

## Supplemental Material for Chapter 7

### S7-1. Fixed Versus Random Factors in the Analysis of Variance

In chapter 3, we present the standard analysis of variance (ANOVA) for a single-factor experiment, assuming that the factor is a **fixed** factor. By a fixed factor, we mean that all levels of the factor of interest were studied in the experiment. Sometimes the levels of a factor are selected at random from a large (theoretically infinite) population of factor levels. This leads to a **random effects** ANOVA model.

In the single factor case, there are only modest differences between the fixed and random models. The model for a random effects experiment is still written as

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

but now the treatment effects  $\tau_i$  are random variables, because the treatment levels actually used in the experiment have been chosen at random. The population of treatments is assumed to be normally and independently distributed with mean zero and variance  $\sigma_\tau^2$ . Note that the variance of an observation is

$$\begin{aligned} V(y_{ij}) &= V(\mu + \tau_i + \varepsilon_{ij}) \\ &= \sigma_\tau^2 + \sigma^2 \end{aligned}$$

We often call  $\sigma_\tau^2$  and  $\sigma^2$  **variance components**, and the random model is sometimes called the **components of variance** model. All of the computations in the random model are the same as in the fixed effects model, but since we are studying an entire population of treatments, it doesn't make much sense to formulate hypotheses about the individual factor levels selected in the experiment. Instead, we test the following hypotheses about the variance of the treatment effects:

$$\begin{aligned} H_0 &: \sigma_\tau^2 = 0 \\ H_1 &: \sigma_\tau^2 > 0 \end{aligned}$$

The test statistic for these hypotheses is the usual  $F$ -ratio,  $F = MS_{Treatments}/MS_E$ . If the null hypothesis is not rejected, there is no variability in the population of treatments, while if the null hypothesis is rejected, there is significant variability among the treatments in the entire population that was sampled. Notice that the conclusions of the ANOVA extend to the entire population of treatments.

The expected mean squares in the random model are different from their fixed effects model counterparts. It can be shown that

$$\begin{aligned} E(MS_{Treatments}) &= \sigma^2 + n\sigma_\tau^2 \\ E(MS_E) &= \sigma^2 \end{aligned}$$

Frequently, the objective of an experiment involving random factors is to estimate the variance components. A logical way to do this is to equate the expected values of the mean squares to their observed values and solve the resulting equations. This leads to

$$\hat{\sigma}_{\tau}^2 = \frac{MS_{Treatments} - MS_E}{n}$$

$$\hat{\sigma}^2 = MS_E$$

A typical application of experiments where some of the factors are random is in a measurement systems capability study, as discussed in Chapter 7. The model used there is a factorial model, so the analysis and the expected mean squares are somewhat more complicated than in the single factor model considered here.

### S7-2. Analysis of Variance Methods for Measurement Systems Capability Studies

In Chapter 7 an analysis of variance model approach to measurement systems studies is presented. This method replaces the tabular approach that was presented along with the ANOVA method in earlier editions of the book. The tabular approach is a relatively simple method, but it is not the most general or efficient approach to conducting gage studies. Gauge and measurement systems studies are **designed experiments**, and often we find that the gauge study must be conducted using an experimental design that does not nicely fit into the tabular analysis scheme. For example, suppose that the operators used with each instrument (or gauge) are different because the instruments are in different physical locations. Then operators are nested within instruments, and the experiment has been conducted as a nested design.

As another example, suppose that the operators are not selected at random, because the specific operators used in the study are the only ones that actually perform the measurements. This is a mixed model experiment, and the random effects approach that the tabular method is based on is inappropriate. The random effects model analysis of variance approach in the text is also inappropriate for this situation. Dolezal, Burdick, and Birch (1998), Montgomery (2001), and Burdick, Borror, and Montgomery (2003) discuss the mixed model analysis of variance for gauge R & R studies.

The tabular approach does not lend itself to constructing confidence intervals on the variance components or functions of the variance components of interest. For that reason we do not recommend the tabular approach for general use. There are three general approaches to constructing these confidence intervals: (1) the Satterthwaite method, (2) the maximum likelihood large-sample method, and (3) the modified large sample method. Montgomery (2001) gives an overview of these different methods. Of the three approaches, there is good evidence that the modified large sample approach is the best in the sense that it produces confidence intervals that are closest to the stated level of confidence.

Hamada and Weerahandi (2000) show how generalized inference can be applied to the problem of determining confidence intervals in measurement systems capability studies. The technique is somewhat more involved than the three methods referenced above. Either numerical integration or simulation must be used to find the desired confidence intervals. Burdick, Borror, and Montgomery (2003) discuss this technique.

While the tabular method should be abandoned, the control charting aspect of measurement systems capability studies should be used more consistently. All too often

a measurement study is conducted and analyzed via some computer program without adequate graphical analysis of the data. Furthermore, some of the advice in various quality standards and reference sources regarding these studies is just not very good and can produce results of questionable validity. The most reliable measure of gauge capability is the probability that parts are misclassified.