BIOMETRICS INFORMATION

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

PAMPHLET NO. # 25 DATE: March 29, 1990

SUBJECT: ANOVA: The Mean Within Sums of Squares as an Average Variance

The Mean Within Sums of Squares (MSW) is used to test for treatment differences in completely randomized analysis of variance designs. The Mean Within Sums of Squares is the Sums of Squares Within (SSW) divided by its degrees of freedom. The MSW is also known as the Mean Error Sums of Squares (MSE) or the Mean Residual Sums of Squares (MSR). This Mean Sums of Squares is the weighted mean of the variances for each treatment in the ANOVA, and it's use as the error term depends on the assumption of homogeneity of variance.

The assumption of homogeneity of variance for completely randomized ANOVA designs[1] means that the data for each treatment[2] has a constant variance or standard error. The specific value of this constant variance is usually unknown and must be estimated from the data. The MSW provides the "best" estimate for this variance since it uses all the data available in the ANOVA.

Calculation of the MSW should follow these steps:

1. First check the homogeneity of variance assumption. There are several tests available (Milliken and Johnson, pgs 18-24), but most are complicated to calculate. The simplest test available is Hartley's F_{max}-test, and it is useful for a quick check. The observed F_{max}-value is determined by:

   i) Calculating the variance for each treatment cell (recall that the variance is the square of the standard deviation for the data in each cell).

   ii) Choosing the largest and the smallest variances and calculating the following ratio:

$$F_{max} = \frac{\text{largest variance}}{\text{smallest variance}}$$

   This value is then compared with the entry in the F_{max} Tables. To use the tables you must also know the number of treatments (k) in the ANOVA and the degrees of freedom for an individual variance[3]. Tables are available in Winer, and Milliken and Johnson. If the observed F_{max} value is small compared to the critical values in the table, then the null hypothesis is not rejected, suggesting that the variances are reasonably constant.

---

[1]These include the one-way ANOVA and factorial designs such as two-way and three-way ANOVA's.

[2]A treatment is defined here as either one level of a treatment for a one-way ANOVA or the combination of treatment levels which define a unique cell in a factorial ANOVA.

[3]This assumes an equal number of observations (n) for each treatment. If this is not the case, and the sample sizes are not too different, then one approach is to use the largest sample size (Winer, pg 207).

BRITISH COLUMBIA

Ministry of Forests
Research Program

2. If the variances can be considered constant[4], then the MSW is calculated as the weighted sum of the cell variances (Var) using the degrees of freedom (n_i-1) for each treatment variance as the weight.

$$MSW = \frac{SSW}{df_W} = \frac{(n_1-1)Var_1 + (n_2-1)Var_2 + \cdots + (n_k-1)Var_k}{(n_1-1) + (n_2-1) + \cdots + (n_k-1)}$$

with degrees of freedom $df_W = (n_1\text{-}1) + (n_2\text{-}1) + \cdots + (n_k\text{-}1)$

$= k(n\text{-}1)$, if all samples sizes are equal.

If the samples sizes are equal then MSW is simply the average of the cell variances (see Biometrics Information pamphlet # 22).

This method of pooling variances is not restricted to ANOVA situations. It can also be used in sample size calculations. For instance, if a nursery has data from past batches of PSB313 Douglas fir seedlings, then the previously observed variances can be pooled to get a "best" estimate of expected variability. This estimate can then be used to determine sample sizes for the next batch or for a forthcoming experimental trial. Note that this pooling of variances is only appropriate when there is good reason to believe that all the variances are estimating the same "true" but unknown variance.

References:

Winer, B. J., 1971, *Statistical principles in experimental design*, 2nd edition, McGraw-Hill Book Company, New York, New York

Milliken, George A., and Johnson, Dallas E., 1984, *Analysis of Messy Data, Volume I: Designed Experiments*, Lifetime Learning Publication, Belmont, CA.

CONTACT: Wendy Bergerud
387-5676

NEW PROBLEM

Calculate the F_{max}-test and the pooled variance for the results of a sample trial:

Treatment	1:	Mean = 15.2	SD = 20	n = 10
	2:	Mean = 16.5	SD = 25	n = 8
	3:	Mean = 18.7	SD = 23	n = 13

Note: SD stands for Standard Deviation.

[4]If the variances are not similar in value, then consider if the variance differences could be a treatment effect.