Chapter 3

Analysis of Variance

Introduction

Experiments are essential to the development and improvement of engineering and scientific methods. Only through experimentation can different variants of a method be compared to see which are most effective. To be useful, an experiment must be designed properly, and the data it produces must be analyzed correctly. In this chapter we will discuss the design of and the analysis of data from a class of experiments known as **factorial experiments**.

3.1 One-Factor Experiments

3.1.1 Basic Ideas

We begin with an example. The article "An Investigation of the $CaCO_3$-CaF_2-K_2SiO_3-SiO_2-Fe Flux System Using the Submerged Arc Welding Process on HSLA-100 and AISI-1081 Steels" (G. Fredrickson, M.S. Thesis, Colorado School of Mines, 1992) describes an experiment in which welding fluxes with differing chemical compositions were prepared. Several welds using each flux were made on AISI-1018 steel base metal. The results of hardness measurements, on the Brinell scale, of five welds using each of four fluxes are presented in Table 3.1.

Table 3.1: Brinell Hardness of Welds Using Four Different Fluxes.

Flux	Sample Values	Sample Mean	Sample Standard Deviation
A	250, 264, 256, 260, 239	253.8	9.7570
B	263, 254, 267, 265, 267	263.2	5.4037
C	257, 279, 269, 273, 277	271.0	8.7178
D	253, 258, 262, 264, 273	262.0	7.4498

Figure 3.1 presents dotplots for the hardnesses using the four fluxes. Each sample mean is marked with an "X." It is clear that the sample means differ. In particular, the welds made using flux C have the largest sample mean and those using flux A have the smallest. Of course, there is uncertainty in the sample means, and the question is whether the sample means differ from each other by a greater amount than could be

accounted for by uncertainty alone. Another way to phrase the question is this: Can we conclude that there are differences in the population means among the four flux types?

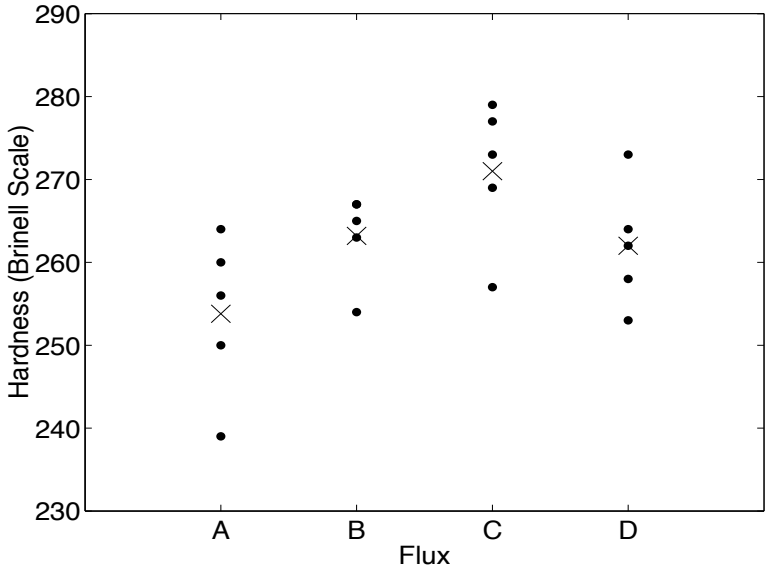


Figure 3.1: Dotplots for each sample in Table 3.1. Each sample mean is marked with an "X." The sample means differ somewhat, but the sample values overlap considerably.

This is an example of a factorial experiment. In general a factorial experiment involves several variables. One variable is the **response variable**, which is sometimes called the **outcome variable** or the **dependent variable**. The other variables are called **factors**. The question addressed by a factorial experiment is whether varying the levels of the factors produces a difference in the mean of the response variable. In the experiment described in Table 3.1, the hardness is the response, and there is one factor: flux type. Since there is only one factor, this is a **one-factor experiment**. There are four different values for the flux type factor in this experiment. These different values are called the **levels** of the factor, and can also be called **treatments**. Finally, the objects upon which measurements are made are called **experimental units**. The units assigned to a given treatment are called **replicates**. In the experiment above, the welds are the experimental units, and there are five replicates for each treatment.

In this welding experiment, the four particular flux compositions were chosen deliberately by the experimenter, rather than at random from a larger population of fluxes. Such an experiment is said to follow a **fixed effects model**. In some experiments, treatments are chosen at random from a population of possible treatments. In this case the experiment is said to follow a **random effects model**. The methods of analysis for these two models are essentially the same, although the conclusions to be drawn from 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.

3.1.2 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, then generate a random ordering of the integers from 1 to 20. The plates whose numbers correspond to the first 5 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.

3.1.3 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

$$\mu_1, ..., \mu_I$$

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

$$J_1, ..., J_I$$

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

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

If there were only two samples, we might use the two-sample t test 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 $\overline{X}_{i\cdot}$:

$$\overline{X}_{i\cdot} = \frac{\sum_{j=1}^{J_i} X_{ij}}{J_i} \tag{3.1}$$

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

$$\overline{X}_{..} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J_i} X_{ij}}{N} \tag{3.2}$$

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

$$\overline{X}_{..} = \frac{\sum_{i=1}^{I} J_i \overline{X}_{i.}}{N} \tag{3.3}$$

Example 3.1

For the data in Table 3.1, find I, $J_1, ..., J_I$, N, X_{23}, $\overline{X}_{3.}$, $\overline{X}_{..}$.

Solution: There are 4 samples, so $I = 4$. Each sample contains 5 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 presented in Table 3.1, and is $\overline{X}_{3.} = 271.0$. Finally, we use Equation (3.3) 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 3.2 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.

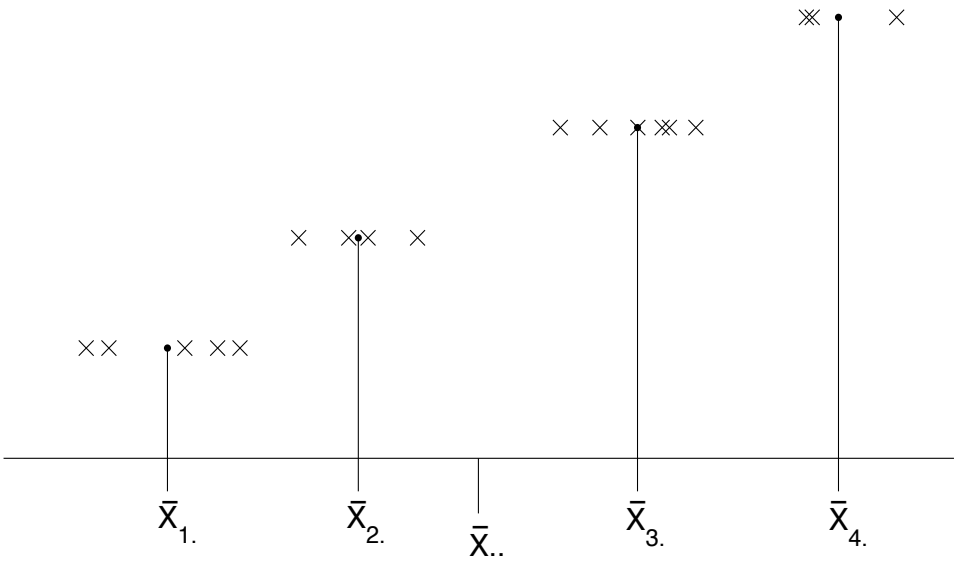


Figure 3.2: The variation of the sample means around the sample grand mean can be due both to random uncertainty and to differences among the treatment means. The variation within a given sample around its own sample mean is due only to random uncertainty.

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

$$SSTr = \sum_{i=1}^{I} J_i(\overline{X}_{i.} - \overline{X}_{..})^2 \tag{3.4}$$

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

$$SSTr = \sum_{i=1}^{I} J_i \overline{X}_{i.}^{\,2} - N\overline{X}_{..}^{\,2} \tag{3.5}$$

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

$$SSE = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (X_{ij} - \overline{X}_{i.})^2 \tag{3.6}$$

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

$$SSE = \sum_{i=1}^{I} \sum_{j=1}^{J_i} X_{ij}^2 - \sum_{i=1}^{I} J_i \overline{X}_{i.}^{\,2} \tag{3.7}$$

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

$$s_i^2 = \frac{\sum_{j=1}^{J_i} (X_{ij} - \overline{X}_{i.})^2}{J_i - 1} \tag{3.8}$$

It follows from Equation (3.8) that $\sum_{j=1}^{J_i} (X_{ij} - \overline{X}_{i.})^2 = (J_i - 1)s_i^2$. Substituting into Equation (3.6) yields

$$SSE = \sum_{i=1}^{I} (J_i - 1)s_i^2 \tag{3.9}$$

Example 3.2

For the data in Table 3.1, compute $SSTr$ and SSE.

Solution: The sample means are presented in Table 3.1. They are:

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

The sample grand mean was computed in Example 3.1 to be $\overline{X}_{\cdot\cdot} = 262.5$. We now use Equation (3.4) to calculate $SSTr$.

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

To compute SSE we will use Equation (3.9), since the sample standard deviations s_i have already been presented in Table 3.1.

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

We can use $SSTr$ and SSE to construct a test statistic, provided 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.

We will present the test statistic below, but first we will explain how it works. If the two assumptions above 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

$$E(SSTr) = (I-1)\sigma^2 \quad \text{when } H_0 \text{ is true} \tag{3.10}$$

$$E(SSTr) > (I-1)\sigma^2 \quad \text{when } H_0 \text{ is false} \tag{3.11}$$

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

$$E(SSE) = (N-I)\sigma^2 \quad \text{whether or not } H_0 \text{ is true} \tag{3.12}$$

Derivations of Equations (3.10) and (3.12) 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

$$MSTr = \frac{SSTr}{I-1} \qquad MSE = \frac{SSE}{N-I} \tag{3.13}$$

It follows from Equations (3.10), (3.11), (3.12), and (3.13) that

$$E(MSTr) = \sigma^2 \quad \text{when } H_0 \text{ is true} \tag{3.14}$$

$$E(MSTr) > \sigma^2 \quad \text{when } H_0 \text{ is false} \tag{3.15}$$

$$E(MSE) = \sigma^2 \quad \text{whether or not } H_0 \text{ is true} \tag{3.16}$$

Equations (3.14) and (3.16) 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 $H_0\colon \mu_1 = \cdots = \mu_I$ is

$$F = \frac{MSTr}{MSE} \tag{3.17}$$

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 $H_0: \mu_1 = \cdots = \mu_I$ vs. $H_1:$ two or more of the μ_i are different:

1. Compute $SSTr = \sum_{i=1}^{I} J_i(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})^2 = \displaystyle\sum_{i=1}^{I} J_i \overline{X}_{i\cdot}{}^2 - N\overline{X}_{\cdot\cdot}{}^2$
2. Compute $SSE = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \overline{X}_{i\cdot})^2 = \displaystyle\sum_{i=1}^{I}\sum_{j=1}^{J_i} X_{ij}^2 - \sum_{i=1}^{I} J_i \overline{X}_{i\cdot}{}^2$
   $= \sum_{i=1}^{I}(J_i - 1)s_i^2$
3. Compute $MSTr = \dfrac{SSTr}{I-1}$ and $MSE = \dfrac{SSE}{N-I}$
4. Compute the test statistic: $F = \dfrac{MSTr}{MSE}$
5. Find the P-value by consulting the F table 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 3.3

For the data in Table 3.1, 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 3.2, $SSTr = 743.4$ and $SSE = 1023.6$. We have $I = 4$ samples and $N = 20$ observations in all the samples taken together. Using Equation (3.13),

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

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 3, beginning on page 192). 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 3.3; 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, so that flux composition does affect hardness.

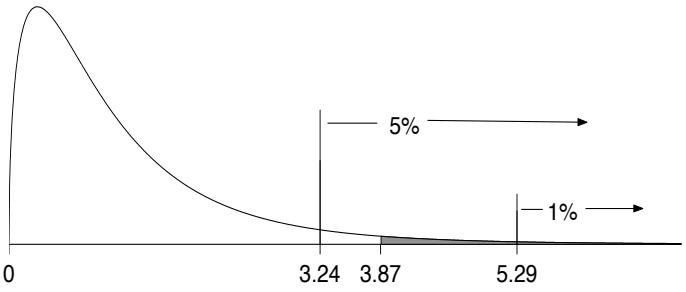


Figure 3.3: The observed value of the test statistic is 3.87. The upper 5% point of the $F_{3,16}$ distribution is 3.24. The upper 1% point of the $F_{3,16}$ distribution is 5.29. Therefore the P-value is between 0.01 and 0.05. A computer software package gives a value of 0.029. ∎

3.1.4 Confidence Intervals for the Treatment Means

The observations on the ith treatment are assumed to be a simple random sample from a normal population with mean μ_i and variance σ^2. To construct a confidence interval for μ_i, the first step is to estimate the population variance σ^2. One way to do this would be to use the sample variance s_i^2 of the observations on the ith treatment. However, since we assume that all observations for all treatments have the same variance, it is better to combine all the sample variances into one "pooled" estimate. To do this, note that SSE is a weighted sum of the sample variances (Equation 3.9) and MSE is the weighted average (Equation 3.13). The quantity MSE is therefore the pooled estimate of the variance σ^2. Since $\overline{X}_{i\cdot}$ is the sample mean of J_i observations, the variance of $\overline{X}_{i\cdot}$ is σ^2/J_i, estimated with MSE/J_i. The number of degrees of freedom for MSE is $N - I$. The quantity

$$\frac{\overline{X}_{i\cdot} - \mu_i}{\sqrt{MSE/J_i}}$$

has a Student's t distribution with $N - I$ degrees of freedom. A confidence interval for μ_i can therefore be constructed using the Student's t distribution.

A level $100(1-\alpha)$ confidence interval for μ_i is given by

$$\overline{X}_{i\cdot} \pm t_{N-I,\,\alpha/2}\sqrt{\frac{MSE}{J_i}} \tag{3.18}$$

Example 3.4

Find a 95% confidence interval for the mean hardness of welds produced with flux A.

Solution: From Table 3.1 on page 127, $\overline{X}_{1\cdot} = 253.8$. The value of MSE was computed in Example 3.3 to be 63.975. There are $I = 4$ treatments, $J_1 = 5$ observations for flux A, and $N = 20$ observations altogether. From the Student's t table we obtain $t_{16,\,.025} = 2.120$. The 95% confidence interval is therefore

$$253.8 \pm 2.120\sqrt{63.975/5} = 253.8 \pm 7.6$$ ∎

3.1.5 The ANOVA Table

The results of an analysis of variance are usually summarized in an analysis of variance (ANOVA) table. This table is much like the analysis of variance table produced in multiple regression. Below is R output showing the analysis of variance for the weld data presented in Table 3.1.

```
> Hardness = c(250, 264, 256, 260, 239, 263, 254, 267, 265, 267, 257, 279, 269,
273, 277, 253, 258, 262, 264, 273)
> Flux = c("A","A","A","A","A","B","B","B","B","B","C","C","C","C","C","D","D","
D","D","D")
> weld.lm = lm(Hardness~Flux, data=weld)
> weld.aov = anova(weld.lm)
> weld.aov

           Df  Sum Sq Mean Sq F value  Pr(>F)
Flux        3  743.40  247.80  3.8734 0.02944
Residuals  16 1023.60   63.98
```

In the ANOVA table, the column labeled "DF" presents the number of degrees of freedom for both the treatment ("Flux") and error ("Residuals") sum of squares. The column labeled "Sum Sq" presents $SSTr$ (in the row labeled "Flux") and SSE (in the row labeled "Residuals"). The column labeled "Mean Sq" presents the mean squares $MSTr$ and MSE. The column labeled "F value" presents the F statistic for testing the null hypothesis that all the population means are equal. Finally, the column labeled "Pr(>F)" presents the P-value for the F test.

Example 3.5

In the article "Review of Development and Application of CRSTER and MPTER Models" (R. Wilson, *Atmospheric Environment*, 1993:41–57), several measurements of the maximum hourly concentrations (in $\mu g/m^3$) of SO_2 are presented for each of four power plants. The results are presented below (two outliers have been deleted).

Plant 1: 438, 619, 732, 638
Plant 2: 857, 1014, 1153, 883, 1053
Plant 3: 925, 786, 1179, 786
Plant 4: 893, 891, 917, 695, 675, 595

R output for a one-way ANOVA is presented below. Can you conclude that the maximum hourly concentrations differ among the plants?

```
        Df  Sum Sq Mean Sq F value   Pr(>F)
Plant    3  378610  126203    6.21 0.005917
Error   15  304838   20323
```

Solution: In the ANOVA table, the P-value for the null hypothesis that all treatment means are equal is 0.0059. Therefore we conclude that not all the treatment means are equal. ∎

3.1.6 Checking the Assumption of Equal Variances

The assumption of equal variances can be difficult to check, because with only a few observations in each sample, the sample standard deviations can differ greatly (by a factor of 2 or more) even when the assumption holds. For the weld data, the sample standard deviations range from 5.4037 to 9.7570. It is reasonable to proceed as though the variances were equal.

The spreads of the observations within the various samples can be checked visually by making a residual plot. This is done by plotting the residuals $X_{ij} - \overline{X}_{i.}$ vs. the fitted values, which are the sample means $\overline{X}_{i.}$. If the spreads differ considerably among the samples, the assumption of equal variances is suspect. If one or more of the samples contain outliers, the assumption of normality is suspect as well. Figure 3.4 presents a residual plot for the weld data. There are no serious outliers, and the spreads do not differ greatly among samples.

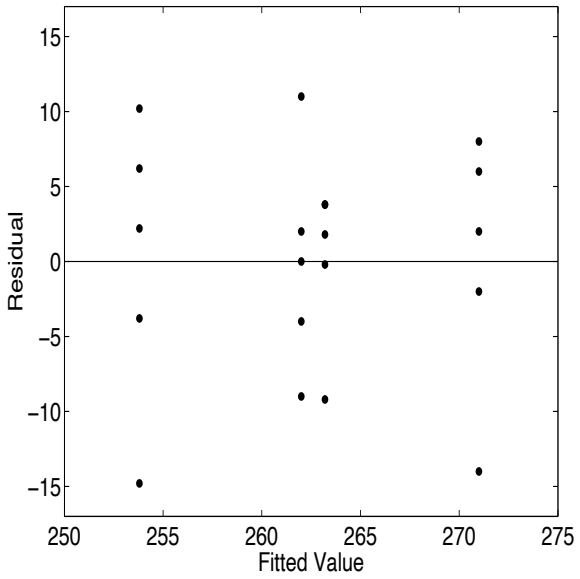


Figure 3.4: Residual plot of the values $X_{ij} - \overline{X}_{i.}$ vs. $\overline{X}_{i.}$ for the weld data. The spreads do not differ greatly from sample to sample, and there are no serious outliers.

3.1.7 Balanced vs. Unbalanced Designs

When equal numbers of units are assigned to each treatment, the design is said to be **balanced**. Although one-way analysis of variance can be used with both balanced and unbalanced designs, balanced designs offer a big advantage. A balanced design is much less sensitive to violations of the assumption of equality of variance than an unbalanced one. Since moderate departures from this assumption can be difficult to detect, it is best to use a balanced design whenever possible, so that undetected violations of the assumption will not seriously compromise the validity of the results. When a balanced design is impossible to achieve, a slightly unbalanced design is preferable to a severely unbalanced one.

Summary

- With a balanced design, the effect of unequal variances is generally not great.
- With an unbalanced design, the effect of unequal variances can be substantial.
- The more unbalanced the design, the greater the effect of unequal variances.

3.1.8 The Analysis of Variance Identity

In both linear regression and analysis of variance, a quantity called the total sum of squares is obtained by subtracting the sample grand mean from each observation, squaring these deviations, then summing them. An analysis of variance identity is an equation that expresses the total sum of squares as a sum of other sums of squares. We have presented analysis of variance identities for simple linear regression and for multiple regression.

The total sum of squares for one-way ANOVA is given by

$$SST = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (X_{ij} - \overline{X}_{\cdot\cdot})^2 \tag{3.19}$$

An equivalent formula is given by

$$SST = \sum_{i=1}^{I} \sum_{j=1}^{J_i} X_{ij}^2 - N\overline{X}_{\cdot\cdot}^{\,2} \tag{3.20}$$

Examining Equations (3.20), (3.7), and (3.5) shows that the total sum of squares is equal to the treatment sum of squares plus the error sum of squares. This is the analysis of variance identity for one-way analysis of variance.

The Analysis of Variance Identity

$$SST = SSTr + SSE \tag{3.21}$$

3.1.9 The Means Parameterization

Our presentation of one-way analysis of variance, as a method to compare several treatment means by using random samples drawn from each treatment population, is one natural way to view the subject. There is another way to express these same ideas, in somewhat different notation, that is sometimes useful.

For each observation X_{ij}, define $\varepsilon_{ij} = X_{ij} - \mu_i$, the difference between the observation and its mean. By analogy with linear regression, the quantities ε_{ij} are called **errors**. It is clearly true that

$$X_{ij} = \mu_i + \varepsilon_{ij} \tag{3.22}$$

Now since X_{ij} is normally distributed with mean μ_i and variance σ^2, it follows that ε_{ij} is normally distributed with mean 0 and variance σ^2.

Sometimes it is of interest to consider one treatment to be a "control" or "baseline" treatment, and to compare the means of the other treatments to it. For example, assume the first treatment is the baseline.

Then we are interested in μ_1 and in the differences $\mu_i - \mu_1$ for each treatment $i = 2, ..., I$. Specifically, define parameters $\nu_2 = \mu_2 - \mu_1, ..., \nu_I = \mu_I - \mu_1$. Define $\nu_1 = 0$. We can then write the means parameterization as

$$X_{ij} = \mu_1 + \nu_i + \varepsilon_{ij}$$

The null hypothesis $H_0 \colon \mu_1 = \cdots = \mu_I$ is equivalent to $H_0 \colon \nu_2 = \cdots = \nu_I = 0$. The estimates are $\widehat{\mu}_1 = \overline{X}_{1.}$ and $\widehat{\nu}_i = \overline{X}_{i.} - \overline{X}_{1.}$ for $i = 2, ..., I$.

Example 3.6

Using Flux A as the baseline, compute the estimates $\widehat{\mu}_1$, $\widehat{\nu}_2$, $\widehat{\nu}_3$, and $\widehat{\nu}_4$ for the flux data in Table 3.1.

Solution: The cell means are $\overline{X}_{1.} = 253.8$. $\overline{X}_{2.} = 263.2$, $\overline{X}_{3.} = 271.0$, and $\overline{X}_{4.} = 262.0$. The parameter estimates are $\widehat{\mu_1} = \overline{X}_{1.} = 253.8$, $\widehat{\nu_2} = \overline{X}_{2.} - \overline{X}_{1.} = 263.2 - 253.8 = 9.4$, $\widehat{\nu_3} = \overline{X}_{3.} - \overline{X}_{1.} = 271.0 - 253.8 = 17.2$, and $\widehat{\nu_4} = \overline{X}_{4.} - \overline{X}_{1.} = 262.0 - 253.8 = 8.2$, ∎

3.1.10 The Effects Parameterization

In a single-factor experiment, we are interested in determining whether the treatment means are all equal. Given treatment means $\mu_1, ...\mu_I$, the quantity

$$\mu = \frac{1}{I} \sum_{i}^{I} \mu_i \tag{3.23}$$

is the average of all the treatment means. The quantity μ is called the **population grand mean**. The ith **treatment effect**, denoted α_i, is the difference between the ith treatment mean and the population grand mean:

$$\alpha_i = \mu_i - \mu \tag{3.24}$$

It follows from the definition of α_i that $\sum_i^I \alpha_i = 0$.

We can now decompose the treatment means as follows:

$$\mu_i = \mu + \alpha_i \tag{3.25}$$

We can write the effects parameterization as

$$X_{ij} = \mu + \alpha_1 + \varepsilon_{ij}$$

The null hypothesis $H_0 \colon \mu_1 = \cdots = \mu_I$ is equivalent to $H_0 \colon \alpha_1 = \cdots = \alpha_I = 0$.

To compute the parameter estimates, we estimate μ_i with $\overline{X}_{i.}$. Therefore we estimate μ with $\widehat{\mu} = \dfrac{1}{I} \displaystyle\sum_{i=1}^{I} \overline{X}_{i.}$, and α_i with $\widehat{\alpha}_i = \overline{X}_{i.} - \widehat{\mu}$.

Example 3.7

Compute estimates of the parameters μ, α_1, α_2, α_3, and α_4 for the weld data in Table 3.1.

Solution: The sample means are $\overline{X}_{1.} = 253.8$. $\overline{X}_{2.} = 263.2$, $\overline{X}_{3.} = 271.0$, and $\overline{X}_{4.} = 262.0$. The parameter estimates are

$$\widehat{\mu} = \frac{1}{4}(253.8 + 263.2 + 271.0 + 262.0) = 262.5$$
$$\widehat{\alpha}_1 = \overline{X}_{1.} - \widehat{\mu} = 253.8 - 262.5 = -8.7$$
$$\widehat{\alpha}_2 = \overline{X}_{2.} - \widehat{\mu} = 263.2 - 262.5 = 0.7$$
$$\widehat{\alpha}_3 = \overline{X}_{3.} - \widehat{\mu} = 271.0 - 262.5 = 8.5$$
$$\widehat{\alpha}_4 = \overline{X}_{4.} - \widehat{\mu} = 262.0 - 262.5 = -0.5$$

Note that $\widehat{\alpha}_1 + \widehat{\alpha}_2 + \widehat{\alpha}_3 + \widehat{\alpha}_4 = 0$, so we could compute $\widehat{\alpha}_4 = -\widehat{\alpha}_1 - \widehat{\alpha}_2 - \widehat{\alpha}_3$.

3.1.11 Analysis of Variance is Just Linear Regression

Analysis of variance is just a special case of linear regression. To see this, assume there are I treatments with means $\mu_1, ..., \mu_I$. Let X_{ij} be the jth measurement for the ith treatment. Define $\varepsilon_{ij} = X_{ij} - \mu_i$. It follows that for all i, j:

$$X_{ij} = \mu_i + \varepsilon_{ij}$$

Because the X_{ij} are independent with mean μ_i, and all have the same variance σ^2, the ε_{ij} are independent with mean 0 and variance σ^2. In particular, if the X_{ij} are normally distributed the ε_{ij} are i.i.d. $N(0, \sigma^2)$.

Fitting the Means Model with Linear Regression

Now for the means model, the parameters are μ_1, ν_2, ..., ν_I, where $\nu_i = \mu_i - \mu_1$. Therefore the one-way analysis of variance model can be written

$$\begin{aligned} X_{1j} &= \mu_1 + \varepsilon_{1j} \\ X_{ij} &= \mu_1 + \nu_i + \varepsilon_{ij} \quad \text{for } i = 2, ..., I \end{aligned}$$

To estimate the parameters with linear regression, write the values X_{ij} as a column vector. We will call this vector $\mathbf{Y}$ in keeping with the usual linear regression notation. The vector of parameters is $\boldsymbol{\beta} = (\mu_1, \nu_2, ..., \nu_I)^T$. The design matrix $\mathbf{X}$ has one column for each parameter.

Following are the vector $\mathbf{Y}$, the matrix $\mathbf{X}$ for the weld data in Table 3.1. The columns of $\mathbf{X}$ have been

labeled with the parameters they represent.

$$\mathbf{Y} = \begin{pmatrix} 250 \\ 264 \\ 256 \\ 260 \\ 239 \\ 263 \\ 254 \\ 267 \\ 265 \\ 267 \\ 257 \\ 279 \\ 269 \\ 273 \\ 277 \\ 253 \\ 258 \\ 262 \\ 264 \\ 273 \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mu_1 & \nu_2 & \nu_3 & \nu_4 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix}$$

The analysis of variance model is the linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The effects can be estimated with least squares. The null hypothesis is tested with the F-test for $H_0 \colon \nu_1 = \nu_2 = \nu_3 = 0$.

Example 3.8

Use linear regression to estimate the parameters for the means parameterization of the weld data. Use the F-test to test the null hypothesis that the treatment means are equal.

Solution: We use R.

```
#Construct the columns of X. No need to construct the intercept.
> FluxB = c(0,0,0,0,0,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0)
> FluxC = c(0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,0,0,0,0,0)
> FluxD = c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1)
> weld.lm = lm(Hardness~FluxB+FluxC+FluxD)
> summary(weld.lm)

Call:
lm(formula = Hardness ~ FluxB + FluxC + FluxD)

Residuals:
   Min     1Q Median     3Q    Max
-14.80  -3.85   1.90   4.35  11.00

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  253.800      3.577  70.953  < 2e-16
FluxB          9.400      5.059   1.858  0.08164
```

```
FluxC         17.200      5.059   3.400  0.00366
FluxD          8.200      5.059   1.621  0.12456
---

Residual standard error: 7.998 on 16 degrees of freedom
Multiple R-squared:  0.4207,Adjusted R-squared:  0.3121
F-statistic: 3.873 on 3 and 16 DF,  p-value: 0.02944
```

The results agree with those computed by hand in Example 3.6.

Fitting the Effects Model with Linear Regression
Now we will express the effects model as a linear regression model. We begin with

$$X_{ij} = \mu_i + \varepsilon_{ij}$$

Recall that the population grand mean is $\mu = \frac{1}{I}\sum_{i=1}^{I} \mu_i$, and the ith treatment effect is $\alpha_i = \mu_i - \mu$. It follows that

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Now the effects are constrained to sum to 0, that is $\sum_{i=1}^{I} \alpha_i = 0$. So we replace α_I with $-\alpha_1 - \alpha_2 - \cdots - \alpha_{i-1}$. The model is

$$\begin{aligned} X_{ij} &= \mu + \alpha_i + \varepsilon_{ij} \quad \text{for } j = 1, \ldots I-1 \\ X_{Ij} &= \mu + -\alpha_1 - \alpha_2 - \cdots - \alpha_{i-1} + \varepsilon_{Ij} \end{aligned}$$

As with the means model, we write the X_{ij} as a column vector and call it $\mathbf{Y}$. The parameter vector is $\boldsymbol{\beta} = (\mu, \alpha_1, ..., \alpha_{i-1})^T$. The design matrix has one column for each parameter.

Following are the vector $\mathbf{Y}$, the matrix $\mathbf{X}$, and the parameter vector $\boldsymbol{\beta}$ for the weld data in Table 3.1. The columns of $\mathbf{X}$ have been labeled with the parameters they represent.

$$\mathbf{Y} = \begin{pmatrix} 250 \\ 264 \\ 256 \\ 260 \\ 239 \\ 263 \\ 254 \\ 267 \\ 265 \\ 267 \\ 257 \\ 279 \\ 269 \\ 273 \\ 277 \\ 253 \\ 258 \\ 262 \\ 264 \\ 273 \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mu & \alpha_1 & \alpha_2 & \alpha_3 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

The analysis of variance model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The effects can be estimated with least squares. The null hypothesis is tested with the F-test for $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$.

Example 3.9

Estimate the treatment effects for the weld data. Test the hypothesis that the effects are all equal to 0.

Solution: We use R.

```
# First we construct the columns of X. No need to construct the intercept.
> FluxA = c(1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,-1,-1,-1,-1,-1)
> FluxB = c(0,0,0,0,0,1,1,1,1,1,0,0,0,0,0,-1,-1,-1,-1,-1)
> FluxC = c(0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,-1,-1,-1,-1,-1)

# Fit the linear model
> weld.effects = lm(Hardness~FluxA+FluxB+FluxC)
> summary(weld.effects)

Call:
lm(formula = Hardness ~ FluxA + FluxB + FluxC)

Residuals:
   Min     1Q Median     3Q    Max
-14.80  -3.85   1.90   4.35  11.00

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  262.500      1.788 146.771   <2e-16
FluxA         -8.700      3.098  -2.808   0.0126
FluxB          0.700      3.098   0.226   0.8241
FluxC          8.500      3.098   2.744   0.0144
---

Residual standard error: 7.998 on 16 degrees of freedom
Multiple R-squared: 0.4207,Adjusted R-squared: 0.3121
F-statistic: 3.873 on 3 and 16 DF,  p-value: 0.02944
```

The results agree with those computed by hand in Example 3.7.

When the Design is Unbalanced, Linear Regression Must Be Used

For unbalanced models, the usual ANOVA commands give the wrong answers in most software packages. However, linear regression will always give the right answer. We illustrate with the plant data first shown in Example 3.5.

Hourly concentrations (in μg/m^3) of SO$_2$

Plant 1: 438, 619, 732, 638
Plant 2: 857, 1014, 1153, 883, 1053
Plant 3: 925, 786, 1179, 786
Plant 4: 893, 891, 917, 695, 675, 595

We compute the effects: The treatment means are: $\overline{X}_{1.} = 606.75$, $\overline{X}_{2.} = 992.0$, $\overline{X}_{3.} = 919.0$, $\overline{X}_{4.} = 777.67$. The estimates are $\widehat{\mu} = 823.25 \quad \widehat{\alpha}_1 = -217.1 \quad \widehat{\alpha}_2 = 168.15 \quad \widehat{\alpha}_3 = 95.15$

Now fit the model using the **aov** command:

```
> SO2 = c(438,619,732,638,857,1014,1153,883,1053,925,786,1179,786,893,891,917,
695,675,595)
> Plant = c("A","A","A","A","B","B","B","B","B","C","C","C","C","D","D","D","D",
"D","D")
> Plant = as.factor(Plant)
> plant.lm = lm(SO2 ~ Plant)
> model.tables(aov(plant.lm))
Tables of effects

 Plant
         A     B     C      D
    -221.1 164.2 91.16 -50.18
rep    4.0   5.0  4.00   6.00
```

These effect estimates are incorrect. Now we will use linear regresssion.

```
> a1 = c(1,1,1,1,0,0,0,0,0,0,0,0,-1,-1,-1,-1,-1,-1)
> a2 = c(0,0,0,0,1,1,1,1,1,0,0,0,0,-1,-1,-1,-1,-1,-1)
> a3 = c(0,0,0,0,0,0,0,0,0,1,1,1,1,-1,-1,-1,-1,-1,-1)
> plant.lm = lm(SO2 ~ a1+a2+a3)
> summary(plant.lm)

Call:
lm(formula = SO2 ~ a1 + a2 + a3)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   823.85      33.18  24.831 1.35e-13
a1           -217.10      60.34  -3.598  0.00264
a2            168.15      55.97   3.004  0.00890
a3             95.15      60.34   1.577  0.13570
---

Residual standard error: 142.6 on 15 degrees of freedom
Multiple R-squared: 0.554,      Adjusted R-squared: 0.4648
F-statistic:  6.21 on 3 and 15 DF,  p-value: 0.005917
```

These estimates are correct.

The reason that the **aov** command does not work correctly with unbalanced designs is that it estimates $\widehat{\mu} = \overline{X}_{..}$ rather than $\widehat{\mu} = \frac{1}{I}\sum_{i=1}^{I} \overline{X}_{i.}$. When the design is balanced these two estimates are equal.

3.1.12 Random Effects Models

In many factorial experiments, the treatments are chosen deliberately by the experimenter. These experiments are said to follow a **fixed effects model**. In some cases, the treatments are chosen at random from a population of possible treatments. In these cases the experiments are said to follow a **random effects model**. In a fixed effects model, the interest is on the specific treatments chosen for the experiment. In a random effects model, the interest is in the whole population of possible treatments, and there is no particular interest in the ones that happened to be chosen for the experiment.

The article describing the weld experiment states that the treatments were chosen deliberately, and do not represent a random sample from a larger population of flux compositions. This experiment therefore follows a fixed effects model. The four power plants in Example 3.5 are a sample of convenience; they are plants at which measurements were readily available. In some cases it is appropriate to treat a sample of convenience as if it were a simple random sample. If these conditions hold, then the power plant experiment may be considered to follow a random effects model, otherwise it must be treated as a fixed effects model.

There is an important difference in interpretation between the results of a fixed effects model and those of a random effects model. In a fixed effects model, the only conclusions that can be drawn are conclusions about the treatments actually used in the experiment. In a random effects model, however, since the treatments are a simple random sample from a population of treatments, conclusions can be drawn concerning the whole population, including treatments not actually used in the experiment.

This difference in interpretations results in a difference in the null hypotheses to be tested. In the fixed effects model, the null hypothesis of interest is $H_0 \colon \mu_1 = \cdots = \mu_I$. In the random effects model, the null hypothesis of interest is

$$H_0 \colon \text{the treatment means are equal for every level in the population.}$$

In the random effects model, the assumption is made that the population of treatment means is normal.

Interestingly enough, although the null hypothesis for the random effects model differs from that of the fixed effects model, the hypothesis test is exactly the same. The F test described above is used for the random effects model as well as for the fixed effects model.

Example 3.10

In Example 3.5 on page 135, assume that it is reasonable to treat the four power plants as a random sample from a large population of power plants, and furthermore, assume that the SO_2 concentrations in the population of plants are normally distributed. Can we conclude that there are differences in SO_2 concentrations among the power plants in the population?

Solution: This is a random effects model, so we can use the F test to test the null hypothesis that all the treatment means in the population are the same. The results of the F test are shown in Example 3.5. The P-value is 0.006. We therefore reject the null hypothesis, and conclude that there are differences in mean SO_2 concentrations among the power plants in the population. ∎

Derivations

Derivations of Equations (3.10) and (3.12)

In what follows it will be easier to use the notation $E(\ \ )$ to denote the mean of a quantity and $V(\ \ )$ to denote the variance. So for example $E(SSE) = \mu_{SSE}$, $E(SSTr) = \mu_{SSTr}$, and $V(X_{ij})$ denotes the variance of X_{ij}.

We will show that $E(SSE) = E[\sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \overline{X}_{i.})^2] = (N - I)\sigma^2$, whether or not the population means are equal. This is Equation (3.12).

We begin by adding and subtracting the treatment mean μ_i from each term in $\sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \overline{X}_{i.})^2$ to obtain

$$SSE = \sum_{i=1}^{I}\sum_{j=1}^{J_i}[(X_{ij} - \mu_i) - (\overline{X}_{i.} - \mu_i)]^2$$

Multiplying out yields

$$SSE = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \mu_i)^2 - \sum_{i=1}^{I}\sum_{j=1}^{J_i}2(X_{ij} - \mu_i)(\overline{X}_{i.} - \mu_i) + \sum_{i=1}^{I}\sum_{j=1}^{J_i}(\overline{X}_{i.} - \mu_i)^2 \qquad (3.26)$$

Now $\sum_{j=1}^{J_i}(X_{ij} - \mu_i) = J_i(\overline{X}_{i.} - \mu_i)$. Substituting into the middle term of the right hand side of (3.26) yields

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

Since $\sum_{i=1}^{I}\sum_{j=1}^{J_i}(\overline{X}_{i.} - \mu_i)^2 = \sum_{i=1}^{I}J_i(\overline{X}_{i.} - \mu_i)^2$, this simplifies to

$$SSE = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \mu_i)^2 - \sum_{i=1}^{I}J_i(\overline{X}_{i.} - \mu_i)^2 \qquad (3.27)$$

Taking means of both sides of (3.27) yields

$$E(SSE) = \sum_{i=1}^{I}\sum_{j=1}^{J_i}E(X_{ij} - \mu_i)^2 - \sum_{i=1}^{I}J_iE(\overline{X}_{i.} - \mu_i)^2 \qquad (3.28)$$

Now $E(X_{ij}) = E(\overline{X}_{i.}) = \mu_i$. The population variances are all equal; denote their common value by σ^2. It follows that

$$E(X_{ij} - \mu_i)^2 = V(X_{ij}) = \sigma^2$$
$$E(\overline{X}_{i.} - \mu_i)^2 = V(\overline{X}_{i.}) = \sigma^2/J_i$$

Substituting into (3.28) yields

$$E(SSE) = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\sigma^2 - \sum_{i=1}^{I}J_i\sigma^2/J_i = N\sigma^2 - I\sigma^2 = (N - I)\sigma^2$$

This completes the derivation of $E(SSE)$.

We now show that $E(SSTr) = E[\sum_{i=1}^{I}J_i(\overline{X}_{i.} - \overline{X}_{..})^2] = (I - 1)\sigma^2$ under the assumption that the treatment means are all equal to a common value denoted by μ. This is Equation (3.12).

We begin by adding and subtracting the common treatment mean μ from each term in $\sum_{i=1}^{I}J_i(\overline{X}_{i.} - \overline{X}_{..})^2$ to obtain

$$SSTr = \sum_{i=1}^{I}J_i[(\overline{X}_{i.} - \mu) - (\overline{X}_{..} - \mu)]^2$$

Multiplying out, we obtain

$$SSTr = \sum_{i=1}^{I} J_i(\overline{X}_{i\cdot} - \mu)^2 - 2\sum_{i=1}^{I} J_i(\overline{X}_{i\cdot} - \mu)(\overline{X}_{\cdot\cdot} - \mu) + \sum_{i=1}^{I} J_i(\overline{X}_{\cdot\cdot} - \mu)^2 \tag{3.29}$$

Now $\overline{X}_{\cdot\cdot} = \sum_{i=1}^{I} J_i\overline{X}_{i\cdot}/N$, so $\overline{X}_{\cdot\cdot} - \mu = \sum_{i=1}^{I} J_i(\overline{X}_{i\cdot} - \mu)/N$, and $\sum_{i=1}^{I} J_i(\overline{X}_{i\cdot} - \mu) = N(\overline{X}_{\cdot\cdot} - \mu)$. Substituting into the middle term of the right hand side of (3.29), we obtain

$$SSTr = \sum_{i=1}^{I} J_i(\overline{X}_{i\cdot} - \mu)^2 - 2N(\overline{X}_{\cdot\cdot} - \mu)^2 + \sum_{i=1}^{I} J_i(\overline{X}_{\cdot\cdot} - \mu)^2$$

Since $\sum_{i=1}^{I} J_i = N$, we obtain

$$SSTr = \sum_{i=1}^{I} J_i(\overline{X}_{i\cdot} - \mu)^2 - N(\overline{X}_{\cdot\cdot} - \mu)^2$$

Taking means of both sides yields

$$E(SSTr) = \sum_{i=1}^{I} J_i E(\overline{X}_{i\cdot} - \mu)^2 - NE(\overline{X}_{\cdot\cdot} - \mu)^2 \tag{3.30}$$

Now $E(\overline{X}_{i\cdot}) = E(\overline{X}_{\cdot\cdot}) = \mu$, so

$$E(\overline{X}_{i\cdot} - \mu)^2 = V(\overline{X}_{i\cdot}) = \sigma^2/J_i$$

$$E(\overline{X}_{\cdot\cdot} - \mu)^2 = V(\overline{X}_{\cdot\cdot}) = \sigma^2/N$$

Substituting into (3.30) yields

$$E(SSTr) = \sum_{i=1}^{I} J_i\sigma^2/J_i - N\sigma^2/N = (I-1)\sigma^2$$

3.2 Pairwise Comparisons in One-Factor Experiments

3.2.1 Introduction

In a one-way ANOVA, an F test is used to test the null hypothesis that all the treatment means are equal. If this hypothesis is rejected, we can conclude that the treatment means are not all the same. But the test does not tell us which ones are different from the rest. Sometimes an experimenter has in mind two specific treatments, i and j, and wants to study the difference $\mu_i - \mu_j$. In this case a method known as Fisher's Least Significant Difference (LSD) method is appropriate, and can be used to construct confidence intervals for $\mu_i - \mu_j$, or to test the null hypothesis that $\mu_i - \mu_j = 0$. At other times, an experimenter may want to determine all the pairs of means that can be concluded to differ from each other. In this case a type of procedure called a **multiple comparisons method** must be used. We will discuss two methods of multiple comparisons, the Bonferroni method and the Tukey-Kramer method.

3.2.2 Fisher's Least Significant Difference (LSD) Method

We begin by describing Fisher's LSD method for constructing confidence intervals. The confidence interval for the difference $\mu_i - \mu_j$ is centered at the difference in sample means $\overline{X}_{i.} - \overline{X}_{j.}$. To determine how wide to make the confidence interval, it is necessary to estimate the standard deviation of $\overline{X}_{i.} - \overline{X}_{j.}$. Let J_i and J_j be the sample sizes at levels i and j, respectively. Since by assumption all observations are normally distributed with variance σ^2, it follows that $\overline{X}_{i.} - \overline{X}_{j.}$ is normally distributed with mean $\mu_i - \mu_j$ and variance $\sigma^2(1/J_i + 1/J_j)$. The variance σ^2 is estimated with MSE, for reasons explained above in the discussion about confidence intervals for the treatment means (page 134). Now the quantity

$$\frac{(\overline{X}_{i.} - \overline{X}_{j.}) - (\mu_i - \mu_j)}{\sqrt{MSE(1/J_i + 1/J_j)}}$$

has a Student's t distribution with $N - I$ degrees of freedom. (The value $N - I$ is the number of degrees of freedom used in computing MSE; see Equation 3.13.) The quantity $t_{N-I,\,\alpha/2}\sqrt{MSE(1/J_i + 1/J_j)}$ is called the least significant difference. This quantity forms the basis for confidence intervals and hypothesis tests.

Fisher's Least Significant Difference Method for Confidence Intervals and Hypothesis Tests

The Fisher's Least Significant Difference confidence interval, at level $100(1-\alpha)\%$, for the difference $\mu_i - \mu_j$ is

$$\overline{X}_{i.} - \overline{X}_{j.} \pm t_{N-I,\,\alpha/2}\sqrt{MSE\left(\frac{1}{J_i} + \frac{1}{J_j}\right)} \tag{3.31}$$

To test the null hypothesis $H_0\colon \mu_i - \mu_j = 0$, the test statistic is

$$\frac{\overline{X}_{i.} - \overline{X}_{j.}}{\sqrt{MSE\left(\frac{1}{J_i} + \frac{1}{J_j}\right)}} \tag{3.32}$$

If H_0 is true, this statistic has a Student's t distribution with $N - I$ degrees of freedom. Specifically, if

$$|\overline{X}_{i.} - \overline{X}_{j.}| > t_{N-I,\,\alpha/2}\sqrt{MSE\left(\frac{1}{J_i} + \frac{1}{J_j}\right)} \tag{3.33}$$

then H_0 is rejected at level α.

The reason that the quantity $t_{N-I,\,\alpha/2}\sqrt{MSE(1/J_i + 1/J_j)}$ is called the least significant difference is that the null hypothesis of equal means is rejected at level α whenever the difference in sample means $|\overline{X}_{i.} - \overline{X}_{j.}|$ exceeds this value. When the design is balanced, with all sample sizes equal to J, the least significant difference is equal to $t_{N-I,\,\alpha/2}\sqrt{2MSE/J}$ for all pairs of means.

Example 3.11

In the weld experiment discussed in Section 3.1.1, hardness measurements were made for five welds from each of four fluxes A, B, C, and D. The sample mean hardness values were $\overline{X}_{A.} = 253.8$, $\overline{X}_{B.} = 263.2$, $\overline{X}_{C.} = 271.0$, and $\overline{X}_{D.} = 262.0$. The ANOVA table is reproduced below.

Source	DF	SS	MS	F	P
Flux	3	743.40	247.800	3.87	0.029
Residual	16	1023.60	63.975		

Before the experiment was performed, the carbon contents of the fluxes were measured. Flux B had the lowest carbon content (2.67% by weight), and flux C had the highest (5.05% by weight). The experimenter is therefore particularly interested in comparing the hardnesses obtained with these two fluxes. Find a 95% confidence interval for the difference in mean hardness between welds produced with flux B and those produced with flux C. Can we conclude that the two means differ?

Solution: We use expression (3.31). The sample means are 271.0 for flux C and 263.2 for flux B. The R output above gives the quantity MSE as 63.975. (This value was also computed in Example 3.3 on page 133). The sample sizes are both equal to 5. There are $I = 4$ levels and $N = 20$ observations in total. For a 95% confidence interval, we consult the t table to find the value $t_{16,\,0.25} = 2.120$. The 95% confidence interval is therefore $271.0 - 263.2 \pm 2.120\sqrt{63.975(1/5 + 1/5)}$ or $(-2.92, 18.52)$.

To perform a test of the null hypothesis that the two treatment means are equal, we compute the value of the test statistic (expression 3.32) and obtain

$$\frac{271.0 - 263.2}{\sqrt{63.975(1/5 + 1/5)}} = 1.54$$

Consulting the t table with $N - I = 16$ degrees of freedom, we find that P is between 2(0.05) = 0.10 and 2(0.10) = 0.20 (note that this is a two-tailed test). We cannot conclude that the treatment means differ.

If it is desired to perform a fixed-level test at level $\alpha = 0.05$ as an alternative to computing the P-value, the critical t value is $t_{16,.025} = 2.120$. The left hand side of the inequality (3.33) is $|271.0 - 263.2| = 7.8$. The right hand side is $2.120\sqrt{63.975(1/5 + 1/5)} = 10.72$. Since 7.8 does not exceed 10.72, we do not reject H_0 the 5% level. ∎

In Example 3.11, a single test was performed on the difference between two specific means. What if we wanted to test every pair of means, to see which ones we could conclude to be different? It might seem reasonable to perform the LSD test on each pair. However, this is not appropriate, because when several tests are performed, the likelihood of rejecting a true null hypothesis increases. This is the multiple testing problem. When several confidence intervals or hypothesis tests are to be considered simultaneously, the confidence intervals must be wider, and the criterion for rejecting the null hypotheses more strict, than in situations where only a single interval or test is involved. In these situations, multiple comparisons methods are used to produce **simultaneous confidence intervals** and **simultaneous hypothesis tests**. If level $100(1 - \alpha)\%$ simultaneous confidence intervals are constructed for differences between every pair of means, then we are confident at the $100(1 - \alpha)\%$ level that *every* confidence interval contains the true difference. If simultaneous hypothesis tests are conducted for all null hypotheses of the form $H_0 \colon \mu_i - \mu_j = 0$, then we may reject, at level α, every null hypothesis whose P-value is less than α.

3.2.3 The Bonferroni Method of Multiple Comparisons

The Bonferroni method is a general method, valid anytime that several confidence intervals or tests are considered simultaneously. The method is simple to apply. Let C be the number of pairs of differences to be compared. For example, if there are I treatments, and all pairs of differences are to be compared, then $C = I(I - 1)/2$. The Bonferroni method is the same as the LSD method, except that α is replaced with α/C.

The Bonferroni Method for Simultaneous Confidence Intervals and Hypothesis Tests

Assume that C differences of the form $\mu_i - \mu_j$ are to be considered. The Bonferroni simultaneous confidence intervals, at level $100(1-\alpha)\%$, for the C differences $\mu_i - \mu_j$ are

$$\overline{X}_{i.} - \overline{X}_{j.} \pm t_{N-I,\,\alpha/(2C)} \sqrt{MSE\left(\frac{1}{J_i} + \frac{1}{J_j}\right)} \tag{3.34}$$

We are $100(1-\alpha)\%$ confident that the Bonferroni confidence intervals contain the true value of the difference $\mu_i - \mu_j$ for all C pairs under consideration.

To test C null hypotheses of the form $H_0 \colon \mu_i - \mu_j = 0$, the test statistics are

$$\frac{\overline{X}_{i.} - \overline{X}_{j.}}{\sqrt{MSE\left(\frac{1}{J_i} + \frac{1}{J_j}\right)}}$$

To find the P-value for each test, consult the Student's t table with $N - I$ degrees of freedom, and multiply the P-value found there by C.

Specifically, if

$$|\overline{X}_{i.} - \overline{X}_{j.}| > t_{N-I,\,\alpha/(2C)} \sqrt{MSE\left(\frac{1}{J_i} + \frac{1}{J_j}\right)}$$

then H_0 is rejected at level α.

Example 3.12

For the weld data discussed in Example 3.11, use the Bonferroni method to determine which pairs of fluxes, if any, can be concluded, at the 5% level, to differ in their effect on hardness.

Solution: There are $I = 4$ levels, with $J = 5$ observations at each level, for a total of $N = 20$ observations in all. With four levels, there are a total of $C = (4)(3)/2 = 6$ pairs of means to compare.

To test at the $\alpha = 5\%$ level, we compute $\alpha/(2C) = 0.004167$. The critical t value is $t_{16,\,.004167}$. This value is not in the table; it is between $t_{16,\,.005} = 2.921$ and $t_{16,\,.001} = 3.686$. Using computer software, we calculated $t_{16,\,.004167} = 3.0083$. Without software, one could roughly approximate this value by interpolation. Now $MSE = 63.975$ (see Example 3.11), so $t_{N-I,\,\alpha/(2C)} \sqrt{MSE(1/J_i + 1/J_j)} = 3.0083\sqrt{63.975(1/5 + 1/5)} = 15.22$. The four sample means are

Flux	A	B	C	D
Mean Hardness	253.8	263.2	271.0	262.0

There is only one pair of sample means, 271.0 and 253.8, whose difference is greater than 15.22. We therefore conclude that welds produced with flux A have different mean hardness than welds produced with flux C. None of the other differences are significant at the 5% level. ∎

Below is R output presenting P-values for testing the null hypothesis of equality for each difference between treatment means in the weld experiment. The first set of P-values use Fisher's LSD method, which involves no adjustment of P-values, and the second uses the Bonferroni adjustment.

```
> pairwise.t.test(weld$Hardness, weld$Flux, p.adj = "none")

Pairwise comparisons using t tests with pooled SD

data:  weld$Hardness and weld$Flux

  A      B      C
B 0.0816 -      -
C 0.0037 0.1426 -
D 0.1246 0.8155 0.0942

P value adjustment method: none

> pairwise.t.test(weld$Hardness, weld$Flux, p.adj = "bonferroni")

Pairwise comparisons using t tests with pooled SD

data:  weld$Hardness and weld$Flux

  A     B     C
B 0.490 -     -
C 0.022 0.856 -
D 0.747 1.000 0.565

P value adjustment method: bonferroni
```

Although easy to use, the Bonferroni method has the disadvantage that as the number of tests becomes large, the confidence intervals become very wide, and the hypothesis tests have low power. The reason for this is that the Bonferroni method is a general method, not specifically designed for analysis of variance or for normal populations. In many cases the number of tests is fairly large, in particular it is often desired to compare all pairs of means. In these cases, a method called the **Tukey-Kramer method** is superior, because it is designed for multiple comparisons of means of normal populations. We now describe this method.

3.2.4 The Tukey-Kramer Method of Multiple Comparisons

The Tukey-Kramer method is based on a distribution called the Studentized Range distribution, rather than on the Student's t distribution. The Studentized Range distribution has two values for degrees of freedom, which for the Tukey-Kramer method are I and $N - I$. (In comparison, the F test uses $I - 1$ and $N - I$ degrees of freedom.) The Tukey-Kramer method uses the $1 - \alpha$ quantile of the Studentized Range distribution with I and $N - I$ degrees of freedom; this quantity is denoted $q_{I,N-I,\alpha}$. Table 4 on page 200 presents values of $q_{I,N-I,\alpha}$ for various values of I, N, and α. The mechanics of the Tukey-Kramer method are the same as those for the LSD method, except that $t_{N-I,\,\alpha/2}\sqrt{MSE(1/J_i + 1/J_j)}$ is replaced with $q_{I,N-I,\alpha}\sqrt{(MSE/2)(1/J_i + 1/J_j)}$. The quantity $q_{I,N-I,\alpha}\sqrt{(MSE/2)(1/J_i + 1/J_j)}$ is sometimes called the **honestly significant difference** (HSD), in contrast to Fisher's least significant difference.

The Tukey-Kramer Method for Simultaneous Confidence Intervals and Hypothesis Tests

The Tukey-Kramer level $100(1-\alpha)\%$ simultaneous confidence intervals for all differences $\mu_i - \mu_j$ are

$$\overline{X}_{i.} - \overline{X}_{j.} \pm q_{I,N-I,\alpha}\sqrt{\frac{MSE}{2}\left(\frac{1}{J_i}+\frac{1}{J_j}\right)} \tag{3.35}$$

We are $100(1-\alpha)\%$ confident that the Tukey-Kramer confidence intervals contain the true value of the difference $\mu_i - \mu_j$ for every i and j.

To test all null hypotheses $H_0\colon \mu_i - \mu_j = 0$ simultaneously, the test statistics are

$$\frac{\overline{X}_{i.} - \overline{X}_{j.}}{\sqrt{\frac{MSE}{2}\left(\frac{1}{J_i}+\frac{1}{J_j}\right)}}$$

The P-value for each test is found by consulting the Studentized Range table (Table 4 on page 200) with I and $N-I$ degrees of freedom.

For every pair of levels i and j for which

$$|\overline{X}_{i.} - \overline{X}_{j.}| > q_{I,N-I,\alpha}\sqrt{\frac{MSE}{2}\left(\frac{1}{J_i}+\frac{1}{J_j}\right)}$$

the null hypothesis $H_0\colon \mu_i - \mu_j = 0$ is rejected at level α.

A note on terminology: When the design is balanced, with all sample sizes equal to J, the quantity $\sqrt{(MSE/2)(1/J_i + 1/J_j)}$ is equal to $\sqrt{MSE/J}$ for all pairs of levels. In this case, the method is often simply called Tukey's method.

Example 3.13

For the weld data in Table 3.1, which pairs of fluxes, if any, can be concluded, at the 5% level, to differ in their effect on hardness?

Solution: There are $I = 4$ levels, with $J = 5$ observations at each level, for a total of $N = 20$ observations in all. To test at level $\alpha = 0.05$, we consult the Studentized Range table on page 200 to find $q_{4,16,.05} = 4.05$.

The value of MSE is 63.975 (see Example 3.11). Therefore $q_{I,N-I,\alpha}\sqrt{MSE/J} = 4.05\sqrt{63.975/5} = 14.49$. The four sample means are

Flux	A	B	C	D
Mean Hardness	253.8	263.2	271.0	262.0

There is only one pair of sample means, 271.0 and 253.8, whose difference is greater than 14.49. We therefore conclude that welds produced with flux A have different mean hardness than welds produced with flux C. None of the other differences are significant at the 5% level. ∎

Comparing the results of Example 3.13 with those of Example 3.12 shows that in this case the Tukey-Kramer is slightly more powerful than the Bonferroni method, since its critical value is only 14.49 while that of the Bonferroni method was 15.22. When all possible pairs are compared, as in this example, the Tukey-Kramer method is always more powerful than the Bonferroni method. When only a few of the possible pairs are to be compared, the Bonferroni method is sometimes more powerful.

Sometimes only a single test is performed, but the difference that is tested is chosen by examining the sample means and choosing two whose difference is large. In these cases a multiple comparisons method should be used, even though only one test is being performed. The following example illustrates the idea.

Example 3.14

An engineer examines the weld data in Table 3.1, and notices that the two treatments with the largest difference in sample means are flux A and flux C. He decides to test the null hypothesis that the mean hardness for welds produced with flux A differs from that for welds produced with flux C. Since he will only perform one test, he uses the Fisher LSD method rather than the Bonferroni or Tukey-Kramer method. Explain why this is wrong.

Solution: The engineer has examined every pair of means and has chosen the two whose difference is largest. Although he is formally performing only one test, he has chosen that test by comparing every pair of sample means. For this reason he should use a multiple comparisons procedure, such as the Bonferroni or Tukey-Kramer method. ∎

Below is R output presenting the Tukey-Kramer 95% simultaneous confidence intervals for the weld data. **Note: The TukeyHSD command may not work for unbalanced designs**.

```
> TukeyHSD(aov(weld.lm))
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Hardness ~ Flux, data = weld)

$Flux
    diff        lwr       upr     p adj
B-A  9.4  -5.072915 23.872915 0.2839920
C-A 17.2   2.727085 31.672915 0.0172933
D-A  8.2  -6.272915 22.672915 0.3953011
C-B  7.8  -6.672915 22.272915 0.4372295
D-B -1.2 -15.672915 13.272915 0.9951084
D-C -9.0 -23.472915  5.472915 0.3185074
```

We are 95% confident that every one of these confidence intervals contains the true difference in treatment means. The rightmost column, labeled "p adj," presents the P-value for the null hypothesis that the treatment means are equal.

3.3 Two-Factor Experiments

3.3.1 Basic Ideas

In one-factor experiments, discussed in the previous two sections, the purpose is to determine whether varying the level of a single factor affects the response. Many experiments involve varying several factors, each of which may affect the response. In this section, we will discuss the case in which there are two factors. The experiments, naturally enough, are called **two-factor experiments**. We illustrate with an example.

A chemical engineer is studying the effects of various reagents and catalysts on the yield of a certain process. Four runs of the process were made for each combination of three reagents and four catalysts. The results are presented in Table 3.2. In this experiment there are two factors, the catalyst and the reagent. The catalyst is called the **row factor**, since its value varies from row to row in the table, while the reagent is called the **column factor**. These designations are arbitrary, in that the table could just as easily have been presented with the rows representing the reagents and the columns representing the catalysts.

Table 3.2: Yields for Runs of a Chemical Process with Various Combinations of Reagent and Catalyst. Units are Percent of a Theoretical Maximum

		Reagent		
		1	2	3
Catalyst	A	86.8, 82.4, 86.7, 83.5	93.4, 85.2, 94.8, 83.1	77.9, 89.6, 89.9, 83.7
	B	71.9, 72.1, 80.0, 77.4	74.5, 87.1, 71.9, 84.1	87.5, 82.7, 78.3, 90.1
	C	65.5, 72.4, 76.6, 66.7	66.7, 77.1, 76.7, 86.1	72.7, 77.8, 83.5, 78.8
	D	63.9, 70.4, 77.2, 81.2	73.7, 81.6, 84.2, 84.9	79.8, 75.7, 80.5, 72.9

In general, there are I levels of the row factor and J levels of the column factor. (In Table 3.2, $I = 4$ and $J = 3$.) There are therefore IJ different combinations of the two factors. The terminology for these factor combinations is not standardized. We will refer to each combination of factors as a **treatment**, but some authors use the term **treatment combination**. Recall that the units assigned to a given treatment are called replicates. When the number of replicates is the same for each treatment, we will denote this number by K. Thus in Table 3.2, $K = 4$.

When observations are taken on every possible treatment, the design is called a **complete design** or a **full factorial design**. Incomplete designs, in which there are no data for one or more treatments, can be difficult to interpret, except for some special cases. When possible, complete designs should be used. When the number of replicates is the same for each treatment, the design is said to be **balanced**. For one-factor experiments, we did not need to assume that the design was balanced. With two-factor experiments, unbalanced designs are much more difficult to analyze than balanced designs. We will restrict our discussion to balanced designs. As with one-factor experiments, the factors may be fixed or random. The methods that we will describe apply to models where both effects are fixed. Later we will briefly describe models where one or both factors are random.

In a completely randomized design, each treatment represents a population, and the observations on that treatment are a simple random sample from that population. We will denote the sample values for the treatment corresponding to the ith level of the row factor and the jth level of the column factor by $X_{ij1}, ..., X_{ijK}$. We will denote the population mean outcome for this treatment by μ_{ij}. The values μ_{ij} are often called the **treatment means**. In general, the purpose of a two-factor experiment is to determine whether the treatment means are affected by varying either the row factor, the column factor, or both. The method of analysis appropriate for two-factor experiments is called **two-way analysis of variance** (ANOVA).

3.3.2 Parameterization for Two-Way Analysis of Variance

In a two-way analysis of variance, we wish to determine whether varying the level of the row or column factors changes the value of μ_{ij}. To do this, we must express μ_{ij} in terms of parameters that describe the row and column factors separately. We'll begin this task by describing some notation for the averages of the treatment means for the different levels of the row and column factors.

For any level i of the row factor, the average of all the treatment means μ_{ij} in the ith row is denoted $\overline{\mu}_{i.}$. We express $\overline{\mu}_{i.}$ in terms of the treatment means as follows:

$$\overline{\mu}_{i.} = \frac{1}{J}\sum_{j=1}^{J}\mu_{ij} \tag{3.36}$$

Similarly, for any level j of the column factor, the average of all the treatment means μ_{ij} in the jth column is denoted $\overline{\mu}_{.j}$. We express $\overline{\mu}_{.j}$ in terms of the treatment means as follows:

$$\overline{\mu}_{.j} = \frac{1}{I}\sum_{i=1}^{I}\mu_{ij} \tag{3.37}$$

Finally, we define the **population grand mean**, denoted by μ, which represents the average of all the treatment means μ_{ij}. The population grand mean can also be expressed as the average of the quantities $\overline{\mu}_{i.}$ or of the quantities $\overline{\mu}_{.j}$:

$$\mu = \frac{1}{I}\sum_{i=1}^{I}\overline{\mu}_{i.} = \frac{1}{J}\sum_{j=1}^{J}\overline{\mu}_{.j} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\mu_{ij} \tag{3.38}$$

Table 3.3 illustrates the relationships among μ_{ij}, $\overline{\mu}_{i.}$, $\overline{\mu}_{.j}$, and μ.

Table 3.3: Treatment Means and Their Averages Across Rows and Down Columns.

	Column Level				
Row Level	1	2	$\cdots$	J	Row Mean
1	μ_{11}	μ_{12}	$\cdots$	μ_{1J}	$\overline{\mu}_{1.}$
2	μ_{21}	μ_{22}	$\cdots$	μ_{2J}	$\overline{\mu}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
I	μ_{I1}	μ_{I2}	$\cdots$	μ_{IJ}	$\overline{\mu}_{I.}$
Column Mean	$\overline{\mu}_{.1}$	$\overline{\mu}_{.2}$	$\cdots$	$\overline{\mu}_{.J}$	μ

Using the quantities $\overline{\mu}_{i.}$, $\overline{\mu}_{.j}$, and μ, we can decompose the treatment mean μ_{ij} as follows:

$$\mu_{ij} = \mu + (\overline{\mu}_{i.} - \mu) + (\overline{\mu}_{.j} - \mu) + (\mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \mu) \tag{3.39}$$

Equation (3.39) expresses the treatment mean μ_{ij} as a sum of four terms. In practice, simpler notation is used for the three rightmost terms in Equation (3.39):

$$\alpha_i = \overline{\mu}_{i.} - \mu \tag{3.40}$$

$$\beta_j = \overline{\mu}_{.j} - \mu \tag{3.41}$$

$$\gamma_{ij} = \mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \mu \tag{3.42}$$

Each of quantities μ, α_i, β_j, and γ_{ij} has an important interpretation:

- The quantity μ is the population grand mean, which is average of all the treatment means.
- The quantity $\alpha_i = \overline{\mu}_{i.} - \mu$ is called the ith **row effect**. It is the difference between the average treatment mean for the ith level of the row factor and the population grand mean. The value of α_i indicates the degree to which the ith level of the row factor tends to produce outcomes that are larger or smaller than the population grand mean.
- The quantity $\beta_j = \overline{\mu}_{.j} - \mu$ is called the jth **column effect**. It is the difference between the average treatment mean for the jth level of the column factor and the population grand mean. The value of β_j indicates the degree to which the jth level of the column factor tends to produce outcomes that are larger or smaller than the population grand mean.
- The quantity $\gamma_{ij} = \mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \mu$ is called the ij **interaction**. The effect of a level of the row (or column) factor may depend on which level of the column (or row) factor it is paired with. The interaction terms measure the degree to which this occurs. For example, assume that level 1 of the row factor tends to produce a large outcome when paired with column level 1, but a small outcome when paired with column level 2. In this case $\gamma_{1,1}$ would be positive, and $\gamma_{1,2}$ would be negative.

Both row effects and column effects are called **main effects**, to distinguish them from the interactions. Note that there are I row effects, one for each level of the row factor, J column effects, one for each level of the column factor, and IJ interactions, one for each treatment. Furthermore, it follows from the definitions of quantities $\overline{\mu}_{i.}$, $\overline{\mu}_{.j}$, and μ in Equations (3.36) – (3.38) that the row effects, column effects, and interactions satisfy the following constraints:

$$\sum_{i=1}^{I} \alpha_i = 0 \qquad \sum_{j=1}^{J} \beta_j = 0 \qquad \sum_{i=1}^{I} \gamma_{ij} = \sum_{j=1}^{J} \gamma_{ij} = 0 \tag{3.43}$$

We now can express the treatment means μ_{ij} in terms of α_i, β_j, and γ_{ij}. From Equation (3.39) it follows that

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \tag{3.44}$$

For each observation X_{ijk}, define $\varepsilon_{ijk} = X_{ijk} - \mu_{ij}$, the difference between the observation and its treatment mean. The quantities ε_{ijk} are called **errors**. It follows that

$$X_{ijk} = \mu_{ij} + \varepsilon_{ijk} \tag{3.45}$$

Combining Equation (3.45) with Equation (3.44) yields the two-way ANOVA model:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \tag{3.46}$$

When the interactions γ_{ij} are all equal to 0, the **additive model** is said to apply. Under the additive model, Equation (3.44) becomes

$$\mu_{ij} = \mu + \alpha_i + \beta_j \tag{3.47}$$

and Equation (3.46) becomes

$$X_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk} \tag{3.48}$$

Under the additive model, the treatment mean μ_{ij} is equal to the population grand mean μ, plus an amount α_i that results from using row level i plus an amount β_j that results from using column level j. In other words, the combined effect of using row level i along with column level j is found by adding the individual

main effects of the two levels. When some or all of the interactions are not equal to 0, the additive model does not hold, and the combined effect of a row level and a column level cannot be determined from their individual main effects.

We will now show how to estimate the parameters for the full two-way model (3.46). The procedure for the additive model is exactly the same, except that the interactions γ_{ij} are not estimated. The procedure is straightforward. We first define some notation for various averages of the data X_{ijk}, using the data in Table 3.2 as an example. Table 3.4 presents the average yield for the four runs for each reagent and catalyst in Table 3.2.

Table 3.4: Average Yields $\overline{X}_{ij.}$ for Runs of a Chemical Process Using Different Combinations of Reagent and Catalyst

		Reagent			
		1	2	3	Row Mean $\overline{X}_{i..}$
Catalyst	A	84.85	89.13	85.28	86.42
	B	75.35	79.40	84.65	79.80
	C	70.30	76.65	78.20	75.05
	D	73.18	81.10	77.23	77.17
	Column Mean $\overline{X}_{.j.}$	75.92	81.57	81.34	Sample Grand Mean $\overline{X}_{...}$ 79.61

Each number in the body of Table 3.4 is the average of the four numbers in the corresponding cell of Table 3.2. These are called the **cell means**. They are denoted $\overline{X}_{ij.}$ and are defined by

$$\overline{X}_{ij.} = \frac{1}{K}\sum_{k=1}^{K} X_{ijk} \tag{3.49}$$

Averaging the cell means across the rows produces the **row means** $\overline{X}_{i..}$

$$\overline{X}_{i..} = \frac{1}{J}\sum_{j=1}^{J} \overline{X}_{ij.} = \frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K} X_{ijk} \tag{3.50}$$

Averaging the cell means down the columns produces the **column means** $\overline{X}_{.j.}$

$$\overline{X}_{.j.} = \frac{1}{I}\sum_{i=1}^{I} \overline{X}_{ij.} = \frac{1}{IK}\sum_{i=1}^{I}\sum_{k=1}^{K} X_{ijk} \tag{3.51}$$

The sample grand mean $\overline{X}_{...}$ can be found by computing the average of the row means, the average of the column means, the average of the cell means, or the average of all the observations:

$$\overline{X}_{...} = \frac{1}{I}\sum_{i=1}^{I} \overline{X}_{i..} = \frac{1}{J}\sum_{j=1}^{J} \overline{X}_{.j.} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} \overline{X}_{ij.} = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} X_{ijk} \tag{3.52}$$

Now we describe how to estimate the parameters in the two-way ANOVA model. The fundamental idea is that the best estimate of the treatment mean μ_{ij} is the cell mean $\overline{X}_{ij.}$, which is the average of the sample observations having that treatment. It follows that the best estimate of the quantity $\overline{\mu}_{i.}$ is the row mean $\overline{X}_{i..}$, the best estimate of the quantity $\overline{\mu}_{.j}$ is the column mean $\overline{X}_{.j.}$, and the best estimate of the population grand mean μ is the sample grand mean $\overline{X}_{...}$. We estimate the row effects α_i, the column effects β_j, and the interactions γ_{ij} by substituting these estimates into Equations (3.40)–(3.42).

$$\widehat{\alpha}_i = \overline{X}_{i..} - \overline{X}_{...} \tag{3.53}$$

$$\widehat{\beta}_j = \overline{X}_{.j.} - \overline{X}_{...} \tag{3.54}$$

$$\widehat{\gamma}_{ij} = \overline{X}_{ij.} - \overline{X}_{i..} - \overline{X}_{.j.} + \overline{X}_{...} \tag{3.55}$$

The row effects, column effects, and interactions satisfy constraints given in Equation (3.43). By performing some algebra, it can be shown that their estimates satisfy the same constraints:

$$\sum_{i=1}^{I} \widehat{\alpha}_i = 0 \quad \sum_{j=1}^{J} \widehat{\beta}_j = 0 \quad \sum_{i=1}^{I} \widehat{\gamma}_{ij} = \sum_{j=1}^{J} \widehat{\gamma}_{ij} = 0 \tag{3.56}$$

Example 3.15

Compute the estimated row effects, column effects, and interactions for the data in Table 3.2.

Solution: Using the quantities in Table 3.4 and Equations (3.53)–(3.55), we compute

$$\widehat{\alpha}_1 = 86.42 - 79.61 = 6.81 \qquad \widehat{\alpha}_2 = 79.80 - 79.61 = 0.19$$

$$\widehat{\alpha}_3 = 75.05 - 79.61 = -4.56 \qquad \widehat{\alpha}_4 = 77.17 - 79.61 = -2.44$$

$$\widehat{\beta}_1 = 75.92 - 79.61 = -3.69 \quad \widehat{\beta}_2 = 81.57 - 79.61 = 1.96 \quad \widehat{\beta}_3 = 81.34 - 79.61 = 1.73$$

$$\begin{array}{rcrrcrrcr} \gamma_{11} & = & 2.12 & \gamma_{12} & = & 0.75 & \gamma_{13} & = & -2.87 \\ \gamma_{21} & = & -0.76 & \gamma_{22} & = & -2.36 & \gamma_{23} & = & 3.12 \\ \gamma_{31} & = & -1.06 & \gamma_{32} & = & -0.36 & \gamma_{33} & = & 1.42 \\ \gamma_{41} & = & -0.30 & \gamma_{42} & = & 1.97 & \gamma_{43} & = & -1.67 \end{array}$$

3.3.3 Using Two-Way ANOVA to Test Hypotheses

A two-way analysis of variance is designed to address three main questions:

1. Does the additive model hold?
2. If so, is the mean outcome the same for all levels of the row factor?
3. If so, is the mean outcome the same for all levels of the column factor?

In general, we ask questions 2 and 3 only when we believe that the additive model may hold. We will discuss this further below. The three questions are addressed by performing hypothesis tests. The null hypotheses for these tests are as follows:

To test whether the additive model holds, we test the null hypothesis that all the interactions are equal to 0:

$$H_0 : \gamma_{11} = \gamma_{12} = \cdots = \gamma_{IJ} = 0$$

If this null hypothesis is true, the additive model holds.

To test whether the mean outcome is the same for all levels of the row factor, we test the null hypothesis that all the row effects are equal to 0:

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

If this null hypothesis is true, then the mean outcome is the same for all levels of the row factor.

To test whether the mean outcome is the same for all levels of the column factor, we test the null hypothesis that all the column effects are equal to 0:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_J = 0$$

If this null hypothesis is true, then the mean outcome is the same for all levels of the column factor.

We now describe the standard tests for these null hypotheses. For the tests to be valid, the following conditions must hold:

Assumptions for Two-Way ANOVA

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

1. The design must be complete.
2. The design must be balanced.
3. The number of replicates per treatment, K, must be at least 2.
4. Within any treatment, the observations $X_{ij1}, ..., X_{ijK}$ are a simple random sample from a normal population.
5. The population variance is the same for all treatments. We denote this variance by σ^2.

Just as in one-way ANOVA, the standard tests for these null hypotheses are based on sums of squares. Specifically, they are the row sum of squares (SSA), the column sum of squares (SSB), the interaction sum of squares ($SSAB$), and the error sum of squares (SSE). Also of interest is the total sum of squares (SST), which is equal to the sum of the others. Formulas for these sums of squares are

$$SSA = JK\sum_{i=1}^{I} \widehat{\alpha}_i^2 = JK\sum_{i=1}^{I} (\overline{X}_{i..} - \overline{X}_{...})^2 = JK\sum_{i=1}^{I} \overline{X}_{i..}{}^2 - IJK\overline{X}_{...}{}^2 \quad (3.57)$$

$$SSB = IK\sum_{j=1}^{J} \widehat{\beta}_j^2 = IK\sum_{j=1}^{J} (\overline{X}_{.j.} - \overline{X}_{...})^2 = IK\sum_{j=1}^{J} \overline{X}_{.j.}{}^2 - IJK\overline{X}_{...}{}^2 \quad (3.58)$$

$$SSAB = K\sum_{i=1}^{I}\sum_{j=1}^{J} \widehat{\gamma}_{ij}^2 = K\sum_{i=1}^{I}\sum_{j=1}^{J} (\overline{X}_{ij.} - \overline{X}_{i..} - \overline{X}_{.j.} + \overline{X}_{...})^2$$

$$= K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{X}_{ij.}{}^2 - JK\sum_{i=1}^{I}\overline{X}_{i..}{}^2 - IK\sum_{j=1}^{J}\overline{X}_{.j.}{}^2 + IJK\overline{X}_{...}{}^2 \tag{3.59}$$

$$SSE = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(X_{ijk} - \overline{X}_{ij.})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk}^2 - K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{X}_{ij.}{}^2 \tag{3.60}$$

$$SST = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(X_{ijk} - \overline{X}_{...})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk}^2 - IJK\overline{X}_{...}{}^2 \tag{3.61}$$

It can be seen from the rightmost expressions in Equations (3.57) through (3.61) that the total sum of squares, SST is equal to the sum of the others. This is the Analysis of Variance identity for two-way ANOVA.

The Analysis of Variance Identity

$$SST = SSA + SSB + SSAB + SSE \tag{3.62}$$

Along with each sum of squares is a quantity known as its degrees of freedom. The sums of squares and their degrees of freedom are generally presented in an ANOVA table. Table 3.5 presents the degrees of freedom for each sum of squares, along with the computationally most convenient formula.

Table 3.5: ANOVA Table for Two-Way ANOVA

Source	Degrees of Freedom	Sum of Squares
Rows (SSA)	$I-1$	$JK\sum_{i=1}^{I}\widehat{\alpha}_i^2 = JK\sum_{i=1}^{I}\overline{X}_{i..}{}^2 - IJK\overline{X}_{...}{}^2$
Columns (SSB)	$J-1$	$IK\sum_{j=1}^{J}\widehat{\beta}_j^2 = IK\sum_{j=1}^{J}\overline{X}_{.j.}{}^2 - IJK\overline{X}_{...}{}^2$
Interactions ($SSAB$)	$(I-1)(J-1)$	$K\sum_{i=1}^{I}\sum_{j=1}^{J}\widehat{\gamma}_{ij}^2$ $= K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{X}_{ij.}{}^2 - JK\sum_{i=1}^{I}\overline{X}_{i..}{}^2 - IK\sum_{j=1}^{J}\overline{X}_{.j.}{}^2 + IJK\overline{X}_{...}{}^2$
Error (SSE)	$IJ(K-1)$	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(X_{ijk} - \overline{X}_{ij.})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk}^2 - K\sum_{i=1}^{I}\sum_{j=1}^{J}\overline{X}_{ij.}{}^2$
Total (SST)	$IJK-1$	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(X_{ijk} - \overline{X}_{...})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk}^2 - IJK\overline{X}_{...}{}^2$

We point out that the degrees of freedom for SST is the sum of the degrees of freedom for the other sums of squares.

Note that the magnitude of SSA depends on the magnitude of the *estimated* row effects $\widehat{\alpha}_i$. Therefore when the *true* row effects α_i are equal to 0, SSA will tend to be smaller, and when some of the true row effects are not equal to 0, SSA will tend to be larger. We will therefore reject $H_0\colon \alpha_1 = \cdots = \alpha_I = 0$ when SSA is sufficiently large. Similarly, SSB will tend to be smaller when the true column effects β_j are all equal to 0 and larger when some column effects are not zero, and $SSAB$ will tend to be smaller when the true interactions γ_{ij} are all equal to 0 and larger when some interactions are not zero. We will therefore reject $H_0\colon \beta_1 = \cdots = \beta_J = 0$ when SSB is sufficiently large, and we will reject $H_0\colon \gamma_{11} = \cdots \gamma_{IJ} = 0$ when $SSAB$ is sufficiently large.

We can determine whether SSA, SSB, and $SSAB$ are sufficiently large by comparing them to the error sum of squares, SSE. As in one-way ANOVA (Section 3.1.1), SSE depends only on the distances between the observations and their own cell means. SSE therefore measures only the random variation inherent in the process, and is not affected by the values of the row effects, column effects, or interactions. To compare SSA, SSB, and $SSAB$ with SSE, we first divide each sum of squares by its degrees of freedom, producing quantities known as **mean squares**. The mean squares, denoted MSA, MSB, $MSAB$, and MSE, are defined as follows:

$$MSA = \frac{SSA}{I-1} \qquad MSB = \frac{SSB}{J-1} \qquad MSAB = \frac{SSAB}{(I-1)(J-1)} \qquad MSE = \frac{SSE}{IJ(K-1)} \tag{3.63}$$

The test statistics for the three null hypotheses are the quotients of MSA, MSB, and $MSAB$ with MSE. The null distributions of these test statistics are F distributions. Specifically,

- Under $H_0\colon \alpha_1 = \cdots = \alpha_I = 0$, the statistic $\dfrac{MSA}{MSE}$ has an $F_{I-1,\,IJ(K-1)}$ distribution.
- Under $H_0\colon \beta_1 = \cdots = \beta_J = 0$, the statistic $\dfrac{MSB}{MSE}$ has an $F_{J-1,\,IJ(K-1)}$ distribution.
- Under $H_0\colon \gamma_{11} = \cdots = \gamma_{IJ} = 0$, the statistic $\dfrac{MSAB}{MSE}$ has an $F_{(I-1)(J-1),\,IJ(K-1)}$ distribution.

In practice, the sums of squares, mean squares, and test statistics are usually calculated with the use of a computer. Below is R output showing the ANOVA table for the data in Table 3.2.

```
> chem.lm = lm(Yield ~ Catalyst*Reagent, data=chem)
> chem.aov = anova(Yield ~ Catalyst*Reagent, data=chem)
> chem.aov
                 Df    Sum Sq   Mean Sq  F value     Pr(>F)
Catalyst          3    877.56   292.521   9.3579  0.0001040
Reagent           2    327.14   163.570   5.2327  0.0101183
Catalyst:Reagent  6    156.98    26.164   0.8370  0.5496030
Residuals        36   1125.33    31.259
```

The labels Df, Sum Sq, Mean Sq, F value, and Pr(>F) refer to degrees of freedom, sum of squares, mean square, F statistic, and P-value, respectively. As in one-way ANOVA, the mean square for error (MSE) is an estimate of the error variance σ^2. This value is 31.259.

Example 3.16

Use the ANOVA table above to determine whether the additive model is plausible for the yield data. If the additive model is plausible, can we conclude that either the catalyst or the reagent affects the yield?

Solution: We first check to see if the additive model is plausible. The P-value for the interactions is 0.55, which is not small. We therefore do not reject the null hypothesis that all the interactions are equal to 0, and we conclude that the additive model is plausible. Since the additive model is plausible, we now ask whether the row or column factors affect the outcome. We see from the table that the P-value for the row effects (Catalyst) is approximately 0, so we conclude that the catalyst does affect the yield. Similarly, the P-value for the column effects (Reagent) is small (0.010), so we conclude that the reagent affects the yield as well. ∎

Example 3.17

The article "Uncertainty in Measurements of Dermal Absorption of Pesticides" (W. Navidi and A. Bunge, *Risk Analysis*, 2002:1175–1182) describes an experiment in which a pesticide was applied to skin at various concentrations and for various lengths of time. The outcome is the amount of the pesticide that was absorbed into the skin. The R output below presents the ANOVA table. Is the additive model plausible? If so, do either the concentration or the duration affect the amount absorbed?

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Concent	2	49.991	24.996	107.99	0.000
Duration	2	19.157	9.579	41.38	0.000
Concent:Duration	4	0.337	0.084	0.36	0.832
Residuals	27	6.250	0.231		

Solution: The P-value for the interaction is 0.832, so we conclude that the additive model is plausible. The P-values for both concentration and dose are very small. Therefore we can conclude that both concentration and duration affect the amount absorbed. ∎

3.3.4 Checking the Assumptions

A residual plot can be used to check the assumption of equal variances. The residual plot plots the residuals $X_{ijk} - \overline{X}_{ij.}$ vs. the fitted values, which are the sample means $\overline{X}_{ij.}$. Figure 3.5 presents a residual plot for the yield data found in Table 3.2 on page 153. The assumptions appear to be well satisfied.

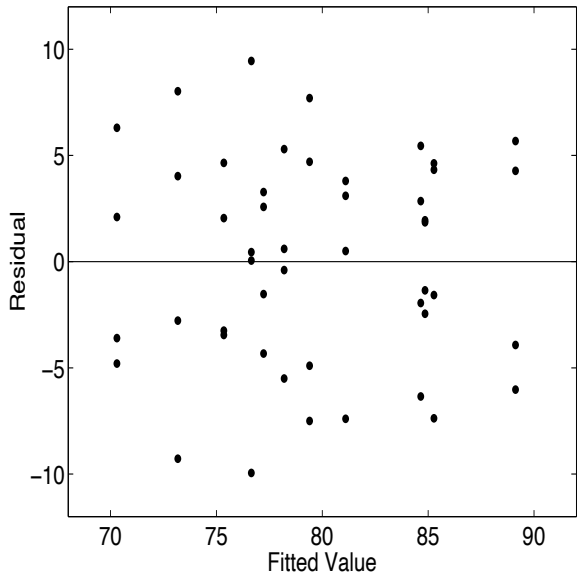


Figure 3.5: Residual plot for the yield data. There is no evidence against the assumption of equal variances.

3.3.5 Don't Interpret the Main Effects When the Additive Model Doesn't Hold

When the interactions are small enough so that the additive model is plausible, interpretation of the main effects is fairly straightforward, as shown in Examples 3.16 and 3.17. When the additive model does not hold, however, it is not always easy to interpret the main effects. Here is a hypothetical example to illustrate the point. Assume that a process is run under conditions obtained by varying two factors at two levels each. Two runs are made at each of the four combinations or row and column levels. The yield of the process is measured each time, with the results presented below:

	Column Level	
	1	2
Row Level 1	51, 49	43, 41
Row Level 2	43, 41	51, 49

Clearly, if it is desired to maximize yield, the row and column factors matter — we want either row level 1 paired with column level 1 or row level 2 paired with column level 2.

Now look at the ANOVA table, presented below:

```
             Df     Sum Sq    Mean Sq  F value   Pr(>F)
Row           1     0.0000     0.0000     0.00    1.000
Column        1     0.0000     0.0000     0.00    1.000
Row:Column    1     128.00     128.00    64.00    0.001
Residuals     4     8.0000     2.0000
```

The main effects sum of squares for both row and column main effects are equal to 0, and their P-values are equal to 1, which is as large as a P-value can be. If we follow the procedure used in Examples 3.16 and 3.17, we would conclude that neither the row factor nor the column factor affects the yield. But it is clear from the data that the row and column factors do affect the yield. What is happening is that the row and column factors do not matter *on the average*. Level 1 of the row factor is better if level 1 of the column factor is used, and level 2 of the row factor is better if level 2 of the column factor is used. When averaged over the two levels of the column factor, the levels of the row factor have the same mean yield. Similarly, the column levels have the same mean yield when averaged over the levels of the row factor. When the effects of the row levels depend on which column levels they are paired with, and vice versa, the main effects can be misleading.

It is the P-value for the interactions that tells us not to try to interpret the main effects. This P-value is quite small, so we reject the additive model. Then we know that some of the interactions are non-zero, so the effects of the row levels depend on the column levels and vice versa. For this reason, when the additive model is rejected, we should not try to interpret the main effects. We need to look at the cell means themselves in order to determine how various combinations of row and column levels affect the outcome.

Summary

In a two-way analysis of variance:

- If the additive model *is not* rejected, then hypothesis tests for the main effects can be used to determine whether the row or column factors affect the outcome.
- If the additive model *is* rejected, then hypothesis tests for the main effects should not be used. Instead, the cell means must be examined to determine how various combinations of row and column levels affect the outcome.

Example 3.18

The thickness of the silicon dioxide layer on a semiconductor wafer is crucial to its performance. In the article "Virgin Versus Recycled Wafers for Furnace Qualification: Is the Expense Justified?" (V. Czitrom and J. Reece, *Statistical Case Studies for Process Improvement*, SIAM-ASA, 1997:87–103), oxide layer thicknesses were measured for three types of wafers: virgin wafers, wafers recycled in-house, and wafers recycled by an external supplier. In addition, several furnace locations were used to grow the oxide layer. A two-way ANOVA for three runs at one wafer site for the three types of wafers at three furnace locations was performed. The data, followed by R output, are presented below.

Furnace Location	WaferType	Oxide Layer Thickness (Angstroms)
1	Virgin	90.1, 90.7, 89.4
1	In-House	90.4, 88.8, 90.6
1	External	92.6, 90.0, 93.3
2	Virgin	91.9, 88.6, 89.7
2	In-House	90.3, 91.9, 91.5
2	External	88.3, 88.2, 89.4
3	Virgin	88.1, 90.2, 86.6
3	In-House	91.0, 90.4, 90.2
3	External	91.5, 89.8, 89.8

```
> wafer.lm=lm(Thickness~Location*WaferType, data=wafer)
```

```
> wafer.aov=anova(Thickness~Location*WaferType, data=wafer)
> wafer.aov
                     Df  Sum Sq Mean Sq F value  Pr(>F)
Location              2  4.1089  2.0544  1.4460 0.26159
WaferType             2  5.8756  2.9378  2.0678 0.15547
Location:WaferType    4 21.3489  5.3372  3.7566 0.02162
Residuals            18 25.5733  1.4207
```

Since recycled wafers are cheaper, the company hopes that there is no difference in the oxide layer thickness among the three types of chips. If possible, determine whether the data are consistent with the hypothesis of no difference. If not possible, explain why not.

Solution: The P-value for the interactions is 0.022, which is small. Therefore the additive model is not plausible, so we cannot interpret the main effects. A good thing to do is to make a table of the cell means. Table 3.6 presents the sample mean for each treatment.

Table 3.6: Sample Means for Each Treatment

Furnace Location	Wafer Type			
	Virgin	In-House	External	Row Mean
1	90.067	89.933	91.967	90.656
2	90.067	91.233	88.633	89.978
3	88.300	90.533	90.367	89.733
Column Mean	89.478	90.566	90.322	

From Table 3.6, it can be seen that the thicknesses do vary among wafer types, but no one wafer type consistently produces the thickest, or the thinnest, oxide layer. For example, at furnace location 1 the externally recycled wafers produce the thickest layer while the in-house recycled wafers produce the thinnest. At furnace location 2 the order is reversed: the in-house wafers produce the thickest layer while the external ones produce the thinnest. This is due to the interaction of furnace location and wafer type. ∎

3.3.6 Analysis of Variance is Just Linear Regression

Like one-way ANOVA, two-way ANOVA is just a special case of linear regression. We will construct the appropriate regression model for the yield data. The parameters we need in the model are:
Grand Mean: μ
Row effects (Catalyst): α_1, α_2 α_3
Column effects (Reagent): β_1, β_2
Interactions: γ_{11}, γ_{12}, γ_{21}, γ_{22}, γ_{31}, γ_{32}
The remaining parameters are functions of these. Specifically:
$\alpha_4 = -\alpha_1 - \alpha_2 - \alpha_3$
$\beta_3 = -\beta_1 - \beta_2$
$\gamma_{13} = -\gamma_{11} - \gamma_{12}$
$\gamma_{23} = -\gamma_{21} - \gamma_{22}$
$\gamma_{33} = -\gamma_{31} - \gamma_{32}$
$\gamma_{41} = -\gamma_{11} - \gamma_{21} - \gamma_{31}$
$\gamma_{42} = -\gamma_{12} - \gamma_{22} - \gamma_{32}$
$\gamma_{43} = -\gamma_{13} - \gamma_{23} - \gamma_{33} = -(-\gamma_{11} - \gamma_{12}) - (-\gamma_{21} - \gamma_{22}) - (-\gamma_{31} - \gamma_{32}) = \gamma_{11} + \gamma_{12} + \gamma_{21} + \gamma_{22} + \gamma_{31} + \gamma_{32}$

To estimate the parameters with linear regression, we write the X_{ijk} as a column vector. We call this vector $\mathbf{Y}$ in keeping with the usual linear regression notation.
The vector of parameters is $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \gamma_{31}, \gamma_{32})$.

Following are the vector $\mathbf{Y}$, the matrix $\mathbf{X}$, and the parameter vector $\boldsymbol{\beta}$. The columns of $\mathbf{X}$ are labeled with the parameters they represent, for convenience.

$$\mathbf{Y} = \begin{pmatrix} 86.8 \\ 82.4 \\ 86.7 \\ 83.5 \\ 93.4 \\ 85.2 \\ 94.8 \\ 83.1 \\ 77.9 \\ 89.6 \\ 89.9 \\ 83.7 \\ 71.9 \\ 72.1 \\ 80.0 \\ 77.4 \\ 74.5 \\ 87.1 \\ 71.9 \\ 84.1 \\ 87.5 \\ 82.7 \\ 78.3 \\ 90.1 \\ 65.5 \\ 72.4 \\ 76.6 \\ 66.7 \\ 66.7 \\ 77.1 \\ 76.7 \\ 86.1 \\ 72.7 \\ 77.8 \\ 83.5 \\ 78.8 \\ 63.9 \\ 70.4 \\ 77.2 \\ 81.2 \\ 73.7 \\ 81.6 \\ 84.2 \\ 84.9 \\ 79.8 \\ 75.7 \\ 80.5 \\ 72.9 \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} \mu & \alpha_1 & \alpha_2 & \alpha_3 & \beta_1 & \beta_2 & \gamma_{11} & \gamma_{12} & \gamma_{21} & \gamma_{22} & \gamma_{31} & \gamma_{32} \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 1 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 1 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 1 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 \\ 1 & -1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 \\ 1 & -1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 \\ 1 & -1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{31} \\ \gamma_{32} \end{pmatrix}$$

It is easier than you might think to construct the matrix $\mathbf{X}$. For each observation, put a 1 for the appropriate row effect, and a -1 for all three row effects if the observation is in row 4. Then put a 1 for the appropriate column effect, and a -1 for both column effects if the observation is in column 3. Finally, the interactions are found by multiplying the corresponding row and column effects. For example, the entries for γ_{11} are the product of the entries for α_1 and β_1.

Here is the linear regression in R. We name the parameters a1, a2, a3, b1, b2, g11, g12, g21, g22, g31, and g32.

```
> fullmodel = lm(yield~a1+a2+a3+b1+b2+g11+g12+g21+g22+g31+g32)
> summary(fullmodel}
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  79.6083     0.8070  98.648  < 2e-16
a1            6.8083     1.3977   4.871 2.23e-05
a2            0.1917     1.3977   0.137  0.89170
a3           -4.5583     1.3977  -3.261  0.00243
b1           -3.6896     1.1413  -3.233  0.00262
b2            1.9604     1.1413   1.718  0.09443
g11           2.1229     1.9767   1.074  0.28999
g12           0.7479     1.9767   0.378  0.70738
g21          -0.7604     1.9767  -0.385  0.70273
g22          -2.3604     1.9767  -1.194  0.24024
g31          -1.0604     1.9767  -0.536  0.59494
g32          -0.3604     1.9767  -0.182  0.85635
---

Residual standard error: 5.591 on 36 degrees of freedom
Multiple R-squared:  0.5475,Adjusted R-squared:  0.4093
F-statistic:  3.96 on 11 and 36 DF,  p-value: 0.0008251
```

The output above presents the parameter estimates and their standard deviations. Now we will test for significance of interactions.

```
#Fit the additive model
> addmodel = lm(yield~a1+a2+a3+b1+b2)
#Test significance of interactions
> anova(addmodel, fullmodel)
Analysis of Variance Table

Model 1: yield ~ a1 + a2 + a3 + b1 + b2
Model 2: yield ~ a1 + a2 + a3 + b1 + b2 + g11 + g12 + g21 + g22 + g13 + g23
  Res.Df    RSS Df Sum of Sq     F Pr(>F)
1     42 1282.3
2     36 1125.3  6    156.98 0.837 0.5496

#P-value is 0.5496; we may interpret main effects

#Now test whether row effects are significant
> norowmodel = lm(yield~b1+b2+g11+g12+g21+g22+g31+g32)
```

```
> anova(norowmodel, fullmodel)
Analysis of Variance Table

Model 1: yield ~ b1 + b2 + g11 + g12 + g21 + g22 + g13 + g23
Model 2: yield ~ a1 + a2 + a3 + b1 + b2 + g11 + g12 + g21 + g22 + g13 + g23
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     39 2002.9
2     36 1125.3  3    877.56 9.3579 0.000104

#P-value is 0.000104; row effects are significant

#Now test whether row effects are significant
> nocolumnmodel = lm(yield~a1+a2=a3+g11+g12+g21+g22+g31+g32)
> anova(nocolumnmodel, fullmodel)

Model 1: yield ~ a1 + a2 + a3 + g11 + g12 + g21 + g22 + g13 + g23
Model 2: yield ~ a1 + a2 + a3 + b1 + b2 + g11 + g12 + g21 + g22 + g13 + g23
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     38 1452.5
2     36 1125.3  2    327.14 5.2327 0.01012

#P-value is 0.01012; column effects are significant
```

3.3.7 Unbalanced Designs

An unbalanced design is one in which there are unequal numbers of observations in the cells. Unbalanced designs cause great consternation among the less knowledgeable. The formulas for sums of squares are invalid, and many computer package ANOVA programs give incorrect answers. Fortunately, there is a simple solution: **When the design is unbalanced, use linear regression.** This works every time. Following is an example.

Example 3.19

A less skillful chemist has attempted to replicate the yield experiment. Unfortunately, many of the runs did not produce useful results. Here are the results.

Table 3.7: Yields for Runs of a Chemical Process with Various Combinations of Reagent and Catalyst. Units are Percent of a Theoretical Maximum

		Reagent		
		1	2	3
Catalyst	A	86.8, 82.4, 86.7	93.4, 85.2	83.7
	B	71.9, 72.1, 80.0, 77.4	84.1	87.5, 82.7
	C	76.6, 66.7	66.7, 76.7, 86.1	72.7
	D	81.2	73.7, 81.6, 84.2, 84.9	75.7, 80.5, 72.9

Estimate the main effects and interactions. Test for significance of interactions. If not significant, test for significance of row and column effects.

Solution:
Following is the vector of responses $\mathbf{Y}$ and the design matrix $\mathbf{X}$.

$$\mathbf{Y} = \begin{bmatrix} 86.8 \\ 82.4 \\ 86.7 \\ 93.4 \\ 85.2 \\ 83.7 \\ 71.9 \\ 72.1 \\ 80.0 \\ 77.4 \\ 84.1 \\ 87.5 \\ 82.7 \\ 76.6 \\ 66.7 \\ 66.7 \\ 76.7 \\ 86.1 \\ 72.7 \\ 81.2 \\ 73.7 \\ 81.6 \\ 84.2 \\ 84.9 \\ 75.7 \\ 80.5 \\ 72.9 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \mu & \alpha_1 & \alpha_2 & \alpha_3 & \beta_1 & \beta_2 & \gamma_{11} & \gamma_{12} & \gamma_{21} & \gamma_{22} & \gamma_{31} & \gamma_{32} \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & -1 & 0 & 0 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0 & -1 & 0 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

We fit the full model. R output is below.

```
> fullmodel = lm(yield~a1+a2+a3+b1+b2+g11+g12+g21+g22+g31+g32)
> summary(fullmodel)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 80.19722    1.21016  66.270  < 2e-16
a1           5.90278    2.12908   2.772  0.01423
a2           1.31944    2.09606   0.629  0.53850
a3          -6.58056    2.12908  -3.091  0.00746
b1          -1.82222    1.66496  -1.094  0.29102
b2           2.55278    1.66496   1.533  0.14604
g11          1.02222    2.74122   0.373  0.71443
g12          0.64722    2.88983   0.224  0.82581
g21         -4.34444    2.63749  -1.647  0.12030
g22          0.03056    3.27448   0.009  0.99268
g31         -0.14444    2.88983  -0.050  0.96079
g32          0.33056    2.74122   0.121  0.90562
---

Residual standard error: 5.489 on 15 degrees of freedom
Multiple R-squared:  0.6078,Adjusted R-squared:  0.3202
```

```
F-statistic: 2.113 on 11 and 15 DF,  p-value: 0.08907
```

Now we fit the additive model to test for significance of the interactions.

```
> addmodel = lm(yield~a1+a2+a3+b1+b2)
> anova(addmodel, fullmodel)
Analysis of Variance Table

Model 1: yield ~ a1 + a2 + a3 + b1 + b2
Model 2: yield ~ a1 + a2 + a3 + b1 + b2 + g11 + g12 + g21 + g22 + g31 + g32
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     21 577.26
2     15 451.90  6    125.36 0.6935 0.6587

#The P-value is 0.6587; interactions are not significant.
#Now test for significance of row effects. Create a model without row effects
#and compare to full model.

> norowmodel = lm(yield~b1+b2+g11+g12+g21+g22+g31+g32)
> anova(norowmodel, fullmodel)
Analysis of Variance Table

Model 1: yield ~ b1 + b2 + g11 + g12 + g21 + g22 + g31 + g32
Model 2: yield ~ a1 + a2 + a3 + b1 + b2 + g11 + g12 + g21 + g22 + g31 + g32
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     18 846.82
2     15 451.90  3    394.92 4.3695 0.0212

#The P-value is 0.0212. Row effects are significant at 5% level.
#Now test for significance of column effects. Create a model without column
#effects and compare to full model.

> nocolumnmodel = lm(yield~a1+a2+a3+g11+g12+g21+g22+g31+g32)
> anova(nocolumnmodel, fullmodel)
Analysis of Variance Table

Model 1: yield ~ a1 + a2 + a3 + g11 + g12 + g21 + g22 + g31 + g32
Model 2: yield ~ a1 + a2 + a3 + b1 + b2 + g11 + g12 + g21 + g22 + g31 + g32
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     17 530.36
2     15 451.90  2    78.458 1.3021  0.301

#The P-value is 0.301. Column effects are not significant.
```

3.3.8 A Two-Way ANOVA is Not the Same as Two One-Way ANOVAs

Example 3.18 presented a two-way ANOVA with 3 row levels and 3 column levels, for a total of 9 treatments. If separate one-way ANOVAs were run on the row and column factors separately, there would be only 6 treatments. This means that in practice, running separate one-way ANOVAs on each factor may be less costly than running a two-way ANOVA. Unfortunately, this "one-at-a-time" design is sometimes used in practice for this reason. It is important to realize that running separate one-way analyses on the individual factors

can give results that are misleading when interactions are present. To see this, look at Table 3.6. Assume that an engineer is trying to find the combination of furnace and location that will produce the thinnest oxide layer. He first runs the process once at each furnace location, using in-house recycled wafers, because those wafers are the ones currently being used in production. Furnace location 1 produces the thinnest layer for in-house wafers. Now the engineer runs the process once for each wafer type, all at location 1, which was the best for the in-house wafers. Of the three wafer types, in-house wafers produce the thinnest layer at location 1. So the conclusion drawn from the one-at-a-time analysis is that the thinnest layers are produced by the combination of in-house wafers at furnace location 1. A look at Table 3.6 shows that the conclusion is false. There are two combinations of furnace location and wafer type that produce thinner layers than this.

The one-at-a-time method assumes that the wafer that produces the thinnest layers at one location will produce the thinnest layers at all locations, and that the location that produces the thinnest layers for one wafer type will produce the thinnest layers for all types. This is equivalent to assuming that there are no interactions between the factors, which in the case of the wafers and locations is incorrect. In summary, the one-at-a-time method fails because it cannot detect interactions between the factors.

Summary

- When there are two factors, a two-factor design must be used.
- Examining one factor at a time cannot reveal interactions between the factors.

3.3.9 Interaction Plots

Interaction plots can help to visualize interactions. Figure 3.6 presents an interaction plot for the wafer data. We describe the method by which this plot was constructed. The vertical axis represents the response, which is layer thickness. One factor is chosen to be represented on the horizontal axis. We chose furnace location; it would have been equally acceptable to have chosen wafer type. Now we proceed through the levels of the wafer type factor. We'll start with external wafers. The three cell means for external wafers, as shown in Table 3.6, are 91.967, 88.633, and 90.367, corresponding to furnace locations 1, 2, and 3, respectively. These values are plotted above their respective furnace locations, and connected with line segments. This procedure is repeated for the other two wafer types to complete the plot.

For the wafer data, the means for external wafers follow a substantially different pattern than those for the other two wafer types. This is the source of the significant interaction, and is the reason that the main effects of wafer and furnace type cannot be easily interpreted. In comparison, for perfectly additive data, for which the interaction estimates $\widehat{\gamma}_{ij}$ are equal to 0, the line segments in the interaction plot are parallel. Figure 3.7 illustrates this hypothetical case.

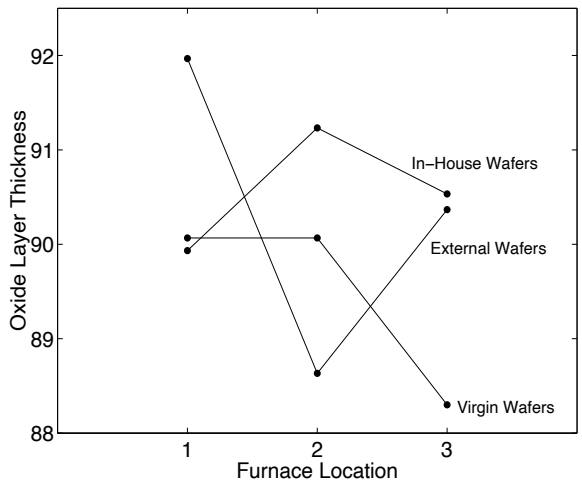


Figure 3.6: Interaction plot for the wafer data. The lines are far from parallel, indicating substantial interaction between the factors.

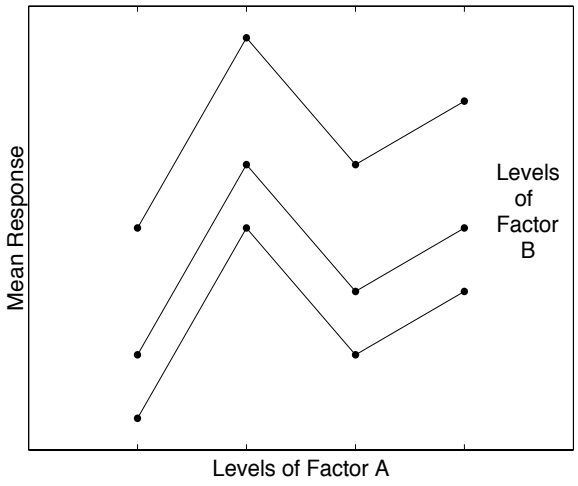


Figure 3.7: Interaction plot for hypothetical data with interaction estimates $\widehat{\gamma}_{ij}$ equal to 0. The line segments are parallel.

Figure 3.8 presents an interaction plot for the yield data. The cell means were presented in Table 3.4. The lines are not parallel, but their slopes match better than those for the wafer data. This indicates that the interaction estimates are nonzero, but are smaller than those for the wafer data. In fact, the P-value for the test of the null hypothesis of no interaction was 0.550 (see page 160). The deviation from parallelism exhibited in Figure 3.8 is therefore small enough to be consistent with the hypothesis of no interaction.

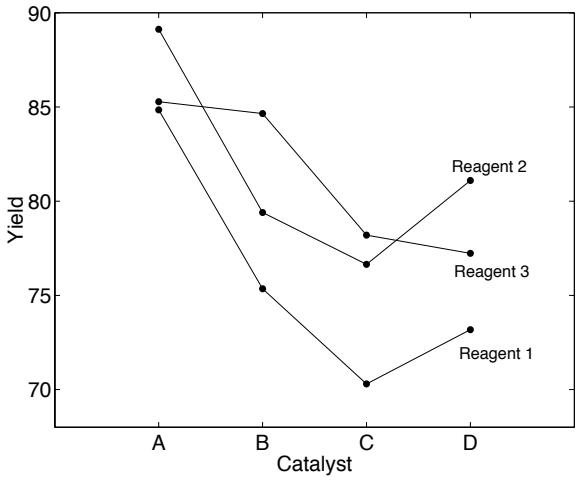


Figure 3.8: Interaction plot for yield data.

3.3.10 Multiple Comparisons in Two-Way ANOVA

An F test is used to test the null hypothesis that all the row effects (or all the column effects) are equal to 0. If the null hypothesis is rejected, we can conclude that some of the row effects (or column effects)

differ from each other. But the hypothesis test does not tell us which ones are different from the rest. If the additive model is plausible, then a method of multiple comparisons known as Tukey's method (related to the Tukey-Kramer method described in Section 3.2.1) can be applied to determine for which pairs the row effects or column effects can be concluded to differ from one another. The method is described below:

Tukey's Method for Simultaneous Confidence Intervals and Hypothesis Tests in Two-way ANOVA

Let I be the number of levels of the row factor, J be the number of levels of the column factor, and K be the sample size for each treatment. **Then, if the additive model is plausible:**

The Tukey level $100(1-\alpha)\%$ simultaneous confidence intervals for all differences $\alpha_i - \alpha_j$, (or all differences $\beta_i - \beta_j$) are

$$\widehat{\alpha}_i - \widehat{\alpha}_j \pm q_{I,IJ(K-1),\alpha}\sqrt{\frac{MSE}{JK}} \qquad \widehat{\beta}_i - \widehat{\beta}_j \pm q_{J,IJ(K-1),\alpha}\sqrt{\frac{MSE}{IK}}$$

We are $100(1-\alpha)\%$ confident that the Tukey confidence intervals contain the true value of the difference $\alpha_i - \alpha_j$ (or $\beta_i - \beta_j$) for every i and j.

For every pair of levels i and j for which $|\widehat{\alpha}_i - \widehat{\alpha}_j| > q_{I,IJ(K-1),\alpha}\sqrt{\frac{MSE}{JK}}$, the null hypothesis $H_0\colon \alpha_i - \alpha_j = 0$ is rejected at level α.

For every pair of levels i and j for which $|\widehat{\beta}_i - \widehat{\beta}_j| > q_{J,IJ(K-1),\alpha}\sqrt{\frac{MSE}{IK}}$, the null hypothesis $H_0\colon \beta_i - \beta_j = 0$ is rejected at level α.

Example 3.20

In Example 3.15 (page 157), the main effects and interactions were computed for the yield data in Table 3.2 (page 153). An ANOVA table for these data was presented on page 160. If appropriate, use Tukey's method to determine which pairs of catalysts and which pairs of reagents can be concluded to differ, at the 5% level, in their effect on yield.

Solution: From the ANOVA table, the P-value for interactions is 0.550. Therefore the additive model is plausible, so it is appropriate to use Tukey's method. Catalyst is the row factor and reagent is the column factor, so $I = 4$, $J = 3$, and $K = 4$. From the ANOVA table, $MSE = 31.259$.

We first find all pairs for which the row effects differ at the 5% level. For the row effects, we use the value of $q_{4,24,.05} = 3.90$, found in the Studentized range table (Table 4 on page 200). We compute $q_{4,24,.05}\sqrt{MSE/JK} = 3.90\sqrt{31.259/12} = 6.29$.

In Example 3.15, the estimated row effects were computed to be

$$\widehat{\alpha}_1 = 6.81 \quad \widehat{\alpha}_2 = 0.19 \quad \widehat{\alpha}_3 = -4.56 \quad \widehat{\alpha}_4 = -2.44$$

The pairs of row effects whose differences are greater than 6.29 are $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$, $\widehat{\alpha}_1$ and $\widehat{\alpha}_3$, and $\widehat{\alpha}_1$ and $\widehat{\alpha}_4$. We conclude that the mean yield of Catalyst A differs from the mean yields of Catalysts B, C, and D, but we cannot conclude that the mean yields of Catalysts B, C, and D differ from each other.

We now find all pairs for which the column effects differ at the 5% level. For the column effects, we use the value $q_{3,24,.05} = 3.53$. We compute $q_{4,40,.05}\sqrt{MSE/IK} = 3.53\sqrt{31.259/16} = 4.93$.

In Example 3.15, the estimated column effects were computed to be

$$\widehat{\beta}_1 = -3.69 \quad \widehat{\beta}_2 = 1.96 \quad \widehat{\beta}_3 = 1.73$$

The pairs of column effects whose differences are greater than 4.93 are $\widehat{\beta}_1$ and $\widehat{\beta}_2$ and $\widehat{\beta}_1$ and $\widehat{\beta}_3$. We conclude that the mean yield of Reagent 1 differs from the mean yields of Reagents 2 and 3, but we cannot conclude that the mean yields of Reagents 2 and 3 differ from each other. ∎

3.3.11 Two-Way ANOVA when $K = 1$

The F tests we have presented require the assumption that the sample size K for each treatment be at least 2. The reason for this is that when $K = 1$, the error sum of squares, SSE, will be equal to 0, since $X_{ijk} = \overline{X}_{ij.}$ for each i and j. In addition, the degrees of freedom for SSE, which is $IJ(K-1)$, is equal to 0 when $K = 1$.

When $K = 1$, a two-way ANOVA cannot be performed unless it is certain that the additive model holds. In this case, since the interactions are assumed to be zero, the mean square for interaction ($MSAB$; see Equation 3.63 on page 160) and its degrees of freedom can be used in place of MSE to test the main row and column effects.

3.4 Randomized Complete Block Designs

In some experiments, there are factors that vary, and that have an effect on the response, but whose effects are not of interest to the experimenter. For example, in one commonly occurring situation, it is impossible to complete the experiment in a single day, so the observations have to be spread out over several days. If conditions that can affect the outcome vary from day to day, then the day becomes a factor in the experiment, even though there may be no interest in estimating its effect.

For a more specific example, imagine that three types of fertilizer are to be evaluated for their effect on yield of fruit in an orange grove, and that three replicates will be performed, for a total of nine observations. An area is divided into nine plots, in three rows of three plots each. Now assume there is a water gradient along the plot area, so that the rows receive differing amounts of water. The amount of water is now a factor in the experiment, even though there is no interest in estimating the effect of water amount on the yield of oranges.

If the water factor is ignored, a one-factor experiment could be carried out with fertilizer as the only factor. Each of the three fertilizers would be assigned to three of the plots. In a completely randomized experiment, the treatments would be assigned to the plots at random. Figure 3.9 presents two possible random arrangements.

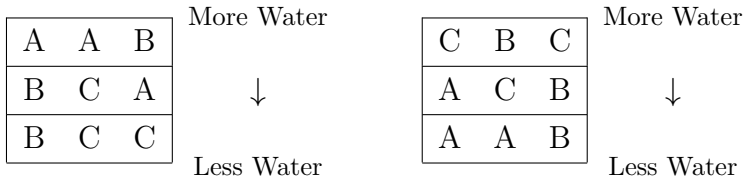


Figure 3.9: Two possible arrangements for three fertilizers, A, B, and C, assigned to nine plots completely at random. It is likely that the amounts of water will differ for the different fertilizers.

In the arrangement on the left, the plots with fertilizer A get more water than those with the other two fertilizers. In the plot on the right, the plots with fertilizer C get the most water. When the treatments for one factor are assigned completely at random, it is likely that they will not be distributed evenly over the levels of another factor.

If the amount of water in fact has a negligible effect on the response, then the completely randomized one-factor design is appropriate. There is no reason to worry about a factor that does not affect the response. But now assume that the water level does have a substantial impact on the response. Then Figure 3.9 shows that in any one experiment, the estimated effects of the treatments are likely to be thrown off the mark, or biased, by the differing levels of water. Different arrangements of the treatments bias the estimates in different directions. If the experiment is repeated several times, the estimates are likely to vary greatly from repetition to repetition. For this reason, the completely randomized one-factor design produces estimated effects that have large uncertainties.

A better design for this experiment is a two-factor design, with water as the second factor. Since the effects of water are not of interest, water is called a **blocking factor**, rather than a treatment factor. In the two-factor experiment, there are nine treatment/block combinations, corresponding to the three fertilizer treatment levels and the three water block levels. With nine experimental units (the nine plots), it is necessary to assign one plot to each combination of fertilizer and water. Figure 3.10 presents two possible arrangements.

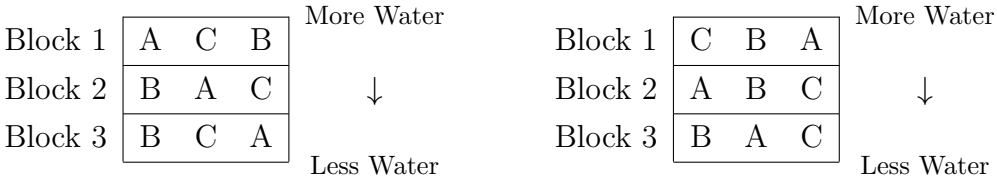


Figure 3.10: Two possible arrangements for three fertilizers, A, B, and C, with the restriction that each fertilizer must appear once at each water level (block). The distribution of water levels is always the same for each fertilizer.

In the two-factor design, each treatment appears equally often (once, in this example) in each block. As a result, the effect of the blocking factor does not contribute to uncertainty in the estimate of the main effects of the treatment factor. Because each treatment must appear equally often in each block, the only randomization in the assignment of treatments to experimental units is the order in which the treatments appear in each block. This is not a completely randomized design; it is a design in which the treatments are **randomized within blocks**. Since every possible combination of treatments and blocks is included in the experiment, the design is **complete**. For this reason the design is called a **randomized complete block design**.

Randomized complete block designs can be constructed with several treatment factors and several blocking factors. We will restrict our discussion to the case where there is one treatment factor and one blocking factor. The data from a randomized complete block design are analyzed with a two-way ANOVA, in the same way that data from any complete, balanced two-factor design would be. There is one important consideration, however. The only effects of interest are the main effects of the treatment factor. In order to interpret these main effects, **there must be no interaction between treatment and blocking factors**.

Example 3.21

Three fertilizers are studied for their effect on yield in an orange grove. Nine plots of land are used, divided into blocks of three plots each. A randomized complete block design is used, with each fertilizer applied once in each block. The results, in pounds of harvested fruit, are presented below, followed by R output for the two-way ANOVA. Can we conclude that the mean yields differ among fertilizers? What assumption is made about interactions between fertilizer and plot? How is the sum of squares for error computed?

	Plot		
Fertilizer	1	2	3
A	430	542	287
B	367	463	253
C	320	421	207

```
> yield.lm = lm(Pounds~Fertilizer+Block, data=yield)
> anova(yield.lm)
            Df Sum Sq Mean Sq F value    Pr(>F)
Fertilizer   2  16214    8107  49.752  0.001494
Block        2  77047   38523 236.421 7.037e-05
Residuals    4    652     163
```

Solution: The P-value for the fertilizer factor is 0.000, so we conclude that fertilizer does have an effect on yield. The assumption is made that there is no interaction between fertilizer and the blocking factor (plot), so that the main effects of fertilizer can be interpreted. Since there is only one observation for each treatment/block combination (i.e., $K = 1$), the sum of squares for error (SSE) reported in the R output is really $SSAB$, the sum of squares for interaction, and the error mean square (MSE) is actually $MSAB$. (See the discussion on page 173.) ∎

A closer look at the ANOVA table above shows that in this experiment, blocking was necessary to to detect the fertilizer effect. To see this, consider the experiment to be a one-factor experiment. The sum of squares for error, SSE, would then be the sum of SSE for the blocked design plus the sum of squares for blocks, or $651.778 + 77046.9 = 77698.7$. The degrees of freedom for error would be equal to the sum of the degrees of freedom for error in the blocked design plus the degrees of freedom for blocks, or $2 + 4 = 6$. The error mean square MSE would then be $77698.7/6 \approx 12950$ rather than 162.9444, and the F statistic for the fertilizer effect would be less than 1, which would result in failure to detect an effect.

In general, using a blocked design reduces the degrees of freedom for error, which by itself tends to reduce the power to detect an effect. However, unless the blocking factor has very little effect on the response, this will usually be more than offset by a reduction in the sum of squares for error. Failing to include a blocking factor that affects the response can reduce the power greatly, while including a blocking factor that does not affect the response reduces the power only modestly in most cases. For this reason it is a good idea to use a blocked design whenever it is thought to be possible that the blocking factor is related to the response.

Summary

- A two-factor randomized complete block design is a complete balanced two-factor design in which the effects of one factor (the treatment factor) are of interest, while the effects of other factor (the blocking factor) are not of interest. The blocking factor is included to reduce the uncertainty in the main effect estimates of the treatment factor.
- Since the object of a randomized complete block design is to estimate the main effects of the treatment factor, there must be no interaction between the treatment factor and the blocking factor.
- A two-way analysis of variance is used to estimate effects and to perform hypothesis tests on the main effects of the treatment factor.
- A randomized complete block design provides a great advantage over a completely randomized design when the blocking factor strongly affects the response, and a relatively small disadvantage when the blocking factor has little or no effect. Therefore, when in doubt, it is a good idea to use a blocked design.

Example 3.22

The article "Experimental Design for Process Settings in Aircraft Manufacturing" (R. Sauter and R. Lenth, *Statistical Case Studies: A Collaboration Between Academe and Industry*, SIAM–ASA, 1998:151–157) describes an experiment in which the quality of holes drilled in metal aircraft parts was studied. One important indicator of hole quality is "excess diameter," which is the difference between the diameter of the drill bit and the diameter of the hole. Small excess diameters are better than large ones. Assume we are interested in the effect of the rotational speed of the drill on the excess diameter of the hole. Holes will be drilled in six test pieces (coupons), at three speeds: 6000, 10,000, and 15,000 rpm. The excess diameter can be affected not only by the speed of the drill, but also by the physical properties of the test coupon. Describe an appropriate design for this experiment.

Solution: A randomized complete block design is appropriate, with drill speed as the treatment factor, and test coupon as the blocking factor. Since six observations can be made in each block, each drill speed should be used twice in each block. The order of the speeds within each block should be chosen at random. ∎

Example 3.23

The design suggested in Example 3.22 has been adopted, and the experiment has been carried out. R output is presented below. Does the output indicate any violation of necessary assumptions? What do you conclude regarding the effect of drill speed on excess diameter?

```
Two-way ANOVA: Excess Diameter versus Block, Speed

             Df   Sum Sq    Mean Sq F value    Pr(>F)
Block         5  0.20156  0.0403117    1.08    0.404
Speed         2  0.07835  0.0391750    1.05    0.370
Block:Speed  10  0.16272  0.0162717    0.44    0.909
Residuals    18  0.67105  0.0372806
```

Solution: In a randomized complete block design, there must be no interaction between the treatment factor and the blocking factor, so that the main effect of the treatment factor may be

interpreted. The P-value for interactions is 0.909, which is consistent with the hypothesis of no interactions. Therefore there is no indication in the output of any violation of assumptions. The P-value for the main effect of speed is 0.370, which is not small. Therefore we cannot conclude that excess hole diameter is affected by drill speed. ∎

The next example shows that a paired design in which a t test is used to compare two population means, is a special case of a randomized block design.

Example 3.24

A tire manufacturer wants to compare the tread wear of tires made from a new material with that of tires made from a conventional material. There are 10 tires of each type. Each tire will be mounted on the front wheel of a front-wheel drive car and driven for 40,000 miles. Then the tread wear will be measured for each tire. Describe an appropriate design for this experiment.

Solution: The response is the tread wear after 40,000 miles. There is one factor of interest: the type of tire. Since the cars may differ in the amounts of wear they produce, the car is a factor as well, but its effect is not of interest. A randomized complete block design is appropriate, in which one tire of each type is mounted on the front wheels of each car. ∎

You may note that the randomized complete block design in Example 3.24 is the same design that is used when comparing two population means with a paired t test. The paired design described there is a special case of a randomized complete block design, in which the treatment factor has only two levels, and each level appears once in each block. In fact, a two-way analysis of variance applied to data from such a design is equivalent to the paired t test.

Multiple Comparisons in Randomized Complete Block Designs

Once an ANOVA table has been constructed, then if the F test shows that the treatment main effects are not all the same, a method of multiple comparisons may be used to determine which pairs of effects may be concluded to differ. We describe Tukey's method, which is a special case of the Tukey-Kramer method described in Section 3.2.1. The degrees of freedom, and the mean square used, differ depending on whether each treatment appears only once, or more than once, in each block.

Tukey's Method for Multiple Comparisons in Randomized Complete Block Designs

In a randomized complete block design, with I treatment levels, J block levels, and treatment main effects $\alpha_1, ..., \alpha_I$:

- If each treatment appears only once in each block, then the null hypothesis $H_0 : \alpha_i - \alpha_j = 0$ is rejected at level α for every pair of treatments i and j for which

$$|\overline{X}_{i.} - \overline{X}_{j.}| > q_{I,(I-1)(J-1),\alpha}\sqrt{\frac{MSAB}{J}}$$

  where $MSAB$ is the mean square for interaction.

- If each treatment appears $K > 1$ times in each block, then the null hypothesis $H_0 : \alpha_i - \alpha_j = 0$ is rejected at level α for every pair of treatments i and j for which

$$|\overline{X}_{i..} - \overline{X}_{j..}| > q_{I,IJ(K-1),\alpha}\sqrt{\frac{MSE}{JK}}$$

  where MSE is the mean square for error.

For more information on randomized block designs, a text on design of experiments, such as Montgomery (2001a), can be consulted.

3.5 2^p Factorial Experiments

3.5.1 Introduction and Notation

When an experimenter wants to study several factors simultaneously, the number of different treatments can become quite large. In these cases, preliminary experiments are often performed in which each factor has only two levels. One level is designated as the "high" level, and the other is designated as the "low" level. If there are p factors, there are then 2^p different treatments. Such experiments are called **2^p factorial experiments**. Often, the purpose of a 2^p experiment is to determine which factors have an important effect on the outcome. Once this is determined, more elaborate experiments can be performed, in which the factors previously found to be important are varied over several levels. We will begin by describing 2^3 factorial experiments.

In a 2^3 factorial experiment, there are three factors and $2^3 = 8$ treatments. The **main effect** of a factor is defined to be the difference between the mean response when the factor is at its high level and the mean response when the factor is at its low level. It is common to denote the main effects by A, B, and C. As with any factorial experiment, there can be interactions between the factors. With three factors, there are three two-way interactions, one for each pair of factors, and one three-way interaction. The two-way interactions are denoted by AB, AC, and BC, and the three-way interaction by ABC. The treatments are traditionally denoted with lower case letters, with a letter indicating that a factor is at its high level. For example, ab denotes the treatment in which the first two factors are at their high level, and the third factor is at its low level. The symbol "1" is used to denote the treatment in which all factors are at their low levels.

3.5.2 Estimating Effects in a 2^3 Factorial Experiment

Assume that there are K replicates for each treatment in a 2^3 factorial experiment. For each treatment, the cell mean is the average of the K observations for that treatment. The formulas for the effect estimates can be easily obtained from the 2^3 **sign table**, presented as Table 3.8.

Table 3.8: Sign Table for a 2^3 Factorial Experiment

Treatment	Cell Mean	A	B	C	AB	AC	BC	ABC
1	$\overline{X}_1$	−	−	−	+	+	+	−
a	$\overline{X}_a$	+	−	−	−	−	+	+
b	$\overline{X}_b$	−	+	−	−	+	−	+
ab	$\overline{X}_{ab}$	+	+	−	+	−	−	−
c	$\overline{X}_c$	−	−	+	+	−	−	+
ac	$\overline{X}_{ac}$	+	−	+	−	+	−	−
bc	$\overline{X}_{bc}$	−	+	+	−	−	+	−
abc	$\overline{X}_{abc}$	+	+	+	+	+	+	+

The signs are placed in the table as follows. For the main effects A, B, C, the sign is + for treatments in which the factor is at its high level, and − for treatments where the factor is at its low level. So for the main effect A, the sign is + for treatments a, ab, ac, and abc, and − for the rest. For the interactions, the signs

are computed by taking the product of the signs in the corresponding main effect columns. For example, the signs for the two-way interaction AB are the products of the signs in columns A and B, and the signs for the three-way interaction ABC are the products of the signs in columns A and B and C.

Estimating main effects and interactions is done with the use of the sign table. We illustrate how to estimate the main effect of factor A. Factor A is at its high level in the rows of the table where there is a "+" sign in column A. Each of the cell means $\overline{X}_a$, $\overline{X}_{ab}$, $\overline{X}_{ac}$, and $\overline{X}_{abc}$ is an average response for runs made with factor A at its high level. We estimate the mean response for factor A at its high level to be the average of these cell means.

$$\text{Estimated mean response for } A \text{ at high level} = \frac{1}{4}(\overline{X}_a + \overline{X}_{ab} + \overline{X}_{ac} + \overline{X}_{abc})$$

Similarly, each row with a "−" sign in column A represents a treatment with factor A set to its low level. We estimate the mean response for factor A at its low level to be the average of the cell means in these rows.

$$\text{Estimated mean response for } A \text{ at low level} = \frac{1}{4}(\overline{X}_1 + \overline{X}_b + \overline{X}_c + \overline{X}_{bc})$$

The estimate of the main effect of factor A is the difference in the estimated mean response between its high and low levels.

$$A \text{ effect estimate} = \frac{1}{4}(-\overline{X}_1 + \overline{X}_a - \overline{X}_b + \overline{X}_{ab} - \overline{X}_c + \overline{X}_{ac} - \overline{X}_{bc} + \overline{X}_{abc})$$

The quantity inside the parentheses is called the **contrast** for factor A. It is computed by adding and subtracting the cell means, using the signs in the appropriate column of the sign table. Note that the number of plus signs is the same as the number of minus signs, so the sum of the coefficients is equal to 0. The effect estimate is obtained by dividing the contrast by half the number of treatments, which is $2^3/2$, or 4. Estimates for other main effects and interactions are computed in an analogous manner. To illustrate, we present the effect estimates for the main effect C and for the two-way interaction AB:

$$\begin{aligned} C \text{ effect estimate} &= \frac{1}{4}(-\overline{X}_1 - \overline{X}_a - \overline{X}_b - \overline{X}_{ab} + \overline{X}_c + \overline{X}_{ac} + \overline{X}_{bc} + \overline{X}_{abc}) \\ AB \text{ interaction estimate} &= \frac{1}{4}(\overline{X}_1 - \overline{X}_a - \overline{X}_b + \overline{X}_{ab} + \overline{X}_c - \overline{X}_{ac} - \overline{X}_{bc} + \overline{X}_{abc}) \end{aligned}$$

Summary

The **contrast** for any main effect or interaction is obtained by adding and subtracting the cell means, using the signs in the appropriate column of the sign table.

For a 2^3 factorial experiment,

$$\text{Effect estimate} = \frac{\text{Contrast}}{4} \tag{3.64}$$

Example 3.25

A 2^3 factorial experiment was conducted to estimate the effects of three factors on the yield of a chemical reaction. The factors were: A: Catalyst concentration (low or high), B: Reagent (standard formulation or new formulation), and C: Stirring rate (slow or fast). Three replicates were obtained for each treatment. The yields are measured as a percent of a theoretical maximum. Estimate all effects and interactions.

Treatment	Yield	Cell Mean
1	71.67, 70.55, 67.40	69.8733
a	78.46, 75.42, 81.77	78.5500
b	77.14, 78.25, 78.33	77.9067
ab	79.72, 76.17, 78.41	78.1000
c	72.65, 71.03, 73.54	72.4067
ac	80.10, 73.91, 74.81	76.2733
bc	80.20, 73.49, 74.86	76.1833
abc	75.58, 80.28, 71.64	75.8333

Solution: We use the sign table (Table 3.8) to find the appropriate sums and differences of the cell means. We present the calculations for the main effect A, the two-way interaction BC, and the three-way interaction ABC:

$$\begin{aligned} A \text{ effect estimate} &= (1/4)(-69.8733 + 78.5500 - 77.9067 + 78.1000 \\ &\quad -72.4067 + 76.2733 - 76.1833 + 75.8333) = 3.0967 \end{aligned}$$

$$\begin{aligned} BC \text{ interaction estimate} &= (1/4)(69.8733 + 78.5500 - 77.9067 - 78.1000 \\ &\quad -72.4067 - 76.2733 + 76.1833 + 75.8333) = -1.0617 \end{aligned}$$

$$\begin{aligned} ABC \text{ interaction estimate} &= (1/4)(-69.8733 + 78.5500 + 77.9067 - 78.1000 \\ &\quad +72.4067 - 76.2733 - 76.1833 + 75.8333) = 1.0667 \end{aligned}$$

We present all the estimated effects in the table below, rounded off to the same precision as the data:

Variables	Effect
A	3.10
B	2.73
C	−0.93
AB	−3.18
AC	−1.34
BC	−1.06
ABC	1.07

∎

For each effect, we can test the null hypothesis that the effect is equal to 0. When the null hypothesis is rejected, this provides evidence that the factors involved actually affect the outcome. To test these null hypotheses, an ANOVA table is constructed containing the appropriate sums of squares. For the tests we present to be valid, the number of replicates must be the same for each treatment, and must be at least 2. In addition, the observations in each treatment must constitute a random sample from a normal population, and the populations must all have the same variance.

We compute the error sum of squares, SSE, by adding the sums of squared deviations from the sample means for all the treatments. To express this in an equation, let $s_1^2, ..., s_8^2$ denote the sample variances of the observations in each of the eight treatments, and let K be the number of replicates per treatment. Then

$$SSE = (K-1)\sum_{i=1}^{8} s_i^2 \tag{3.65}$$

Each main effect and interaction has its own sum of squares as well. These are easy to compute. The sum of squares for any effect or interaction is computed by squaring its contrast, multiplying by the number of replicates K, and dividing by the total number of treatments, which is $2^3 = 8$.

$$\text{Sum of Squares for an effect} = \frac{K(\text{Contrast})^2}{8} \tag{3.66}$$

When using Equation (3.66), it is best to keep as many digits in the effect estimates as possible, in order to obtain maximum precision in the sum of squares. For presentation in a table, effect estimates and sums of squares may be rounded to the same precision as the data.

The sums of squares for the effects and interactions have 1 degree of freedom each. The error sum of squares has $8(K-1)$ degrees of freedom. The method for computing mean squares and F statistics is the same as the one presented in Section 3.3.1 for a two-way ANOVA table. Each mean square is equal to its sum of squares divided by its degrees of freedom. The test statistic for testing the null hypothesis that an effect or interaction is equal to 0 is computed by dividing the mean square for the effect estimate by the mean square for error. When the null hypothesis is true, the test statistic has an $F_{1,\,8(K-1)}$ distribution.

Example 3.26

Refer to Example 3.25. Construct an ANOVA table. For each effect and interaction, test the null hypothesis that it is equal to 0. Which factors, if any, seem most likely to have an effect on the outcome?

Solution: The ANOVA table is presented below. The sums of squares for the effects and interactions were computed by using Equation (3.66). The error sum of squares was computed by applying Equation (3.65) to the data in Example 3.25. Each F statistic is the quotient of the mean square with the mean square for error. Each F statistic has 1 and 16 degrees of freedom.

Source	Effect	Sum of Squares	df	Mean Square	F	P
A	3.10	57.54	1	57.54	7.34	0.015
B	2.73	44.72	1	44.72	5.70	0.030
C	−0.93	5.23	1	5.23	0.67	0.426
AB	−3.18	60.48	1	60.48	7.71	0.013
AC	−1.34	10.75	1	10.75	1.37	0.259
BC	−1.06	6.76	1	6.76	0.86	0.367
ABC	1.07	6.83	1	6.83	0.87	0.365
Error		125.48	16	7.84		
Total		317.78	23			

The main effects of factors A and B, as well as the AB interaction, have fairly small P-values. This suggests that these effects are not equal to 0, and that factors A and B do affect the outcome. There is no evidence that the main effect of factor C, or any of its interactions, differ from 0. Further experiments might focus on factors A and B. Perhaps a two-way ANOVA would be conducted, with each of the factors A and B evaluated at several levels, to get more detailed information about their effects on the outcome. ∎

3.5.3 Interpreting Computer Output

In practice, analyses of factorial designs are usually carried out on a computer, using a software package such as R. Below is R output for the analysis described in Examples 3.25 and 3.26.

```
> reaction.lm = lm(Yield~A*B*C, data=reaction)
> anova(reaction.lm)
            Df  Sum Sq Mean Sq F value  Pr(>F)
A            1  57.536  57.536  7.3367 0.01550
B            1  44.717  44.717  5.7021 0.02962
C            1   5.227   5.227  0.6665 0.42627
A:B          1  60.484  60.484  7.7126 0.01346
A:C          1  10.747  10.747  1.3704 0.25889
B:C          1   6.763   6.763  0.8624 0.36688
A:B:C        1   6.827   6.827  0.8705 0.36468
Residuals   16 125.476   7.842
```

3.5.4 Estimating Effects in a 2^p Factorial Experiment

A sign table can be used to obtain the formulas for computing effect estimates in any 2^p factorial experiment. The method is analogous to the 2^3 case. The treatments are listed in a column. The sign for any main effect is $+$ in the rows corresponding to treatments where the factor is at its high level, and $-$ in rows corresponding to treatments where the factor is at its low level. Signs for the interactions are found by multiplying the signs corresponding to the factors in the interaction. The estimate for any effect or interaction is found by adding and subtracting the cell means for the treatments, using the signs in the appropriate columns, to compute a contrast. The contrast is then divided by half the number of treatments, or 2^{p-1}, to obtain the effect estimate.

Summary

For a 2^p factorial experiment:

$$\text{Effect estimate} = \frac{\text{Contrast}}{2^{p-1}} \tag{3.67}$$

As an example, Table 3.9 presents a sign table for a 2^5 factorial experiment. We list the signs for the main effects and for selected interactions.

Table 3.9: Sign Table for the Main Effects and Selected Interactions for a 2^5 Factorial Experiment.

Treatment	A	B	C	D	E	AB	CDE	$ABDE$	$ABCDE$
1	−	−	−	−	−	+	−	+	−
a	+	−	−	−	−	−	−	−	+
b	−	+	−	−	−	−	−	−	+
ab	+	+	−	−	−	+	−	+	−
c	−	−	+	−	−	+	+	+	+
ac	+	−	+	−	−	−	+	−	−
bc	−	+	+	−	−	−	+	−	−
abc	+	+	+	−	−	+	+	+	+
d	−	−	−	+	−	+	+	−	+
ad	+	−	−	+	−	−	+	+	−
bd	−	+	−	+	−	−	+	+	−
abd	+	+	−	+	−	+	+	−	+
cd	−	−	+	+	−	+	−	−	−
acd	+	−	+	+	−	−	−	+	+
bcd	−	+	+	+	−	−	−	+	+
$abcd$	+	+	+	+	−	+	−	−	−
e	−	−	−	−	+	+	+	−	+
ae	+	−	−	−	+	−	+	+	−
be	−	+	−	−	+	−	+	+	−
abe	+	+	−	−	+	+	+	−	+
ce	−	−	+	−	+	+	−	−	−
ace	+	−	+	−	+	−	−	+	+
bce	−	+	+	−	+	−	−	+	+
$abce$	+	+	+	−	+	+	−	−	−
de	−	−	−	+	+	+	−	+	−
ade	+	−	−	+	+	−	−	−	+
bde	−	+	−	+	+	−	−	−	+
$abde$	+	+	−	+	+	+	−	+	−
cde	−	−	+	+	+	+	+	+	+
$acde$	+	−	+	+	+	−	+	−	−
$bcde$	−	+	+	+	+	−	+	−	−
$abcde$	+	+	+	+	+	+	+	+	+

Sums of squares are computed by a method analogous to that for a 2^3 experiment. To compute the error sum of squares, SSE, let $s_1, ..., s_{2^p}$ be the sample variances of the observations in each of the 2^p treatments. Then

$$SSE = (K-1)\sum_{i=1}^{2^p} s_i^2$$

The degrees of freedom for error is $2^p(K-1)$, where K is the number of replicates per treatment. The sum of squares for each effect and interaction is equal to the square of the contrast, multiplied by the number of replicates K and divided by the number of treatments 2^p. The sums of squares for the effects and interactions have one degree of freedom each.

$$\text{Sum of Squares for an effect} = \frac{K(\text{Contrast})^2}{2^p} \tag{3.68}$$

F statistics for main effects and interactions are computed by dividing the sum of squares for the effect by the mean square for error. The null distribution of the F statistic is $F_{1,\,2^p(K-1)}$.

3.5.5 Factorial Experiments Without Replication

When the number of factors p is large, it is often not feasible to perform more than one replicate for each treatment. In this case, it is not possible to compute SSE, so the hypothesis tests described above cannot be performed. If it is reasonable to assume that some of the higher order interactions are equal to 0, then the sums of squares for those interactions can be added together and treated like an error sum of squares. Then the main effects and lower order interactions can be tested.

Example 3.27

A 2^5 factorial experiment was conducted to estimate the effects of five factors on the quality of light bulbs manufactured by a certain process. The factors were: A: Plant (#1 or #2) B: Machine type (low or high speed), C: Shift (day or evening), D: Lead wire material (standard or new), and E: Method of loading materials into the assembler (manual or automatic). One replicate was obtained for each treatment. The table below presents the results. Compute estimates of the main effects and interactions, and their sums of squares. Assume that the third-, fourth-, and fifth-order interactions are negligible, and add their sums of squares to use as a substitute for an error sum of squares. Use this substitute to test hypotheses concerning the main effects and second-order interactions.

Treatment	Outcome	Treatment	Outcome	Treatment	Outcome	Treatment	Outcome
1	32.07	*d*	35.64	*e*	25.10	*de*	40.60
a	39.27	*ad*	35.91	*ae*	39.25	*ade*	37.57
b	34.81	*bd*	47.75	*be*	37.77	*bde*	47.22
ab	43.07	*abd*	51.47	*abe*	46.69	*abde*	56.87
c	31.55	*cd*	33.16	*ce*	32.55	*cde*	34.51
ac	36.51	*acd*	35.32	*ace*	32.56	*acde*	36.67
bc	28.80	*bcd*	48.26	*bce*	28.99	*bcde*	45.15
abc	43.05	*abcd*	53.28	*abce*	48.92	*abcde*	48.72

Solution: We compute the effects, using the rules for adding and subtracting observations given by the sign table, and the sums of squares, using Equation (3.68).

Variable	Effect	Sum of Squares	Variable	Effect	Sum of Squares
A	6.33	320.05	ABD	-0.29	0.67
B	9.54	727.52	ABE	0.76	4.59
C	-2.07	34.16	ACD	0.11	0.088
D	6.70	358.72	ACE	-0.69	3.75
E	0.58	2.66	ADE	-0.45	1.60
AB	2.84	64.52	BCD	0.76	4.67
AC	0.18	0.27	BCE	-0.82	5.43
AD	-3.39	91.67	BDE	-2.17	37.63
AE	0.60	2.83	CDE	-1.25	12.48
BC	-0.49	1.95	$ABCD$	-2.83	63.96
BD	4.13	136.54	$ABCE$	0.39	1.22
BE	0.65	3.42	$ABDE$	0.22	0.37
CD	-0.18	0.26	$ACDE$	0.18	0.24
CE	-0.81	5.23	$BCDE$	-0.25	0.52
DE	0.24	0.46	$ABCDE$	-1.73	23.80
ABC	1.35	14.47			

Note that none of the three-, four-, or five-way interactions are among the larger effects. If some of them were, it would not be wise to combine their sums of squares. As it is, we add the sums of squares of the three-, four-, and five-way interactions. The results are presented in the R output below:

```
> bulb.lm = lm(Outcome ~ .*., data=bulb)
> anova(bulb.lm)
            Df Sum Sq Mean Sq F value    Pr(>F)
A            1 320.05  320.05 29.1800 5.875e-05
B            1 727.52  727.52 66.3315 4.393e-07
C            1  34.16   34.16  3.1141  0.096696
D            1 358.72  358.72 32.7060 3.167e-05
E            1   2.66    2.66  0.2422  0.629301
A:B          1  64.52   64.52  5.8830  0.027486
A:C          1   0.27    0.27  0.0243  0.878090
A:D          1  91.67   91.67  8.3576  0.010639
A:E          1   2.83    2.83  0.2582  0.618276
B:C          1   1.95    1.95  0.1778  0.678865
B:D          1 136.54  136.54 12.4488  0.002792
B:E          1   3.42    3.42  0.3117  0.584346
C:D          1   0.26    0.26  0.0233  0.880574
C:E          1   5.23    5.23  0.4771  0.499645
D:E          1   0.46    0.46  0.0416  0.840999
Residuals   16 175.49   10.97

> bulb.aov$coef
(Intercept)           A           B           C           D           E
  39.658125    3.162500    4.768125   -1.033125    3.348125    0.288125
        A:B         A:C         A:D         A:E         B:C         B:D
   1.420000    0.091250   -1.692500    0.297500   -0.246875    2.065625
        B:E         C:D         C:E         D:E
   0.326875   -0.089375   -0.404375    0.119375
```

The estimates have not changed for the main effects or two-way interactions. The residual error

sum of squares (175.49) in the analysis of variance table is found by adding the sum of squares for all the higher-order interactions that were dropped from the model. The number of degrees of freedom (16) is equal to the sum of the degrees of freedom (1 each) for the 16 higher order interactions.

We conclude from the output that factors A, B, and D are likely to affect the outcome. There seem to be interactions between some pairs of these factors as well. It might be appropriate to plan further experiments to focus on factors A, B, and D. ∎

3.5.6 Fractional Factorial Experiments

When the number of factors is large enough, it may not be feasible to perform even one replicate for each treatment. In these cases, observations may be taken only for some fraction of the treatments. If these treatments are chosen correctly, it is still possible to obtain information about the factors.

When each factor has two levels, the fraction must always be a power of 2, e.g., one-half, one-quarter, etc. An experiment in which half the treatments are used is called a **half-replicate**, if one-quarter of the treatments are used it is a **quarter-replicate**, and so on. A half-replicate of a 2^p experiment is often denoted 2^{p-1}, to indicate that while there are p factors, there are only 2^{p-1} treatments being considered. Similarly, a quarter-replicate is often denoted 2^{p-2}. We will focus on half-replicate experiments.

We present a method for choosing a half-replicate of a 2^5 experiment. Such an experiment will have 16 treatments chosen from the 32 in the 2^5 experiment. To choose the 16 treatments, start with a sign table for a 2^4 design that shows the signs for the main effects and the highest order interaction. This is presented as Table 3.10.

Table 3.10: Sign Table for the Main Effects and Four-Way Interaction in a 2^4 Factorial Experiment.

Treatment	A	B	C	D	$ABCD$
1	−	−	−	−	+
a	+	−	−	−	−
b	−	+	−	−	−
ab	+	+	−	−	+
c	−	−	+	−	−
ac	+	−	+	−	+
bc	−	+	+	−	+
abc	+	+	+	−	−
d	−	−	−	+	−
ad	+	−	−	+	+
bd	−	+	−	+	+
abd	+	+	−	+	−
cd	−	−	+	+	+
acd	+	−	+	+	−
bcd	−	+	+	+	−
$abcd$	+	+	+	+	+

This table has the right number of treatments (16), but only 4 factors. To transform it into a half-replicate for a 2^5 design, we must introduce a fifth factor, E. We do this by replacing the highest order interaction by E. This establishes the signs for the main effect of E. Then in each row where the sign for E is +, we add

the letter e to the treatment, indicating that factor E is to be set to its high level for that treatment. Where the sign for E is $-$, factor E is set to its low level. The resulting design is called the **principal fraction** of the 2^5 design. Table 3.11 presents the signs for the main effects and selected interactions of this design.

Table 3.11: Sign Table for the Main Effects and Selected Interactions for the Principal Fraction of a 2^5 Factorial Experiment.

Treatment	A	B	C	D	$E = ABCD$	AB	CDE	$ACDE$
e	$-$	$-$	$-$	$-$	$+$	$+$	$+$	$-$
a	$+$	$-$	$-$	$-$	$-$	$-$	$-$	$-$
b	$-$	$+$	$-$	$-$	$-$	$-$	$-$	$+$
abe	$+$	$+$	$-$	$-$	$+$	$+$	$+$	$+$
c	$-$	$-$	$+$	$-$	$-$	$+$	$+$	$-$
ace	$+$	$-$	$+$	$-$	$+$	$-$	$-$	$-$
bce	$-$	$+$	$+$	$-$	$+$	$-$	$-$	$+$
abc	$+$	$+$	$+$	$-$	$-$	$+$	$+$	$+$
d	$-$	$-$	$-$	$+$	$-$	$+$	$+$	$-$
ade	$+$	$-$	$-$	$+$	$+$	$-$	$-$	$-$
bde	$-$	$+$	$-$	$+$	$+$	$-$	$-$	$+$
abd	$+$	$+$	$-$	$+$	$-$	$+$	$+$	$+$
cde	$-$	$-$	$+$	$+$	$+$	$+$	$+$	$-$
acd	$+$	$-$	$+$	$+$	$-$	$-$	$-$	$-$
bcd	$-$	$+$	$+$	$+$	$-$	$-$	$-$	$+$
$abcde$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$

There is a price to be paid for using only half of the treatments. To see this, note that in Table 3.11 the AB interaction has the same signs as the CDE interaction, and the $ACDE$ interaction has the same signs as the main effect for B. When two effects have the same signs, they are said to be **aliased**. In fact, the main effects and interactions in a half-fraction form pairs in which each member of the pair is aliased with the other. The alias pairs for this half-fraction of the 2^5 design are:

$$\{A, BCDE\} \quad \{B, ACDE\} \quad \{C, ABDE\} \quad \{D, ABCE\} \quad \{E, ABCD\}$$
$$\{AB, CDE\} \quad \{AC, BDE\} \quad \{AD, BCE\} \quad \{AE, BCD\} \quad \{BC, ADE\}$$
$$\{BD, ACE\} \quad \{BE, ACD\} \quad \{CD, BCD\} \quad \{CE, ADE\} \quad \{DE, ACE\}$$

When two effects are aliased, their effect estimates are the same, because they involve the same signs. In fact, when the principal fraction of a design is used, the estimate of any effect actually represents the sum of that effect and its alias. Therefore for the principal fraction of a 2^5 design, each main effect estimate actually represents the sum of the main effect plus its aliased four-way interaction, and each two-way interaction estimate represents the sum of the two-way interaction and its aliased three-way interaction.

In many cases, it is reasonable to assume that the higher-order interactions are small. In the 2^5 half-replicate, if the four-way interactions are negligible, the main effect estimates will be accurate. If in addition the three-way interactions are negligible, the two-way interaction estimates will be accurate as well.

In a fractional design without replication, there is often no good way to compute an error sum of squares, and therefore no rigorous way to test the hypotheses that the effects are equal to 0. In many cases, the purpose of a fractional design is simply to identify a few factors that appear to have the greatest impact

on the outcome. This information may then be used to design more elaborate experiments to investigate these factors. For this purpose, it may be enough simply to choose those factors whose effects or two-way interactions are unusually large, without performing hypothesis tests. This can be done by listing the estimates in decreasing order, then looking to see if there are a few that are noticeably larger than the rest. Another method is to plot the effect and interaction estimates on a normal probability plot, as discussed above.

Example 3.28

In an emulsion liquid membrane system, an emulsion (internal phase) is dispersed into an external liquid medium containing a contaminant. The contaminant is removed from the external liquid through mass transfer into the emulsion. Internal phase leakage occurs when portions of the extracted material spill into the external liquid. In the article "Leakage and Swell in Emulsion Liquid Membrane Systems: Batch Experiments" (R. Pfeiffer, W. Navidi, and A. Bunge, *Separation Science and Technology*, 2003:519–539), the effects of five factors were studied to determine the effect on leakage in a certain system. The five factors were A: Surfactant concentration, B: Internal phase lithium hydroxide concentration, C: Membrane phase, D: Internal phase volume fraction, and E: Extraction vessel stirring rate. A half-fraction of a 2^5 design was used. The data are below (in the actual experiment, each point actually represented the average of two measurements). Leakage is measured in units of percent. Assume that the third-, fourth-, and fifth-order interactions are negligible. Estimate the main effects and two-way interactions. Which, if any, stand out as being noticeably larger than the rest?

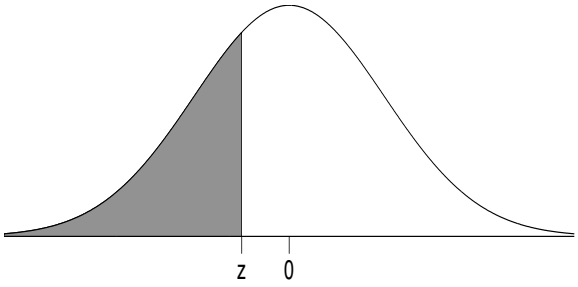
Treatment	Leakage	Treatment	Leakage	Treatment	Leakage	Treatment	Leakage
e	0.61	c	0.35	d	2.03	cde	1.45
a	0.13	ace	0.075	ade	0.64	acd	0.31
b	2.23	bce	7.31	bde	11.72	bcd	1.33
abe	0.095	abc	0.080	abd	0.56	$abcde$	6.24

Solution: Using the sign table, (Table 3.11), we compute estimates for the main effects and two-way interactions, shown in the table below.

Variable	Effect	Variable	Effect
A	-2.36	AE	-1.15
B	3.00	BC	0.20
C	-0.11	BD	0.86
D	1.68	BE	2.65
E	2.64	CD	-1.30
AB	-1.54	CE	0.61
AC	1.43	DE	1.32
AD	0.17		

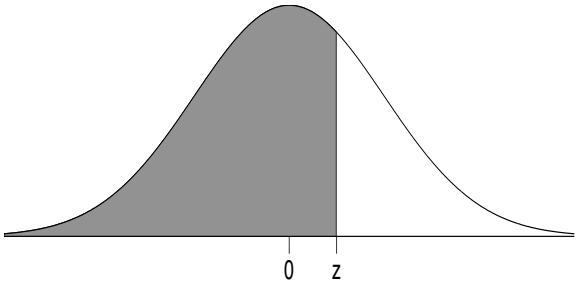
Note that we do not bother to compute sums of squares for the estimates, because we have no SSE to compare them to. To determine informally which effects may be most worthy of further investigation, we rank the estimates in order of their absolute values: B: 3.00, BE: 2.65, E: 2.64, A: -2.36, D: 1.68, and so forth. It seems reasonable to decide that there is a fairly wide gap between the A and D effects, and therefore that factors A, B, and E are most likely to be important. ∎

Table 1: Cumulative Normal Distribution



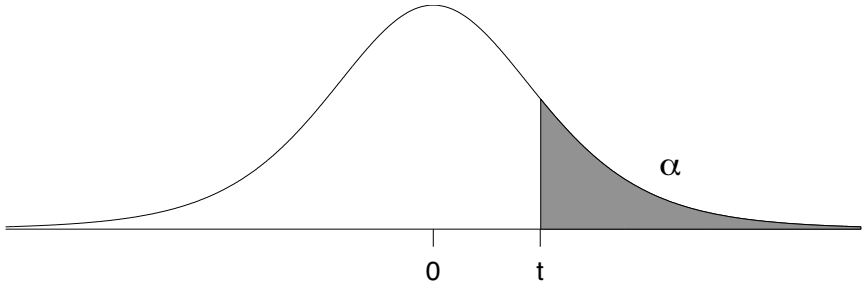
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table 1: Cumulative Normal Distribution (continued)



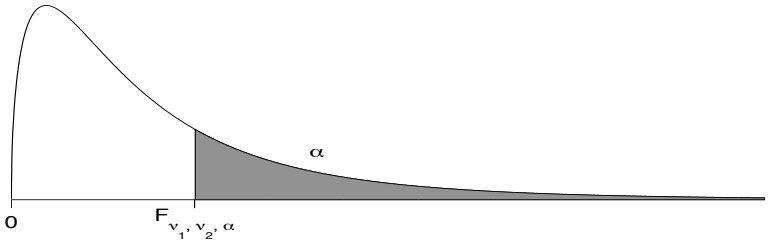
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

Table 2: Upper Percentage Points for the Student's t Distribution



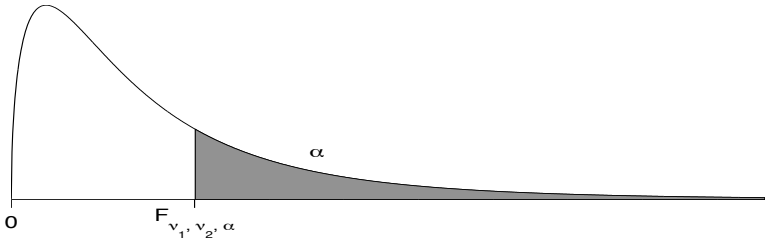
	α								
ν	.40	.25	.10	.05	.025	.01	.005	.001	.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.385	3.646
35	0.255	0.682	1.306	1.690	2.030	2.438	2.724	3.340	3.591
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Table 3: Upper Percentage Points for the F Distribution



		ν_1								
ν_2	α	1	2	3	4	5	6	7	8	9
1	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
1	.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
1	.010	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47
1	.001	405284	500012	540382	562501	576405	585938	592874	598144	603040
2	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
2	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
2	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
2	.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
3	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
3	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
3	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
3	.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
4	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
4	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
4	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
4	.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
5	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
5	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
5	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
5	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24
6	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
6	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
6	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
6	.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69
7	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
7	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
7	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
7	.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33
8	.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
8	.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
8	.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
8	.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77
9	.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
9	.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
9	.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
9	.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11

Table 3: Upper Percentage Points for the F Distribution (continued)



		ν_1								
ν_2	α	10	12	15	20	25	30	40	50	60
1	.100	60.19	60.71	61.22	61.74	62.05	62.26	62.53	62.69	62.79
1	.050	241.88	243.91	245.95	248.01	249.26	250.10	251.14	251.77	252.20
1	.010	6055.85	6106.32	6157.29	6208.73	6239.83	6260.65	6286.78	6302.52	6313.03
1	.001	606316	611276	616292	621362	624430	626486	659725	660511	6610390
2	.100	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.47
2	.050	19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48	19.48
2	.010	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.48
2	.001	999.40	999.42	999.43	999.45	999.46	999.47	999.47	999.48	999.48
3	.100	5.23	5.22	5.20	5.18	5.17	5.17	5.16	5.15	5.15
3	.050	8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.58	8.57
3	.010	27.23	27.05	26.87	26.69	26.58	26.50	26.41	26.35	26.32
3	.001	129.25	128.32	127.37	126.42	125.84	125.45	124.96	124.66	124.47
4	.100	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.80	3.79
4	.050	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.70	5.69
4	.010	14.55	14.37	14.20	14.02	13.91	13.84	13.75	13.69	13.65
4	.001	48.05	47.41	46.76	46.10	45.70	45.43	45.09	44.88	44.75
5	.100	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.15	3.14
5	.050	4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.44	4.43
5	.010	10.05	9.89	9.72	9.55	9.45	9.38	9.29	9.24	9.20
5	.001	26.92	26.42	25.91	25.39	25.08	24.87	24.60	24.44	24.33
6	.100	2.94	2.90	2.87	2.84	2.81	2.80	2.78	2.77	2.76
6	.050	4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.75	3.74
6	.010	7.87	7.72	7.56	7.40	7.30	7.23	7.14	7.09	7.06
6	.001	18.41	17.99	17.56	17.12	16.85	16.67	16.44	16.31	16.21
7	.100	2.70	2.67	2.63	2.59	2.57	2.56	2.54	2.52	2.51
7	.050	3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.32	3.30
7	.010	6.62	6.47	6.31	6.16	6.06	5.99	5.91	5.86	5.82
7	.001	14.08	13.71	13.32	12.93	12.69	12.53	12.33	12.20	12.12
8	.100	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.35	2.34
8	.050	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.02	3.01
8	.010	5.81	5.67	5.52	5.36	5.26	5.20	5.12	5.07	5.03
8	.001	11.54	11.19	10.84	10.48	10.26	10.11	9.92	9.80	9.73
9	.100	2.42	2.38	2.34	2.30	2.27	2.25	2.23	2.22	2.21
9	.050	3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.80	2.79
9	.010	5.26	5.11	4.96	4.81	4.71	4.65	4.57	4.52	4.48
9	.001	9.89	9.57	9.24	8.90	8.69	8.55	8.37	8.26	8.19

Table 3: Upper Percentage Points for the F Distribution (continued)

		ν_1								
ν_2	α	1	2	3	4	5	6	7	8	9
10	0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
10	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
10	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
10	0.001	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.96
11	0.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
11	0.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
11	0.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
11	0.001	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12
12	0.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
12	0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
12	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
12	0.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48
13	0.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
13	0.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
13	0.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
13	0.001	17.82	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98
14	0.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
14	0.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
14	0.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
14	0.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.58
15	0.100	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
15	0.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
15	0.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
15	0.001	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26
16	0.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
16	0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
16	0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
16	0.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.19	5.98
17	0.100	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
17	0.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
17	0.010	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
17	0.001	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75
18	0.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
18	0.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
18	0.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
18	0.001	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56
19	0.100	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
19	0.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
19	0.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
19	0.001	15.08	10.16	8.28	7.27	6.62	6.18	5.85	5.59	5.39
20	0.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
20	0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
20	0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
20	0.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24

Table 3: Upper Percentage Points for the F Distribution (continued)

		ν_1								
ν_2	α	10	12	15	20	25	30	40	50	60
10	0.100	2.32	2.28	2.24	2.20	2.17	2.16	2.13	2.12	2.11
10	0.050	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.64	2.62
10	0.010	4.85	4.71	4.56	4.41	4.31	4.25	4.17	4.12	4.08
10	0.001	8.75	8.45	8.13	7.80	7.60	7.47	7.30	7.19	7.12
11	0.100	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.04	2.03
11	0.050	2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.51	2.49
11	0.010	4.54	4.40	4.25	4.10	4.01	3.94	3.86	3.81	3.78
11	0.001	7.92	7.63	7.32	7.01	6.81	6.68	6.52	6.42	6.35
12	0.100	2.19	2.15	2.10	2.06	2.03	2.01	1.99	1.97	1.96
12	0.050	2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.40	2.38
12	0.010	4.30	4.16	4.01	3.86	3.76	3.70	3.62	3.57	3.54
12	0.001	7.29	7.00	6.71	6.40	6.22	6.09	5.93	5.83	5.76
13	0.100	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.92	1.90
13	0.050	2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.31	2.30
13	0.010	4.10	3.96	3.82	3.66	3.57	3.51	3.43	3.38	3.34
13	0.001	6.80	6.52	6.23	5.93	5.75	5.63	5.47	5.37	5.30
14	0.100	2.10	2.05	2.01	1.96	1.93	1.91	1.89	1.87	1.86
14	0.050	2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.24	2.22
14	0.010	3.94	3.80	3.66	3.51	3.41	3.35	3.27	3.22	3.18
14	0.001	6.40	6.13	5.85	5.56	5.38	5.25	5.10	5.00	4.94
15	0.100	2.06	2.02	1.97	1.92	1.89	1.87	1.85	1.83	1.82
15	0.050	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.18	2.16
15	0.010	3.80	3.67	3.52	3.37	3.28	3.21	3.13	3.08	3.05
15	0.001	6.08	5.81	5.54	5.25	5.07	4.95	4.80	4.70	4.64
16	0.100	2.03	1.99	1.94	1.89	1.86	1.84	1.81	1.79	1.78
16	0.050	2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.12	2.11
16	0.010	3.69	3.55	3.41	3.26	3.16	3.10	3.02	2.97	2.93
16	0.001	5.81	5.55	5.27	4.99	4.82	4.70	4.54	4.45	4.39
17	0.100	2.00	1.96	1.91	1.86	1.83	1.81	1.78	1.76	1.75
17	0.050	2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.08	2.06
17	0.010	3.59	3.46	3.31	3.16	3.07	3.00	2.92	2.87	2.83
17	0.001	5.58	5.32	5.05	4.78	4.60	4.48	4.33	4.24	4.18
18	0.100	1.98	1.93	1.89	1.84	1.80	1.78	1.75	1.74	1.72
18	0.050	2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.04	2.02
18	0.010	3.51	3.37	3.23	3.08	2.98	2.92	2.84	2.78	2.75
18	0.001	5.39	5.13	4.87	4.59	4.42	4.30	4.15	4.06	4.00
19	0.100	1.96	1.91	1.86	1.81	1.78	1.76	1.73	1.71	1.70
19	0.050	2.38	2.31	2.23	2.16	2.11	2.07	2.03	2.00	1.98
19	0.010	3.43	3.30	3.15	3.00	2.91	2.84	2.76	2.71	2.67
19	0.001	5.22	4.97	4.70	4.43	4.26	4.14	3.99	3.90	3.84
20	0.100	1.94	1.89	1.84	1.79	1.76	1.74	1.71	1.69	1.68
20	0.050	2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.97	1.95
20	0.010	3.37	3.23	3.09	2.94	2.84	2.78	2.69	2.64	2.61
20	0.001	5.08	4.82	4.56	4.29	4.12	4.00	3.86	3.77	3.70

Table 3: Upper Percentage Points for the F Distribution (continued)

		ν_1								
ν_2	α	1	2	3	4	5	6	7	8	9
21	0.100	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
21	0.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
21	0.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
21	0.001	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11
22	0.100	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
22	0.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
22	0.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
22	0.001	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99
23	0.100	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
23	0.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
23	0.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
23	0.001	14.20	9.47	7.67	6.70	6.08	5.65	5.33	5.09	4.89
24	0.100	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
24	0.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
24	0.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
24	0.001	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80
25	0.100	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
25	0.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
25	0.010	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
25	0.001	13.88	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71
26	0.100	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
26	0.050	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
26	0.010	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
26	0.001	13.74	9.12	7.36	6.41	5.80	5.38	5.07	4.83	4.64
27	0.100	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
27	0.050	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
27	0.010	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
27	0.001	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57
28	0.100	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
28	0.050	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
28	0.010	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
28	0.001	13.50	8.93	7.19	6.25	5.66	5.24	4.93	4.69	4.50
29	0.100	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
29	0.050	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
29	0.010	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
29	0.001	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45
30	0.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
30	0.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
30	0.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
30	0.001	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39
31	0.100	2.87	2.48	2.27	2.14	2.04	1.97	1.92	1.88	1.84
31	0.050	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20
31	0.010	7.53	5.36	4.48	3.99	3.67	3.45	3.28	3.15	3.04
31	0.001	13.20	8.70	6.99	6.07	5.48	5.07	4.77	4.53	4.34

Table 3: Upper Percentage Points for the F Distribution (continued)

		ν_1								
ν_2	α	10	12	15	20	25	30	40	50	60
21	0.100	1.92	1.87	1.83	1.78	1.74	1.72	1.69	1.67	1.66
21	0.050	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.94	1.92
21	0.010	3.31	3.17	3.03	2.88	2.79	2.72	2.64	2.58	2.55
21	0.001	4.95	4.70	4.44	4.17	4.00	3.88	3.74	3.64	3.58
22	0.100	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.65	1.64
22	0.050	2.30	2.23	2.15	2.07	2.02	1.98	1.94	1.91	1.89
22	0.010	3.26	3.12	2.98	2.83	2.73	2.67	2.58	2.53	2.50
22	0.001	4.83	4.58	4.33	4.06	3.89	3.78	3.63	3.54	3.48
23	0.100	1.89	1.84	1.80	1.74	1.71	1.69	1.66	1.64	1.62
23	0.050	2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.88	1.86
23	0.010	3.21	3.07	2.93	2.78	2.69	2.62	2.54	2.48	2.45
23	0.001	4.73	4.48	4.23	3.96	3.79	3.68	3.53	3.44	3.38
24	0.100	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.62	1.61
24	0.050	2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.86	1.84
24	0.010	3.17	3.03	2.89	2.74	2.64	2.58	2.49	2.44	2.40
24	0.001	4.64	4.39	4.14	3.87	3.71	3.59	3.45	3.36	3.29
25	0.100	1.87	1.82	1.77	1.72	1.68	1.66	1.63	1.61	1.59
25	0.050	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.84	1.82
25	0.010	3.13	2.99	2.85	2.70	2.60	2.54	2.45	2.40	2.36
25	0.001	4.56	4.31	4.06	3.79	3.63	3.52	3.37	3.28	3.22
26	0.100	1.86	1.81	1.76	1.71	1.67	1.65	1.61	1.59	1.58
26	0.050	2.22	2.15	2.07	1.99	1.94	1.90	1.85	1.82	1.80
26	0.010	3.09	2.96	2.81	2.66	2.57	2.50	2.42	2.36	2.33
26	0.001	4.48	4.24	3.99	3.72	3.56	3.44	3.30	3.21	3.15
27	0.100	1.85	1.80	1.75	1.70	1.66	1.64	1.60	1.58	1.57
27	0.050	2.20	2.13	2.06	1.97	1.92	1.88	1.84	1.81	1.79
27	0.010	3.06	2.93	2.78	2.63	2.54	2.47	2.38	2.33	2.29
27	0.001	4.41	4.17	3.92	3.66	3.49	3.38	3.23	3.14	3.08
28	0.100	1.84	1.79	1.74	1.69	1.65	1.63	1.59	1.57	1.56
28	0.050	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.79	1.77
28	0.010	3.03	2.90	2.75	2.60	2.51	2.44	2.35	2.30	2.26
28	0.001	4.35	4.11	3.86	3.60	3.43	3.32	3.18	3.09	3.02
29	0.100	1.83	1.78	1.73	1.68	1.64	1.62	1.58	1.56	1.55
29	0.050	2.18	2.10	2.03	1.94	1.89	1.85	1.81	1.77	1.75
29	0.010	3.00	2.87	2.73	2.57	2.48	2.41	2.33	2.27	2.23
29	0.001	4.29	4.05	3.80	3.54	3.38	3.27	3.12	3.03	2.97
30	0.100	1.82	1.77	1.72	1.67	1.63	1.61	1.57	1.55	1.54
30	0.050	2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.76	1.74
30	0.010	2.98	2.84	2.70	2.55	2.45	2.39	2.30	2.25	2.21
30	0.001	4.24	4.00	3.75	3.49	3.33	3.22	3.07	2.98	2.92
31	0.100	1.81	1.77	1.71	1.66	1.62	1.60	1.56	1.54	1.53
31	0.050	2.15	2.08	2.00	1.92	1.87	1.83	1.78	1.75	1.73
31	0.010	2.96	2.82	2.68	2.52	2.43	2.36	2.27	2.22	2.18
31	0.001	4.19	3.95	3.71	3.45	3.28	3.17	3.03	2.94	2.87

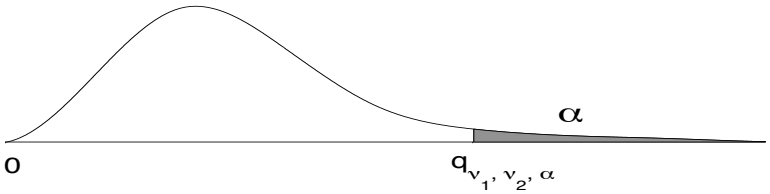
Table 3: Upper Percentage Points for the F Distribution (continued)

		ν_1								
ν_2	α	1	2	3	4	5	6	7	8	9
32	0.100	2.87	2.48	2.26	2.13	2.04	1.97	1.91	1.87	1.83
32	0.050	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19
32	0.010	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02
32	0.001	13.12	8.64	6.94	6.01	5.43	5.02	4.72	4.48	4.30
33	0.100	2.86	2.47	2.26	2.12	2.03	1.96	1.91	1.86	1.83
33	0.050	4.14	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18
33	0.010	7.47	5.31	4.44	3.95	3.63	3.41	3.24	3.11	3.00
33	0.001	13.04	8.58	6.88	5.97	5.38	4.98	4.67	4.44	4.26
34	0.100	2.86	2.47	2.25	2.12	2.02	1.96	1.90	1.86	1.82
34	0.050	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17
34	0.010	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98
34	0.001	12.97	8.52	6.83	5.92	5.34	4.93	4.63	4.40	4.22
35	0.100	2.85	2.46	2.25	2.11	2.02	1.95	1.90	1.85	1.82
35	0.050	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16
35	0.010	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96
35	0.001	12.90	8.47	6.79	5.88	5.30	4.89	4.59	4.36	4.18
36	0.100	2.85	2.46	2.24	2.11	2.01	1.94	1.89	1.85	1.81
36	0.050	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15
36	0.010	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95
36	0.001	12.83	8.42	6.74	5.84	5.26	4.86	4.56	4.33	4.14
37	0.100	2.85	2.45	2.24	2.10	2.01	1.94	1.89	1.84	1.81
37	0.050	4.11	3.25	2.86	2.63	2.47	2.36	2.27	2.20	2.14
37	0.010	7.37	5.23	4.36	3.87	3.56	3.33	3.17	3.04	2.93
37	0.001	12.77	8.37	6.70	5.80	5.22	4.82	4.53	4.30	4.11
38	0.100	2.84	2.45	2.23	2.10	2.01	1.94	1.88	1.84	1.80
38	0.050	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14
38	0.010	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92
38	0.001	12.71	8.33	6.66	5.76	5.19	4.79	4.49	4.26	4.08
39	0.100	2.84	2.44	2.23	2.09	2.00	1.93	1.88	1.83	1.80
39	0.050	4.09	3.24	2.85	2.61	2.46	2.34	2.26	2.19	2.13
39	0.010	7.33	5.19	4.33	3.84	3.53	3.30	3.14	3.01	2.90
39	0.001	12.66	8.29	6.63	5.73	5.16	4.76	4.46	4.23	4.05
40	0.100	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
40	0.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
40	0.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
40	0.001	12.61	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02
50	0.100	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76
50	0.050	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
50	0.010	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78
50	0.001	12.22	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.82
60	0.100	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
60	0.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
60	0.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
60	0.001	11.97	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69
120	0.100	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
120	0.050	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
120	0.010	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
120	0.001	11.38	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38

Table 3: Upper Percentage Points for the F Distribution (continued)

		ν_1								
ν_2	α	10	12	15	20	25	30	40	50	60
32	0.100	1.81	1.76	1.71	1.65	1.62	1.59	1.56	1.53	1.52
32	0.050	2.14	2.07	1.99	1.91	1.85	1.82	1.77	1.74	1.71
32	0.010	2.93	2.80	2.65	2.50	2.41	2.34	2.25	2.20	2.16
32	0.001	4.14	3.91	3.66	3.40	3.24	3.13	2.98	2.89	2.83
33	0.100	1.80	1.75	1.70	1.64	1.61	1.58	1.55	1.53	1.51
33	0.050	2.13	2.06	1.98	1.90	1.84	1.81	1.76	1.72	1.70
33	0.010	2.91	2.78	2.63	2.48	2.39	2.32	2.23	2.18	2.14
33	0.001	4.10	3.87	3.62	3.36	3.20	3.09	2.94	2.85	2.79
34	0.100	1.79	1.75	1.69	1.64	1.60	1.58	1.54	1.52	1.50
34	0.050	2.12	2.05	1.97	1.89	1.83	1.80	1.75	1.71	1.69
34	0.010	2.89	2.76	2.61	2.46	2.37	2.30	2.21	2.16	2.12
34	0.001	4.06	3.83	3.58	3.33	3.16	3.05	2.91	2.82	2.75
35	0.100	1.79	1.74	1.69	1.63	1.60	1.57	1.53	1.51	1.50
35	0.050	2.11	2.04	1.96	1.88	1.82	1.79	1.74	1.70	1.68
35	0.010	2.88	2.74	2.60	2.44	2.35	2.28	2.19	2.14	2.10
35	0.001	4.03	3.79	3.55	3.29	3.13	3.02	2.87	2.78	2.72
36	0.100	1.78	1.73	1.68	1.63	1.59	1.56	1.53	1.51	1.49
36	0.050	2.11	2.03	1.95	1.87	1.81	1.78	1.73	1.69	1.67
36	0.010	2.86	2.72	2.58	2.43	2.33	2.26	2.18	2.12	2.08
36	0.001	3.99	3.76	3.51	3.26	3.10	2.98	2.84	2.75	2.69
37	0.100	1.78	1.73	1.68	1.62	1.58	1.56	1.52	1.50	1.48
37	0.050	2.10	2.02	1.95	1.86	1.81	1.77	1.72	1.68	1.66
37	0.010	2.84	2.71	2.56	2.41	2.31	2.25	2.16	2.10	2.06
37	0.001	3.96	3.73	3.48	3.23	3.07	2.95	2.81	2.72	2.66
38	0.100	1.77	1.72	1.67	1.61	1.58	1.55	1.52	1.49	1.48
38	0.050	2.09	2.02	1.94	1.85	1.80	1.76	1.71	1.68	1.65
38	0.010	2.83	2.69	2.55	2.40	2.30	2.23	2.14	2.09	2.05
38	0.001	3.93	3.70	3.45	3.20	3.04	2.92	2.78	2.69	2.63
39	0.100	1.77	1.72	1.67	1.61	1.57	1.55	1.51	1.49	1.47
39	0.050	2.08	2.01	1.93	1.85	1.79	1.75	1.70	1.67	1.65
39	0.010	2.81	2.68	2.54	2.38	2.29	2.22	2.13	2.07	2.03
39	0.001	3.90	3.67	3.43	3.17	3.01	2.90	2.75	2.66	2.60
40	0.100	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.48	1.47
40	0.050	2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.66	1.64
40	0.010	2.80	2.66	2.52	2.37	2.27	2.20	2.11	2.06	2.02
40	0.001	3.87	3.64	3.40	3.14	2.98	2.87	2.73	2.64	2.57
50	0.100	1.73	1.68	1.63	1.57	1.53	1.50	1.46	1.44	1.42
50	0.050	2.03	1.95	1.87	1.78	1.73	1.69	1.63	1.60	1.58
50	0.010	2.70	2.56	2.42	2.27	2.17	2.10	2.01	1.95	1.91
50	0.001	3.67	3.44	3.20	2.95	2.79	2.68	2.53	2.44	2.38
60	0.100	1.71	1.66	1.60	1.54	1.50	1.48	1.44	1.41	1.40
60	0.050	1.99	1.92	1.84	1.75	1.69	1.65	1.59	1.56	1.53
60	0.010	2.63	2.50	2.35	2.20	2.10	2.03	1.94	1.88	1.84
60	0.001	3.54	3.32	3.08	2.83	2.67	2.55	2.41	2.32	2.25
120	0.100	1.65	1.60	1.55	1.48	1.44	1.41	1.37	1.34	1.32
120	0.050	1.91	1.83	1.75	1.66	1.60	1.55	1.50	1.46	1.43
120	0.010	2.47	2.34	2.19	2.03	1.93	1.86	1.76	1.70	1.66
120	0.001	3.24	3.02	2.78	2.53	2.37	2.26	2.11	2.02	1.95

Table 4: Upper Percentage Points for the Studentized Range q_{ν_1,ν_2}



		ν_1													
ν_2	α	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.10	8.93	13.44	16.36	18.49	20.15	21.51	22.64	23.62	24.48	25.24	25.92	26.54	27.10	27.62
1	0.05	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07	50.59	51.96	53.20	54.33	55.36
1	0.01	90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6	253.2	260.0	266.2	271.8	277.0
2	0.10	4.13	5.73	6.77	7.54	8.14	8.63	9.05	9.41	9.72	10.01	10.26	10.49	10.70	10.89
2	0.05	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99	14.39	14.75	15.08	15.38	15.65
2	0.01	14.04	19.02	22.29	24.72	26.63	28.20	29.53	30.68	31.69	32.59	33.40	34.13	34.81	35.43
3	0.10	3.33	4.47	5.20	5.74	6.16	6.51	6.81	7.06	7.29	7.49	7.67	7.83	7.98	8.12
3	0.05	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72	9.95	10.15	10.35	10.52
3	0.01	8.26	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69	17.13	17.53	17.89	18.22	18.52
4	0.10	3.01	3.98	4.59	5.03	5.39	5.68	5.93	6.14	6.33	6.49	6.65	6.78	6.91	7.02
4	0.05	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.66
4	0.01	6.51	8.12	9.17	9.96	10.58	11.10	11.55	11.93	12.27	12.57	12.84	13.09	13.32	13.53
5	0.10	2.85	3.72	4.26	4.66	4.98	5.24	5.46	5.65	5.82	5.97	6.10	6.22	6.34	6.44
5	0.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72
5	0.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48	10.70	10.89	11.08	11.24
6	0.10	2.75	3.56	4.07	4.44	4.73	4.97	5.17	5.34	5.50	5.64	5.76	5.87	5.98	6.07
6	0.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14
6	0.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.48	9.65	9.81	9.95
7	0.10	2.68	3.45	3.93	4.28	4.55	4.78	4.97	5.14	5.28	5.41	5.53	5.64	5.74	5.83
7	0.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76
7	0.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55	8.71	8.86	9.00	9.12
8	0.10	2.63	3.37	3.83	4.17	4.43	4.65	4.83	4.99	5.13	5.25	5.36	5.46	5.56	5.64
8	0.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48
8	0.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86	8.03	8.18	8.31	8.44	8.55
9	0.10	2.59	3.32	3.76	4.08	4.34	4.54	4.72	4.87	5.01	5.13	5.23	5.33	5.42	5.51
9	0.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98	6.09	6.19	6.28
9	0.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49	7.65	7.78	7.91	8.03	8.13
10	0.10	2.56	3.27	3.70	4.02	4.26	4.47	4.64	4.78	4.91	5.03	5.13	5.23	5.32	5.40
10	0.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11
10	0.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.49	7.60	7.71	7.81
11	0.10	2.54	3.23	3.66	3.96	4.20	4.40	4.57	4.71	4.84	4.95	5.05	5.15	5.23	5.31
11	0.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98
11	0.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25	7.36	7.46	7.56
12	0.10	2.52	3.20	3.62	3.92	4.16	4.35	4.51	4.65	4.78	4.89	4.99	5.08	5.16	5.24
12	0.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61	5.71	5.80	5.88
12	0.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06	7.17	7.26	7.36

Table 4: Upper Percentage Points for the Studentized Range q_{ν_1,ν_2} (continued)

		ν_1													
ν_2	α	2	3	4	5	6	7	8	9	10	11	12	13	14	15
13	0.10	2.50	3.18	3.59	3.88	4.12	4.30	4.46	4.60	4.72	4.83	4.93	5.02	5.10	5.18
13	0.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79
13	0.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	6.90	7.01	7.10	7.19
14	0.10	2.49	3.16	3.56	3.85	4.08	4.27	4.42	4.56	4.68	4.79	4.88	4.97	5.05	5.12
14	0.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.71
14	0.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77	6.87	6.96	7.05
15	0.10	2.48	3.14	3.54	3.83	4.05	4.23	4.39	4.52	4.64	4.75	4.84	4.93	5.01	5.08
15	0.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65
15	0.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66	6.76	6.84	6.93
16	0.10	2.47	3.12	3.52	3.80	4.03	4.21	4.36	4.49	4.61	4.71	4.81	4.89	4.97	5.04
16	0.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59
16	0.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56	6.66	6.74	6.82
17	0.10	2.46	3.11	3.50	3.78	4.00	4.18	4.33	4.46	4.58	4.68	4.77	4.86	4.93	5.01
17	0.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.54
17	0.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48	6.57	6.66	6.73
18	0.10	2.45	3.10	3.49	3.77	3.98	4.16	4.31	4.44	4.55	4.65	4.75	4.83	4.90	4.98
18	0.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27	5.35	5.43	5.50
18	0.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41	6.50	6.58	6.65
19	0.10	2.45	3.09	3.47	3.75	3.97	4.14	4.29	4.42	4.53	4.63	4.72	4.80	4.88	4.95
19	0.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23	5.31	5.39	5.46
19	0.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34	6.43	6.51	6.58
20	0.10	2.44	3.08	3.46	3.74	3.95	4.12	4.27	4.40	4.51	4.61	4.70	4.78	4.85	4.92
20	0.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43
20	0.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.28	6.37	6.45	6.52
24	0.10	2.42	3.05	3.42	3.69	3.90	4.07	4.21	4.34	4.44	4.54	4.63	4.71	4.78	4.85
24	0.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32
24	0.01	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33
30	0.10	2.40	3.02	3.39	3.65	3.85	4.02	4.16	4.28	4.38	4.47	4.56	4.64	4.71	4.77
30	0.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92	5.00	5.08	5.15	5.21
30	0.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85	5.93	6.01	6.08	6.14
40	0.10	2.38	2.99	3.35	3.60	3.80	3.96	4.10	4.21	4.32	4.41	4.49	4.56	4.63	4.69
40	0.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82	4.90	4.98	5.04	5.11
40	0.01	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60	5.69	5.76	5.83	5.90	5.96
60	0.10	2.36	2.96	3.31	3.56	3.75	3.91	4.04	4.16	4.25	4.34	4.42	4.49	4.56	4.62
60	0.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00
60	0.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.78
120	0.10	2.34	2.93	3.28	3.52	3.71	3.86	3.99	4.10	4.19	4.28	4.35	4.42	4.48	4.54
120	0.05	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90
120	0.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.37	5.44	5.50	5.56	5.61
∞	0.10	2.33	2.90	3.24	3.48	3.66	3.81	3.93	4.04	4.13	4.21	4.28	4.35	4.41	4.47
∞	0.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80
∞	0.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45