NUMBER OF REPLICATES IN EXPERIMENTAL RESEARCH

Victor Chew

Biometrics Unit, SR, ARS, USDA
412 Rolfs Hall, University of Florida
Gainesville, Florida 32611

ABSTRACT

Before a consulting biometrician can advise his client on the number of replicates needed for testing differences among treatments, the experimenter must provide the following information: (a) common standard deviation of data from same treatment; (b) smallest difference between treatments that is of interest; (c) probability of detecting above difference; and (d) probability of spuriously finding a difference. For interval estimation of the difference between treatment means, (b) above is replaced by the required precision of the estimate; and (c) and (d) are replaced by the probability that the estimate will have the prescribed precision. After a brief discussion of hypothesis testing and interval estimation, formulas are given for both testing and estimating means of one or more normal and binomial populations.

INTRODUCTION

The question most frequently asked of a consulting biometrician by research scientists is "How many replicates should I take?" This question cannot be answered until the experimenter supplies three quantities. The first is the standard deviation (σ) of observations from the same "treatment;" obviously, if $\sigma=0$, one observation per treatment will suffice. It also depends inversely on the magnitude of the treatment effects. If two treatments differ considerably in their effects, we need fewer replicates to establish that difference than will be the case if the two treatments are more alike. Finally, it depends on the desired probability of drawing the correct conclusion from the experiment. A random sample may, by chance, be nonrepresentative of the population from which it has been selected. To insure a higher degree of confidence of a correct inference, we have to pay the premium of a larger sample size.

The phrases "number of replicates" and "sample size" will be used synonymously. The former is more often used in connection with experiments and the latter with surveys. The number of replicates is not necessarily the same as the total number of observations per treatment. Suppose each treatment unit consists of a plot of land (or of 4 rows, say, of cotton plants) and is replicated r times. If 4 soil samples are taken from each plot (or if each row of plants is sampled) and the 4 subsamples are individually analysed (instead of being composited), the number of replicates is still r, although the total number of observations per treatment is 4r. Variation among subsamples from the same plot is smaller than that among plots, and use of the former or a combination of both will lead to an underestimation of the error variance. See Misuse #6 in Nelson and Rawlings (1983).

The sample size question is of economic importance. An experiment that is larger than necessary to give the required precision is wasteful; and so is an experiment that is not large enough (due to manpower and/or financial constraints) to achieve the prescribed objectives. If these specifications cannot be relaxed, it is better to cancel or postpone the experiment until adequate resources are available. Before giving the sample size formulas for the various situations, we need first to introduce some statistical terminology.

STATISTICAL INFERENCE

Statistical (or inductive) inference is of two kinds: hypothesis testing and interval estimation. We will discuss them in the above traditional order, although interval estimation is much more useful and informative; in fact, Chew (1977a) regards hypothesis testing as a academic exercise in futility. This is because no two true (or population) treatment means will be exactly equal (to billions of decimal places). It is absurd therefore to test a hypothesis that we know beforehand is not strictly true. It is more logical to assume that there is a true difference, and perform the experiment to estimate this difference. Chew (1980) gives still other alternatives to hypothesis testing. We will use the case of two normal populations with means μ_1 and μ_2 and common standard deviation σ for illustration.

Hypothesis Testing. In comparing two treatments, the hypothesis to be tested (called the null hypothesis and denoted by H_o) is that the two population means μ_1 and μ_2 are equal. We write it as H_o: $\mu_1 - \mu_2 = 0$ or H_o: $\mu_1 = \mu_2$. The classical test allows only one of the following two conclusions: "accept H_o" or "reject H_o." (In sequential testing, we have a third option: collect more data.) Whichever conclusion we draw, we do not know whether it is right or wring. There are two possible errors: rejecting H_o when H_o is in fact true (i.e., saying $\mu_1 \neq \mu_2$ when in fact $\mu_1 = \mu_2$, a Type 1 error) or accepting H_o when H_o is in fact false (Type 2 error). The conditional probabilities of a test making these two errors are denoted by α and β, respectively; i.e., α is the probability of a test rejecting H_o conditional on (or given that) H_o is true, and β is the probability of the test accepting H_o given that H_o is false. The quantity α (usually expressed as a percent) is called the significance level of the test. The quantity $(1-\beta)$ is thus the probability of the test giving the correct decision of rejecting H_o when H_o is false, and is called the power of the test. The power of a test increases with larger differences.

The ideal but unattainable test of course is the infallible one, for which $\alpha=\beta=0$. In a good experiment, both α and β will be small (α=.05 or .01 and β=.10 or .20, say). Requiring both α and β to be equal to .01 or even .05 usually requires a prohibitively large experiment. We can decrease α only at the expense of increasing β, and vice-versa. To decrease both α and β simultaneously, we have to decrease σ (by refining the experimental technique, using more homogeneous experimental material, etc.) and/or increase the sample size. The null hypothesis H_o is unique, so that we know the probability distribution of the statistic for testing H_o. We can thus make α as small as we like by suitably choosing the so-called rejection region for the test statistic, but remember that the smaller the α, the bigger the β.

The Type 2 error probability β is harder to control. If H_o is false, the alternative hypothesis H_a: $\mu_1 \neq \mu_2$ is not uniquely defined. We need to know the actual magnitude of the difference to derive the probability distribution of the test statistic when H_a is true, and consequently we cannot control β by suitably choosing the rejection region, as we can with α. This is why very often many experimenters only control α (at .05, say) and ignore the Type 2 error probability. If β is ignored, one may just as well simulate the experiment by randomly selecting a card from a deck of 20 index cards (one of which is marked), and rejecting H_o (whatever it is) if the marked card is drawn. At a savings of hundreds of dollars, Type 1 error is controlled at 5%.

One important objective in statistically designing an experiment is to ensure (by a proper choice of sample size) that it has a reasonable chance (at least 75%, say) of detecting a difference d between μ_1 and μ_2 that is of economic

importance, where d is to be numerically specified by the experimenter. Acceptance of H_o in such an experiment then implies (with high probability) that at most a small difference exists between μ_1 and μ_2. If H_o is accepted in a statistically unplanned experiment, the power should be calculated, to help decide whether the acceptance of H_o is due to a small difference or to insufficient replication.

For added power at no extra cost, we should always perform a so-called one-tail test, provided of course that it is justifiable. Comparing two new insecticides will require a two-tail test (H_a: $\mu_1 \neq \mu_2$), but a one-tail test (H_a: $\mu_1 > \mu_2$) can be defended when comparing an insecticide with control (no insecticide) or when comparing two doses of an insecticide.

Interval Estimation. The unknown difference $(\mu_1 - \mu_2)$ may be estimated by a point (the single number $(\bar{x}_1 - \bar{x}_2)$ or difference between the two sample means) or by an interval (two numbers L and U, called the lower and upper confidence limits). The numbers L and U are of the form $(\bar{x}_1 - \bar{x}_2) \pm k$ [standard error of $(\bar{x}_1 - \bar{x}_2)$], where k depends on the confidence (probability) that we would like the interval (L, U) to contain (include) the unknown difference $(\mu_1 - \mu_2)$. From a large sample, k=1.96 corresponds to 95% probability, and (L,U) is then called a 95% confidence interval (CI) for $(\mu_1 - \mu_2)$.

Interval estimation is more informative than hypothesis testing. If the signs of L and U are the same (either both positive or both negative), then without doing any testing, we know that H_o: $\mu_1 - \mu_2 = 0$ will be rejected. H_o will be accepted only if L is negative and U is positive. Thus, the interval estimate contains the information from hypothesis testing, plus more. If the CI for $(\mu_1 - \mu_2)$ is (0.7, 0.9), say, H_o will be rejected, but there is nothing to be excited about (assuming a difference of about 0.8 unit is of no practical importance). A CI of 17 to 19 gives not only a statistically significant difference between the two means but also one of practical importance, if a difference of about 18 units is important. If the CI is (-10, 30), H_o may be accepted in error because of insufficient replication, and the experiment should be repeated, as there may be an important difference.

SAMPLE SIZE FOR TESTING MEANS OF NORMAL POPULATIONS

One Mean. We wish to test H_o: $\mu=\mu_0$ (numerically specified) against a one-tail or two-tail alternative, depending on circumstances. Here, μ_0 may be the mean of an old treatment (assumed known exactly from long experience with it) or the mean of a new treatment claimed by a supplier. If we fix Type 1 error probability at α and we wish to have at least $100(1-\beta)\%$ probability of rejecting H_o if $\mu-\mu_0=d$ (specified difference) or larger, Zar (1981) gives the necessary sample size as

$$n = (\sigma/d)^2 (t_{\alpha(i),n-1} + t_{\beta(1),n-1})^2, \qquad (1)$$

where i=1 for a one-tail and i=2 for a two-tail test, and (n-1) is the number of degrees of freedom (d.f.). For illustration, $t_{0.05(1),15} = 1.753$ and $t_{0.05(2),15} = 2.131$, from tables of the t-distribution.

From the formula, we see that n increases as σ increases, decreases as d increases, and increases as α and β decrease, since $t_{0.01} > t_{.05}$. Note that we do not really need to know the population standard deviation σ, but only its ratio to d; e.g., we may be interested in a difference d that is 50% of σ, say. However, if we are interested in a difference of a specified magnitude, we have to

estimate σ from previous experience with similar material, or from a small preliminary sample. After the experiment has been performed, the posterior or experimental estimate of σ should be compared with the prior estimate. If the prior estimate is smaller, the experiment may not have the power that it was designed to have.

Formula (1) can only be used iteratively, because the t-values depend on the unknown n. A practical approximation is to take infinite d.f. for the t-values, in which case the t-distribution becomes the z-distribution (standardized normal). Equation (1) becomes

$$n = (\sigma/d)^2 (z_{\alpha(i)} + z_{\beta(1)})^2. \qquad (2)$$

For example, $z_{0.05(2)} = t_{0.05(2),\infty} = 1.96$ and $z_{0.05(1)} = t_{0.05(1),\infty} = 1.645$. Since the t-value decreases as the d.f. increases, Equation (2) underestimates n, so add a few more observations to compensate for this bias.

Equation (2) is very easy to use for calculating n, but charts are available (see later). Equation (2) may also be used in reverse. Suppose H_o is accepted in an unplanned experiment (i.e., one where no prior consideration was given to the requisite sample size). Is H_o accepted because the difference d is small or because of inadequate replication? We can answer the following questions: (a) What is the probability (power) of the experiment detecting a given difference (d=1.8, say)? We have an estimate of σ from the data. We know n, d, and $z_{\alpha(i)}$, from which we can calculate $z_{\beta(1)}$, and hence obtain β from tables of the normal distribution. (b) What size difference can the experiment detect with a given power (80% or β=0.20, say)? From the known values of n, σ, $z_{\alpha(i)}$, and $z_{\beta(1)}$, we can calculate d.

Two Means. For Type 1 error probability of α and 100(1-β)% power of detecting a difference of specified magnitude $(\mu_1 - \mu_2) = d$, Cochran and Cox (1957) and Zar (1981) give the following sample size per treatment for testing H_o: $\mu_1 = \mu_2$.

$$n = 2(\sigma/d)^2 (t_{\alpha(i),n-1} + t_{\beta(1)n-1})^2. \qquad (3)$$

The sample size from Equation (3) is twice that from Equation (1), due to the fact that the variance of the difference between two means is twice that of one mean. As before, we may approximate (3) by

$$n = 2(\sigma/d)^2 (z_{\alpha(i)} + z_{\beta(1)})^2. \qquad (4)$$

Also, as before, we may use Equation (4) in reverse, if H_o is accepted, to calculate the power of the experiment to detect a difference of a given magnitude (d=2.0, say), and/or to calculate the size of the difference that can be detected with a given power (80% or β=0.2, say).

To facilitate the use of Equation (3), tables are given in Cochran and Cox (1957) and, more extensively, in Cohen (1977), Pages 26-37 and 52-53. Croarkin (1962) gives 5 charts with n on the vertical axis and $\delta = |\mu_1 - \mu_2|/\sigma = |d|/\sigma$ on the horizontal axis, corresponding to 10%, 50%, 90%, 95%, and 99% power, for each value of $\alpha(2) = 2\alpha(1) = .20, .10, .05, .02$, and .01. (This paper also gives similar charts for the one-mean case, with $\delta = |\mu - \mu_0|/\sigma$.) We can use the charts or tables directly to find n (for planning purposes), given α, β, and δ. Conversely, for post-experiment analysis if H_o is accepted, we can use them to determine the magnitude of the difference that can be detected with a given power and/or the power of the experiment to detect a given difference. Cohen (1977) considers δ = .2, .5, and .8 as small, medium, and large effects, respectively. These values may be used if we cannot specify $(\mu_1 - \mu_2)$ and σ separately. For the number of replicates with paired samples, see corresponding section on

estimation.

Three or More Means. To control β, we need to know the probability distribution of the test statistic under the alternative hypothesis H_a. With two means, it is easy to specify H_a; namely, H_a: $\mu_1 - \mu_2 = d$. We can then find β or power $(1-\beta)$ for various assigned values of d. With k (3 or more) means, it is more difficult to specify H_a, as there are (k-1) independent differences. It turns out mathematically that the power does not depend directly on these (k-1) differences but only on the so-called non-centrality parameter $\Sigma(t_i - \bar{t})^2/\sigma$, where t_i is the "effect" of the i-th treatment (i = 1, 2, ..., k) and $\bar{t}$ is the average of the k treatment effects. Charts for obtaining sample size corresponding to a given power or vice-versa are given in Pearson and Hartley (1972).

It is too much to expect experimenters to be able to specify the value of the noncentrality parameter given above or to specify the (k-1) independent differences (from which the noncentrality parameter can be calculated, assuming as usual that σ is known). Kastenbaum et al. (1970) give tables (for both Completely Randomized and Randomized Complete Block designs) using the simpler noncentrality parameter $\delta = (\mu_{max} - \mu_{min})/\sigma$, the largest of the (k-1) standardized differences. If k=2, the above reduces to the noncentrality parameter used in the previous two-mean case.

SAMPLE SIZE FOR ESTIMATING MEANS OF NORMAL POPULATIONS

One Mean. We assume a normal population with unknown mean μ and known standard deviation σ. We wish to know the number n of observations to take such that there is a large probability $(100(1-\alpha)\%)$ that the sample mean $\bar{x}$ will be "close" to μ. We can define closeness either (a) absolutely or (b) relatively (as a percentage of the mean).

(a) To be $100(1-\alpha)\%$ confident that the sample mean will not be in error (too large or too small) by more than d units, the required sample size is

$$n = (z_{\alpha(2)}\sigma/d)^2. \tag{5}$$

Values of n are given in Table 10.3 (Page 142) in Marks (1982).

(b) To specify that the estimate $\bar{x}$ is to be in error by no more than d=3 units say, will be satisfactory if μ is about 100; but if μ is about 10, an error of 3 units is a 30% error. Another way to define the maximum allowable error in $\bar{x}$, therefore, is as a percentage of μ. To be $100(1-\alpha)\%$ confident that the sample mean will not be in error (in either direction) by more than 100p% of the mean, the required sample size is

$$n = (z_{\alpha(2)}/p)^2(\sigma/\mu)^2. \tag{6}$$

Now, not only do we need σ but also μ, the quantity we are estimating! However, we do not need σ and μ individually, but only their ratio or coefficient of variation. Of course, μ and σ may be estimated from a preliminary sample or from experience with similar material.

Karandinos (1976) also applies Equations (5) and (6) to other distributions (binomial, negative binomial, and Poisson), by invoking the Central Limit Theorem that $\bar{x}$ will be approximately normally distributed for large n. When applied to these other distributions, the known relationships between μ and σ may be utilized (e.g., $\mu=\sigma^2$ for the Poisson).

Two Means. Let $x_{11}, x_{12}, \ldots, x_{1n}$ be a sample from Treatment 1 with mean μ_1, and let $x_{21}, x_{22}, \ldots, x_{2n}$ be a sample from Treatment 2 with mean μ_2. There are two cases to be considered. Usually, the two samples are independent of each other, but there are instances where the samples are correlated in pairs; e.g., two legs from the same chicken, two leaves from the same plant, two measurements at different times on the same experimental unit, etc.

(a) Dependent (Paired) Samples. Let $d_i = (x_{1i} - x_{2i})$, $i = 1, 2, \ldots, n$. Then, μ_d = expected mean of $d = (\mu_1 - \mu_2)$. Any testing or interval estimation of $(\mu_1 - \mu_2)$ is thus reduced to the one-mean case of testing or estimation of the mean of the population of d's.

(b) Independent Samples. We want the number of observations n from each sample to be large enough such that the probability is at least $100(1-\alpha)\%$ that our estimate $(\bar{x}_1 - \bar{x}_2)$ will be within d units of $(\mu_1 - \mu_2)$. The necessary sample size is

$$n = 2(z_{\alpha(2)}\sigma/d)^2, \tag{7}$$

which is double the sample size in Equation (5). The percentage error criterion of precision is not as important here, but if needed, n is simply twice that given in Equation (6), where μ will be replaced by $(\mu_1 - \mu_2)$.

Three or More Means. After the data have been collected, multiple comparison methods (Chew (1976)) are available for getting simultaneous confidence intervals for all possible differences among the k means. Tukey's HSD (honestly significant difference) is to declare the means of Treatments i and j to be different if $|\bar{x}_i - \bar{x}_j|$ exceeds HSD; similarly, the confidence interval for $(\mu_i - \mu_j)$ is $(\bar{x}_i - \bar{x}_j) \pm$ HSD, where

$$\text{HSD} = q(\alpha, k, \nu)\sqrt{s^2/n}. \tag{8}$$

In the above equation, s^2 is the error mean square with ν d.f., and $q(\alpha, k, \nu)$ is the α-point of the so-called studentized range distribution. Using large sample approximation, we may replace the above equation by

$$\text{HSD} \simeq q(\alpha, k, \infty)\sqrt{\sigma^2/n}, \tag{9}$$

just as we approximated (1) by (2). Selected values of $q(\alpha, k, \infty)$ are:

		Number of Treatments (k)								
		2	3	4	5	6	7	8	9	10
α	.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47
	.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16

See Chew (1977b) for values of $q(\alpha, k, \infty)$ for more than 10 treatments. We can solve Equation (9), by setting HSD = d, to find the required number of observations n per sample such that the probability is $100(1-\alpha)\%$ that $(\bar{x}_i - \bar{x}_j)$ will be within d units of $(\mu_i - \mu_j)$ for all pairs (i, j).

$$n = \lceil q(\alpha, k, \infty)\sigma/d \rceil^2. \tag{10}$$

If k=2, Equation (10) reduces to (7).

SAMPLE SIZE FOR BINOMIAL POPULATIONS

Research data are sometimes quantal in nature. Instead of a measurement, each experimental unit (e.u.) gives only a "yes" or a "no" (e.g., dead or alive). In general, denote the two outcomes from an e.u. by S(success) or F(failure). A good general reference is Fleiss (1981).

One Population. Let π be the proportion of success in the population. (E.g., π may be the true proportion of insects that will survive when exposed to a certain dose of some insecticide.) Denote the sample data by $x_1, x_2, \ldots, x_n$. Each x will be either 0(failure) or 1(success). The ordinary mean $\bar{x}$ is the sample proportion of successes and will estimate π. From theory, the expected mean of x is $\mu=\pi$ and the standard deviation is $\sigma = \sqrt{\pi(1-\pi)}$. We can use the large sample normal approximation to the binomial for testing H_o: $\pi=\pi_o$ (specified)

using Equation (2), and use Equations (5) and (6) for interval estimation of π, after replacing σ by $\sqrt{\pi(1-\pi)}$ and μ by π. To apply these equations, we need to have an estimate of π.

(a) Sample size for testing H_o: $\pi=\pi_o$ (specified) with Type 1 error probability α and power at least $(1-\beta)$ if $|\pi-\pi_o| \geq d$.

$$n = (\pi(1-\pi)/d^2)(z_{\alpha(1)} + z_{\beta(1)})^2 \qquad (11)$$

(b) Sample size for estimating π such that the probability is $100(1-\alpha)\%$ that the estimate will not be in error by more than d in absolute magnitude.

$$n = (z_{\alpha(2)}/d)^2\pi(1-\pi) \qquad (12)$$

(c) Sample size for estimating π such that the probability is $100(1-\alpha)\%$ that the estimate will not be in error by more than $(100p)\%$ of π.

$$n = (z_{\alpha(2)}/p)^2(1-\pi)/\pi \qquad (13)$$

Values of n in Equation (12) are given in Table 10.1 (Page 132) in Marks (1982). For an improved approximation to n in Equation (11), see Guenther (1974). Power and sample size tables are given in Chapter 5 in Cohen (1977) for testing H_o: $\pi = 0.5$. Exact 95% and 99% confidence intervals for π for n = 1, 2, ..., 30 are given in Blyth and Still (1983). For a multinomial population (quantal with 3 or more outcomes - dead, sick, or healthy), see Angers (1974) and Tortora (1978).

Two Populations. Let π_1 and π_2 be the true proportions of successes in the 2 populations. As examples, we may be comparing two insecticides (with the same concentration of active ingredient), two rates of an insecticide, etc.

(a) Testing H_o: $(\pi_1 - \pi_2) = 0$. Power and sample sizes tables are given in Cohen (1977), Chapter 6. This problem is more difficult than testing the equality of the means of two normal populations. Besides specifying α and β, we also need to specify both π_1 and π_2 and not just their difference $d = \pi_1 - \pi_2$. Differentiating between $\pi_1 = 0.65$ and $\pi_2 = 0.45$ is more difficult than that between $\pi_1 = 0.25$ and $\pi_2 = 0.05$, although both differences are equal to $d = 0.20$. In a two-tail test with $\alpha = 0.05$ and $\beta = 0.20$, Table 6.4.1 in Cohen (1977) shows that the above 2 problems require n = 98 and 44, respectively. Tables are also given in Marks (1982), Pages 149-164 and in Cochran and Cox (1957), Pages 24-25. Exact tables are given in Casagrande et al. (1978). Aleong and Bartlett (1979) give graphs for the sample size. Fleiss et al. (1980) have generalized it to the case of unequal sample sizes. For optimum allocation, see Brittain and Schlesselman (1982).

(b) Estimating $(\pi_1 - \pi_2)$. The sample size for estimating $\pi_1 - \pi_2$ such that the probability is $100(1-\alpha)\%$ that the estimate $(\bar{x}_1 - \bar{x}_2)$ will not be in error by more than d in absolute magnitude is

$$n = (z_{\alpha(2)}/d)^2[\pi_1(1-\pi_1) + \pi_2(1-\pi_2)] \qquad (14)$$

To use the equation, we need to know π_1 and π_2 (at least approximately). The values of n are given in Table 10.2 (Pages 135-141) in Marks (1982).

Three or More Populations. For sample size for testing H_o: $\pi_1 = \pi_2 = \ldots = \pi_k$, see Bruvold and Murphy (1978).

CONCLUSION

We have only considered simple random sampling. In many instances, sampling is multi-stage; e.g., we select two or more fields from a locality, two or more trees from a field, two or more branches from a tree, two or more leaves from a

branch, and two or more analyses from a leaf. The costs of taking a sample from the different stages obviously differ. Optimal sample size per stage can be calculated to either maximize precision for a given total cost, or to minimize cost for a prescribed precision. See Cochran (1977) for details and for other sampling plans.

LITERATURE CITED

Aleong, J. and D.E. Bartlett. 1979. Improved graphs for calculating sample sizes when comparing two independent binomial distributions. Biometrics 35:875-81.

Angers, C. 1974. A graphical method to evaluate sample sizes for the multinomial distribution. Technometrics 16:469-71.

Blyth, C.R. and H.A. Still. 1983. Binomial confidence intervals. J. Amer. Stat. Assoc. 78:108-16.

Brittain, E. and J.J. Schlesselman. 1982. Optimal allocation for the comparison of proportions. Biometrics 38:1003-9.

Bruvold, N.T. and R.A. Murphy. 1978. Sample sizes for comparison or proportions. Technometrics 20:437-40.

Casagrande, J.T., M.C. Pike and P.G. Smith. 1978. The power function of the "exact" test for comparing two binomial distributions. Applied Statistics (Series C, J. Royal Stat. Soc.) 27:176-80.

Chew, V. 1976. Comparing treatment means: a compendium. HortScience 11:348-57.

Chew V. 1977a. Statistical hypothesis testing: an academic exercise in futility. Proc. Fla. State Hort. Soc. 90:214-15.

Chew, V. 1977b. Comparisons among treatment means in an analysis of variance. ARS, USDA Technical Bulletin ARS/H/6 (available from author). 64 pp.

Chew, V. 1980. Testing differences among means: correct interpretation and some alternatives. HortScience 15:467-70.

Cochran, W.G. 1977. Sampling Techniques, 3rd ed. Wiley, NY. 413 pp.

Cochran, W.G. and G.M. Cox. 1957. Experimental Designs, 2nd ed. Wiley, NY. 611 pp.

Cohen, J. 1977. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Academic Press, NY. 415 pp.

Croarkin, C. 1962. Graphs for determining the power of Student's t-test. J. Res. Nat. Bur. Standards (B) 66B:59-70.

Fleiss, J.L. 1981. Statistical Methods for Rates and Proportions, 2nd ed. Wiley, NY. 321 pp.

Fleiss, J.L., A. Tytun, and H.K. Ury. 1980. A simple approximation for calculating sample sizes for comparing independent proportions. Biometrics 36:343-46.

Guenther, W.C. 1974. Sample size formulas for some binomial type problems. Technometrics 16:465-67.

Karandinos, M.G. 1976. Optimum sample size and comments on some published formulae. Entomol. Soc. Amer. Bull. 22:417-21.

Kastenbaum, M.A., D.G. Hoel and K.O. Bowman. 1970. Sample size requirements. Biometrika 57:421-30 and 573-77.

Marks, R.G. 1982. Designing a Research Project. Wadsworth, Belmont, CA. 192 pp.

Nelson, L.A. and J.O. Rawlings. 1983. Ten common misuses of statistics in agronomic research and reporting. J. Agron. Educ. 12:100-5.

Pearson, E.S. and H.O. Hartley. 1972. Biometrika Tables for Statisticians. Volume 2. University Press, Cambridge. 355 pp.

Tortora, R.D. 1978. A note on sample size estimation for multinomial population. Amer. Statistician 32:100-2.

Zar, J.H. 1981. Power of statistical testing: hypotheses about means. Amer. Lab. 13:102-7.