BIOMETRICS INFORMATION

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

PAMPHLET NO. # 49 DATE: September 29, 1995

SUBJECT: Power Analysis and Sample Sizes for Completely Randomized Designs with Subsampling

Most research studies in forestry subsample experimental units to obtain response measures. A common example occurs when rows or plots of trees are randomly assigned one of several treatments. While the row is the experimental unit, the response variables of height, diameter, etc. must be measured on individual trees. These subsamples (trees) provide an estimate of the response of that experimental unit to the treatment (most commonly by averaging the individual responses). Increasing the number of subsamples per experimental unit is often cheaper than including more experimental units, and is desirable if this increases the power of the experiment. On the other hand, collecting many subsamples may not increase the power enough to be worth the effort because power is typically increased more by increasing the number of experimental units. To balance these considerations it is helpful to calculate the power for a range of numbers of experimental units and subsamples. This pamphlet will graphically demonstrate the above statements for the Completely Randomized Design and will discuss some of the considerations in using these graphs to choose suitable sample sizes. The next pamphlet (BI #50) will extend the results to the Randomized Block Design, while the following pamphlet (BI #51) will briefly describe how to create these graphs in general and provide an example SAS program that produces the plots shown in this pamphlet.

For discussion purposes, let us assume a study with four treatments (factor T with t = 4 levels) that will each be randomly assigned to p plots (factor P with p levels). Each plot is the experimental unit and will be subsampled e times to obtain an estimate of the plot response (subsamples will be factor E and if there were 10 subsamples per plot then e = 10). The ANOVA table for this one-way completely randomized design with subsamples[1] is:

Source of Variation		Degrees of freedom	Expected Mean Squares	Mean Square	Error
Treatment	T	t-1	$\sigma_e^2 + e\sigma_p^2 + pe\phi_T$	MST	MSP
Plots	P(T)	t(p-1)	$\sigma_e^2 + e\sigma_p^2$	MSP	MSE
Subsamples	E(PT)	pt(e-1)	σ_e^2	MSE	-

[1] This is a nested ANOVA and has the same components of variation as a multi-staged sampling plan where samples are taken from large populations so the finite population correction factor (fpc) is not needed. See Snedecor and Cochran, Sections 13.3 and 21.10, Cochran, Chap. 10 and Wetherill, Chap. 14 for more discussion.

In this case, the estimation of power requires information on two components of variation, namely, σ_e^2 , the variation between subsamples within each plot, and σ_p^2 , the variation between plots or experimental units. As usual, the null hypothesis of the ANOVA is that the treatment means μ_i all have the same value as the grand mean, μ. The third term, ϕ_T [which is $\sum(\mu_i-\mu)^2/(t-1)$], is a measure of the treatment effect and is zero if the null hypothesis is correct. To determine sample sizes or power we must look at:

1. The alternate hypothesis, H_A. For discussion purposes, we shall assume that for the alternate hypothesis the four treatments have mean values of μ_i = 10, 15, 20, 25 . This H_A can be characterized by the Sums of Squares of the Means: SSM = $\sum (\mu_i - \mu)^2$ = 125. The hypothesis sums of squares (SSH) = sample size * SSM (for balanced ANOVA's only)[2] ;
2. Our choices for: α = prob(rejecting H_o if it is true): the usual value is α = 0.05; and
   $1 - \beta$ = prob(rejecting H_o when H_A is true);
3. Number of subsamples, e and their component of variation, σ_e^2 ;
4. Number of plots, p and their component of variation, σ_p^2 .

Obtaining estimates for σ_e^2 and σ_p^2 can be the most difficult part of a sample size/power analysis. If a similar study has already been conducted, possibly as a pre-trial, or a post-hoc analysis is desired, then the variance components can be estimated from the obtained data[3]. From the above ANOVA table, we note that the expected MSP = $\sigma_e^2 + e\sigma_p^2$ and expected MSE = σ_e^2 . By rearranging, we can estimate these components by:

$$\hat{\sigma}_e^2 = \text{MSE}, \text{ and } \hat{\sigma}_p^2 = \frac{\text{MSP} - \text{MSE}}{e}.$$ [4]

To examine the influence of the sample numbers e and p and their components of variation, σ_e^2 and σ_p^2, we will graph the power for a range of values for e and p and allow σ_e^2 to have the values 100 and 1000 and σ_p^2 to have the values of 100 and 500. These graphs are shown on the next two pages and have been produced using α = 0.05 and the alternate hypothesis noted above.

---

[2] Note that the definition of sample size is clear for designs without subsamples since, in that case, it is simply the number of experimental units for each treatment. In this situation I will use the term to refer to the number of numbers used to calculate a treatment mean. For the balanced design considered here that will be number of experimental units times the number of subsamples, i.e. p * e.

[3] Also see BI # 25 for some suggestions.

[4] It is possible to get negative estimates with this equation. This is non-sensical and it may be appropriate to set negative values to zero. There are other methods of estimating variance components so as to avoid negative estimates but these are refinements we can ignore since only rough estimates are needed for power calculations.

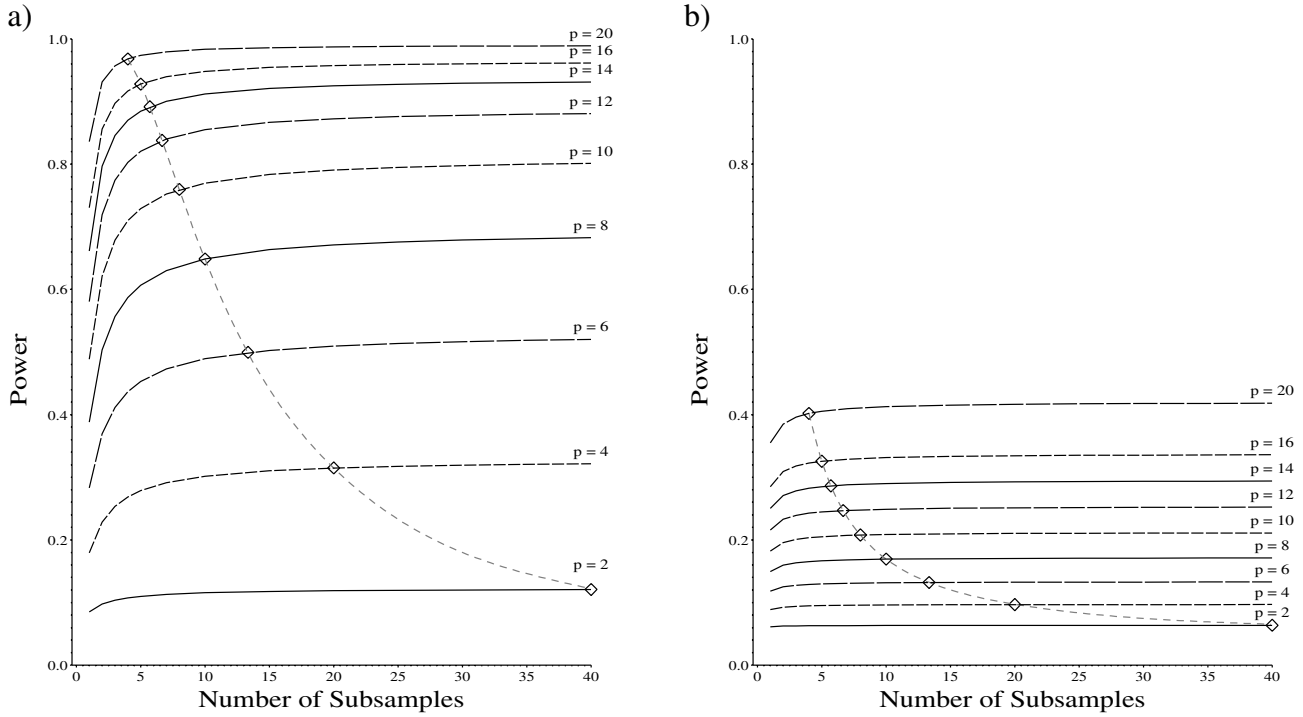

Figure 1. Power curves for $\alpha = 0.05$, SSM = 125, $\sigma_e^2 = 100$ and a) $\sigma_p^2 = 100$ and b) $\sigma_p^2 = 500$. The diamonds show where the total number of subsamples is 320.

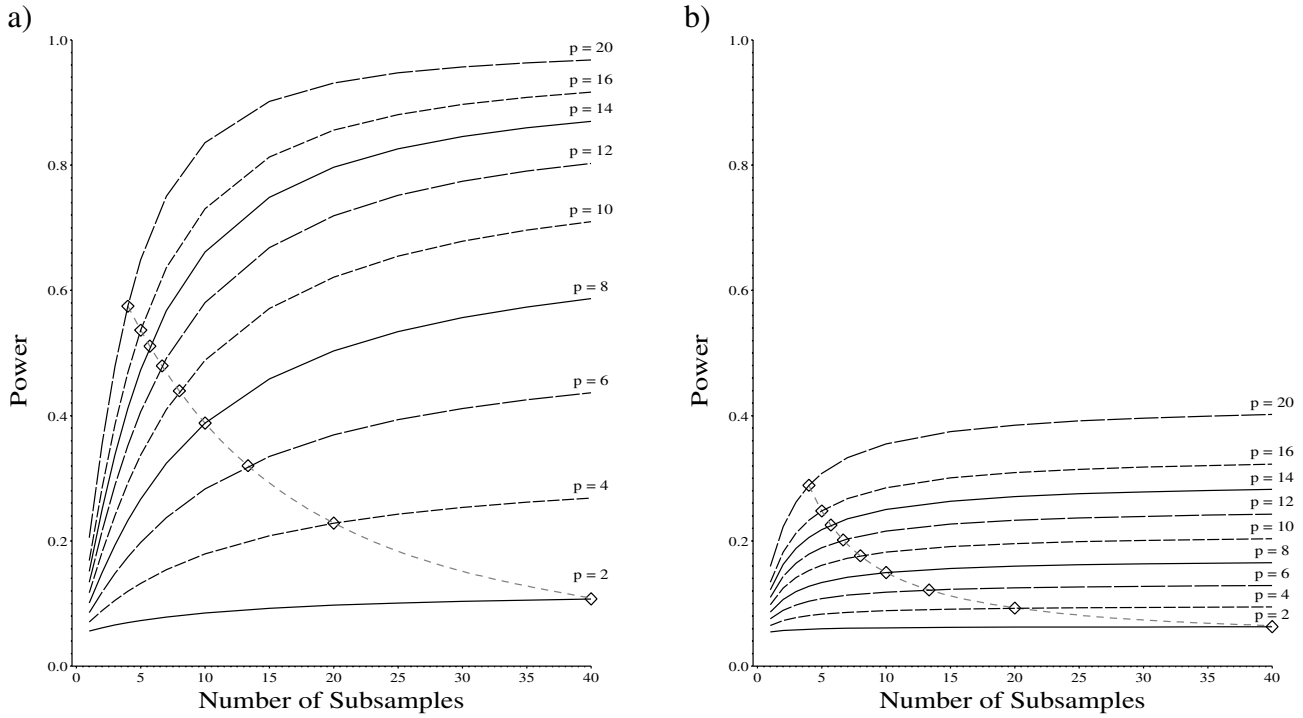

Figure 2. Power curves for $\alpha = 0.05$, SSM = 125, $\sigma_e^2 = 1000$ and a) $\sigma_p^2 = 100$ and b) $\sigma_p^2 = 500$. The diamonds show where the total number of subsamples is 320.

The diamonds connected by a dashed line represent the power for studies with a total of t*p*e = 320 subsamples. If the total cost of the experiment was entirely due to the number of subsamples so that plots did not cost more to establish, then the curve described by these diamonds could be used to help decide the most effective way to arrange the experiment so as to get the most power for a constant cost (total number of subsamples).

Cox's Rule of Thumb: When determining subsample numbers it is useful to note that for the completely randomized design ***there is not much increase in power when the number of subsamples, e, is greater than 4(σ_e^2/σ_p^2)*** (from Cox, 1958, page 181). The values of σ_e^2 and σ_p^2 used in the graphs and the corresponding value of the ratio are:

σ_e^2	σ_p^2	Cox's Ratio: $4(\sigma_e^2/\sigma_p^2)$
100	100	4
100	500	4/5 ≈ 1
1000	100	40
1000	500	8

Note that the power curves have flattened out by the time the subsample number, e, reaches Cox's Ratio and that the most efficient number of subsamples per plot is less than or equal to this Ratio.

These graphs show that increasing the number of subsamples, e, per plot has a limited ability to increase the power and that the size of this ability depends upon how large σ_e^2 is relative to σ_p^2. The study designs in Figure 1a have a Cox's ratio of 4 and each of the curves is largely flat for numbers of subsamples greater than 4 so that increasing the number of subsamples beyond four increases the power by very little. This is even more true for the study designs represented by Figure 1b where Cox's ratio is less than one. One the other hand, for the study designs represented by Figure 2a, Cox's ratio is 40 and increasing the number of subsamples from 5 to 20 for designs with six or more plots substantially increases the power. Notice that Cox's ratio provides an upper limit for a useful number of subsamples. For example, increasing the number of subsamples from 20 to 40 for the design with 20 plots in Figure 2a will improve the power very little, while this might be worthwhile for a study with 10 plots.

In general, increasing the number of plots increases the power more than increasing the number of subsamples per plot. This can be understood by noticing that increasing the number of plots increases the degrees of freedom for the error term of the F-test (thus decreasing the critical F-value[5]), while increasing the number of subsamples does not. Further, increasing the number of experimental units directly reduces the standard error of the treatment means (since the standard error is a function of 1/p), thus reducing the size of treatment differences that can be detected.

[5] There are limits on this too. The critical F-value and corresponding power changes little for error degrees of freedom greater than 30.

Increasing the number of subsamples improves the precision of the estimate of the experimental unit mean and, while this may help reduce the variability between experimental unit means, it does so indirectly. As we have seen, increasing numbers of subsamples is more effective when the subsample variability (σ_e^2) is large relative to the between experimental unit variability (σ_p^2). In general, increasing the number of plots is preferable to increasing the number of subsamples per plot and this has the further advantage of testing the treatments over a wider variety of conditions.

A large number of plots would be needed to get reasonable power for the study designs represented in Figures 1b and 2b. This is because the Sums of Squares of the alternate hypothesis (SSH) is small relative to a σ_p^2 of 500. A larger SSM (implying larger differences between the means) might need to be considered as attainable. How could we calculate this value? As an example, suppose that our best estimates for σ_e^2 and σ_p^2 are 1000 and 500. It is clear from Figure 2b that even with twenty plots per treatment level, the power of our study for an SSM of 125 is no greater than 0.40. Further, suppose that ten plots per treatment (p = 10) with twenty subsamples each (e = 20) is the most that can be managed. To determine the SSM value required to obtain the traditional minimum power of 0.80, we could 1) create a graph with SSM on the x-axis instead of plot or subsample numbers, or 2) do some quick calculations using the graphs we already have.

The quick calculations require that we first determine what observed F-value would have about 80% power. Since F-tests with the same degrees of freedom and observed F-value will have the same power, we can look at the curves in the figures where p = 10 and e = 20 to see if any have sufficient power. In Figure 1a, with $\sigma_e^2 = 100$ and $\sigma_p^2 = 100$, the power is just below 0.80. The expected F-value for that point is calculated by:

$$F = \frac{e * p * SSM/(t-1)}{\sigma_e^2 + e * \sigma_p^2} = \frac{20 * 10 * SSM/3}{100 + 20 * 100}.$$

With SSM = 125 the F-value is 3.97 or about 4.0. Now we can calculate the value of SSM that will give us an F-value of 4.0 (for a power of about 80 %) when $\sigma_e^2 = 1000$ and $\sigma_p^2 = 500$. This is done by rearranging the above equation to get:

$$SSM = \frac{(t-1) * F * (\sigma_e^2 + e * \sigma_p^2)}{e * p} = \frac{3 * 4 * (1000 + 20 * 500)}{20 * 10} = 660.$$

This is quite a large increase in SSM and corresponds roughly to means of 10, 21.5, 33, and 44.5 (SSM = 661.25), instead of 10, 15, 20, and 25.[6]

[6] Help in interpreting this SSM can be found in Handbook #2, the Power Analysis Workshop Notes and the appendix of BI #52.

While the calculations and graphs in this pamphlet were produced using SAS and SAS/Graph, any software that can calculate 1) the critical F-value given a probability, degrees of freedom, and non-centrality parameter (`finv` in SAS) and 2) the cumulative probability function for the F-distribution given an F-value, degrees of freedom, and non-centrality parameter (`probf` in SAS) can generate the data needed for the graphs. See pamphlet #51 for more information on how to create graphs for specific situations.

Contact: Wendy Bergerud
387-5676

References:

Bergerud, W. A., and V. Sit, 1992, Power Analysis Workshop Notes. Minstry of Forests, Research Branch.

Cochran, W.G., 1977, *Sampling Techniques*, 3rd ed., John Wiley.

Cox, D. R., 1958, *Planning of Experiments*, John Wiley.

Nemec, A. F. L., 1991. Biometrics Information Handbook #2: Power analysis handbook for the design and analysis of forestry trials.

Snedecor, G.W., and W. G. Cochran, 1980. *Statistical Methods*, 7th ed., The Iowa State University Press.

Wetherill, G.B., 1981, *Intermediate Statistical Methods*, Chapman and Hall.