Chapter 2: Simple Comparative Experiments

"Many of the most useful designs are extremely simple."

Sir Ronald Fisher

We now look at a method for making simple comparisons of two or more "treatments." It's called the "F-test" after Sir Ronald Fisher, a geneticist who developed the technique for application to agricultural experiments. The F-test compares the variance among the treatment means versus the variance of individuals within the specific treatments. High values of F indicate that one or more of the means differ from another. This can be very valuable information when, for example, you must select from several suppliers, or materials, or levels of a process factor. The F-test is a vital tool for any kind of DOE, not just simple comparisons, so it's important to understand as much as you can about it.

> *SIDEBAR: Put on your knee-length boots!*
>
> *In his landmark paper on DOE entitled "The Differential Effect of Manures on Potatoes," Fisher analyzes the impact of varying types of animal waste on yield of spuds. The agricultural heritage explains some of the farm jargon you see in writings about DOE: blocks, plots, treatments, environmental factors, etc. It's a lot to wade through for non-statisticians, but worth the effort.*

The F-Test as Simple as Possible

Without getting into all the details, the following formula for F can be derived from part two of the central limit theorem:

$$F \equiv \frac{ns_{\bar{Y}}^2}{s_{pooled}^2}$$

This formula assumes that all samples are of equal size n. You might think of F as a ratio of signal (differences caused by the treatments) versus noise. The F-ratio increases as the treatment differences become larger. It becomes more sensitive to a given treatment difference as the sample size (n) increases. Thus, if something really does differ, you will eventually find it if you collect more data. On the other hand, the F-ratio decreases as variation (s_{pooled}^2) increases. This noise is your enemy. Before you even begin doing experimentation, do what you can to dampen system variability via statistical process control (SPC), quality assurance on your response measurement, and control of environmental factors.

If the treatments have no effect, then the F-ratio will be near a value of 1. As the F-ratio increases, it becomes less and less likely that this could occur by chance. With the use of statistical tables such as those provided in the Appendix, you can quantify this probability ("p"). The "p-value" becomes a good 'bottom-line' indicator of significance. When the F-ratio gets so high that the p-value falls below 0.05, then you can say with 95% confidence that one or more of the treatments is having an effect on the measured response. This still leaves a 5% "risk" that noise is the culprit. If 5% risk is too much, you can set a standard of 1%, thus ensuring 99% confidence. Conversely, you may want to live dangerously by taking a 10% risk (90% confidence). It's your choice. Our choice for examples shown in this book will be 95% confidence for all tests and intervals.

SIDEBAR: ANOVA: Not a car, not an exploding star (or your brain from statistics)

The F-test uses variance as its underlying statistic. Therefore, statisticians call the overall procedure "analysis of variance," or "ANOVA". When this is applied to simple comparisons, it's called a "one-way" ANOVA. With the advent of built-in spreadsheet functions and dedicated statistical software, ANOVA can be accomplished very easily. However, it still can be somewhat scary to look at for non-statisticians.