BIOMETRICS INFORMATION

(You're 95% likely to need this information)

PAMPHLET NO. # 11 DATE: August 22, 1988

SUBJECT: Sample Sizes: for one mean

This pamphlet discusses how to determine the required sample size to adequately estimate a mean. The following example will illustrate the method.

Suppose that a nursery manager wants to check some attribute of a crop of seedlings, say average height or dry weight. Obviously, not all seedlings can be examined so the manager will select a simple random sample of, say, n seedlings from N, the total number of seedlings in the crop. The mean of the sample, $\bar{y}$, will be used to infer plausible values of $\bar{Y}$, the mean for the whole crop. One of the first questions the manager will ask is: "How many seedlings should one collect? i.e. What value should n have?"

The manager must know or choose the following quantities before a required sample size can be calculated:

1) a prior estimate of variability, as either:
   i) the variance or standard error
   or *ii*) the coefficient of variation, CV.
2) the maximum acceptable error, as either:
   i) a specific value or confidence interval, CI
   or *ii*) a percent error, PE.
3) the desired confidence level $(1 - \alpha)$ for the CI or PE.

Recall that:

SAMPLE MEAN: $\bar{y} = \sum y_i / n$, y_i = individual measurement

VARIANCE OF $\bar{y}$: $\text{var}(\bar{y}) = [SE(\bar{y})]^2 = (1 - f)\, s^2/n$

where $f = n/N$ = sampling fraction

and $s^2 = \sum (y_i - \bar{y})^2 / (n-1)$ = variance of the observed data.

Thus, if seedling height or dry weight are normally distributed or, $n > 30$, then the mean may be normally distributed so that confidence limits for the population value Y, are:

$$\bar{y} \pm t \sqrt{\text{var}(y)} = \bar{y} \pm t^*SE = \bar{y} \pm CI = \bar{y} \pm PE$$

where t is obtained from the t-tables with:

i) degrees of freedom, df = n - 1

and *ii*) probability level $\alpha/2$ (for 95% confidence limits $\alpha = 1 - 0.95$)

The above equations can be solved for sample size:

$$n = \frac{t^2 s^2 \; (1 - f)}{CI^2} = \frac{t^2 \; CV^2 \; (1 - f)}{PE^2}$$

BRITISH COLUMBIA Ministry of Forests Research Program

Usually, the population size is much greater than the sample size so that the sampling fraction, f, is effectively zero. In this case,

$$n = \frac{t^2\ s^2}{CI^2} = \frac{t^2\ CV^2}{PE^2}$$

The solution to these equations requires iteration since t depends on n (through its df). Stauffer (1983) discusses this and provides a variety of sample size tables (based on CV and PE) so that the work is done for you.

The mean can be tested against a standard. For instance, to test whether the population mean height of a crop was not different from some standard height, h, the following t-test could be used.

$$t = \frac{\bar{y}\ -\ h}{SE},\ \ df = n - 1$$

CAUTION: Stauffer's tables are only appropriate for determining the sample size required to estimate ONE mean. They are NOT appropriate when the main purpose of the data collection is to compare two or more treatments or if other sampling methods, such as multi-stage, are used. In these cases the procedures must be modified. They may also have to be modified if the data is categorical, such as when survival, dead or alive, is the response variable.

EXAMPLE: Suppose the nursery manager wants to sample a crop so that the crop mean height will be known within 10% error with a 95% confidence. Previous sampling indicates a CV of 50%. From Table 5, (Stauffer, page 22) it is found that n = 99. Thus a simple random sample of about 100 seedlings would be required.

Further, suppose that the sampling results in a mean of $\bar{y}$ = 30 cm. with a standard error, s.e. = 5 cm. The 95% confidence limits would be 30 cm ± (1.982)(5 cm.) = 30 cm. ± 9.9 cm. = 30 cm. ± 33%.

Reference:

Stauffer, Howard B., 1983. *Some sample size tables for forest sampling*. B.C.M.F. Research Note. No. 90.

CONTACT: Wendy Bergerud
387-5676

P.S. While this is a standard method for determining sample sizes for one mean, the estimates will tend to be too small. See a paper by L.L. Kuper and K.B. Hafner, May 1989 in The American Statistician Journal titled ***How appropriate are popular sample size formulas?*** on pages 101 to 105.