19: Sample Size, Precision, and Power

A study that is insufficiently precise or lacks the power to reject a false null hypothesis is a waste of time and money. A study that collects too much data is also wasteful. Therefore, before collecting data, it is essential to determine the sample size requirements of a study. Before calculating the sample size requirements of a study you must address some questions. For example:

- Do you want to learn about a mean? a mean difference? a proportion? a proportion (risk) ratio? an odds ratio? a slope?
- Do you want to *estimate* something with a given precision or do you want to test something with a given power?
- What type of sample(s) will you be working with? A single group? Two or more independent groups? Matched pairs?

Let us first address the problem of estimating a mean or mean difference with given precision.

Sample Size Requirements for Estimating a Mean or Mean Difference

To determine an appropriate sample size you must first declare an acceptable margin of error d. Recall that margin of error d is the wiggle room around the point estimate. This is equal to *half* the confidence interval width. When estimating μ with 95% confidence use

$$n = \frac{4s^2}{d^2} \qquad \qquad (19.1)$$

Example: To obtain a margin of error of 5 for a variable with a standard deviation of 15, $n = (4)(15^2)/(5^2) = 36$.

This method is applied to estimating a mean difference based on *paired samples* (μ_d) by using the standard deviation of the DELTA variable (s_d) in your formula:

$$n = \frac{4s_d^2}{d^2} \qquad \qquad (19.2)$$

With *independent samples*, use the pooled estimate of standard deviation (s_p) as your standard deviation estimate:

$$n = \frac{4s_p^2}{d^2} \qquad \qquad (19.3)$$

You should put considerable effort into getting a good estimate of the standard deviation of the variable you are studying since sample size calculations depend on this fact. Such estimates come from prior studies, pilot studies, and "Gestalt" (a combination of sources that contribute to knowledge about the variable).

EpiCalc 2000: Right click the tree window and select `Sample > Precision > Single mean`. Enter the standard deviation estimate into the `SD` field and the required margin of error (d) into the `Precision` field. Because

EpiCalc 2000 uses a slightly different formula than those above, its sample size calculation will be *slightly* smaller. This should not concern you because the problem with sample size calculations come not from small differences in formulas but from differences in values entered into the formula. *Sample size calculations are always rough.*

Sample Size Requirements for Testing a Mean or Mean Difference

Power $(1-\beta)$ is the probability of avoiding a type II error. A type II error occurs when you retain a false H_0. Conventional practice is to determine the sample size that gives 80% power at the 0.05 level of significance (two-sided). This can be determined with:

$$n = \frac{16s^2}{\Delta^2} + 1$$

(19.4)

(Dallal, 1997, www.tufts.edu/~gdallal/SIZE.HTM) where

> n represents the required sample size *per group*

> Δ represents the expected mean difference (a difference worth detecting). For a one-sample t test $\Delta = \mu - \mu_0$. For a paired-sample t test $\Delta = \mu_d$. For an independent t test $\Delta = \mu_1 - \mu_2$. These values are NOT calculated but come from conjecture.

> σ represents the standard deviation of the variable as estimated by s, s_d, or s_p depending on whether data are from a single sample, paired samples, or independent samples.

Example: If you are trying to detect a mean difference of 18 for a variable with a standard deviation of 30, the

required sample size *per group* $= \frac{(16)(30^2)}{18^2} + 1 \cong 46$.

EpiCalc 2000: Right click the left tree and select `Sample > Size > Single mean`. Enter the mean difference worth detecting (Δ) into the field labeled `Mean`, leave the field labeled `Null hypothesis value` as 0, and enter s, s_d, or s_p into the field labeled `SD`. *EpiCalc 2000* sample size calculations will be *slightly* smaller than those provided by the above formula.

When testing multiple treatments with ANOVA I suggest you focus on the most important post hoc comparison of means and use the above formula, perhaps using the *square root* of the Mean Square Within as your standard deviation.

Power of a t Test (Comparison of Means)

Let $f(z)$ represent the area under the curve to the left of z on a standard normal curve. For example, $f(0) = 50\%$, $f(0.84) = 80\%$, and $f(1.28) = 90\%$. (If this notation confuses you, draw a standard normal curve and place the z value on the x axis. The area under the curve to the left of the Z point represents the probability you want to know.)

The power of a test $(1-\beta)$ comparing means is approximately equal to:

$$1 - b = f\left(-1.96 + \frac{|\Delta|\sqrt{n}}{s}\right) \tag{19.5}$$

where Δ denotes the *expected* mean difference (or difference worth detecting), n denotes the per group sample size, and σ denotes the standard deviation of the variable (e.g., s, s_d, s_{pooled}, s_w, etc., depending on your sampling scheme).

Example: A study of 30 pairs expects a mean difference of 2. The standard deviation of the paired difference is 6. Thus, the power the test is

$f\left(-1.96 + \frac{|2|\sqrt{30}}{6}\right) = f(-0.13) = 0.45$ (Fig. right). This power is inadequate by most

any standard.
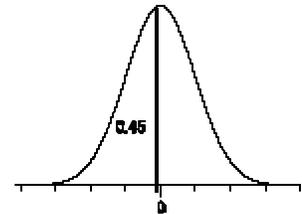
How this works. Formula 19.5 assumes a sampling distribution of no difference (H_0) and an alternative sampling distribution of a difference (H_1). Let critical value c determines the point at which you will reject H_0. The power of the test is the area under the *alternative* curve beyond c (Fig., below).
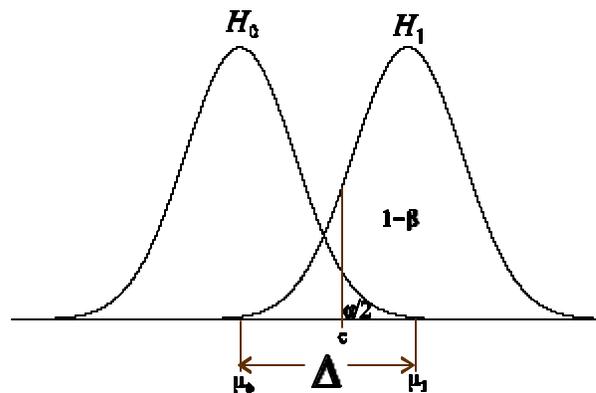
Estimating A Single Proportion

To calculate a 95% confidence interval for proportion p with margin of error d use a sample of size

$$n = \frac{(1.96)^2 \, pq}{d^2} \qquad\qquad \textbf{(19.6)}$$

where $q = 1 - p$.

Examples: To calculate a 95% confidence interval for p that is expected to be about 50% (0.50) with a margin of error (d) no more than 0.05, $n = (1.96^2)(.50)(1-0.50)/(.05^2) \cong 384$. To attain margin of error $d = 0.03$, $n = (1.96^2)(0.50)(1-0.50)/(.03^2) \cong 1067$.

This method requires you to enter a value for p, which is exactly what you want to estimate This bothered me as a student but is merely part of improving the accuracy of your estimate—one estimate begets a better estimate. If no estimate of p is available, assume $p = 0.50$ to obtain a sample that is big enough to ensure precision.

EpiCalc 2000: Right-click the left-hand tree and select `Sample > Precision > Single Proportion`. Enter p in the form of a percentage into the field labeled `Proportion (%)` and enter d in the form of a percentage into the field labeled `Precision (+/-)`.

Testing Two Proportions Based on a Cohort or Cross-Sectional Sample

Use *EpiCalc 2000* `> Size > Two Proportions` when testing H_0: $RR = 1$. Enter the significance level of the test, required power, expected incidence proportion or prevalence in the exposed group (in the form of a `%`) in the field labeled `Proportion 1 (%)`, and the expected incidence proportion or prevalence in the nonexposed group in the field labeled `Proportion 2 (%)`.

Example: In testing H_0: $RR = 1$ at $\alpha = .05$ (two-sided) with power = 80% while assuming an incidence of 10% in the exposed group and the incidence of 5% in the nonexposed group, you'll need 433 people per group.

Testing an Odds Ratio from A Case-Control Study

Use *EpiCalc 2000* `> Size > Case-Control Study` when testing data from a case-control study (H_0: $OR = 1$). Enter the significance level of the test, required power, `Ratio of cases to controls`, the odds ratio worth detecting (`OR to detect`), and expected "prevalence" of exposure in the source population ("controls") in the field labeled `Proportion (%) controls exposed`.

Example: Suppose you test H_0: $OR = 1$ at $\alpha = .05$ (two-sided) with power = 80% using a 1:1 ratio of cases to controls while looking for an odds ratio of 2. You assume the prevalence of exposure in the source population (controls) is 25%. *EpiCalc 2000* determines you'll need 151 cases and 151 controls in your study.