though it is obtained through two different estimates of the common population *variance.* Adding 10 to each value of the first sample causes the three sample means to grow farther apart, with the result that the F test statistic increases and the P-value decreases.

Calculations with Unequal Sample Sizes

While the calculations for cases with equal sample sizes are somewhat reasonable, they become much more complicated when the sample sizes are not all the same, but the same basic reasoning applies. We calculate an F test statistic that is the ratio of two different estimates of the common population variance σ^2. With unequal sample sizes, we must base the calculations on *weighted* measures that take the sample sizes into account. The test statistic is essentially the same as the one given earlier, and its interpretation is also the same as described earlier. The denominator depends only on the sample variances that measure variation within the treatments and is not affected by the differences among the sample means. In contrast, the numerator is affected by differences among the sample means. If the differences among the sample means are very large, those large differences will cause the numerator to be very large, so F will also be very large. Consequently, very large values of F suggest unequal means, and the ANOVA test is therefore right-tailed. Instead of providing the relevant messy formulas required for cases with unequal sample sizes, we wisely and conveniently assume that technology should be used to obtain the P-value for the analysis of variance. We become unencumbered by complex computations and we can focus on checking requirements and interpreting results.

Designing the Experiment With one-way (or single-factor) analysis of variance, we use one factor as the basis for partitioning the data into different categories. If we conclude that the differences among the means are significant, we can't be absolutely sure that the differences can be explained by the factor being used. It is possible that the variation of some other unknown factor is responsible. One way to reduce the effect of the extraneous factors is to design the experiment so that it has a **completely randomized design,** in which each sample value is given the same chance of belonging to the different factor groups. For example, you might assign subjects to two different treatment groups and a third placebo group through a process of random selection equivalent to picking slips of paper from a bowl. Another way to reduce the effect of extraneous factors is to use a **rigorously controlled design,** in which sample values are carefully chosen so that all other factors have no variability. In general, good results require that the experiment be carefully designed and executed.

Identifying Which Means Are Different

After conducting an analysis of variance test, we might conclude that there is sufficient evidence to reject a claim of equal population means, but we cannot conclude from ANOVA that any *particular* means are different from the others. There are several formal and informal procedures that can be used to identify the specific means that are different. Here are two *informal* methods for comparing means:

- Construct boxplots of the different samples to see if one or more of them is very different from the others.
- Construct confidence interval estimates of the means for each of the different samples, then compare those confidence intervals to see if one or more of them does not overlap with the others.