them differ. We will focus on fixed effects models. Later in this section, we will discuss some of the differences between fixed and random effects models.

Completely Randomized Experiments

In this welding experiment, a total of 20 welds were produced, five with each of the four fluxes. Each weld was produced on a different steel base plate. Therefore, to run the experiment, the experimenter had to choose, from a total of 20 base plates, a group of 5 to be welded with flux A, another group of 5 to be welded with flux B, and so on. The best way to assign the base plates to the fluxes is at random. In this way, the experimental design will not favor any one treatment over another. For example, the experimenter could number the plates from 1 to 20, and then generate a random ordering of the integers from 1 to 20. The plates whose numbers correspond to the first five numbers on the list are assigned to flux A, and so on. This is an example of a **completely randomized experiment**.

Definition

A factorial experiment in which experimental units are assigned to treatments at random, with all possible assignments being equally likely, is called a **completely randomized experiment**.

In many situations, the results of an experiment can be affected by the order in which the observations are taken. For example, the performance of a machine used to make measurements may change over time, due, for example, to calibration drift, or to warm-up effects. In cases like this, the ideal procedure is to take the observations in random order. This requires switching from treatment to treatment as observations are taken, rather than running all the observations that correspond to a given treatment consecutively. In some cases changing treatments involves considerable time or expense, so it is not feasible to switch back and forth. In these cases, the treatments should be run in a random order, with all the observations corresponding to the first randomly chosen treatment being run first, and so on.

In a completely randomized experiment, it is appropriate to think of each treatment as representing a population, and the responses observed for the units assigned to that treatment as a simple random sample from that population. The data from the experiment thus consist of several random samples, each from a different population. The population means are called **treatment means**. The questions of interest concern the treatment means—whether they are all equal, and if not, which ones are different, how big the differences are, and so on.

One-Way Analysis of Variance

To make a formal determination as to whether the treatment means differ, a hypothesis test is needed. We begin by introducing the notation. We have I samples, each from a different treatment. The treatment means are denoted

$$\mu_1, \ldots, \mu_I$$

It is not necessary that the sample sizes be equal, although it is desirable, as we will discuss later in this section. The sample sizes are denoted

$$J_1, \ldots, J_I$$

The total number in all the samples combined is denoted by N.

$$N = J_1 + J_2 + \cdots + J_I$$

The hypotheses we wish to test are

$$H_0 : \mu_1 = \cdots = \mu_I \quad \text{versus} \quad H_1 : \text{two or more of the } \mu_i \text{ are different}$$

If there were only two samples, we could use the two-sample t test (Section 6.7) to test the null hypothesis. Since there are more than two samples, we use a method known as **one-way analysis of variance** (ANOVA). To define the test statistic for one-way ANOVA, we first develop the notation for the sample observations. Since there are several samples, we use a double subscript to denote the observations. Specifically, we let X_{ij} denote the jth observation in the ith sample. The sample mean of the ith sample is denoted $\overline{X}_{i.}$.

$$\overline{X}_{i.} = \frac{\sum_{j=1}^{J_i} X_{ij}}{J_i} \tag{9.1}$$

The **sample grand mean**, denoted $\overline{X}_{..}$, is the average of all the sampled items taken together:

$$\overline{X}_{..} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J_i} X_{ij}}{N} \tag{9.2}$$

With a little algebra, it can be shown that the sample grand mean is also a weighted average of the sample means:

$$\overline{X}_{..} = \frac{\sum_{i=1}^{I} J_i \overline{X}_{i.}}{N} \tag{9.3}$$

Example 9.1

For the data in Table 9.1, find I, $J_1, \ldots, J_I$, N, X_{23}, $\overline{X}_{3.}$, $\overline{X}_{..}$.

Solution

There are four samples, so $I = 4$. Each sample contains five observations, so $J_1 = J_2 = J_3 = J_4 = 5$. The total number of observations is $N = 20$. The quantity X_{23} is the third observation in the second sample, which is 267. The quantity $\overline{X}_{3.}$ is the sample mean of the third sample. This value is $\overline{X}_{3.} = 271.0$. Finally, we use Equation (9.3) to compute the sample grand mean $\overline{X}_{..}$.

$$\begin{aligned}\overline{X}_{..} &= \frac{(5)(253.8) + (5)(263.2) + (5)(271.0) + (5)(262.0)}{20} \\ &= 262.5\end{aligned}$$

Figure 9.2 presents the idea behind one-way ANOVA. The figure illustrates several hypothetical samples from different treatments, along with their sample means and the sample grand mean. The sample means are spread out around the sample grand mean. One-way ANOVA provides a way to measure this spread. If the sample means are highly spread out, then it is likely that the treatment means are different, and we will reject H_0.

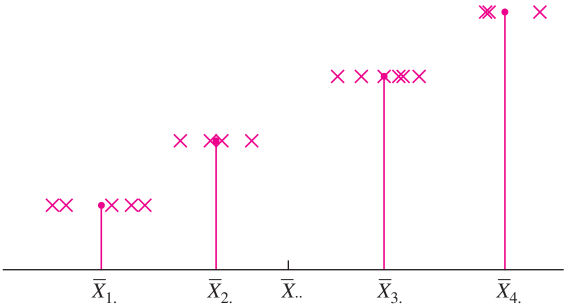

FIGURE 9.2 The variation of the sample means around the sample grand mean can be due both to random uncertainty and to differences among the treatment means. The variation within a given sample around its own sample mean is due only to random uncertainty.

The variation of the sample means around the sample grand mean is measured by a quantity called the **treatment sum of squares** (SSTr for short), which is given by

$$\text{SSTr} = \sum_{i=1}^{I} J_i(\overline{X}_{i.} - \overline{X}_{..})^2 \tag{9.4}$$

Each term in SSTr involves the distance from the sample means to the sample grand mean. Note that each squared distance is multiplied by the sample size corresponding to its sample mean, so that the means for the larger samples count more. SSTr provides an indication of how different the treatment means are from each other. If SSTr is large, then the sample means are spread out widely, and it is reasonable to conclude that the treatment means differ and to reject H_0. If on the other hand SSTr is small, then the sample means are all close to the sample grand mean and therefore to each other, so it is plausible that the treatment means are equal.

An equivalent formula for SSTr, which is a bit easier to compute by hand, is

$$\text{SSTr} = \sum_{i=1}^{I} J_i\overline{X}_{i.}^{\,2} - N\overline{X}_{..}^{\,2} \tag{9.5}$$

In order to determine whether SSTr is large enough to reject H_0, we compare it to another sum of squares, called the **error sum of squares** (SSE for short). SSE measures the variation in the individual sample points around their respective sample means. This variation is measured by summing the squares of the distances from each point to its own sample mean. SSE is given by

$$\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \overline{X}_{i.})^2 \tag{9.6}$$

The quantities $X_{ij} - \overline{X}_{i.}$ are called the **residuals**, so SSE is the sum of the squared residuals. SSE, unlike SSTr, depends only on the distances of the sample points from their own means and is not affected by the location of treatment means relative to one another. SSE therefore measures only the underlying random variation in the process being studied. It is analogous to the error sum of squares in regression.

An equivalent formula for SSE, which is a bit easier to compute by hand, is

$$\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J_i} X_{ij}^2 - \sum_{i=1}^{I} J_i \overline{X}_{i.}{}^2 \tag{9.7}$$

Another equivalent formula for SSE is based on the sample variances. Let s_i^2 denote the sample variance of the ith sample. Then

$$s_i^2 = \frac{\sum_{j=1}^{J_i}(X_{ij} - \overline{X}_{i.})^2}{J_i - 1} \tag{9.8}$$

It follows from Equation (9.8) that $\sum_{j=1}^{J_i}(X_{ij} - \overline{X}_{i.})^2 = (J_i - 1)s_i^2$. Substituting into Equation (9.6) yields

$$\text{SSE} = \sum_{i=1}^{I}(J_i - 1)s_i^2 \tag{9.9}$$

Example 9.2

For the data in Table 9.1, compute SSTr and SSE.

Solution

The sample means are presented in Table 9.1. They are

$$\overline{X}_{1.} = 253.8 \qquad \overline{X}_{2.} = 263.2 \qquad \overline{X}_{3.} = 271.0 \qquad \overline{X}_{4.} = 262.0$$

The sample grand mean was computed in Example 9.1 to be $\overline{X}_{..} = 262.5$. We now use Equation (9.4) to calculate SSTr.

$$\begin{aligned}\text{SSTr} &= 5(253.8-262.5)^2+5(263.2-262.5)^2+5(271.0-262.5)^2+5(262.0-262.5)^2 \\ &= 743.4\end{aligned}$$

To compute SSE we will use Equation (9.9), since the sample standard deviations s_i have already been presented in Table 9.1.

$$\begin{aligned}\text{SSE} &= (5-1)(9.7570)^2 + (5-1)(5.4037)^2 + (5-1)(8.7178)^2 + (5-1)(7.4498)^2 \\ &= 1023.6\end{aligned}$$

We can use SSTr and SSE to construct a test statistic, provided the following two assumptions are met.

Assumptions for One-Way ANOVA

The standard one-way ANOVA hypothesis tests are valid under the following conditions:

1. The treatment populations must be normal.
2. The treatment populations must all have the same variance, which we will denote by σ^2.

Before presenting the test statistic, we will explain how it works. If the two assumptions for one-way ANOVA are approximately met, we can compute the means of SSE and SSTr. The mean of SSTr depends on whether H_0 is true, because SSTr tends to be smaller when H_0 is true and larger when H_0 is false. The mean of SSTr satisfies the condition

$$\mu_{\text{SSTr}} = (I-1)\sigma^2 \quad \text{when } H_0 \text{ is true} \tag{9.10}$$

$$\mu_{\text{SSTr}} > (I-1)\sigma^2 \quad \text{when } H_0 \text{ is false} \tag{9.11}$$

The likely size of SSE, and thus its mean, does not depend on whether H_0 is true. The mean of SSE is given by

$$\mu_{\text{SSE}} = (N-I)\sigma^2 \quad \text{whether or not } H_0 \text{ is true} \tag{9.12}$$

Derivations of Equations (9.10) and (9.12) are given at the end of this section.

The quantities $I - 1$ and $N - I$ are the **degrees of freedom** for SSTr and SSE, respectively. When a sum of squares is divided by its degrees of freedom, the quantity obtained is called a **mean square**. The **treatment mean square** is denoted MSTr, and the **error mean square** is denoted MSE. They are defined by

$$\text{MSTr} = \frac{\text{SSTr}}{I-1} \qquad \text{MSE} = \frac{\text{SSE}}{N-I} \tag{9.13}$$

It follows from Equations (9.10) through (9.13) that

$$\mu_{\text{MSTr}} = \sigma^2 \quad \text{when } H_0 \text{ is true} \tag{9.14}$$

$$\mu_{\text{MSTr}} > \sigma^2 \quad \text{when } H_0 \text{ is false} \tag{9.15}$$

$$\mu_{\text{MSE}} = \sigma^2 \quad \text{whether or not } H_0 \text{ is true} \tag{9.16}$$

Equations (9.14) and (9.16) show that when H_0 is true, MSTr and MSE have the same mean. Therefore, when H_0 is true, we would expect their quotient to be near 1. This quotient is in fact the test statistic. The test statistic for testing $H_0 : \mu_1 = \cdots = \mu_I$ is

$$F = \frac{\text{MSTr}}{\text{MSE}} \tag{9.17}$$

When H_0 is true, the numerator and denominator of F are on average the same size, so F tends to be near 1. In fact, when H_0 is true, this test statistic has an F distribution with $I - 1$ and $N - I$ degrees of freedom, denoted $F_{I-1,N-I}$. When H_0 is false, MSTr tends to be larger, but MSE does not, so F tends to be greater than 1.

Summary

The *F* test for One-Way ANOVA

To test $H_0: \mu_1 = \cdots = \mu_I$ versus H_1: two or more of the μ_i are different:

1. Compute $\text{SSTr} = \sum_{i=1}^{I} J_i(\overline{X}_{i.} - \overline{X}_{..})^2 = \sum_{i=1}^{I} J_i\overline{X}_{i.}^{\,2} - N\overline{X}_{..}^{\,2}$.
2. Compute $\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \overline{X}_{i.})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J_i}X_{ij}^2 - \sum_{i=1}^{I} J_i\overline{X}_{i.}^{\,2}$
   $= \sum_{i=1}^{I}(J_i - 1)s_i^2$.
3. Compute $\text{MSTr} = \dfrac{\text{SSTr}}{I-1}$ and $\text{MSE} = \dfrac{\text{SSE}}{N-I}$.
4. Compute the test statistic: $F = \dfrac{\text{MSTr}}{\text{MSE}}$.
5. Find the P-value by consulting the F table (Table A.8 in Appendix A) with $I - 1$ and $N - I$ degrees of freedom.

We now apply the method of analysis of variance to the example with which we introduced this section.

Example 9.3

For the data in Table 9.1, compute MSTr, MSE, and F. Find the P-value for testing the null hypothesis that all the means are equal. What do you conclude?

Solution

From Example 9.2, SSTr $= 743.4$ and SSE $= 1023.6$. We have $I = 4$ samples and $N = 20$ observations in all the samples taken together. Using Equation (9.13),

$$\text{MSTr} = \frac{743.4}{4-1} = 247.8 \qquad \text{MSE} = \frac{1023.6}{20-4} = 63.975$$

The value of the test statistic F is therefore

$$F = \frac{247.8}{63.975} = 3.8734$$

To find the P-value, we consult the F table (Table A.8). The degrees of freedom are $4 - 1 = 3$ for the numerator and $20 - 4 = 16$ for the denominator. Under H_0, F has an $F_{3,16}$ distribution. Looking at the F table under 3 and 16 degrees of freedom, we find that the upper 5% point is 3.24 and the upper 1% point is 5.29. Therefore the P-value is between 0.01 and 0.05 (see Figure 9.3 on page 666; a computer software package gives a value of 0.029 accurate to two significant digits). It is reasonable to conclude that the population means are not all equal, and, thus, that flux composition does affect hardness.