BIOMETRICS INFORMATION

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

PAMPHLET NO. # 37 DATE: March 14, 1992

SUBJECT: A general description of hypothesis testing and power analysis

This pamphlet will describe the concepts of hypothesis testing and power analysis. When conducting a statistical test there are two hypotheses under consideration. The first is the null hypothesis, H_o, which is the simple hypothesis of no difference between treatments or of no response to the variable(s) of interest. The second is the alternate hypothesis, H_a (or H_1), which describes a difference or effect of the treatment or variable(s) of interest. This second hypothesis is usually vaguely worded for example, that the means are different or the slope of a line is not zero. These alternate hypotheses need to be more precisely worded before a power analysis can be conducted.

If H_o is rejected by a statistical test there are two possibilities:

1) **the decision is incorrect**. The frequency of making this mistake is known as the Type I error rate and is usually denoted by α. It is the probability of rejecting H_o when it is, in fact, true.

2) **the decision is correct**. The frequency of making this correct decision is known as the power of the test and is denoted by $1 - \beta$. The probability of rejecting H_o when it should not be rejected is known as the Type II error rate and is denoted by β.

The general conceptual procedure required for hypothesis testing is:

H1) **Describe the null hypothesis, H_o, and relevant statistical properties of the data.** For example, in a simple completely randomized design with four treatments, we might hypothesize that the four treatment means are the same, namely, H_o: $\mu_1=\mu_2=\mu_3=\mu_4$. The relevant statistical properties of the data might be that the the responses measured for each experimental unit (e.u.) were independent of each other (by virtue of the random assignment of treatments to e.u.'s), and follow a normal distribution with the same variance, σ^2.

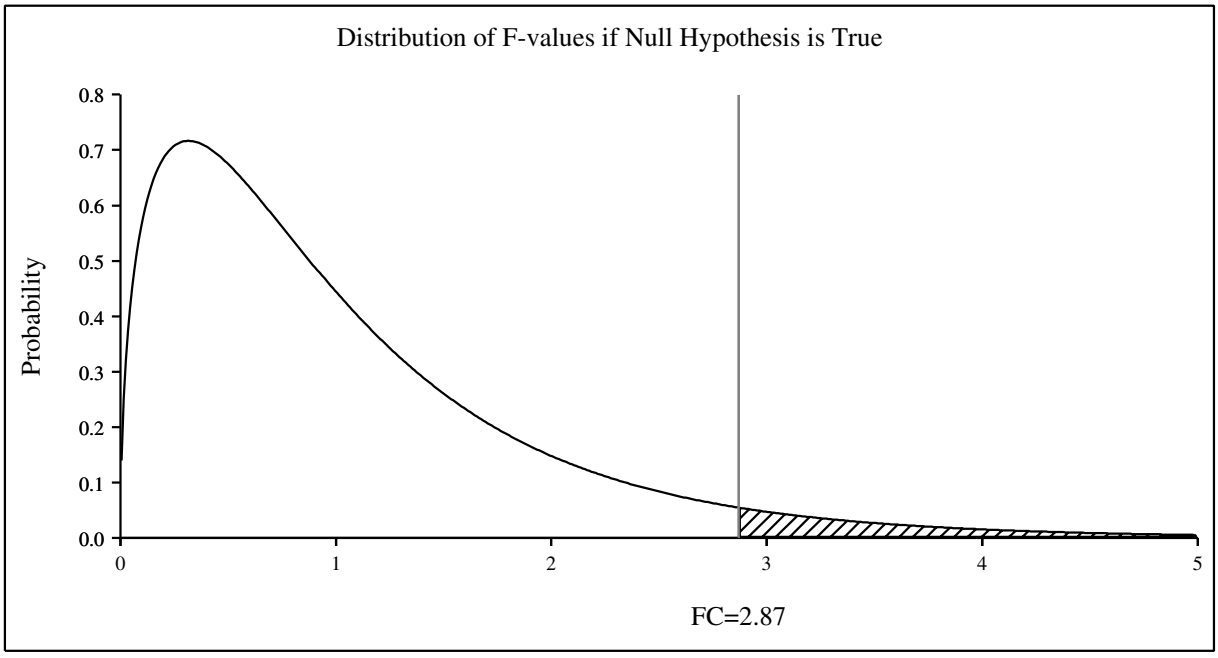
H2) **Assuming that the null hypothesis is true, determine a useful test statistic and its distribution.** While this may be quite difficult mathematically, a thought experiment can help us understand the basic process. Suppose that H_o is true and we ran our experiment 100,000 times. For each experiment the test statistic is calculated and the frequency of values obtained is plotted in a frequency curve. This frequency curve would be the distribution of the test statistic, and could be used to determine a range of unlikely values for the test statistic if H_o were, in fact, true. For the example ANOVA, the ratio of the Mean Square

Between to the Mean Square Within has an F-distribution which can be described mathematically and is tabled in many texts. This ratio is:

$$F = \frac{\text{sample size * variance of the means}}{\text{average variance of data for each mean}} = \frac{\text{Mean Square Between}}{\text{Mean Square Within}} \quad \dagger$$

A picture of its distribution is shown below after step H3. If the distribution of the test statistic cannot be described mathematically, it may be possible to determine it empirically by simulating it with the help of computers.

H3) **Determine a decision rule based on the test statistic's distribution.** A typical decision rule is to reject H_o if the observed test statistic falls within some specific range of values (called the critical range). This range is usually determined such that observed values as great or greater than a critical value would be expected to occur less than $100*\alpha\%$ of the time given a true H_o. Typically, α is set at 0.05 or 5%. For the ANOVA example, if the sample size is ten for each mean, then the df for the F-distribution are 3, 36 and its distribution is shown in the graph below. The critical range is described by the F-value, FC, which divides the area under the curve into 5% on the right and 95% on the left (for $\alpha = 0.05$). For the example, this occurs at FC = 2.87 and is indicated by the vertical line on the graph below. We might state the decision rule as "H_o will be rejected when the observed F-value is in the critical range of $F_{obs} > FC = 2.87$".



H4) **Calculate the test statistic for your data and make a decision based on the decision rule determined in the previous step.** Another approach is to calculate the p-value associated with

---

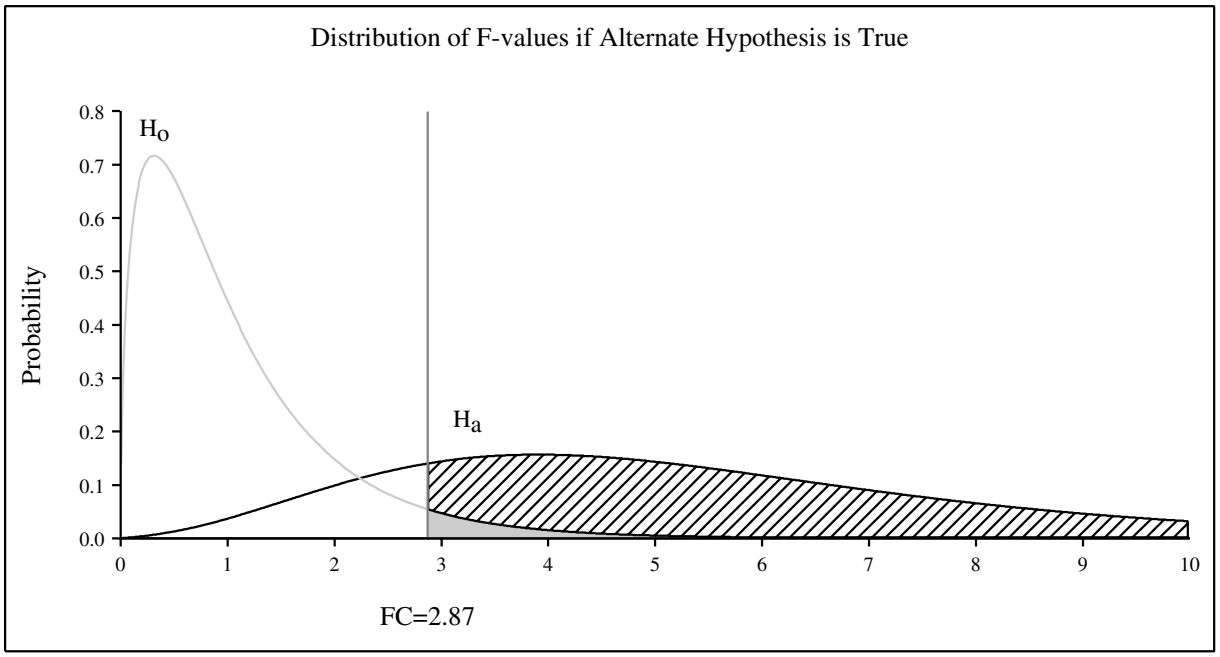
† See Biometrics Information pamphlets #22 and #25 for more discussion of this ratio.

the observed test statistic. For the F-distribution this means calculating the area to the right of the observed F-value. For the ANOVA example, if the observed F-value is 3.81 then 1.8% of the area is to the right of this value and the observed probability is 0.018. According to the decision rule, H_o is rejected.

The above procedure is used to control the Type I Error rate (α). The Type II Error rate (β) is also important and should be controlled or assessed, especially if H_o is not rejected (see Peterman). The following conceptual steps are required for a power calculation. The first two steps repeat those above except that H_a is now assumed.

P1) **Describe the alternate hypothesis, H_a, and relevant statistical properties of the data.** For the example of a simple completely randomized design with four treatments, we might hypothesize that the four treatment means had the following values, namely, H_a: μ_1=10, μ_2=15, μ_3=20, μ_4=25. The statistical properties of the data remain the same as in step H1.

P2) **Assuming that the alternate hypothesis is true, determine the new distribution of the test statistic.** The thought experiment can again help us understand the basic process. Suppose that H_a is true and we reran our experiment 100,000 times. For each experiment the test statistic is calculated and the frequency of values obtained is plotted in a frequency curve. The same test statistic would be used but its distribution would now be different. For the example ANOVA, this new (non-central) F-distribution would be (note the change in scale):



P3) **Using the decision rule from step H3, determine how often the test statistic would occur in the critical range. This will be the power 1-β, and is the probability of rejecting H_o when H_a is, in fact, true.** For the example ANOVA, this value is 1-β = 0.81 and is shown by the shaded area in the graph above (this example is taken from the Power Analysis Workbook and the technical details involved in calculating this power can be found there).

P4) **Make a decision about the experimental results based on both α and β.**

Power analysis requires the following four basic ingredients:

1) the Type I Error rate, α, or the observed p-value;
2) the Type II Error rate, β, which may be preset as is α or which may be observed as is the p-value above;
3) an estimate of the variability of the data;

and 4) a specific alternate hypothesis or a range of means if maximum and minimum means can be described.

Once these basic ingredients are understood, power analysis can be looked at in several different ways. For instance, it may be possible to generate a plot of power versus a range of alternate hypotheses after an experiment has been conducted and H_o has not been rejected. This plot (or the corresponding table of values) can then be used to determine what sort of differences the experiment **could** have detected. If these differences are larger than the practical differences of concern then it would be unwise to accept H_o. On the other hand, if they are quite a bit smaller, then it might be quite reasonable to act on the assumption that H_o is true. **In general, it is not justifiable to accept H_o if a power calculation to determine β or 1-β has not been done.** See Peterman's paper for an excellent discussion of this point.

References:

Bergerud, W.A. and V. Sit. 1992. Power analysis workshop notes.

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

Peterman, R.M., 1990. Statistical power analysis can improve fisheries research and management. Can. J. Fisheries and Aquatic Sciences, 47: 2-15.

Contact: Wendy Bergerud
387-5676