								
	r								
Power	.10	.20	.30	.40	.50	.60	.70	.80	.90
.25	99	24	12	8	6	4	4	3	3
.50	277	69	30	17	11	8	6	5	4
.60	368	92	40	22	14	10	7	5	4
2/3	430	107	47	26	16	11	8	6	4
.70	470	117	51	28	18	12	8	6	4
.75	537	133	58	32	20	13	9	7	5
.80	618	153	68	37	22	15	10	7	5
.85	727	180	78	43	26	17	12	8	6
.90	864	213	93	50	31	20	13	9	6
.95	1105	272	118	64	39	25	16	11	7
.99	1585	389	168	91	55	35	23	15	10

	a_1 = .10 (a_2 = .20)								
	r								
Power	.10	.20	.30	.40	.50	.60	.70	.80	.90
.25	39	11	6	4	3	3	3	3	3
.50	165	42	19	11	7	5	4	3	3
.60	236	59	27	15	10	7	5	4	3
2/3	293	73	33	18	12	8	6	4	4
.70	326	81	36	20	13	9	6	5	4
.75	383	95	42	23	14	10	7	5	4
.80	450	112	49	27	17	11	8	6	4
.85	536	133	58	32	19	13	9	6	4
.90	655	162	71	39	24	16	11	7	5
.95	864	213	93	50	31	20	13	9	6
.99	1296	319	138	75	45	29	19	13	8

Table 2 (continued)

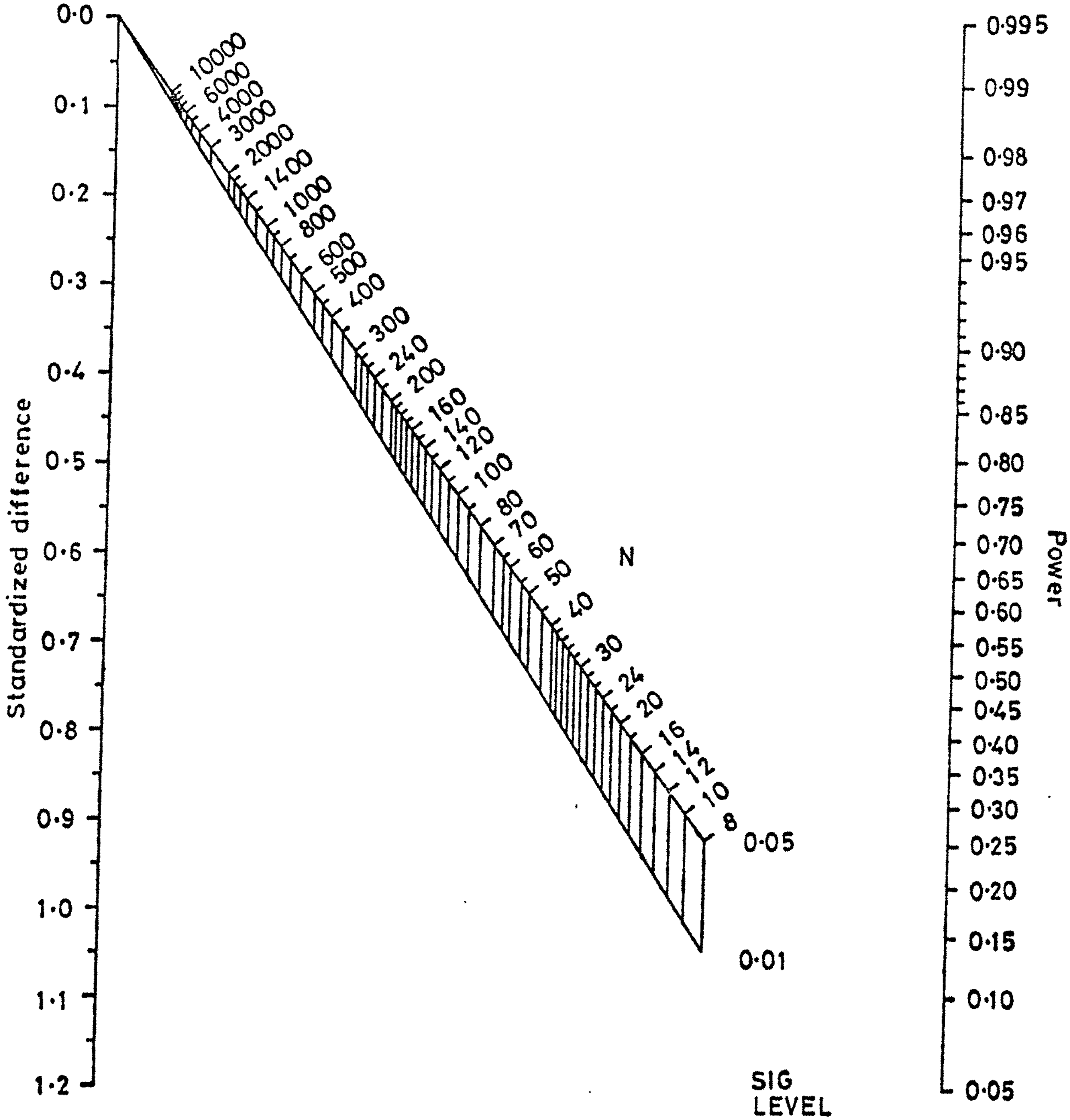
$a_2 = .01$ ($a_1 = .005$)

Power	r .10	.20	.30	.40	.50	.60	.70	.80	.90
.25	362	90	40	23	15	11	8	6	5
.50	662	164	71	39	24	16	12	8	6
.60	797	197	86	47	29	19	13	9	7
2/3	901	222	96	53	32	21	15	10	7
.70	957	236	102	56	34	23	15	11	7
.75	1052	259	112	61	37	25	17	11	8
.80	1163	286	124	67	41	27	18	12	8
.85	1299	320	138	75	45	30	20	13	9
.90	1480	364	157	85	51	34	22	15	9
.95	1790	440	190	102	62	40	26	17	11
.99	2390	587	253	136	82	52	34	23	13

$a_2 = .05$ ($a_1 = .025$)

Power	r .10	.20	.30	.40	.50	.60	.70	.80	.90
.25	166	42	20	12	8	6	5	4	3
.50	384	95	42	24	15	10	7	6	4
.60	489	121	53	29	18	12	9	6	5
2/3	570	141	62	34	21	14	10	7	5
.70	616	152	66	37	23	15	10	7	5
.75	692	171	74	41	25	17	11	8	6
.80	783	193	84	46	28	18	12	9	6
.85	895	221	96	52	32	21	14	10	6
.90	1046	258	112	61	37	24	15	11	7
.95	1308	322	139	75	46	30	19	13	8
.99	1828	449	194	104	63	40	27	18	11

Figure 1


Standardized difference
0·0
0·1
0·2
0·3
0·4
0·5
0·6
0·7
0·8
0·9
1·0
1·1
1·2
N
10000
6000
4000
3000
2000
1400
1000
800
600
500
400
300
240
200
160
140
120
100
80
70
60
50
40
30
24
20
16
14
12
10
8
0·05
0·01
SIG LEVEL
Power
0·995
0·99
0·98
0·97
0·96
0·95
0·90
0·85
0·80
0·75
0·70
0·65
0·60
0·55
0·50
0·45
0·40
0·35
0·30
0·25
0·20
0·15
0·10
0·05