BIOMETRICS INFORMATION

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

PAMPHLET NO. # 19 DATE: June 19, 1989

SUBJECT: ANOVA: Approximate or Pseudo F-tests

Analysis of variance designs will occasionally have two or more random factors that are crossed with a treatment. This complicates the determination of the proper error term for the F-tests since no single source in the design is appropriate. Thus a pseudo† F-test, using a combination of sources in the ANOVA, must be used to test for treatment effects.

I will use the following example as an illustration. Suppose a researcher wants to test the storability of pollen at three different moisture contents. Several pollen lots are selected to represent all possible pollen lots. The pollen is dried to moisture contents of either 4%, 8%, and 12% and then stored for some time. In a following Spring, the viability of the pollen is tested by applying the pollen to receptive flowers on ten trees in a nearby seed orchard. The trees are blocks in a randomized block design since they receive all the treatments. Two branches per tree are assigned one pollen lot which had been stored at one of the moisture levels. Thus block, pollen lot, and moisture content are completely **crossed**. Since both pollen lot and trees may respond quite differently to each other and to the treatments, it is important to include several examples of each in the experiment. This is why both pollen lot and trees are considered **random** factors. The **Expected Mean Squares** (EMS) for the design are as follows:

Source of Variation	Degrees of Freedom	Expected Mean Squares
Block	9	$\sigma^2_e + 6\sigma^2_{BL} + 24\sigma^2_B$
Pollen L	3	$\sigma^2_e + 6\sigma^2_{BL} + 60\sigma^2_L$
B x L	27	$\sigma^2_e + 6\sigma^2_{BL}$
Moisture M	2	$\sigma^2_e + 2\sigma^2_{BLM} + 8\sigma^2_{BM} + 20\sigma^2_{LM} + 80\sigma^2_M$
L x M	6	$\sigma^2_e + 2\sigma^2_{BLM} + \quad + 20\sigma^2_{LM}$
B x M	18	$\sigma^2_e + 2\sigma^2_{BLM} + 8\sigma^2_{BM}$
B x L x M	54	$\sigma^2_e + 2\sigma^2_{BLM}$
Error	120	σ^2_e

† Pseudo F-tests are also called approximate F-tests.

BRITISH COLUMBIA Ministry of Forests Research Program

As you can see from this table the error terms for block and pollen is their interaction, B x L, and the error terms for their interactions with the moisture treatment is B x L x M. This is determined by noting, for example, that if σ^2_B is zero then the EMS for block becomes $\sigma^2_e + 6\sigma^2_{BL}$, which is the EMS for B x L. Thus B x L is the error term, because if σ^2_B is zero then the F-statistic will be the ratio of two identical expected mean squares.

This simple situation does not occur for testing the moisture treatment. If σ^2_M is zero there is no other source in the table that has an EMS the same as $\sigma^2_e + 2\sigma^2_{BLM} + 8\sigma^2_{BM} + 20\sigma^2_{LM}$. Nevertheless, if we add some EMS's together we can obtain an approximate F-ratio of two identical terms. Specifically:

$$F = \frac{EMS(BLM) + EMS(M)}{EMS(BM) + EMS(LM)}$$

$$= \frac{\sigma^2_e + 2\sigma^2_{BLM} + \sigma^2_e + 2\sigma^2_{BLM} + 8\sigma^2_{BM} + 20\sigma^2_{LM} + 80\sigma^2_M}{\sigma^2_e + 2\sigma^2_{BLM} + 8\sigma^2_{BM} + \sigma^2_e + 2\sigma^2_{BLM} + 20\sigma^2_{LM}} .$$

You will note that except for σ^2_M, every term in the above ratio occurs an equal number of times in both the numerator and the denominator. The observed F-value is calculated by:

$$F = \frac{MS(BLM) + MS(M)}{MS(BM) + MS(LM)}$$

and with degrees of freedom calculated by:

$$df_{num} = \frac{\left[MS(BLM) + MS(M)\right]^2}{\left[\frac{MS(BLM)^2}{df_{BLM}} + \frac{MS(M)^2}{df_M}\right]} \quad \text{and} \quad df_{den} = \frac{\left[MS(BM) + MS(LM)\right]^2}{\left[\frac{MS(BM)^2}{df_{BM}} + \frac{MS(LM)^2}{df_{LM}}\right]}$$

NOTE: Another ratio could have been obtained by **subtracting** MS(BLM) from the MS(BM) + MS(LM) instead of **adding** it to MS(M). Zar (1974) uses subtraction in his ANOVA tables in Appendix C, but this is not recommended (Dr. Keith Hastings, pers. comm.) since the subtraction can lead to strange results.

As an example, suppose the following mean squares had been obtained for an experiment:

Source of Variation	Degrees of Freedom	Mean Squares	F-values
Block B	9	192	4.0
Pollen L	3	144	3.0
B x L	27	48	4.0
Moisture M	2	756	see below
L x M	6	162	3.0
B x M	18	108	2.0
B x L x M	54	54	4.5
Error	120	12	-

The pseudo F-value for moisture is calculated by:

$$F = \frac{54 + 756}{108 + 162} = 3.0 \text{ with degrees of freedom calculated by:}$$

$$df_{num} = \frac{\left[\ 54 \ + 756\ \right]^2}{\left[\frac{(54)^2}{54} + \frac{(756)^2}{2}\right]} = 2.295$$

and $$df_{den} = \frac{\left[\ 108 \ + \ 162\ \right]^2}{\left[\frac{(108)^2}{18} + \frac{(162)^2}{6}\right]} = 14.516.$$

Using the SAS program described in Biometrics Information pamphlet # 15, the probability for this F-value is 0.0757. Thus, there is weak evidence of a moisture effect.

References:

Milliken, G.A., and Johnson, D.E., 1984. *Analysis of Messy Data, Volume 1: Designed Experiments*, Lifetime Learning Publications, Belmont, CA. (see page 250, also look in the index under Satterthwaite).

Satterthwaite, F.E., 1946. *An approximate distribution of estimates of variance components.* Biometrics Bulletin, 2: 110-114.

Zar, J.H., 1974. *Biostatistical Analysis*. Prentice-Hall Inc., Englewood Cliffs, N.J.

CONTACT: Wendy Bergerud
387-5676

———————————-PROBLEM FROM BI #18———————————

To test the 3-variable model calculate:

$$F = \frac{0.3965/3}{(1-0.3965)/(192-4)} = 41.169\ , \quad df = 3,\ 188$$

The contribution of X1 and X3 is tested by:

$$F = \frac{(0.39648035-0.3952283)/(3-1)}{(1-0.39648035)/(192-4)} = 0.195\ , \quad df = 2,\ 188$$

The probabilities for these F-values are 0.823 and <0.0001 respectively. Thus it is quite reasonable that X2 does a fine job on its own and variables X1 and X3 are unnecessary in the final model.

———————————NEW PROBLEM———————————

Calculate the F-test for moisture if it's mean square had been 1836 instead of 756.