6 Experiments with Two Crossed Treatment Factors

6.1 Introduction

In this chapter, we discuss the use of completely randomized designs for experiments that involve two crossed treatment factors. We label the treatment factors as A and B, where factor A has a levels coded $1, 2, \ldots, a$, and factor B has b levels coded $1, 2, \ldots, b$. Factors are *crossed* if every combination of levels may be observed. For experiments considered in this chapter, every level of A is observed with every level of B, so the factors are crossed. In total, there are $v = ab$ treatments (treatment combinations), and these are coded as $11, 12, \ldots, 1b, 21, 22, \ldots, 2b, \ldots, ab$.

In the previous three chapters, we recoded the treatment combinations as $1, 2, \ldots, v$ and used the one-way analysis of variance for comparing their effects. In this chapter, we investigate the contributions that each of the factors make individually to the response, and it is more convenient to retain the 2-digit code ij for a treatment combination in which factor A is at level i and factor B is at level j. In Sect. 6.2.1, we define the "interaction" of two treatment factors. Allowing for the possibility of interaction leads one to select a "two-way complete model" to model the data (Sect. 6.4). However, if it is known in advance that the factors do not interact, a "two-way main-effects model" would be selected (Sect. 6.5). Estimation of contrasts, confidence intervals, and analysis of variance techniques are described for these basic models. The calculation of sample sizes is also discussed (Sect. 6.6). The corresponding commands for SAS and R software are described in Sects. 6.8 and 6.9, respectively.

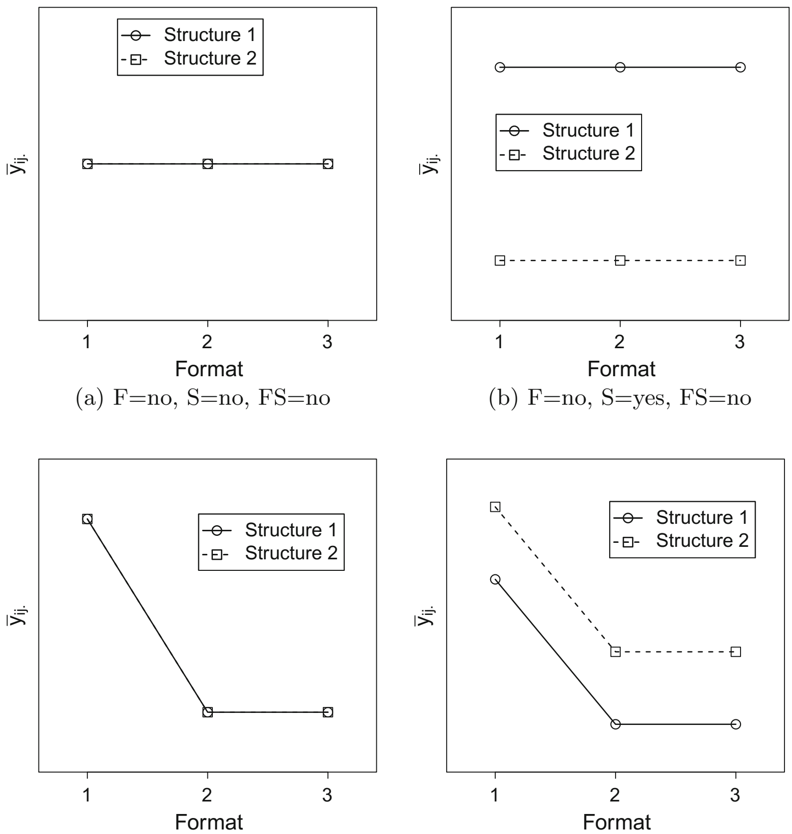
If each of the two factors has a large number of levels, the total number of treatment combinations could be quite large. When observations are costly, it may be necessary to limit the number of observations to one per treatment combination. Analysis for this situation is discussed in Sect. 6.7.

6.2 Models and Factorial Effects

6.2.1 The Meaning of Interaction

In order to understand the meaning of the interaction between two treatment factors, it is helpful to look at possible data sets from a hypothetical experiment. Universities have become increasingly interested in online courses and other nontraditional modes of instruction. While an online course may be offered for a group of students and involve interaction between the students in a common section, consider development of a course that students take independently of one another. Suppose that a hypothetical statistics department wishes to know to what extent student performance in an introductory

A. Dean et al., *Design and Analysis of Experiments*,
Springer Texts in Statistics, DOI 10.1007/978-3-319-52250-0_6



(a) F=no, S=no, FS=no

(b) F=no, S=yes, FS=no

(c) F=yes, S=no, FS=no

(d) F=yes, S=yes, FS=no

Fig. 6.1 Possible configurations of effects present for two factors, presentation format (F) and course structure (S) when the significant interaction effect is absent

online course is affected by the primary presentation format (textbook reading assignments, videotaped lectures, or interactive software) and course structure (structured, with regular deadlines throughout the term; or unstructured, with only a deadline to finish by the end of the term).

There are two treatment factors of interest, namely "presentation format," which has three levels, coded 1, 2, and 3, and "course structure," which has two levels, coded 1 and 2. Both of the treatment factors have fixed effects, since their levels have been specifically chosen (see Sect. 2.2, p. 11, step (f)). The students who enroll in the introductory course are the experimental units and are allocated at random to one of the six treatment combinations in such a way that approximately equal numbers of students are assigned to each combination of presentation format and course structure. Student performance is to be measured by means of a computer-graded multiple-choice examination, and an average exam score $\overline{y}_{ij.}$ for each treatment combination will be obtained, averaging over students for each treatment combination.

There are eight different types of situations that could occur, and these are depicted in Figs. 6.1 and 6.2, where the plotted character indicates the course structure used. The plots are called *interaction plots* and give an indication of how the different format–structure combinations affect the average exam score.

In plots (a)–(d) of Fig. 6.1, the lines joining the average exam scores for the two course structures are parallel (and sometimes coincide). In plot (b), all the presentation formats have obtained higher exam scores with course structure 1 than with structure 2, but the presentation formats themselves look very similar in terms of the average exam scores obtained. Thus there is an effect on the average exam

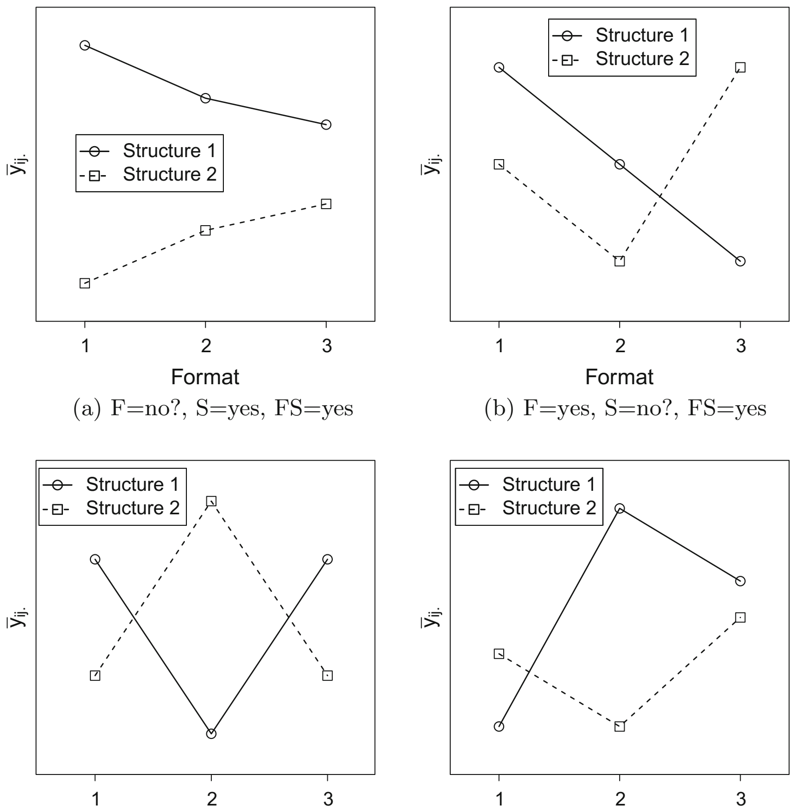


Fig. 6.2 Possible configurations of effects present for two factors, presentation format (F) and course structure (S) when the significant interaction effect is present

(a) F=no?, S=yes, FS=yes

(b) F=yes, S=no?, FS=yes

(c) F=no?, S=no?, FS=yes

(d) F=yes, S=yes, FS=yes

score of course structure (S) but no effect of presentation format (F). Below the plot this is highlighted by the notation "F = no, S = yes." The notation "FS = no" refers to the fact that the lines are parallel, indicating that there is no interaction (see below). In plot (c), no difference can be seen in the average scores obtained from the two course structures for any presentation format, although the presentation formats themselves appear to have achieved different average scores. Thus, the presentation formats have an effect on the average exam score, but the course structures do not (F = yes, S = no). Plot (d) shows the type of plot that might be obtained if there is both a presentation-format effect and a course-structure effect. The plot shows that all three presentation formats have obtained higher average exam scores using structure 1 than using structure 2. But also, presentation format 1 has obtained higher average scores than the other two presentation formats. The individual course-structure effects and presentation-format effects are known as *main effects*.

In plots (a)–(d) of Fig. 6.2, the lines are not parallel. This means that more is needed to explain the differences in exam scores than just course structure and presentation format effects. For example, in plot (a), all presentation formats have obtained higher exam scores using course structure 1 than using structure 2, but the difference is very small for presentation format 3 and very large for presentation format 1. In plot (d), presentation format 1 has obtained higher exam scores with structure 2, while the other two presentation formats have done better with structure 1. In all of plots (a)–(d) the presentation formats have performed differently with the different structures. This is called an effect of *interaction* between presentation format and course structure.

In plot (c), the presentation formats clearly differ. Two do better with structure 1 and one with structure 2. However, if we ignore course structures, the presentation formats appear to have achieved very similar average exam scores overall. So, averaged over the structures, there is little difference between them. In such a case, a standard computer analysis will declare that there is no difference between presentation formats, which is somewhat misleading. We use the notation "FS = yes" to denote an interaction between presentation format and Structure, and "F = no?" to highlight the fact that a conclusion of no difference between presentation formats should be interpreted with caution in the presence of interaction. In general, *if there is an interaction between two treatment factors, then it may not be sensible to examine either of the main effects separately*. Instead, it will often be preferable to compare the effects of the treatment combinations themselves.

While interaction plots are extremely helpful in interpreting the analysis of an experiment, they give no indication of the size of the experimental error. Sometimes a perceived interaction in the plot will not be distinguishable from error variability in the analysis of variance. On the other hand, if the error variability is very small, then an interaction effect may be statistically significant in the analysis, even if it appears negligible in the plot.

6.2.2 Models for Two Treatment Factors

If we use the two-digit codes ij for the treatment combinations in the one-way analysis of variance model (3.3.1), we obtain the model

$$\begin{gathered} Y_{ijt} = \mu + \tau_{ij} + \epsilon_{ijt}\,, \\ \epsilon_{ijt} \sim N(0, \sigma^2)\,, \\ \epsilon_{ijt}\text{'s independent}\,, \\ t = 1, \ldots, r_{ij}; \quad i = 1, \ldots, a; \quad j = 1, \ldots, b, \end{gathered} \tag{6.2.1}$$

where i and j are the levels of A and B, respectively. This model is known as the *cell-means model*. The "cell" refers to the cell of a table whose rows represent the levels of A and whose columns represent the levels of B.

Since the interaction plot arising from a two-factor experiment could be similar to any of the plots of Figs. 6.1 and 6.2, it is often useful to model the effect on the response of treatment combination ij to be the sum of the individual effects of the two factors, together with their interaction; that is,

$$\tau_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}.$$

Here, α_i is the effect (positive or negative) on the response due to the fact that the ith level of factor A is observed, and β_j is the effect (positive or negative) on the response due to the fact that the jth level of factor B is observed, and $(\alpha\beta)_{ij}$ is the extra effect (positive or negative) on the response of observing levels i and j of factors A and B together. The corresponding model, which we call the *two-way complete model,* or the *two-way analysis of variance model,* is as follows:

$$\begin{gathered} Y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}\,, \\ \epsilon_{ijt} \sim N(0, \sigma^2)\,, \\ \epsilon_{ijt}\text{'s are mutually independent}\,, \\ t = 1, \ldots, r_{ij}; \quad i = 1, \ldots, a; \quad j = 1, \ldots, b. \end{gathered} \tag{6.2.2}$$

The phrase "two-way" refers to the fact that there are two primary sources of variation, namely, the two treatment factors. Model (6.2.2) is equivalent to model (6.2.1), since all we have done is to express the effect of the treatment combination in terms of its constituent parts.

Occasionally, an experimenter has sufficient knowledge about the two treatment factors being studied to state with reasonable certainty that the factors do not interact and that an interaction plot similar to one of the plots of Fig. 6.1 will occur. This knowledge may be gleaned from previous similar experiments or from scientific facts about the treatment factors. If this is so, then the interaction term can be dropped from model (6.2.2), which then becomes

$$\begin{aligned} Y_{ijt} &= \mu + \alpha_i + \beta_j + \epsilon_{ijt}\,, \\ \epsilon_{ijt} &\sim N(0, \sigma^2)\,, \\ \epsilon_{ijt}{}'\text{s are} & \text{ mutually independent}\,, \\ t = 1, \ldots, r_{ij}; \quad i &= 1, \ldots, a; \quad j = 1, \ldots, b\,. \end{aligned} \tag{6.2.3}$$

Model (6.2.3) is a "submodel" of the two-way complete model and is called a *two-way main-effects model,* or *two-way additive model*, since the effect on the response of treatment combination ij is modeled as the sum of the individual effects of the two factors. If an additive model is used when the factors really do interact, then inferences on main effects can be very misleading. Consequently, if the experimenter does not have reasonable knowledge about the interaction, then the two-way complete model (6.2.2) or the equivalent cell-means model (6.2.1) should be used.

6.2.3 Checking the Assumptions on the Model

The assumptions implicit in both the two-way complete model (6.2.2) and the two-way main-effects model (6.2.3) are that the error random variables have equal variances, are mutually independent, and are normally distributed. The strategy and methods for checking the error assumptions are the same as those in Chap. 5. The standardized residuals are calculated as

$$z_{ijt} = (y_{ijt} - \hat{y}_{ijt})/\sqrt{ssE/(n-1)}$$

with

$$\hat{y}_{ijt} = \hat{\tau}_{ij} = \hat{\alpha}_i + \hat{\beta}_j + (\widehat{\alpha\beta})_{ij}$$

or

$$\hat{y}_{ijt} = \hat{\tau}_{ij} = \hat{\alpha}_i + \hat{\beta}_j\,,$$

depending upon which model is selected, where the "hat" denotes a least squares estimate. The residuals are plotted against

(i) the order of observation to check independence,
(ii) the levels of each factor and $\hat{y}_{ijt}$ to check for outliers and for equality of variances,
(iii) the normal scores to check the normality assumption.

When the main-effects model is selected, interaction plots of the data, such as those in Figs. 6.2 and 6.1, can be used to check the assumption of no interaction. An alternative way to check for interaction is to plot the standardized residuals against the levels of one of the factors with the plotted labels being the levels of the second factor. An example of such a plot is shown in Fig. 6.3. (For details of the original

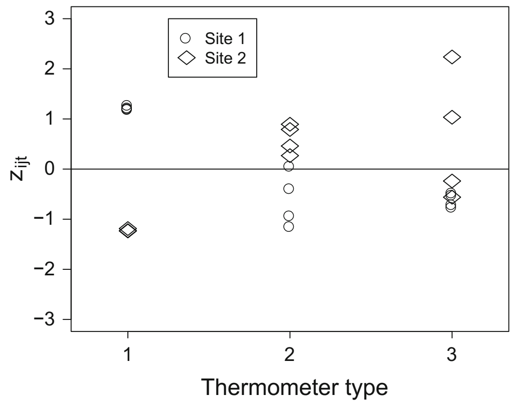


Fig. 6.3 Residual plot for the temperature experiment

experiment, see Exercise 17.9.1, p. 650.) If the main-effects model had represented the data well, then the residuals would have been randomly scattered around zero. However, a pattern can be seen that is reminiscent of the interaction plot (b) of Fig. 6.2 suggesting that a two-way complete model would have been a much better description of the data. If the model is changed based on the data, subsequent stated confidence levels and significance levels will be inaccurate, and analyses must be interpreted with caution.

If there is some doubt about the equality of the variances, the rule of thumb $s^2_{\max}/s^2_{\min} < 3$ can be employed, where $s^2_{\max}$ is the maximum of the variances of the data values within the cells, and $s^2_{\min}$ is the minimum (see Sect. 5.6.1). In a two-way layout, however, there may not be sufficient observations per cell to allow this calculation to be made. Nevertheless, we can at least check that the error variances are the same for each level of any given factor by employing the rule of thumb for the variances of the nonstandardized residuals calculated at each level of the factor.

6.3 Contrasts

6.3.1 Contrasts for Main Effects and Interactions

Since the cell-means model (6.2.1) is equivalent to the one-way analysis of variance model, we know that all contrasts in the treatment effects τ_{ij} are estimable (cf. Sect. 3.4.1, p. 34). Contrasts of interest for a cell-means model are typically of three main types: treatment contrasts, interaction contrasts, and main-effect contrasts.

Treatment contrasts $\Sigma_i \Sigma_j d_{ij}\tau_{ij}$ are no different from the types of contrasts described in Chap. 4. For example, $\tau_{ij} - \tau_{sh}$ is a pairwise difference between treatment combinations ij and sh. All the confidence interval methods of Chap. 4 are directly applicable.

Interaction contrasts are the contrasts that we use in order to measure whether or not the lines on the interaction plots (cf. Figs. 6.1 and 6.2) are parallel. An example of an interaction contrast is

$$(\tau_{sh} - \tau_{(s+1)h}) - (\tau_{sq} - \tau_{(s+1)q})\,. \tag{6.3.4}$$

We can verify that this is, indeed, an interaction contrast by using the equivalent two-way complete model notation with $\tau_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}$. Substituting this into (6.3.4) gives the contrast

$$\big((\alpha\beta)_{sh} - (\alpha\beta)_{(s+1)h}\big) - \big((\alpha\beta)_{sq} - (\alpha\beta)_{(s+1)q}\big)\,, \tag{6.3.5}$$

which is a function of interaction parameters only. Interaction contrasts are always of the form

$$\sum_i \sum_j d_{ij}\tau_{ij} = \sum_i \sum_j d_{ij}(\alpha\beta)_{ij}\,, \tag{6.3.6}$$

where

$$\sum_i d_{ij} = 0 \text{ for each } j \quad \text{and} \quad \sum_j d_{ij} = 0 \text{ for each } i\,.$$

Some, but not all, interaction contrasts have coefficients $d_{ij} = c_i k_j$. For example, if we take $c_s = k_h = 1$ and $c_{s+1} = k_q = -1$ and all other c_i and k_j zero, then, setting $d_{ij} = c_i k_j$ in (6.3.6), we obtain the coefficients in contrast (6.3.5).

If the interaction effect is very small, then the lines on an interaction plot are almost parallel (as in plots (a)–(d) of Fig. 6.1). We can then compare the average effects of the different levels of A (averaging over the levels of B). Thus, contrasts of the form $\Sigma c_i \overline{\tau}_{i.}$, with $\Sigma c_i = 0$, would be of interest. However, if there is an interaction (as in plot (c) of Fig. 6.2), such an average may make little sense. This becomes obvious when we use the two-way complete model formulation, since a main effect contrast in A is

$$\sum_i c_i \overline{\tau}_{i.} = \sum_i c_i(\alpha_i + (\overline{\alpha\beta})_{i.}) \tag{6.3.7}$$

where $(\overline{\alpha\beta})_{i.} = \frac{1}{b}\sum_j (\alpha\beta)_{ij}$, and we can see clearly that we have averaged over any interaction effect that might be present. We will often write

$$\alpha_i^* = \alpha_i + (\overline{\alpha\beta})_{i.} \text{ and } \beta_j^* = \beta_j + (\overline{\alpha\beta})_{.j}$$

for convenience. A contrast in the main effect of A for the two-way complete model is then written as $\Sigma c_i \alpha_i^*$ $(\Sigma c_i = 0)$, and a contrast in the main effect of B is

$$\sum_j k_j \overline{\tau}_{.j} = \sum_j k_j(\beta_j + (\overline{\alpha\beta})_{.j}) = \sum_j k_j \beta_j^*\,, \tag{6.3.8}$$

where $\Sigma k_j = 0$ and $(\overline{\alpha\beta})_{.j} = \frac{1}{a}\sum_i (\alpha\beta)_{ij}$.

Sometimes, it is of interest to compare the effects of the levels of one factor separately at each level of the other factor. Consider a variation on the hypothetical experiment in Sect. 6.2.1. Suppose the hypothetical statistics department also wishes to study the effects on student learning of two pedagogies (traditional lecture, and discovery-based learning) for three instructors teaching an introductory statistics course. Unless the department wants all instructors (factor A, say) to use the same pedagogy (factor B, say) in teaching the course, a natural objective might be to choose a best pedagogy for each instructor separately. If comparison of the effects of levels of factor B for each level of factor A is required, then contrasts of the form

$$\sum_j c_j \tau_{ij}\,, \quad \text{with} \quad \sum_j c_j = 0 \quad \text{for each } i = 1, 2, \ldots, a\,,$$

are of interest. We call such contrasts *simple contrasts* in the levels of B. As a special case, we have the *simple pairwise differences* of factor B:

$$\tau_{ih} - \tau_{ij}, \quad \text{for each } i = 1, \ldots, a\,.$$

These are a subset of the pairwise comparison contrasts. Simple contrasts and simple pairwise differences of factor A are defined in an analogous way.

When it is known in advance of the experiment that factors A and B do not interact, the two-way main-effects model (6.2.3) would normally be used. In this model, there is no interaction term, so $\tau_{ij} = \alpha_i + \beta_j$. The main-effects contrasts for A and B are respectively of the form

$$\sum c_i \overline{\tau}_{i.} = \sum c_i \alpha_i \text{ and } \sum k_j \overline{\tau}_{.j} = \sum k_j \beta_j,$$

with $\sum c_i = 0$ and $\sum k_j = 0$.

6.3.2 Writing Contrasts as Coefficient Lists

Instead of writing out a contrast explicitly, it is sometimes sufficient, and more convenient, to list the contrast coefficients only. For the two-way complete model, we have a choice. We can refer to contrasts as either a list of coefficients of the parameters α_i^*, β_j^*, and $(\alpha\beta)_{ij}$ or as a list of coefficients of the τ_{ij}'s. This is illustrated in the following example.

Example 6.3.1 `Battery experiment, continued`

The four treatment combinations in the battery experiment of Sect. 2.5.2, p. 24, involved two treatment factors, "duty" and "brand," each having two levels (1 for alkaline and 2 for heavy duty; 1 for name brand and 2 for store brand), giving treatment combinations 11, 12, 21, and 22. (These were coded in previous examples as 1, 2, 3, and 4, respectively.) There were $r = 4$ observations on each treatment combination.

The interaction plot in Fig. 6.4 shows a possible interaction between the two factors, since the dotted lines on the plot are not close to parallel. However, we should remember that we cannot be certain whether the nonparallel lines are due to an interaction or to inherent variability in the data, and we will need to investigate the cause in more detail later.

The interaction is measured by the contrast

$$\tau_{11} - \tau_{12} - \tau_{21} + \tau_{22} = (\alpha\beta)_{11} - (\alpha\beta)_{12} - (\alpha\beta)_{21} + (\alpha\beta)_{22}\,,$$

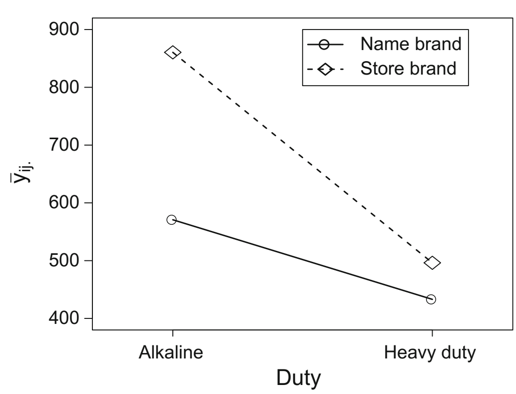


Fig. 6.4 Plot of average life per unit cost against "Duty" level i by "Brand" level j for the battery experiment

which can be written in terms of the coefficient list [1, −1, −1, 1].

The contrast that compares the average lifetimes of heavy duty and alkaline batteries (averaged across brands) is

$$\overline{\tau}_{2.} - \overline{\tau}_{1.} = \frac{1}{2}(\tau_{21} + \tau_{22}) - \frac{1}{2}(\tau_{11} + \tau_{12}) = \alpha_2^* - \alpha_1^* .$$

This has coefficient list [−1, 1] in terms of the effects α_1^*, α_2^* of the levels of duty, but coefficient list $\frac{1}{2}$[−1, −1, 1, 1] in terms of the effects $\tau_{11}, \tau_{12}, \tau_{21}, \tau_{22}$ of the treatment combinations. Similarly, the contrast that compares the average life of store brand with that of name brand (averaged over duty) has coefficient list [−1, 1] in terms of the effects β_j^* of brand, but coefficient list $\frac{1}{2}$[−1, 1, −1, 1] in terms of the τ_{ij}'s.

Since the main-effect contrasts each have divisor 2, the interaction contrast is often divided by 2 also. This has the effect that the least squares estimators of all three contrasts have the same variances (see Example 6.4.1), and their magnitudes are more directly comparable. An alternative way to achieve equal variances is to normalize the contrasts (see Sect. 4.2), in which case all three contrasts would all be divided by $\sqrt{\Sigma c_i^2/r}$. □

Contrast coefficients are often listed as columns in a table. For example, the contrast coefficients of the τ_{ij}'s for the main effect and interaction contrasts of Example 6.3.1 are written as below, with ±1's in the body of the table, and the constants listed as divisors in the last row.

ij	*A*	*B*	*AB*
11	−1	−1	1
12	−1	1	−1
21	1	−1	−1
22	1	1	1
Divisor	2	2	2

The benefit of this representation is that we can see easily that each *AB* interaction coefficient can be obtained by multiplying the corresponding *A* and *B* main-effect coefficients. Most of the interaction contrasts that we shall use have this product form. We will mention the exceptions when they arise.

Example 6.3.2 `Trend contrasts`

Suppose that the two factors, *A* and *B*, have $a = 3$ and $b = 6$ equally spaced quantitative levels, respectively, and that the sample sizes are equal. From Table A.2, we see that A_L, the linear trend contrast for *A*, has contrast coefficient list [−1, 0, 1] in terms of the α_i^*'s, and A_Q, the quadratic trend contrast for *A*, has contrast coefficient list [1, −2, 1]; that is

$$A_L = -\alpha_1^* + \alpha_3^* ,$$
$$A_Q = \alpha_1^* - 2\alpha_2^* + \alpha_3^* .$$

Similarly, in terms of the β_j^*'s, the coefficient lists for the linear and quadratic trends in the effects of the six levels of *B* are also obtained from Table A.2 as [−5, −3, −1, 1, 3, 5] and [5, −1, −4, −4, −1, 5], respectively; that is,

$$B_L = -5\beta_1^* - 3\beta_2^* - \beta_3^* + \beta_4^* + 3\beta_5^* + 5\beta_6^* ,$$
$$B_Q = 5\beta_1^* - \beta_2^* - 4\beta_3^* - 4\beta_4^* - \beta_5^* + 5\beta_6^* .$$

Now,

$$\alpha_i^* = \overline{\tau}_{i.},\ \text{giving}\ \ \Sigma_i c_i \alpha_i^* = \Sigma_i c_i \left(\frac{1}{6}\Sigma_j \tau_{ij}\right) = \frac{1}{6}\Sigma_i \Sigma_j c_i \tau_{ij}\,,$$

and

$$\beta_j^* = \overline{\tau}_{.j},\ \text{giving}\ \ \Sigma_j k_j \beta_j^* = \Sigma_j k_j \left(\frac{1}{3}\Sigma_i \tau_{ij}\right) = \frac{1}{3}\Sigma_i \Sigma_j k_j \tau_{ij}\,,$$

and we can write all of the above trends in terms of contrasts in τ_{ij}, as shown in the columns of Table 6.1. Contrast coefficients are also listed for cubic, quartic, and quintic trends for B. If we wish to compare the A and B trends on the same scale, we can normalize the contrasts (see Sect. 4.2).

In order to model a three-dimensional surface, we need to know not only how the response is affected by the levels of each factor averaged over the levels of the other factor, but also how the response changes as the levels of A and B change together. The linear$A\times$linear$B$ trend ($A_L B_L$) measures whether or not the linear trend in A changes in a linear fashion as the levels of B are increased, and vice versa. This is an interaction contrast whose coefficients are of the form $d_{ij} = c_i k_j$, where c_i are the contrast coefficients for A, and k_j are the contrast coefficients for B. The $A_L B_L$ contrast coefficients are shown in Table 6.1, and it can be verified that they are obtained by multiplying together corresponding main-effect linear trend coefficients in the same row. Coefficients for the linear$A\times$quintic$B$ ($A_L B_{qn}$) contrast is also shown for use later in this chapter. □

Table 6.1 Trend contrasts when A and B have 3 and 6 equally spaced levels, respectively

ij	A_L	A_Q	B_L	B_Q	B_C	B_{qr}	B_{qn}	$A_L B_L$	$A_L B_{qn}$
11	−1	1	−5	5	−5	1	−1	5	1
12	−1	1	−3	−1	7	−3	5	3	−5
13	−1	1	−1	−4	4	2	−10	1	10
14	−1	1	1	−4	−4	2	10	−1	−10
15	−1	1	3	−1	−7	−3	−5	−3	5
16	−1	1	5	5	5	1	1	−5	−1
21	0	−2	−5	5	−5	1	−1	0	0
22	0	−2	−3	−1	7	−3	5	0	0
23	0	−2	−1	−4	4	2	−10	0	0
24	0	−2	1	−4	−4	2	10	0	0
25	0	−2	3	−1	−7	−3	−5	0	0
26	0	−2	5	5	5	1	1	0	0
31	1	1	−5	5	−5	1	−1	−5	−1
32	1	1	−3	−1	7	−3	5	−3	5
33	1	1	−1	−4	4	2	−10	−1	−10
34	1	1	1	−4	−4	2	10	1	10
35	1	1	3	−1	−7	−3	−5	3	−5
36	1	1	5	5	5	1	1	5	1
Divisor	6	6	3	3	3	3	3	1	1

6.4 Analysis of the Two-Way Complete Model

In the analysis of an experiment with two treatment factors that possibly interact, we may proceed with the analysis in two equivalent ways. We may use the cell-means model (6.2.1) together with all the analysis techniques of Chaps. 3 and 4, or we may use the two-way complete model (6.2.2) and isolate the contributions to the response made by each of the two factors and their interaction separately.

A sensible strategy is to start with the two-way complete model and test a hypothesis of no interaction. If the hypothesis is not rejected, we may then continue with the analysis by examining the main effects under the same two-way complete model. We would not change to the two-way main-effects model, since this is not an equivalent model. However, if the hypothesis of no interaction is rejected, then we would normally prefer to change to the equivalent cell-means model and examine differences in the effects of the treatment combinations. We would also use the cell-means model when the objective of the experiment is to find the best treatment combination.

6.4.1 Least Squares Estimators for the Two-Way Complete Model

As in Sect. 3.4.3, p. 35, the least squares estimator of $\mu + \tau_{ij}$ is $\overline{Y}_{ij.}$, so the least squares estimators of the parameters in the cell-means model (6.2.1) and the equivalent two-way complete model (6.2.2) are

$$\hat{\mu} + \hat{\tau}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + (\widehat{\alpha\beta})_{ij} = \overline{Y}_{ij.}\,,$$

and the corresponding variance is σ^2/r_{ij}. Any interaction contrast of the form $\Sigma\Sigma d_{ij}\tau_{ij}$ (with $\Sigma_i d_{ij} = 0$ and $\Sigma_j d_{ij} = 0$) has least squares estimator and associated variance equal to

$$\sum_i \sum_j d_{ij}\overline{Y}_{ij.} \text{ and } \sigma^2 \sum_i \sum_j \left(\frac{d_{ij}^2}{r_{ij}}\right).$$

In particular, the least squares estimator of the interaction contrast

$$(\tau_{sh} - \tau_{uh}) - (\tau_{sq} - \tau_{uq})$$

is

$$\overline{Y}_{sh.} - \overline{Y}_{uh.} - \overline{Y}_{sq.} + \overline{Y}_{uq.} \qquad (6.4.9)$$

with variance

$$\sigma^2\left(\frac{1}{r_{sh}} + \frac{1}{r_{uh}} + \frac{1}{r_{sq}} + \frac{1}{r_{uq}}\right). \qquad (6.4.10)$$

The least squares estimators of main-effect contrasts $\Sigma c_i\alpha_i^*$ and $\Sigma k_j\beta_j^*$ are

$$\sum_i c_i\hat{\alpha}_i^* = \sum_i c_i\left(\frac{1}{b}\sum_j \overline{Y}_{ij.}\right) \text{ and } \sum_j k_j\hat{\beta}_j^* = \sum_j k_j\left(\frac{1}{a}\sum_i \overline{Y}_{ij.}\right) \qquad (6.4.11)$$

with variances

$$\text{Var}(\Sigma c_i\hat{\alpha}_i^*) = \sigma^2\left(\sum_i\sum_j \frac{c_i^2}{b^2 r_{ij}}\right) \text{ and } \text{Var}(\Sigma k_j\hat{\beta}_j^*) = \sigma^2\left(\sum_i\sum_j \frac{k_j^2}{a^2 r_{ij}}\right), \tag{6.4.12}$$

respectively. If the sample sizes are equal, the least squares estimators of $\sum c_i\alpha_i^*$ and $\sum k_j\beta_j^*$ reduce to

$$\sum_i c_i\hat{\alpha}_i^* = \sum_i c_i\overline{Y}_{i..} \text{ and } \sum_j k_j\hat{\beta}_j^* = \sum_j k_j\overline{Y}_{.j.}, \tag{6.4.13}$$

where $\overline{Y}_{i..} = \sum_j\sum_t Y_{ijt}/br$ and $\overline{Y}_{.j.} = \sum_i\sum_t Y_{ijt}/ar$. Thus, for equal sample sizes,

$$\hat{\alpha}_i^* - \hat{\alpha}_s^* = \overline{Y}_{i..} - \overline{Y}_{s..} \text{ and } \hat{\beta}_j^* - \hat{\beta}_q^* = \overline{Y}_{.j.} - \overline{Y}_{.q.} \tag{6.4.14}$$

with associated variances $2\sigma^2/(br)$ and $2\sigma^2/(ar)$, respectively.

Example 6.4.1 `Battery experiment, continued`

The four treatment combinations in the battery experiment of Sect. 2.5.2, p. 24, involved two treatment factors, "duty" and "brand," each having two levels (1 for alkaline and 2 for heavy duty; 1 for name brand and 2 for store brand), giving treatment combinations 11, 12, 21, and 22. There were $r = 4$ observations on each treatment combination. The observed average lifetimes per unit cost for the treatment combinations were

$$\overline{y}_{11.} = 570.75, \quad \overline{y}_{12.} = 860.50, \quad \overline{y}_{21.} = 433.00, \quad \overline{y}_{22.} = 496.25.$$

The interaction contrast

$$\frac{1}{2}(\tau_{11} - \tau_{12} - \tau_{21} + \tau_{22}) = \frac{1}{2}\left((\alpha\beta)_{11} - (\alpha\beta)_{12} - (\alpha\beta)_{21} + (\alpha\beta)_{22}\right)$$

has least squares estimate

$$\frac{1}{2}(\overline{y}_{11.} - \overline{y}_{12.} - \overline{y}_{21.} + \overline{y}_{22.}) = -113.25\,,$$

with associated variance

$$\sigma^2\left(\sum\sum d_{ij}^2/r\right) = \sigma^2\left((\tfrac{1}{2})^2 + (-\tfrac{1}{2})^2 + (-\tfrac{1}{2})^2 + (\tfrac{1}{2})^2\right)/4 = \sigma^2/4\,.$$

The duty contrast,

$$\alpha_1^* - \alpha_2^* = (\alpha_1 + \overline{(\alpha\beta)}_{1.}) - (\alpha_2 + \overline{(\alpha\beta)}_{2.}) = \frac{1}{2}(\tau_{11} + \tau_{12} - \tau_{21} - \tau_{22})\,,$$

has least squares estimate $\overline{y}_{1..} - \overline{y}_{2..} = 251.00$ and associated variance $\sigma^2/4$. The brand contrast,

$$\beta_1^* - \beta_2^* = (\beta_1 + \overline{(\alpha\beta)}_{.1}) - (\beta_2 + \overline{(\alpha\beta)}_{.2}) = \frac{1}{2}(\tau_{11} - \tau_{12} + \tau_{21} - \tau_{22})\,,$$

has least squares estimate $\overline{y}_{.1.} - \overline{y}_{.2.} = -176.50$ and associated variance $\sigma^2/4$. □

6.4.2 Estimation of σ^2 for the Two-Way Complete Model

Since the two-way complete model (6.2.2) is equivalent to the cell-means model (6.2.1), an unbiased estimate of σ^2 is the same as that for the one-way analysis of variance model, apart from an extra subscript j. Thus, the error sum of squares ssE can be obtained from (3.4.4) or (3.4.5), p. 39, that is,

$$ssE = \sum_i \sum_j \sum_t (y_{ijt} - \overline{y}_{ij.})^2 \tag{6.4.15}$$

$$= \sum_i \sum_j \sum_t y_{ijt}^2 - \sum_i \sum_j r_{ij} \overline{y}_{ij.}^2 \tag{6.4.16}$$

An unbiased estimate for σ^2 is obtained as $msE = ssE/(n - v)$, with $v = ab$. An upper $100(1 - \alpha)\%$ confidence bound for σ^2 is given by (3.4.9), p. 40, that is,

$$\sigma^2 \leq \frac{ssE}{\chi^2_{n-ab,1-\alpha}} . \tag{6.4.17}$$

Example 6.4.2 `Reaction time experiment, continued`

The reaction time pilot experiment, run in 1996 by Liming Cai, Tong Li, Nishant, and Andre van der Kouwe, was described in Exercise 4 of Chap. 4. The experiment was run to compare the speed of response of a human subject to audio and visual stimuli. A personal computer was used to present a "stimulus" to a subject, and the time that the subject took to press a key in response was monitored. The subject was warned that the stimulus was forthcoming by means of an auditory or a visual cue. The two treatment factors were "Cue Stimulus" at two levels, "auditory" and "visual" (Factor A, coded 1, 2), and "Cue Time" at three levels, 5, 10, and 15 seconds between cue and stimulus (Factor B, coded 1, 2, 3), giving a total of $v = 6$ treatment combinations (coded 11, 12, 13, 21, 22, 23). Three observations were taken on each treatment combination for a single subject. The reaction times are shown in Table 6.2. It can be verified that $\sum\sum\sum y_{ijt}^2 = 0.96519$. Using (6.4.16) and the sums in Table 6.2, the sum of squares for error is

$$ssE = \sum_i \sum_j \sum_t y_{ijt}^2 - 3 \sum_i \sum_j \overline{y}_{ij.}^2$$
$$= 0.96519 - 3(0.32057) = 0.00347 ,$$

Table 6.2 Data (in seconds) for the reaction time experiment

A: Cue stimulus	B: Cue time	Treatment combination	Reaction time y_{ijt}			Sums $y_{ij.}$
1	1	11	0.204	0.170	0.181	0.555
1	2	12	0.167	0.182	0.187	0.536
1	3	13	0.202	0.198	0.236	0.636
2	1	21	0.257	0.279	0.269	0.805
2	2	22	0.283	0.235	0.260	0.778
2	3	23	0.256	0.281	0.258	0.795

and an unbiased estimate of σ^2 is $msE = ssE/(18-6) = 0.000289$ seconds^2. An upper 95% confidence bound for σ^2 is

$$\frac{ssE}{\chi^2_{12,.95}} = \frac{0.00347}{5.226} = 0.000664 \text{ seconds}^2 ,$$

and taking square roots, an upper 95% confidence bound for σ is 0.0257 seconds. □

6.4.3 Multiple Comparisons for the Complete Model

In outlining the analysis at step (g) of the checklist of Chap. 2, the experimenter should specify which treatment contrasts are of interest, together with overall error rates for hypothesis tests and overall confidence levels for confidence intervals. If the two-way complete model has been selected, comparison of treatment combinations, comparison of main effects of A, and comparison of main effects of B may all be of interest. A possibility in outlining the analysis is to select error rates of α_1, α_2, and α_3 for the three sets of inferences. Then, by the Bonferroni method, the *experimentwise* simultaneous error rate is at most $\alpha_1 + \alpha_2 + \alpha_3$, and the experimentwise confidence level is at least $100(1 - \alpha_1 - \alpha_2 - \alpha_3)\%$. If interaction contrasts are also of interest, then the overall α-level can be divided into four parts instead of three.

Comparing Treatment Combinations

When comparison of treatment combinations is of most interest, the cell-means model (6.2.1) is used. The formulae for the Bonferroni, Scheffé, Tukey, and Dunnett methods can all be used in the same way as was done in Chap. 4, but with ssE given by (6.4.16) and with $v = ab$.

The best treatment combination can be found using Tukey's method of multiple comparisons. The best treatment combination may not coincide with the apparent best levels of A and B separately. For example, in Fig. 6.2(d), p. 141, the apparent best treatment combination occurs with presentation format 2 and structure 1, whereas the best presentation format, on average, appears to be number 3.

Comparing Main Effects

Main-effect contrasts compare the effects of the levels of one factor *averaging* over the levels of the other factor and may not be of interest if the two factors interact. If main-effect contrasts are to be examined, then the Bonferroni, Scheffé, Tukey, and Dunnett methods can be used for each factor separately. The general formula is equivalent to (4.4.20), p. 83. For factor A and *equal sample sizes* the formula is

$$\sum_i c_i \overline{\tau}_{i.} = \sum_i c_i \alpha_i^* \in \left(\sum_i c_i \overline{y}_{i..} \pm w \sqrt{msE \sum_i c_i^2/br} \right) , \tag{6.4.18}$$

where the critical coefficient w for each of the four methods is, respectively,

$$w_B = t_{n-ab,\alpha/2m} \ ; \ \ w_S = \sqrt{(a-1)F_{a-1,n-ab,\alpha}} \ ;$$
$$w_T = q_{a,n-ab,\alpha}/\sqrt{2} \ ; \ \ w_{D2} = |t|^{(0.5)}_{a-1,n-ab,\alpha} \ .$$

The general formula for a confidence interval for a contrast in factor B is

$$\sum_j k_j \overline{\tau}_{.j} = \sum_j k_j \beta_j^* \in \left(\sum_j k_j \overline{y}_{.j.} \pm w \sqrt{msE \sum_j k_j^2/(ar)} \right) \tag{6.4.19}$$

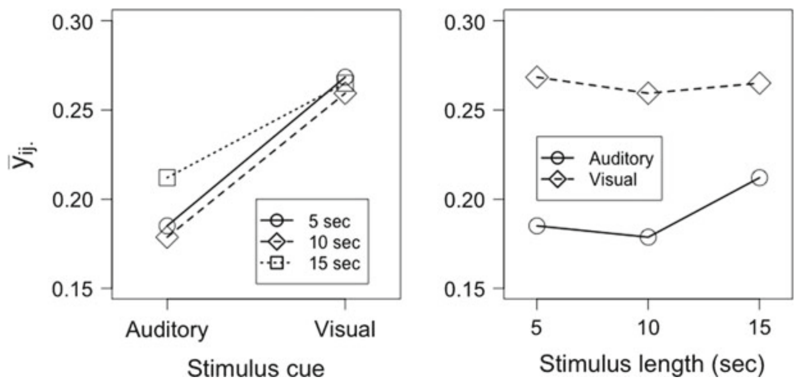


Fig. 6.5 Average times for the reaction time experiment

with critical coefficients as above but interchanging a and b. The error variance estimate is $msE = ssE/(n - ab)$, where ssE is obtained from (6.4.16).

For *unequal sample sizes*, the Bonferroni and Scheffé methods can be used, but the least squares estimates and variances must be replaced by (6.4.11) and (6.4.12), respectively. It has not yet been proved that the other two methods retain an overall confidence level of at least $100(1-\alpha)\%$ for unequal sample sizes, although this is widely believed to be the case for Tukey's method.

Example 6.4.3 `Reaction time experiment, continued`

Suppose the preplanned analysis for the reaction time experiment of Example 6.4.2 (p. 151) had been to use the two-way complete model and to test the null hypothesis of no interaction. If the hypothesis were to be rejected, then the plan was to use Tukey's method at level 99% for the pairwise comparisons of the treatment combinations. Otherwise, Tukey's method would be used at level 99% for the pairwise comparison of the levels of B (cue time), and a single 99% confidence interval would be obtained for comparing the two levels of A (cue stimulus). Then the experimentwise confidence level for the three sets of intervals would have been at least 97%.

After looking at the data plotted in Fig. 6.5, the experimenters might decide that comparison of the levels of cue stimulus (averaged over cue time) is actually the only comparison of interest. However, the experimentwise confidence level remains at least 97%, because two other sets of intervals were planned ahead of time and only became uninteresting after the data were examined.

The sample mean weights for the two cue stimuli (averaged over cue times) are

$$\bar{y}_{1..} = 0.1919, \quad \bar{y}_{2..} = 0.2642.$$

The mean square for error was calculated in Example 6.4.2 to be $msE = 0.000289$. The formula for a 99% confidence interval for the comparison of $a = 2$ treatments and $br = 9$ observations on each treatment is obtained from (6.4.18) with $w = w_B = t_{18-6,0.005} = 3.055$, giving

$$\alpha_2^* - \alpha_1^* \in \left(\bar{y}_{2..} - \bar{y}_{1..} \pm w_B\sqrt{msE\,(1/br + 1/br)}\right)$$
$$= 0.0723 \pm (3.055)\sqrt{0.000289(2/9)} \;=\; (0.0478, 0.0968)\,.$$

Thus, at an experimentwise confidence level of at least 97%, we can conclude that the average reaction time with an auditory cue is between 0.0478 and 0.0968 seconds faster than with a visual cue. □

Multiple Comparisons When Variances are Unequal

When the variances of the error variables are unequal, and no transformation can be found to remedy the problem, Satterthwaite's approximation, introduced in Sect. 5.6.3 (p. 115), can be used. This is illustrated in Example 6.4.4.

Example 6.4.4 `Bleach experiment`

The bleach experiment was run by Annie Autret in 1986 to study the effect of different bleach concentrations (factor A) and the effect of the type of stain (factor B) on the speed of stain removal from a piece of cloth. The bleach concentration was to be observed at levels 3, 5, and 7 teaspoonfuls of bleach per cup of water (coded 1, 2, 3), and three types of stain (blue ink, jam, tomato sauce; coded 1, 2, 3) were of interest, giving $v = 9$ treatment combinations in total. The experimenter calculated that she needed $r = 5$ observations per treatment combination in order to be able to detect, with probability 0.9, at significance level 0.05, a difference of 5 min in the time of stain removal between the levels of either treatment factor.

The data are shown in Table 6.3 together with the sample mean and standard deviation for each treatment combination. The maximum sample standard deviation is about 8.9 times the size of the minimum sample standard deviation, so the ratio of the maximum to the minimum variance is about 80, and a transformation of the data should be contemplated. The reader can verify, using the technique described in Sect. 5.6.2, that a plot of $\ln(s_{ij}^2)$ against $\ln(\overline{y}_{ij.})$ is not linear, so no transformation of the form $h(y_{ijt}) = y_{ijt}^{1-(q/2)}$ will adequately equalize the error variances.

An alternative is to apply Satterthwaite's approximation (Sect. 5.6.3, p. 115). The plan of the analysis was to use Tukey's method with an error rate of 0.01 for each of the main-effect comparisons and for the pairwise differences of the treatment combinations, giving an experimentwise confidence level of at least 97%. For the main effect of B, for example, a pairwise comparison of levels u and h of factor B is of the form

$$\beta_u^* - \beta_h^* = \overline{\tau}_{.u} - \overline{\tau}_{.h} = \frac{1}{3}\left(\tau_{1u} + \tau_{2u} + \tau_{3u} - \tau_{1h} - \tau_{2h} - \tau_{3h}\right) ,$$

which has least squares estimate

$$\widehat{\beta}_u^* - \widehat{\beta}_h^* = \overline{y}_{.u.} - \overline{y}_{.h.} = \frac{1}{3}\left(\overline{y}_{1u.} + \overline{y}_{2u.} + \overline{y}_{3u.} - \overline{y}_{1h.} - \overline{y}_{2h.} - \overline{y}_{3h.}\right) .$$

Table 6.3 Data for the bleach experiment, with treatment factors "concentration" (A) and "stain type" (B)

ij	Time for stain removal (in seconds)					$\overline{y}_{ij.}$	s_{ij}
11	3600	3920	3340	3173	2452	3297.0	550.27
12	495	236	515	573	555	474.8	137.04
13	733	525	793	1026	510	717.4	212.85
21	2029	2271	2156	2493	2805	2350.8	305.94
22	428	432	335	288	376	371.8	61.60
23	880	759	1138	780	1625	1036.4	361.91
31	3660	4105	4545	3569	3342	3844.2	479.85
32	410	225	437	350	140	312.4	126.32
33	539	1354	347	584	781	721.0	386.02

If s^2_{ij} denotes the sample variance of the data for treatment combination ij, the estimated variance of this estimator, as in (5.6.4), p. 115, is

$$\widehat{\text{Var}}\left(\widehat{\beta}^*_u - \widehat{\beta}^*_h\right) = \sum_i \sum_j d^2_{ij} \frac{s^2_{ij}}{r_{ij}} = \frac{1}{9 \times 5}(s^2_{1u} + s^2_{2u} + s^2_{3u} + s^2_{1h} + s^2_{2h} + s^2_{3h}),$$

and since $r = 5$, the approximate number of degrees of freedom for error is

$$df = \frac{(s^2_{1u} + s^2_{2u} + s^2_{3u} + s^2_{1h} + s^2_{2h} + s^2_{3h})^2}{(s^4_{1u}/4) + (s^4_{2u}/4) + (s^4_{3u}/4) + (s^4_{1h}/4) + (s^4_{2h}/4) + (s^4_{3h}/4)}$$

after canceling the factor $r^2 = 25$ in the numerator and denominator.

For Tukey's method of pairwise comparisons for factor B with $b = 3$ levels, the minimum significant difference is

$$msd = w_T \sqrt{\widehat{\text{Var}}\left(\widehat{\beta}^*_u - \widehat{\beta}^*_h\right)},$$

with $w_T = q_{3,df,.01}/\sqrt{2}$. For measurements in seconds, we have the following values:

(u, h)	df	$q_{3,df,0.01}$	$\widehat{\text{Var}}\left(\hat{\beta}^*_u - \hat{\beta}^*_h\right)$	msd	$\overline{y}_{.u.} - \overline{y}_{.h.}$
(1, 2)	11.5	5.09	14,780.6	437.57	2,777.67
(1, 3)	18.6	4.68	21,153.5	481.31	2,339.07
(3, 2)	12.6	4.99	8,084.7	317.26	438.60

The set of 99% simultaneous Tukey confidence intervals for pairwise differences is then

$$\beta^*_1 - \beta^*_2 \in (2777.67 \pm 437.57) = (2340.10, 3215.24)\,,$$

$$\beta^*_1 - \beta^*_3 \in (1857.76, 2820.38)\,, \qquad \beta^*_3 - \beta^*_2 \in (121.34, 755.86)\,.$$

Since none of the intervals contains zero, we can state that all pairs of levels of B (stain types) have different effects on the speed of stain removal, averaged over the three concentrations of bleach. With experimentwise confidence level at least 97%, the mean time to remove blue ink (level 1) is between 1857 and 2820 seconds longer than that for tomato sauce (level 3), and the mean time to remove tomato sauce is between 121 and 755 seconds longer than that for jam (level 2). □

6.4.4 Analysis of Variance for the Complete Model

There are three standard hypotheses that are usually examined when the two-way complete model is used. The first hypothesis is that the interaction between treatment factors A and B is negligible; that is,

$$H^{AB}_0 : \{(\alpha\beta)_{ij} - (\alpha\beta)_{iq} - (\alpha\beta)_{sj} + (\alpha\beta)_{sq} = 0 \text{ for all } i \neq s, j \neq q\}\,,$$

which occurs when the interaction plots show parallel lines. Notice that if all of the contrasts $(\alpha\beta)_{ij} - (\alpha\beta)_{iq} - (\alpha\beta)_{sj} + (\alpha\beta)_{sq}$ are zero, then their averages over s and q are also zero. This leads to an equivalent way to write H^{AB}_0 as

$$H_0^{AB}: \{(\alpha\beta)_{ij} - (\overline{\alpha\beta})_{i.} - (\overline{\alpha\beta})_{.j} + (\overline{\alpha\beta})_{..} = 0 \text{ for all } ij\}.$$

In this form, it appears that H_0^{AB} is based on ab estimable contrasts, but in fact, some of them are redundant, since the ab contrasts add to zero over the subscript $i = 1, 2, \ldots, a$ and also over the subscript $j = 1, 2, \ldots, b$. Consequently, H_0^{AB} is actually based on $(a-1)(b-1)$ estimable contrasts, and the test is based on $(a-1)(b-1)$ degrees of freedom.

The other two standard hypotheses are the main-effect hypotheses

$$H_0^A: \{\alpha_1^* = \alpha_2^* = \ldots = \alpha_a^*\} \text{ and } H_0^B: \{\beta_1^* = \beta_2^* = \ldots = \beta_b^*\},$$

where $\alpha_i^* = \alpha_i + (\overline{\alpha\beta})_{i.}$ and $\beta_j^* = \beta_j + (\overline{\alpha\beta})_{.j}$. However, these main-effect hypotheses may not be of interest if there is a sizable interaction. Each of the main-effect hypotheses can be rephrased in terms of estimable contrasts in the parameters, and so can be tested. As in Chap. 3, the tests will be based on $(a-1)$ and $(b-1)$ degrees of freedom, respectively.

When the sample sizes are unequal, there are no neat algebraic formulae for the decision rules of the hypothesis tests. Therefore, we will obtain the tests for equal sample sizes and postpone discussion of the unequal sample size case to Sects. 6.8 and 6.9, where analysis will be done by computer.

Testing Interactions—Equal Sample Sizes

Since tests for main effects may not be relevant if the two factors interact, the hypothesis of negligible interaction should be tested first. As in Sect. 3.5.1, p. 41, in order to test

$$H_0^{AB}: \{(\alpha\beta)_{ij} - (\overline{\alpha\beta})_{i.} - (\overline{\alpha\beta})_{.j} + (\overline{\alpha\beta})_{..} = 0 \text{ for all } ij\}$$

against the alternative hypothesis H_A^{AB}:{the interaction is not negligible}, we compare the sum of squares for error ssE under the two-way complete model (6.2.2) with the sum of squares for error ssE_0^{AB} under the reduced model obtained when H_0^{AB} is true. The difference

$$ssAB = ssE_0^{AB} - ssE$$

is called the *sum of squares for the interaction* AB, and the test rejects H_0^{AB} in favor of H_A^{AB} if $ssAB$ is large relative to ssE.

We can rewrite the two-way complete model as

$$\begin{aligned} y_{ijt} &= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt} \\ &= \mu^* + \alpha_i^* + \beta_j^* + [(\alpha\beta)_{ij} - (\overline{\alpha\beta})_{i.} - (\overline{\alpha\beta})_{.j} + (\overline{\alpha\beta})_{..}] + \epsilon_{ijt}, \end{aligned}$$

where μ^* is the constant $\mu - (\overline{\alpha\beta})_{..}$. So, when H_0^{AB} is true, the reduced model is

$$y_{ijt} = \mu^* + \alpha_i^* + \beta_j^* + \epsilon_{ijt},$$

which has the same form as the two-way main-effects model.

We will show in Sect. 6.5.1 that the least squares estimate of $\mu+\alpha_i+\beta_j$ for the two-way main-effects model is $\overline{y}_{i..} + \overline{y}_{.j.} - \overline{y}_{...}$, for equal sample sizes. Similarly, the least squares estimate of $\mu^* + \alpha_i^* + \beta_j^*$ in the above reduced model is also $\overline{y}_{i..} + \overline{y}_{.j.} - \overline{y}_{...}$. Hence, the sum of squares for error for the reduced model is

$$ssE_0^{AB} = \sum_i \sum_j \sum_t \left(y_{ijt} - \hat{\mu}^* - \hat{\alpha}_i^* - \hat{\beta}_j^*\right)^2$$
$$= \sum_i \sum_j \sum_t (y_{ijt} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...})^2 .$$

Adding and subtracting a term $\overline{y}_{ij.}$ to this expression, we have

$$ssE_0^{AB} = \sum_i \sum_j \sum_t \left((y_{ijt} - \overline{y}_{ij.}) + (\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...})\right)^2$$
$$= \sum_i \sum_j \sum_t (y_{ijt} - \overline{y}_{ij.})^2 + \sum_i \sum_j \sum_t (\overline{y}_{ij.} - \overline{y}_{i.} - \overline{y}_{.j.} + \overline{y}_{...})^2 .$$

But the first term is just ssE given in (6.4.15). So, for equal sample sizes,

$$ssAB = ssE_0^{AB} - ssE$$
$$= r \sum_i \sum_j (\overline{y}_{ij.} - \overline{y}_{i.} - \overline{y}_{.j.} + \overline{y}_{...})^2 \tag{6.4.20}$$
$$= r \sum_i \sum_j \overline{y}_{ij.}^2 - br \sum_i \overline{y}_{i.}^2 - ar \sum_j \overline{y}_{.j.}^2 + abr\overline{y}_{...}^2 .$$

It can be shown that when H_0^{AB} is true, the corresponding random variable $SS(AB)/\sigma^2$ has a chi-squared distribution with $(a-1)(b-1)$ degrees of freedom. Also, $SSE/\sigma^2 \sim \chi^2_{n-ab}$ and SSE can be shown to be independent of $SS(AB)$. So, when H_0^{AB} is true,

$$\frac{SS(AB)/(a-1)(b-1)\sigma^2}{SSE/(n-ab)\sigma^2} = \frac{MS(AB)}{MSE} \sim F_{(a-1)(b-1),n-ab} .$$

We reject H_0^{AB} for large values of the ratio $msAB/msE$. Thus, the rule for testing the hypothesis H_0^{AB} against the alternative hypothesis that the interaction is not negligible is

$$\text{reject } H_0^{AB} \text{ if } \frac{msAB}{msE} > F_{(a-1)(b-1),n-ab,\alpha} , \tag{6.4.21}$$

where $msAB = ssAB/(a-1)(b-1)$, $msE = ssE/(n-ab)$, $ssAB$ is given in (6.4.20), and ssE is

$$ssE = \sum_i \sum_j \sum_t y_{ijt}^2 - \sum_i \sum_j r\overline{y}_{ij.}^2 .$$

If H_0^{AB} is rejected, it is often preferable to use the equivalent cell-means model and look at contrasts in the treatment combinations. If H_0^{AB} is not rejected, then tests and contrasts for main effects are usually of interest, and the two-way complete model is retained. (We do not change to the inequivalent main-effects model.)

Testing Main Effects of *A*—Equal Sample Sizes

In testing the hypothesis that factor A has no effect on the response, one can either test the hypothesis that the levels of A (averaged over the levels of B) have the same average effect on the response, that is,

$$H_0^A : \{\alpha_1^* = \alpha_2^* = \cdots = \alpha_a^*\} ,$$

or one can test the hypothesis that the response depends only on the level of B, that is

$$H_0^{A+AB} : \{H_0^A \text{ and } H_0^{AB} \text{are both true}\} .$$

The traditional test, which is produced automatically by many computer packages, is a test of the former, and the sum of squares for error ssE under the two-way complete model is compared with the sum of squares for error ssE_0^A under the reduced model

$$Y_{ijt} = \mu^{**} + \beta_j^* + \Big((\alpha\beta)_{ij} - (\overline{\alpha\beta})_{i.} - (\overline{\alpha\beta})_{.j} + (\overline{\alpha\beta})_{..}\Big) + \epsilon_{ijt} .$$

It is, perhaps, more intuitively appealing to test H_0^{A+AB} rather than H_0^A, since the corresponding reduced model is

$$Y_{ijt} = \mu^* + \beta_j^* + \epsilon_{ijt} ,$$

suggesting that A has no effect on the response whatsoever.

In this book, we take the view that the main effect of A would not be tested unless the hypothesis of no interaction were first accepted. If it is true that there is no interaction, then the two hypotheses and corresponding reduced models are the same, and the results of the two tests should be similar. Consequently, we will derive the test of the standard hypothesis H_0^A.

It can be shown that if the sample sizes are equal, the least squares estimate of $E[Y_{ijt}]$ for the reduced model under H_0^A is

$$\overline{y}_{ij.} - \overline{y}_{i..} + \overline{y}_{...} ,$$

and so the sum of squares for error for the reduced model is

$$ssE_0^A = \sum_i \sum_j \sum_t (y_{ijt} - \overline{y}_{ij.} + \overline{y}_{i..} - \overline{y}_{...})^2 .$$

Taking the terms in pairs and expanding the terms in parentheses, we obtain

$$ssE_0^A = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{t=1}^{r} (y_{ijt} - \overline{y}_{ij.})^2 - br \sum_{i=1}^{a} (\overline{y}_{i..} - \overline{y}_{...})^2 .$$

Since the first term is the formula (6.4.15) for ssE, the *sum of squares for treatment factor A* is

$$ssA \;=\; ssE_0^A - ssE \;=\; br \sum_{i=1}^{a} (\overline{y}_{i..} - \overline{y}_{...})^2 \;=\; br \sum_{i=1}^{a} \overline{y}_{i..}^2 - abr\overline{y}_{...}^2 . \tag{6.4.22}$$

Notice that this formula for ssA is similar to the formula (3.5.12), p. 43, for ssT used to test the hypothesis H_0:$\{\tau_1 = \tau_2 = \cdots = \tau_a\}$ in the one-way analysis of variance.

We write SSA for the random variable corresponding to ssA. It can be shown that if H_0^A is true, SSA/σ^2 has a chi-squared distribution with $a - 1$ degrees of freedom, and that SSA and SSE are independent. So, writing $MSA = SSA/(a - 1)$, we have that MSA/MSE has an F-distribution when

Table 6.4 Two-way ANOVA, crossed fixed effects with interaction

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Ratio
Factor A	$a-1$	ssA	$\frac{ssA}{a-1}$	$\frac{msA}{msE}$
Factor B	$b-1$	ssB	$\frac{ssB}{b-1}$	$\frac{msB}{msE}$
AB	$(a-1)(b-1)$	$ssAB$	$\frac{ssAB}{(a-1)(b-1)}$	$\frac{msAB}{msE}$
Error	$n-ab$	ssE	$\frac{ssE}{n-ab}$	
Total	$n-1$	$sstot$		
Computational formulae for equal sample sizes				
$ssE = \sum_i \sum_j \sum_t y_{ijt}^2$		$ssA = br\sum_i \overline{y}_{i..}^2 - n\overline{y}_{...}^2$		
$\quad - r\sum_i \sum_j \overline{y}_{ij.}^2$		$ssB = ar\sum_j \overline{y}_{.j.}^2 - n\overline{y}_{...}^2$		
$sstot = \sum_i \sum_j \sum_t y_{ijt}^2 - n\overline{y}_{...}^2$		$ssAB = r\sum_i \sum_j \overline{y}_{ij.}^2 - br\sum_i \overline{y}_{i..}^2$		
$n = abr$		$\quad - ar\sum_j \overline{y}_{.j.}^2 + n\overline{y}_{...}^2$		

H_0^A is true, and the rule for testing $H_0^A : \{\alpha_1^* = \cdots = \alpha_a^*\}$ against H_A^A : {not all of the α_i^*'s are equal} is

$$\text{reject } H_0^A \text{ if } \frac{msA}{msE} > F_{a-1,n-ab,\alpha}\,, \tag{6.4.23}$$

where $msA = ssA/(a-1)$ and $msE = ssE/(n-ab)$.

Testing Main Effects of *B*—Equal Sample Sizes

Analogous to the test for main effects of A, we can show that the rule for testing $H_0^B : \{\beta_1^* = \beta_2^* = \cdots = \beta_b^*\}$ against H_A^B : {not all of the β_j^*'s are equal} is

$$\text{reject } H_0^B \text{ if } \frac{msB}{msE} > F_{b-1,n-ab,\alpha}\,, \tag{6.4.24}$$

where $msB = ssB/(b-1)$, $msE = ssE/(n-ab)$, and

$$ssB = ar\sum_j (\overline{y}_{.j.} - \overline{y}_{...})^2 = ar\sum_j \overline{y}_{.j.}^2 - abr\overline{y}_{...}^2\,. \tag{6.4.25}$$

Analysis of Variance Table

The tests of the three hypotheses are summarized in a two-way analysis of variance table, shown in Table 6.4. The computational formulae are given for equal sample sizes. The last line of the table is $sstot = \sum_i \sum_j \sum_t (y_{ijt} - \overline{y}_{...})^2$, which is the total sum of squares similar to (3.5.16). It can be verified that

$$ssA + ssB + ssAB + ssE = sstot\,.$$

When the sample sizes are not equal, the formulae for ssA, ssB, and $ssAB$ are more complicated, the corresponding random variables SSA, SSB, and $SS(AB)$ are not independent, and

$$ssA + ssB + ssAB + ssE \neq sstot\,.$$

The analysis of experiments with unequal sample sizes will be discussed in Sects. 6.8 and 6.9 using the software packages SAS and R, respectively.

Example 6.4.5 `Reaction time experiment, continued`

The reaction time experiment was described in Example 6.4.2, p. 151. There were $a = 2$ levels of cue stimulus and $b = 3$ levels of cue time, and $r = 3$ observations per treatment combination. Using the data in Table 6.2, we have

$$sstot = \sum_i \sum_j \sum_t y_{ijt}^2 - abr\overline{y}_{...}^2 \;=\; 0.96519 - 0.93617 = 0.02902\,,$$

$$ssA = br\sum_i \overline{y}_{i..}^2 - abr\overline{y}_{...}^2 \;=\; 9(0.1918^2 + 0.2642^2) - 0.93617 = 0.02354\,,$$

$$ssB = ar\sum_j \overline{y}_{.j.}^2 - abr\overline{y}_{...}^2 \;=\; 6(0.2267^2 + 0.2190^2 + 0.2385^2) - 0.93617 = 0.00116$$

$$\begin{aligned} ssAB &= r\sum_i \sum_j \overline{y}_{ij.}^2 - br\sum_i \overline{y}_{i..}^2 - ar\sum_j \overline{y}_{.j.}^2 + abr\overline{y}_{...}^2 \\ &= 0.96172 - 0.95971 - 0.93733 + 0.93617 = 0.00085\,, \end{aligned}$$

and in Example 6.4.2, ssE was calculated to be 0.00347. It can be seen that $sstot = ssA + ssB + ssAB + ssE$. The analysis of variance table is shown in Table 6.5. The mean squares are the sums of squares divided by their degrees of freedom.

There are three hypotheses to be tested. If the Type I error probability α is selected to be 0.01 for each test, then the probability of incorrectly rejecting at least one hypothesis when it is true is at most 0.03. The interaction plots in Fig. 6.5, p. 153, suggest that there is no interaction between cue stimulus (A) and cue time (B). To test this hypothesis, we obtain from the analysis of variance table

$$msAB/msE = 0.00043/0.00029 = 1.46\,.$$

which is less than $F_{2,12,.01} = 6.93$. Therefore, at individual significance level $\alpha = 0.01$, there is not sufficient evidence to reject the null hypothesis H_0^{AB} that the interaction is negligible. This agrees with the interaction plot.

Now consider the main effects. Looking at Fig. 6.5, if we average over cue stimulus, there does not appear to be much difference in the effect of cue time. If we average over cue time, then auditory cue stimulus (level 1) appears to produce a shorter reaction time than a visual cue stimulus (level 2). From the analysis of variance table, $msA/msE = 0.02354/0.00029 = 81.38$. This is larger than $F_{1,12,.01} = 9.33$, so we reject $H_0^A\!:\{\alpha_1^* = \alpha_2^*\}$, and we would conclude that there is a difference in cue stimulus averaged over the cue times. On the other hand, $msB/msE = 0.00058/0.00029 = 2.0$, which

Table 6.5 Two-way ANOVA for the reaction time experiment

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Ratio	p-value
Cue stimulus	1	0.02354	0.02354	81.38	0.0001
Cue time	2	0.00116	0.00058	2.00	0.1778
Interaction	2	0.00085	0.00043	1.46	0.2701
Error	12	0.00347	0.00029		
Total	17	0.02902			

is less than $F_{2,12,.01} = 6.93$. Consequently, we do not reject $H_0^B : \{\beta_1^* = \beta_2^* = \beta_3^*\}$ and conclude that there is no evidence for a difference in the effects of the cue times averaged over the two cue stimuli.

If the analysis were done by a computer program, the p-values in Table 6.5 would be printed. We would reject any hypothesis whose corresponding p-value is less than the selected individual α^* level. In this example, we selected $\alpha^* = 0.01$, and we would fail to reject H_0^{AB} and H_0^B, but we would reject H_0^A, as in the hand calculations.

This was a pilot experiment, and since the experimenters already believed that cue stimulus and cue time really do not interact, they selected the two-way main-effects model in planning the main experiment. □

6.5 Analysis of the Two-Way Main-Effects Model

6.5.1 Least Squares Estimators for the Main-Effects Model

The two-way main-effects model (6.2.3) is

$$\begin{gathered} Y_{ijt} = \mu + \alpha_i + \beta_j + \epsilon_{ijt}\,, \\ \epsilon_{ijt} \sim N(0, \sigma^2)\,, \\ \epsilon_{ijt}\text{'s are mutually independent}\,, \\ t = 1, \ldots, r_{ij}; \quad i = 1, \ldots, a; \quad j = 1, \ldots, b. \end{gathered}$$

This model is a submodel of the two-way complete model (6.2.2) in the sense that it can only describe situations similar to those depicted in plots (a)–(d) of Fig. 6.1 and cannot describe plots (a)–(d) of Fig. 6.2. When the sample sizes are unequal, the least squares estimators of the parameters in the main-effects model are not easy to obtain, and calculations are best left to a computer (see Sects. 6.8 and 6.9). In the optional subsection below, we show that when the sample sizes are all equal to r, the least squares estimator of $E[Y_{ijt}] = \mu + \alpha_i + \beta_j$ is

$$\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \overline{Y}_{i..} + \overline{Y}_{.j.} - \overline{Y}_{...}\,. \tag{6.5.26}$$

The least squares estimator for the estimable main-effect contrast $\sum_i c_i\alpha_i$ with $\sum_i c_i = 0$ is then

$$\begin{aligned} \sum_i c_i\hat{\alpha}_i = \sum_i c_i(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j) &= \sum_i c_i\left(\overline{Y}_{i..} + \overline{Y}_{.j.} - \overline{Y}_{...}\right) \\ = \sum_i c_i\overline{Y}_{i..}\,, \end{aligned}$$

which has variance

$$\text{Var}\left(\sum_i c_i\hat{\alpha}_i\right) = \text{Var}\left(\sum_i c_i\overline{Y}_{i..}\right) = \frac{\sigma^2}{br}\sum_i c_i^2\,. \tag{6.5.27}$$

For example, $\alpha_p - \alpha_s$, the pairwise comparison of levels p and s of A, has least squares estimator and associated variance

$$\hat{\alpha}_p - \hat{\alpha}_s = \overline{Y}_{p..} - \overline{Y}_{s..} \qquad \text{with} \qquad \text{Var}(\overline{Y}_{p..} - \overline{Y}_{s..}) = \frac{2\sigma^2}{br}\,.$$

These are exactly the same formulas as for the two-way complete model and similar to those for the one-way model. Likewise for B, a main-effect contrast $\sum k_j\beta_j$ with $\sum_j k_j = 0$ has least squares estimator and associated variance

$$\sum_j k_j\hat{\beta}_j = \sum_j k_j\overline{Y}_{.j.} \quad \text{and} \quad \text{Var}\left(\sum_j k_j\overline{Y}_{.j.}\right) = \frac{\sigma^2}{ar}\sum_j k_j^2\,, \tag{6.5.28}$$

and the least squares estimator and associated variance for the pairwise difference $\beta_h - \beta_q$ is

$$\hat{\beta}_h - \hat{\beta}_q = \overline{Y}_{.h.} - \overline{Y}_{.q.} \quad \text{with} \quad \text{Var}(\overline{Y}_{.h.} - \overline{Y}_{.q.}) = \frac{2\sigma^2}{ar}\,.$$

Example 6.5.1 `Nail varnish experiment`

An experiment on the efficacy of nail varnish solvent in removing nail varnish from cloth was run by Pascale Quester in 1986. Two different brands of solvent (factor A) and three different brands of nail varnish (factor B) were investigated. One drop of nail varnish was applied to a piece of cloth (dropped from the applicator 20 cm above the cloth). The cloth was immersed in a bowl of solvent and the time measured (in minutes) until the varnish completely dissolved. There were six treatment combinations 11, 12, 13, 21, 22, 23, where the first digit represents the brand of solvent and the second digit represents the brand of nail varnish used in the experiment. The design was a completely randomized design with $r = 5$ observations on each of the six treatment combinations. The data are listed in Table 6.6 in the order in which they were collected.

The experimenter had run a pilot experiment to estimate the error variance σ^2 and to check that the experimental procedure was satisfactory. The pilot experiment indicated that the interaction between nail varnish and solvent was negligible. The similarity of the chemical composition of the varnishes and solvents, and the verification from the pilot experiment, suggest that the main-effects model (6.2.3) will be a satisfactory model for the main experiment. The data from the main experiment give the interaction plots in Fig. 6.6. Although the lines are not quite parallel, the selected main-effects model would not be a severely incorrect representation of the data.

Using the data in Table 6.6, the average dissolving time (in minutes) for the two brands of solvent are

Table 6.6 Data (minutes) for the nail varnish experiment

Solvent	2	1	1	2	2	2	1	2
Varnish	3	3	3	3	2	2	2	2
Time	32.50	30.20	27.25	24.25	34.42	26.00	22.50	31.08
Solvent	1	2	1	1	2	1	2	2
Varnish	2	1	1	1	1	3	3	2
Time	25.17	29.17	27.58	28.75	31.75	29.75	30.75	29.17
Solvent	1	1	2	1	2	2	1	2
Varnish	2	1	2	2	1	3	3	1
Time	27.75	25.83	24.75	21.50	32.08	29.50	24.50	28.50
Solvent	2	1	1	2	1	1		
Varnish	3	3	1	1	1	2		
Time	28.75	22.75	29.25	31.25	22.08	25.00		

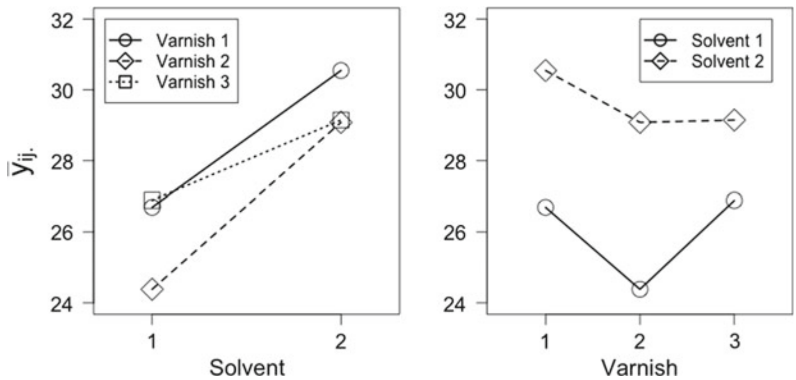


Fig. 6.6 Average dissolving times for the nail varnish experiment

$$\bar{y}_{1..} = 25.9907 \quad \text{and} \quad \bar{y}_{2..} = 29.5947\,.$$

So the least squares estimate of the difference in the dissolving times for the two solvents is

$$\hat{\alpha}_1 - \hat{\alpha}_2 = \bar{y}_{1..} - \bar{y}_{2..} = -3.6040\,,$$

and the variance of the estimator is $2\sigma^2/(rb) = 2\sigma^2/15$. A difference of 3.6 minutes seems quite substantial, but this needs to be compared with the experimental error via a confidence interval to see whether such a difference could have occurred by chance (see Examples 6.5.2 and 6.5.3).

The average dissolving times for the three brands of nail varnish are

$$\bar{y}_{.1.} = 28.624, \quad \bar{y}_{.2.} = 26.734, \text{ and } \bar{y}_{.3.} = 28.020\,,$$

and the least squares estimates of the pairwise comparisons are

$$\hat{\beta}_1 - \hat{\beta}_2 = 1.890\,, \quad \hat{\beta}_1 - \hat{\beta}_3 = 0.604\,, \text{ and } \hat{\beta}_2 - \hat{\beta}_3 = -1.286\,,$$

each with associated variance $2\sigma^2/10$. Since levels 1 and 2 of the nail varnish represented French brands, while level 3 represented an American brand, the difference of averages contrast

$$\frac{1}{2}(\beta_1 + \beta_2) - \beta_3$$

would also be of interest. The least squares estimate of this contrast is

$$\frac{1}{2}(\hat{\beta}_1 + \hat{\beta}_2) - \hat{\beta}_3 \;=\; \frac{1}{2}(\bar{y}_{.1.} + \bar{y}_{.2.}) - \bar{y}_{.3.} \;=\; -0.341\,,$$

with associated variance $6\sigma^2/40$. □

Deriving Least Squares Estimators for Equal Sample Sizes (Optional)

We now sketch the derivation (using calculus) of the least squares estimators for the parameters of the two-way main-effects model (6.2.3), when the sample sizes are all equal to r. A reader without knowledge of calculus may jump to Sect. 6.5.2, p. 165.

As in Sect. 3.4.3, the least squares estimates of the parameters in a model are those estimates that give the minimum value of the sum of squares of the estimated errors. For the two-way main-effects model (6.2.3), the sum of squared errors is

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{t=1}^{r} e_{ijt}^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{t=1}^{r} \left(y_{ijt} - (\mu + \alpha_i + \beta_j)\right)^2 .$$

The least squares estimates are obtained by differentiating the sum of squared errors with respect to each of the parameters μ, α_i $(i = 1, \ldots, a)$, and β_j $(j = 1, \ldots, b)$ in turn and setting the derivatives equal to zero. The resulting set of normal equations is as follows.

$$y_{...} - abr\hat{\mu} - br\sum_i \hat{\alpha}_i - ar\sum_j \hat{\beta}_j = 0, \tag{6.5.29}$$

$$y_{i..} - br\hat{\mu} - br\hat{\alpha}_i - r\sum_j \hat{\beta}_j = 0\,, \quad i = 1, \ldots, a, \tag{6.5.30}$$

$$y_{.j.} - ar\hat{\mu} - r\sum_i \hat{\alpha}_i - ar\hat{\beta}_j = 0, \quad j = 1, \ldots, b. \tag{6.5.31}$$

There are $1 + a + b$ normal equations in $1 + a + b$ unknowns. However, the equations are not all distinct (linearly independent), since the sum of the a equations listed in (6.5.30) is equal to the sum of the b equations listed in (6.5.31), which is equal to (6.5.29). Consequently, there are at most, and, in fact, exactly, $1 + a + b - 2$ distinct equations, and two extra equations are needed in order to obtain a solution. Many computer packages, including the SAS software, use the extra equations $\hat{\alpha}_a = 0$ and $\hat{\beta}_b = 0$, while the R package uses the extra equations $\hat{\alpha}_1 = 0$ and $\hat{\beta}_1 = 0$. However, when working by hand, it is easier to use the equations $\sum_i \hat{\alpha}_i = 0$ and $\sum_j \hat{\beta}_j = 0$, in which case (6.5.29)–(6.5.31) give the following least squares solutions:

$$\begin{aligned} \hat{\mu} &= \overline{y}_{...}\,, \\ \hat{\alpha}_i &= \overline{y}_{i..} - \overline{y}_{...}\,, \quad i = 1, \ldots, a\,, \\ \hat{\beta}_j &= \overline{y}_{.j.} - \overline{y}_{...}\,, \quad j = 1, \ldots, b\,. \end{aligned} \tag{6.5.32}$$

Then the least squares estimate of $\mu + \alpha_i + \beta_j$ is

$$\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \overline{y}_{i..} + \overline{y}_{.j.} - \overline{y}_{...}\,, \quad i = 1, \ldots, a, \quad j = 1, \ldots, b.$$

Deriving Least Squares Estimators for Unequal Sample Sizes (Optional)

If the sample sizes are not equal, then the normal equations for the two-way main-effects model become

$$y_{...} - n\hat{\mu} - \sum_{p=1}^{a} r_{p.}\hat{\alpha}_p - \sum_{q=1}^{b} r_{.q}\hat{\beta}_q = 0, \tag{6.5.33}$$

$$y_{i..} - r_{i.}\hat{\mu} - r_{i.}\hat{\alpha}_i - \sum_{q=1}^{b} r_{iq}\hat{\beta}_q = 0\,, \quad i = 1, \ldots, a, \tag{6.5.34}$$

$$y_{.j.} - r_{.j}\hat{\mu} - \sum_{p=1}^{a} r_{pj}\hat{\alpha}_p - r_{.j}\hat{\beta}_j = 0, \quad j = 1, \ldots, b, \tag{6.5.35}$$

where $n = \sum_i \sum_j r_{ij}$, $r_{p.} = \sum_j r_{pj}$, and $r_{.q} = \sum_i r_{iq}$. As in the equal sample size case, the normal equations represent $a + b - 1$ distinct equations in $1 + a + b$ unknowns, and two extra equations are needed to obtain a particular solution. Looking at (6.5.33), a sensible choice might be $\sum_p r_{p.}\hat{\alpha}_p = 0$ and $\sum_q r_{.q}\hat{\beta}_q = 0$. Then $\hat{\mu} = \overline{y}_{...}$ as in the equal sample size case. However, obtaining solutions for the $\hat{\alpha}_i$'s and $\hat{\beta}_j$'s is not so easy. One can solve for $\hat{\beta}_j$ in (6.5.35) and substitute this into (6.5.34), which gives the following equations in the $\hat{\alpha}_i$'s:

$$\hat{\alpha}_i - \sum_{q=1}^{b} \frac{r_{iq}}{r_{.q} r_{i.}} \sum_{p=1}^{a} r_{pq}\hat{\alpha}_p = \overline{y}_{i..} - \sum_{q=1}^{b} \frac{r_{iq}}{r_{.q} r_{i.}} y_{.q}\,, \quad \text{for } i = 1, \ldots, a. \tag{6.5.36}$$

Equations in the $\hat{\beta}_j$'s can be obtained similarly. Algebraic expressions for the individual parameter estimates are generally complicated, and we will leave the unequal sample size case to a computer analysis (Sects. 6.8 and 6.9).

6.5.2 Estimation of σ^2 in the Main-Effects Model

The minimum value of the sum of squares of the estimated errors for the two-way main-effects model is

$$\begin{aligned} ssE &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{t=1}^{r}(y_{ijt} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2 \\ &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{t=1}^{r}(y_{ijt} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...})^2\,. \end{aligned} \tag{6.5.37}$$

Expanding the terms in parentheses in (6.5.37) yields the following formula useful for direct hand calculation of ssE:

$$ssE = \sum_i\sum_j\sum_t y_{ijt}^2 - br\sum_i \overline{y}_{i..}^2 - ar\sum_j \overline{y}_{.j.}^2 + abr\overline{y}_{...}^2 \tag{6.5.38}$$

Now, ssE is the observed value of

$$SSE = \sum_i\sum_j\sum_t (Y_{ijt} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^2\,.$$

In Exercise 19, the reader will be asked to prove, for the equal sample size case, that

$$E[SSE] = (n - a - b + 1)\sigma^2\,,$$

where $n = abr$, so an unbiased estimator of σ^2 is

$$MSE = SSE/(n - a - b + 1)\,.$$

It can be shown that SSE/σ^2 has a chi-squared distribution with $(n - a - b + 1)$ degrees of freedom. An upper $100(1 - \alpha)\%$ confidence bound for σ^2 is therefore given by

$$\sigma^2 \le \frac{ssE}{\chi^2_{n-a-b+1,1-\alpha}}\,.$$

Example 6.5.2 `Nail varnish experiment, continued`

The data for the nail varnish experiment are given in Table 6.6 of Example 6.5.1, p. 162, and $a = 2$, $b = 3$, $r = 5$, $n = 30$. It can be verified that

$$\sum_i \sum_j \sum_t y_{ijt}^2 = 23,505.7976, \quad \overline{y}_{...} = 27.7927\,,$$

and

$$\overline{y}_{1..} = 25.9907, \qquad \overline{y}_{2..} = 29.5947,$$

$$\overline{y}_{.1.} = 28.624, \qquad \overline{y}_{.2.} = 26.734, \qquad \overline{y}_{.3.} = 28.020.$$

Thus, from (6.5.38),

$$\begin{aligned} ssE &= 23,505.7976 - 23,270.3857 - 23,191.6053 + 23,172.9696 \\ &= 216.7762\,, \end{aligned}$$

and an unbiased estimate of σ^2 is

$$msE = 216.7762/(30 - 2 - 3 + 1) = 8.3375 \text{ minutes}^2\,.$$

A 95% upper confidence bound for σ^2 is

$$\frac{ssE}{\chi^2_{26,.95}} = \frac{216.7762}{15.3791} = 14.096 \text{ minutes}^2\,,$$

and taking square roots, a 95% upper confidence limit for σ is 3.7544 minutes. □

6.5.3 Multiple Comparisons for the Main-Effects Model

When the sample sizes are equal, the Bonferroni, Scheffé, Tukey, and Dunnett methods described in Sect. 4.4 can all be used for obtaining simultaneous confidence intervals for sets of contrasts comparing the levels of A or of B. A set of $100(1 - \alpha)\%$ simultaneous confidence intervals for contrasts comparing the levels of factor A is of the form (4.4.20), which for the two-way model becomes

$$\sum c_i\alpha_i \quad \in \quad \left(\sum c_i\overline{y}_{i..} \pm w\sqrt{msE\sum c_i^2/br}\right), \tag{6.5.39}$$

where the critical coefficients for the various methods are, respectively,

$$w_B = t_{n-a-b+1,\alpha/2m} \;\; ; \;\; w_S = \sqrt{(a-1)F_{a-1,n-a-b+1,\alpha}} \;\; ;$$
$$w_T = q_{a,n-a-b+1,\alpha}/\sqrt{2} \;\; ; \;\; w_{D2} = |t|^{(0.5)}_{a-1,n-a-b+1,\alpha} \; .$$

Similarly, a set of $100(1-\alpha)\%$ confidence intervals for contrasts comparing the levels of factor B is of the form

$$\sum k_j \beta_j \in \left(\sum k_j \overline{y}_{.j.} \pm w \sqrt{msE \sum k_j^2/ar} \right), \tag{6.5.40}$$

and the critical coefficients are as above after interchanging a and b.

We can also obtain confidence intervals for the treatment means $\mu + \alpha_i + \beta_j$ using the least squares estimators $\overline{Y}_{i..} + \overline{Y}_{.j.} - \overline{Y}_{...}$, each of which has a normal distribution and variance $\sigma^2(a+b-1)/(abr)$. We obtain a set of $100(1-\alpha)\%$ simultaneous confidence intervals for the ab treatment means as

$$\mu + \alpha_i + \beta_j \in \left\{ (\overline{y}_{i..} + \overline{y}_{.j.} - \overline{y}_{...}) \pm w \sqrt{msE\left(\frac{a+b-1}{abr}\right)} \right\}, \tag{6.5.41}$$

with critical coefficient

$$w_{BM} = t_{\alpha/(2ab),(n-a-b+1)} \quad \text{or} \quad w_{SM} = \sqrt{(a+b-1)F_{a+b-1,n-a-b+1,\alpha}}$$

for the Bonferroni and Scheffé methods, respectively.

When confidence intervals are calculated for treatment means and for contrasts in the main effects of factors A and B, an experimentwise confidence level should be calculated. For example, if intervals for contrasts for factor A have overall confidence level $100(1-\alpha_1)\%$, and intervals for B have overall confidence level $100(1-\alpha_2)\%$, and intervals for means have overall confidence level $100(1-\alpha_3)\%$, the experimentwise confidence level for all the intervals combined is at least $100(1-(\alpha_1+\alpha_2+\alpha_3))\%$. Alternatively, w_{SM} could be used in (6.5.39) and (6.5.41), and the overall level for all three sets of intervals together would be $100(1-\alpha)\%$.

Example 6.5.3 `Nail varnish experiment, continued`

The least squares estimates for the differences in the effects of the two nail varnish solvents and for the pairwise differences in the effects of the three nail varnishes were calculated in Example 6.5.1, p. 162. From Table 6.8, $msE = 8.3375$ with error degrees of freedom $n-a-b+1 = 26$. There is only $m = 1$ contrast for factor A, and a simple 99% confidence interval of the form (6.5.39) can be used to give

$$\begin{aligned} \alpha_2 - \alpha_1 &\in \left(\overline{y}_{2..} - \overline{y}_{1..} \pm t_{n-a-b+1,\alpha/2}\sqrt{msE(2/br)} \right) \\ &= \left(3.6040 \pm t_{26,0.005}\sqrt{(8.3375/15)} \right) . \end{aligned}$$

From Table A.4, $t_{26,0.005} = 2.779$, so a 99% confidence interval for $\alpha_2 - \alpha_1$ is

$$1.5321 \leq \alpha_2 - \alpha_1 \leq 5.6759 \, .$$

The confidence interval indicates that solvent 2 takes between 1.5 and 5.7 minutes longer, on average, in dissolving the three nail varnishes than does solvent 1.

To compare the nail varnishes in terms of their speed of dissolving, confidence intervals are required for the three pairwise comparisons $\beta_1 - \beta_2$, $\beta_1 - \beta_3$, and $\beta_2 - \beta_3$. If an overall confidence level of 99% is required, Tukey's method gives confidence intervals of the form

$$\beta_j - \beta_p \in \left(\overline{y}_{.j.} - \overline{y}_{.p.} \pm (q_{b,df,0.01}/\sqrt{2})\ \sqrt{msE(2/(ar))}\right).$$

From Table A.8, $q_{3,26,0.01} = 4.54$. Using the least squares estimates computed in Example 6.5.1, p. 162, and $msE = 8.3375$ with $n - a - b + 1 = 26$ as above, the minimum significant difference is $msd = (4.54/\sqrt{2})\ \sqrt{8.3375(2/10)} = 4.145$. A set of 99% confidence intervals for the pairwise comparisons for factor B is

$$\beta_1 - \beta_2 \in (1.890 \pm 4.145) = (-2.255, 6.035),$$

$$\beta_1 - \beta_3 \in (-3.541, 4.749), \qquad \beta_2 - \beta_3 \in (-5.431, 2.859).$$

Each of these intervals includes zero, indicating insufficient evidence to conclude a difference in the speed at which the nail varnishes dissolve. The overall confidence level for the four intervals for factors A and B together is at least 98%. Bonferroni's method could have been used instead for all four intervals. To have obtained an overall level of at least 98%, we could have set $\alpha^* = \alpha/m = 0.02/4 = 0.005$ for each of the four intervals. The critical coefficient in (6.5.39) would then have been $w_B = t_{0.0025,26} = 3.067$. So the Bonferroni method would have given a longer interval for $\alpha_1 - \alpha_2$ but shorter intervals for $\beta_j - \beta_p$. □

6.5.4 Unequal Variances

When the variances of the error variables are unequal and no equalizing transformation can be found, Satterthwaite's approximation can be used. Since the approximation uses the sample variances of the observations for each treatment combination individually, and since the least squares estimates of the main-effect contrasts are the same whether or not interaction terms are included in the model, the procedure is exactly the same as that illustrated for the bleach experiment in Example 6.4.4, p. 154.

6.5.5 Analysis of Variance for Equal Sample Sizes

Testing Main Effects of B—Equal Sample Sizes

The hypothesis that the levels of B all have the same effect on the response is $H_0^B : \{\beta_1 = \beta_2 = \cdots = \beta_b\}$, which can be written in terms of estimable contrasts as H_0^B:$\{\beta_j - \overline{\beta}_. = 0,$ for all $j = 1, \ldots, b\}$. To obtain a test of H_0^B against the alternative hypothesis H_A^B : { at least two of the β_j's differ}, the sum of squares for error for the two-way main-effects model is compared with the sum of squares for error for the reduced model

$$Y_{ijt} = \mu + \alpha_i + \epsilon_{ijt}. \tag{6.5.42}$$

This is identical to the one-way analysis of variance model (3.3.1) with μ replaced by $\mu* = \mu + \overline{\beta}_.$ and with br observations on the ith level of treatment factor A. Thus ssE_0^B is the same as the sum of squares for error in a one-way analysis of variance, and can be obtained from (3.4.4), p. 39, by replacing the subscript t by the pair of subscripts jt, yielding

$$ssE_0^B = \sum_i \sum_j \sum_t (y_{ijt} - \overline{y}_{i..})^2 \,. \tag{6.5.43}$$

The sum of squares for testing H_0^B is $ssE_0^B - ssE$, where ssE was derived in (6.5.37), p. 165. So,

$$\begin{aligned} ssB &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{t=1}^{r}(y_{ijt} - \overline{y}_{i..})^2 - \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{t=1}^{r}\left((y_{ijt} - \overline{y}_{i..}) - (\overline{y}_{.j.} - \overline{y}_{...})\right)^2 \\ &= ar\sum_j (\overline{y}_{.j.} - \overline{y}_{...})^2 \\ &= ar\sum_j \overline{y}_{.j.}^2 - abr\overline{y}_{...}^2 \,. \end{aligned} \tag{6.5.44}$$

Notice that the formula for ssB is identical to the formula (6.4.25) for testing the equivalent main-effect hypothesis in the two-way complete model. It can be shown that when H_0^B is true, the corresponding random variable SSB/σ^2 has a chi-squared distribution with $(b-1)$ degrees of freedom, and SSB and SSE are independent. Therefore, when H_0^B is true,

$$\frac{SSB/(b-1)\sigma^2}{SSE/(n-a-b+1)\sigma^2} = \frac{MSB}{MSE} \sim F_{b-1,n-a-b+1} \,,$$

and the decision rule for testing H_0^B against H_A^B is

$$\text{reject } H_0^B \text{ if } \frac{msB}{msE} > F_{b-1,n-a-b+1,\alpha} \,. \tag{6.5.45}$$

Testing Main Effects of *A*—Equal Sample Sizes

A similar rule is obtained for testing $H_0^A : \{\alpha_1 = \alpha_2 = \cdots = \alpha_a\}$ against the alternative hypothesis H_A^A : {at least two of the α_i's differ}. The decision rule is

$$\text{reject } H_0^A \text{ if } \frac{msA}{msE} > F_{a-1,n-a-b+1,\alpha} \,, \tag{6.5.46}$$

where $msA = ssA/(a-1)$, and

$$ssA = br\sum_i (\overline{y}_{i..} - \overline{y}_{...})^2 = br\sum_i \overline{y}_{i..}^2 - abr\overline{y}_{...}^2 \,. \tag{6.5.47}$$

similar to the formula (6.4.22) for testing the equivalent hypothesis in the two-way complete model.

Analysis of Variance Table

The information for testing H_0^A and H_0^B is summarized in the analysis of variance table shown in Table 6.7. When sample sizes are equal, $ssE = sstot - ssA - ssB$. When the sample sizes are not equal, the formulae for the sums of squares are complicated, and the analysis should be done by computer (Sects. 6.8 and 6.9).

Table 6.7 Two-Way ANOVA, negligible interaction, equal sample sizes

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Ratio
Factor A	$a-1$	ssA	$\frac{ssA}{a-1}$	$\frac{msA}{msE}$
Factor B	$b-1$	ssB	$\frac{ssB}{b-1}$	$\frac{msB}{msE}$
Error	$n-a-b+1$	ssE	$\frac{ssE}{n-a-b+1}$	
Total	$n-1$	$sstot$		

Computational Formulae for Equal Sample Sizes

$ssA = br\sum_i \overline{y}_{i..}^2 - n\overline{y}_{...}^2$ $\qquad ssB = ar\sum_j \overline{y}_{.j.}^2 - n\overline{y}_{...}^2$

$sstot = \sum_i \sum_j \sum_t y_{ijt}^2 - n\overline{y}_{...}^2$ $\qquad ssE = sstot - ssA - ssB$

$n = abr$

Table 6.8 Analysis of variance for the nail varnish experiment

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Ratio	p-value
Solvent	1	97.4161	97.4161	11.68	0.0021
Varnish	2	18.6357	9.3178	1.12	0.3423
Error	26	216.7761	8.3375		
Total	29	332.8279			

Example 6.5.4 `Nail varnish experiment, continued`

The analysis of variance table for the nail varnish experiment of Example 6.5.1, p. 162, is given in Table 6.8. The experimenter selected the Type I error probability as 0.05 for testing each of H_0^A and H_0^B, giving an overall error rate of at most 0.1. The ratio $msA/msE = 11.68$ is larger than $F_{1,26,0.05} \approx 4.0$, and therefore, the null hypothesis can be rejected. It can be concluded at individual significance level 0.05 that there is a difference in dissolving times for the two solvents.

The ratio $msB/msE = 1.12$ is smaller than $F_{2,26,0.05} \approx 3.15$. Therefore, the null hypothesis H_0^B cannot be rejected at individual significance level 0.05, and it is not possible to conclude that there is a difference in dissolving time among the three varnishes. □

6.5.6 Model Building

In some experiments, the primary objective is to find a model that gives an adequate representation of the experimental data. Such experiments are called experiments for *model building*. If there are two crossed, fixed treatment factors, it is legitimate to use the two-way complete model (6.2.2) as a preliminary model. Then, if H_0^{AB} fails to be rejected, the two-way main effects model (6.2.3) can be accepted as a reasonable model to represent the same type of experimental data in *future* experiments.

Note that it is *not legitimate* to adopt the two-way main effects model and to use the corresponding analysis of variance table, Table 6.7, to test further hypotheses or calculate confidence intervals using the *same* set of data. If this is done, the model is changed based on the data, and the quoted significance levels and confidence levels associated with further inferences will not be correct. Model building should be regarded as a completely different exercise from confidence interval calculation. *They should be done using different experimental data.*

6.6 Calculating Sample Sizes

In Chaps. 3 and 4, we showed two methods of calculating sample sizes. The method of Sect. 3.6 aims to achieve a specified power of a hypothesis test, and the method of Sect. 4.5 aims to achieve a specified length of a confidence interval. Both of these techniques rely on knowledge of the largest likely value of σ^2 or msE and can also be used for the two-way complete model.

Alternatively, sample sizes can be calculated to ensure that confidence intervals for main-effect contrasts are no longer than a stated size, using the formulae (6.4.18) and (6.4.19) or, for the two-way main-effects model, the formulae (6.5.39) and (6.5.40).

Similarly, the method of Sect. 3.6 for choosing the sample size to achieve the required power of a hypothesis test can be used for each factor separately, with the modification that the sample size calculation is based on

$$r = 2a\sigma^2\phi^2/(b\Delta_A^2) \tag{6.6.48}$$

for factor A and

$$r = 2b\sigma^2\phi^2/(a\Delta_B^2)$$

for factor B, where Δ_A is the smallest difference in the α_i's (or α_i^*'s) and Δ_B is the smallest difference in the β_j's (or β_j^*'s) that are of interest. The calculation procedure is identical to that in Sect. 3.6, except that the error degrees of freedom are $\nu_2 = n - v$ for the complete model and $\nu_2 = n - a - b + 1$ for the main-effects model (with $n = abr$), and the numerator degrees of freedom are $\nu_1 = a - 1$ for factor A and $\nu_1 = b - 1$ for factor B.

If several different calculations are done and the calculated values of r differ, then the largest value should be selected.

6.7 Small Experiments

6.7.1 One Observation Per Cell

When observations are extremely time-consuming or expensive to collect, an experiment may be designed to have $r = 1$ observation on each treatment combination. Such experiments are called *experiments with one observation per cell* or *single replicate experiments*. Since the ability to choose the sample sizes is lost, it should be recognized that confidence intervals may be wide and hypothesis tests not very powerful.

If it is known in advance that the interaction between the two treatment factors is negligible, then the experiment can be analyzed using the two-way main-effects model (6.2.3). If this information is not available, then the two-way complete model (6.2.2) needs to be used. However, there is a problem. Under the two-way complete model, the number of degrees of freedom for error is $ab(r-1)$. If $r = 1$, then this number is zero, and σ^2 cannot be estimated.

Thus, a single replicate experiment with a possible interaction between the two factors can be analyzed only if one of the following is true:

(i) σ^2 is known in advance.
(ii) The interaction is expected to be of a certain form that can be modeled with fewer than $(a-1)(b-1)$ degrees of freedom.

(iii) The number of treatment combinations is large, and only a few contrasts are likely to be nonnegligible (*effect sparsity*).

If σ^2 is known in advance, formulae for confidence intervals would be based on the normal distribution, and hypothesis tests would be based on the chi-squared distribution. However, this situation is unlikely to occur, and we will not pursue it. The third case tends to occur when the experiment involves a large number of treatment factors and will be discussed in detail in Chap. 7. Here, we look at the second situation and consider two methods of analysis, the first based on orthogonal contrasts, and the second known as Tukey's test for additivity.

6.7.2 Analysis Based on Orthogonal Contrasts

Two estimable contrasts are called *orthogonal contrasts* if and only if their least squares estimators are uncorrelated or, equivalently, have zero covariance. For the moment, we recode the treatment combinations to obtain a single-digit code, as we did in Chap. 3. Two contrasts $\Sigma c_i \tau_i$ and $\Sigma k_s \tau_s$ are orthogonal if and only if

$$\begin{aligned} 0 = \text{Cov}\left(\sum_{i=1}^{v} c_i \overline{Y}_{i.}, \sum_{s=1}^{v} k_s \overline{Y}_{s.}\right) &= \sum_{i=1}^{v}\sum_{s=1}^{v} c_i k_s \text{Cov}(\overline{Y}_{i.}, \overline{Y}_{s.}) \\ = \sum_i c_i k_i \text{Cov}(\overline{Y}_{i.}, \overline{Y}_{i.}) + \sum_i \sum_{s \neq i} c_i k_s \text{Cov}(\overline{Y}_{i.}, \overline{Y}_{s.}) \\ = \sum_i c_i k_i \text{Var}(\overline{Y}_{i.}) \quad + \quad 0 \\ = \sigma^2 \sum_i c_i k_i / r_i \,. \end{aligned}$$

In the above calculation $\text{Cov}(\overline{Y}_{i.}, \overline{Y}_{s.})$ is zero when $s \neq i$, because all the Y_{it}'s are independent of each other in the cell-means model. Thus, two contrasts $\Sigma c_i \tau_i$ and $\Sigma k_i \tau_i$ are orthogonal if and only if

$$\sum_i c_i k_i / r_i = 0 \,. \tag{6.7.49}$$

If the sample sizes are equal, then this reduces to

$$\sum_i c_i k_i = 0 \,.$$

Changing back to two subscripts, we have that two contrasts $\Sigma\Sigma d_{ij}\tau_{ij}$ and $\Sigma\Sigma h_{ij}\tau_{ij}$ are orthogonal if and only if

$$\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij} h_{ij} / r_{ij} = 0 \,, \tag{6.7.50}$$

or, for equal sample sizes, the contrasts are orthogonal if and only if

Table 6.9 Three orthogonal contrasts for the battery experiment

Contrast	Coefficients	$\sum c_i \bar{y}_{i.}$	$\sum c_i^2/r_i$	ssc
Duty	$\frac{1}{2}[\,1,\ \ 1,\ -1,\ -1]$	251.00	$\frac{1}{4}$	252,004.00
Brand	$\frac{1}{2}[\,1,\ -1,\ \ 1,\ -1]$	−176.50	$\frac{1}{4}$	124,609.00
Interaction	$\frac{1}{2}[\,1,\ -1,\ -1,\ \ 1]$	−113.25	$\frac{1}{4}$	51,302.25

$$\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij} h_{ij} = 0\,. \tag{6.7.51}$$

For equal sample sizes, the trend contrasts provide an illustration of orthogonal contrasts. For example, it can be verified that any pair of trend contrasts in Table 6.1, p. 148, satisfy (6.7.51). For the models considered in this book, the contrast estimators are normally distributed, so orthogonality of contrasts implies that their least squares estimators are independent.

For v treatments, or treatment combinations, a set of $v-1$ orthogonal contrasts is called a *complete set of orthogonal contrasts*. It is not possible to find more than $v-1$ contrasts that are mutually orthogonal. We write the sum of squares for the qth orthogonal contrast in a complete set as ssc_q, where

$$ssc_q = (\Sigma\Sigma c_{ij}\bar{y}_{ij.})^2/(\Sigma\Sigma c_{ij}^2/r_{ij})$$

is the square of the normalized contrast estimator (see Sect. 4.3.3, p. 77). The sum of squares for treatments, ssT, can be partitioned into the sums of squares for the $v-1$ orthogonal contrasts in a complete set; that is,

$$ssT = ssc_1 + ssc_2 + \cdots + ssc_{v-1}\,. \tag{6.7.52}$$

Example 6.7.1 `Battery experiment, continued`

Main effect and interaction contrasts for the battery experiment were examined in Example 6.3.1, p. 146 and, following that example, were written as columns in a table. Since the sample sizes are all equal, we need only check that (6.7.51) holds by multiplying corresponding coefficients for any two contrasts and adding their products. The duty, brand, and interaction contrasts form a complete set of $v-1=3$ orthogonal contrasts.

The sums of squares for the three contrasts are shown in Table 6.9. It can be verified that they add to the treatment sum of squares $ssT = 427{,}915.25$ that was calculated in Example 3.5.1, p. 44. □

We can use the same idea to split the interaction sum of squares $ssAB$ into independent pieces. For the two-way complete model (6.2.2) with $r=1$ observation per cell, the sum of squares for testing the null hypothesis that a particular interaction contrast, say $\sum_i \sum_j d_{ij}(\alpha\beta)_{ij}$ (with $\sum_i d_{ij} = 0$ and $\sum_j d_{ij} = 0$), is negligible, against the alternative hypothesis that the contrast is not negligible, is

$$ssc = \frac{(\sum_i \sum_j d_{ij} y_{ij})^2}{\sum_i \sum_j d_{ij}^2}\,. \tag{6.7.53}$$

The interaction has $(a-1)(b-1)$ degrees of freedom. Consequently, there are $(a-1)(b-1)$ orthogonal interaction contrasts in a complete set, and their corresponding sums of squares add to $ssAB$, that is,

Table 6.10 Two-way ANOVA, one observation per cell, e negligible interaction contrasts, and $m = (a-1)(b-1) - e$ interaction degrees of freedom

Source of variation	Degrees of freedom	Sum of squares	Mean square
Factor A	$a-1$	ssA	msA
Factor B	$b-1$	ssB	msB
Interaction	m	$ssAB_m$	$msAB$
Error	e	ssE	msE
Total	$ab-1$	$sstot$	

$$ssAB = \sum_{h=1}^{(a-1)(b-1)} ssc_h \,,$$

where ssc_h is the sum of squares for the hth such contrast.

Suppose it is known *in advance* that e specific orthogonal interaction contrasts are likely to be negligible. Then the sums of squares for these e negligible contrasts can be pooled together to obtain an estimate of error variance, based on e degrees of freedom,

$$ssE = \sum_{h=1}^{e} ssc_h \quad \text{and} \quad msE = ssE/e \,.$$

The sums of squares for the remaining interaction contrasts can be used to test the contrasts individually or added together to obtain an interaction sum of squares

$$ssAB_m = \sum_{h=e+1}^{(a-1)(b-1)} ssc_h \,.$$

Then the decision rule for testing the hypothesis H_0^{AB}:{the interaction AB is negligible} against the alternative hypothesis that the interaction is not negligible is

$$\text{reject } H_0^{AB} \text{ if } \frac{ssAB_m/m}{ssE/e} > F_{m,e,\alpha} \,,$$

where $m = (a-1)(b-1) - e$. Likewise, the main effect test statistics have denominator ssE/e and error degrees of freedom $df = e$. The tests are summarized in Table 6.10, which shows a modified form of the analysis of variance table for the two-way complete model. A worked example is given in Sect. 6.7.4.

To save calculating the sums of squares for all of the contrasts, the error sum of squares is usually obtained by subtraction, that is,

$$ssE = sstot - ssA - ssB - ssAB_m \,.$$

The above technique is most often used when the factors are quantitative, since higher-order interaction trends are often likely to be negligible. The information about the interaction effects must be known prior to running the experiment. If this information is not available, then one of the techniques discussed in Sect. 7.5 must be used instead.

6.7.3 Tukey's Test for Additivity

Tukey's test for additivity uses only one degree of freedom to measure the interaction. It tests the null hypothesis H_0^γ :$\{(\alpha\beta)_{ij} = \gamma\alpha_i\beta_j$ for all $i, j\}$ against the alternative hypothesis that the interaction is not of this form. The test is appropriate only if the size of the interaction effect is expected to increase proportionally to each of the main effects, and it is not designed to measure any other form of interaction. The test requires that the normality assumption be well satisfied. The decision rule is

$$\text{reject } H_0^\gamma \text{ if } \frac{ssAB^*}{ssE/e} > F_{1,e,\alpha}\,, \qquad (6.7.54)$$

where

$$ssAB^* = \frac{ab\left[\sum_i \sum_j y_{ij}\bar{y}_{i.}\bar{y}_{.j} - (ssA + ssB + ab\bar{y}_{..}^2)\bar{y}_{..}\right]^2}{(ssA)(ssB)}$$

and

$$ssE = sstot - ssA - ssB - ssAB^*\,.$$

The analysis of variance table is as in Table 6.10 with $m = 1$ and with $e = (a-1)(b-1) - 1$.

Table 6.11 Data for the air velocity experiment, with factors Rib Height (A) and Reynolds Number (B)

		Reynolds Number, j						
	i	1	2	3	4	5	6	$\bar{y}_{i.}$
Rib	1	−24	−23	1	8	29	23	2.333
Height	2	33	28	45	57	74	80	52.833
	3	37	79	79	95	101	111	83.667
	$\bar{y}_{.j}$	15.333	28.00	41.667	53.333	68.000	71.333	$46.278 = \bar{y}_{..}$

Source Willke (1962). Copyright ©1962 Blackwell Publishers. Reprinted with permission

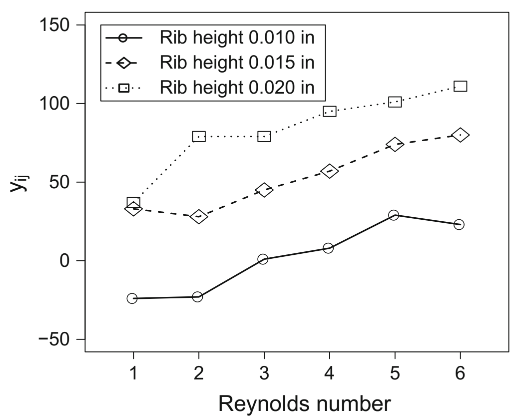


Fig. 6.7 Data for the air velocity experiment

6.7.4 A Real Experiment—Air Velocity Experiment

The data given in Table 6.11, and plotted in Fig. 6.7, form part of an experiment described by D. Wilkie in the 1962 issue of *Applied Statistics* (volume 11, pages 184–195). The experiment was designed to examine the position of maximum velocity of air blown down the space between a roughened rod and a smooth pipe surrounding it. The treatment factors were the height of ribs on the roughened rod (factor A) at equally spaced heights 0.010, 0.015, and 0.020 inches (coded 1, 2, 3) and Reynolds number (factor B) at six levels (coded 1–6) equally spaced logarithmically over the range 4.8–5.3. The responses were measured as $y = (d - 1.4) \times 10^3$, where d is the distance in inches from the center of the rod.

Figure 6.7 shows very little interaction between the factors. However, prior to the experiment, the investigators had thought that the factors would interact to some extent. They wanted to use the set of orthogonal polynomial trend contrasts for the AB interaction and were reasonably sure that the contrasts $A_Q B_{qr}$, $A_L B_{qn}$, $A_Q B_{qn}$ would be negligible. Thus the sum of squares for these three contrasts could be used to estimate σ^2 with 3 degrees of freedom. We are using "L, Q, C, qr, qn" as shorthand notation for linear, quadratic, cubic, quartic, and quintic contrasts, respectively. The coefficients for these three orthogonal polynomial trend contrasts can be obtained by multiplying the corresponding main-effect coefficients shown in Table 6.1, p. 148. The coefficients for $A_L B_{qn}$ are shown in the table as an example. Also shown are the contrast coefficients for the linear$A \times$linear$B$ contrast, $A_L B_L$. These are

$$[\,5,\ 3,\ 1,\ -1,\ -3,\ -5,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ -5,\ -3,\ -1,\ 1,\ 3,\ 5\,]\,.$$

The estimate of $A_L B_L$ is then

Table 6.12 Analysis of variance for the air velocity experiment

Source of variation	Degrees of freedom	Sum of squares	Mean square	Ratio	p-value
Rib height (A)	2	20232.111			
A_L	1	19845.333	19845.333	338.77	0.0003
A_Q	1	386.778	386.778	6.60	0.0825
Reynolds number (B)	5	7386.944			
B_L	1	7262.976	7262.976	123.98	0.0016
B_Q	1	65.016	65.016	1.11	0.3695
B_C	1	36.296	36.296	0.62	0.4887
B_{qr}	1	13.762	13.762	0.23	0.6611
B_{qn}	1	8.894	8.894	0.15	0.7228
Interaction (AB)	7	616.817			
$A_L B_L$	1	20.829	20.829	0.36	0.5930
$A_L B_Q$	1	47.149	47.149	0.80	0.4358
$A_L B_C$	1	265.225	265.225	4.53	0.1233
$A_L B_{qr}$	1	33.018	33.018	0.56	0.5073
$A_Q B_L$	1	15.238	15.238	0.26	0.6452
$A_Q B_Q$	1	170.335	170.335	2.91	0.1867
$A_Q B_C$	1	65.023	65.023	1.11	0.3694
Error	3	175.739	58.580		
Total	17	28411.611			

$$\Sigma\Sigma d_{ij}y_{ij} = 5(-24) + 3(-23) + \cdots + 3(101) + 5(111) = 54\,.$$

Now,

$$\Sigma\Sigma d_{ij}^2 = (5^2 + 3^2 + \cdots + 3^2 + 5^2) = 140\,,$$

so the corresponding sum of squares is

$$ss(A_L B_L) = \frac{54^2}{140} = 20.829\,.$$

The sums of squares for the other contrasts are computed similarly, and the error sum of squares is calculated as the sum of the sums of squares of the three negligible contrasts. The analysis of variance table is given in Table 6.12.

The hypotheses that the individual contrasts are zero can be tested using Scheffé's procedure or Bonferroni's procedure. If Bonferroni's procedure is used, each of the 14 hypotheses should be tested at a very small α-level. Taking $\alpha = 0.005$, so that the overall level is at most 0.07, we have $F_{1,3,0.005} = 55.6$, and only the linear A and linear B contrasts appear to be significantly different from zero. The plot of the data shown in Fig. 6.7 supports this conclusion.

6.8 Using SAS Software

Table 6.13 contains a sample SAS program for analysis of the two-way complete model (6.2.2). For illustration, we use the data of the reaction time experiment shown in Table 4.4, p. 101, but with the last four observati ons missing, so that $r_{11} = r_{21} = 2, r_{12} = r_{22} = r_{23} = 3, r_{13} = 1$. In the data input lines, the levels of each of the two treatment factors A and B are shown together with the response, the order in which the observations were collected, and the treatment factor level `TRTMT`. A two-digit code for each treatment combination `TC` is easily generated by the statement `TC = 10*A + B` following the `INPUT` statement. This way of coding the treatment combinations works well for all applications except for drawing plots with `TC` on one axis. Such a plot would not show numeric codes 11, 12, …, 23 as equally spaced. In the statement `TC = PUT(10* A + B, 2.)`, the function `PUT` converts the created variable from numeric to character.

6.8.1 Analysis of Variance

The `GLM` procedure in Table 6.13 is used to generate the analysis of variance table, to estimate and to test contrasts, and for multiple comparisons. As in the one-way analysis of variance, the treatment factors must be declared as class variables using a `CLASS` statement. The two-way complete model is represented as

```
MODEL Y = A B A*B;
```

with the main effects listed in either order, but before the interaction. The two-way main-effects model (6.2.3) would be represented as

```
MODEL Y = A B;
```

The program also shows the cell-means model (6.2.1) in a second `GLM` procedure, using

```
MODEL Y = TC;
```

Table 6.13 SAS program to illustrate aspects of analysis of a two-way complete model (reaction time experiment)

```
DATA RTIME;
 INPUT ORDER TRTMT A B Y;
 TC = PUT(10*A + B, 2.); * create TC as a character variable for plots;
 LINES;
 1 6 2 3 0.256
 2 6 2 3 0.281
 : : : :   :
 14 4 2 1 0.279
;
PROC GLM;
  CLASS A B;
  MODEL Y = A B A*B;
  LSMEANS A / PDIFF CL ALPHA=0.01;
  LSMEANS B / PDIFF = ALL CL ADJUST = TUKEY ALPHA = 0.01;
  CONTRAST '11-13-21+23'  A*B 1  0 -1 -1  0  1;
  CONTRAST 'B1-B2' B 1 -1  0;
  ESTIMATE 'B1-B2' B 1 -1  0;
  ESTIMATE 'B1-B3' B 1  0 -1;
  ESTIMATE 'B2-B3' B 0  1 -1;
PROC GLM;
  CLASS TC;
  MODEL Y = TC;
  LSMEANS TC / PDIFF = ALL CL ADJUST = TUKEY ALPHA = 0.01;
  LSMEANS TC / PDIFF = CONTROL CL ADJUST = DUNNETT ALPHA = 0.01;
  LSMEANS TC / PDIFF = CONTROLL CL ADJUST = DUNNETT ALPHA = 0.01;
  LSMEANS TC / PDIFF = CONTROLU CL ADJUST = DUNNETT ALPHA = 0.01;
  CONTRAST '11-13-21+23'  TC  1  0 -1  -1  0  1;
  CONTRAST 'B1-B2'        TC  1 -1  0   1 -1  0;
```

The output from the first GLM procedure is shown in Fig. 6.8. The analysis of variance table is organized differently from that in Table 6.4, p. 159. The five "model" degrees of freedom are the treatment degrees of freedom corresponding to the six treatment combinations. Information concerning main effects and interactions is provided underneath the table under the heading "Type I" and "Type III" sums of squares.

The *Type III sums of squares* are the values *ssA*, *ssB*, and *ssAB* and are used for hypothesis testing whether or not the sample sizes are equal. They are calculated by comparing the sum of squares for error in the full and reduced models as in Sect. 6.4.4. The sums of squares listed in the output are always in the same order as the effects in the MODEL statement, but the hypothesis of no interaction should be tested first.

The *Type I sum of squares* for an effect is the additional variation in the data that is explained by adding that effect to a model containing the previously listed sources of variation. For example, in the program output, the Type I sum of squares for *A* is the reduction in the error sum of squares that is achieved by adding the effect of factor *A* to a model containing only an intercept term. The reduction in the error sum of squares is equivalent to the extra variation in the data that is explained by adding *A* to the model. Here, the "full model" contains *A* and the intercept, while the "reduced model" contains only

Fig. 6.8 Some output for the SAS program for a two-way complete model with unequal sample sizes, using data from the reaction time experiment Table 4.4, p. 101, omitting the last 4 observations

Results Viewer - SAS Output

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.02153160	0.00430632	13.38	0.0010
Error	8	0.00257533	0.00032192		
Corrected Total	13	0.02410693			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	0.02101572	0.02101572	65.28	<.0001
B	2	0.00033302	0.00016651	0.52	0.6148
A*B	2	0.00018286	0.00009143	0.28	0.7600

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	0.01682504	0.01682504	52.27	<.0001
B	2	0.00045773	0.00022887	0.71	0.5198
A*B	2	0.00018286	0.00009143	0.28	0.7600

Results Viewer - SAS Output

The GLM Procedure

Dependent Variable: Y

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
11-13-21+23	1	0.00013886	0.00013886	0.43	0.5298
B1-B2	1	0.00017340	0.00017340	0.54	0.4839

Parameter	Estimate	Standard Error	t Value	Pr > |t|
B1-B2	0.00850000	0.01158153	0.73	0.4839
B1-B3	-0.00600000	0.01370346	-0.44	0.6731
B2-B3	-0.01450000	0.01268694	-1.14	0.2861

the intercept. The Type I sum of squares for B is the additional variation in the data that is explained by adding the effect of factor B to a model that already contains the intercept and the effect of A (so that the "full model" contains A, B and the intercept, while the "reduced model" contains only the A and the intercept). The Type I sums of squares (also known as *sequential sums of squares*) depend upon the order in which the effects are listed in the `MODEL` statement. Type I sums of squares are used for model building, not for hypothesis testing under an assumed model. Consequently, we will use only the Type III sums of squares.

The Type I and Type III sums of squares are identical when the sample sizes are equal, since the factorial effects are then estimated independently of one another. But when the sample sizes are unequal, as in the illustrated data set, the Type I and Type III sums of squares differ. In the absence of a sophisticated computer package, each Type I and Type III sum of squares can be calculated as the difference of the error sums of squares obtained from two analysis of variance tables, one for the full model and one for the reduced model.

6.8.2 Contrasts and Multiple Comparisons

In the first GLM procedure in Table 6.13, the two-way complete model is used, and the coefficient lists are entered for each factor separately, rather than for the treatment combinations. The first CONTRAST statement is used to test the hypothesis that the interaction contrast $(\alpha\beta)_{11} - (\alpha\beta)_{13} - (\alpha\beta)_{21} + (\alpha\beta)_{23}$ is negligible, and the second CONTRAST statement is used to test the hypothesis that $\beta_1^* - \beta_2^*$ is negligible. These same contrasts are entered as coefficient lists for the treatment combinations in the second GLM procedure. In either case, the contrast sum of squares is as shown under Contrast SS in Fig. 6.8, and the p-value for the test is as shown under Pr > F.

The statement

```
LSMEANS A / PDIFF CL ALPHA = 0.01;
```

of the first GLM procedure causes generation of a 99% confidence interval for the main effect of A pairwise comparison, $\alpha_1^* - \alpha_2^*$, comparing the effects of A averaged over the levels of B, as well as an individual 99% confidence interval for each of the A means $\mu + \alpha_i + \overline{\beta}_. + \overline{(\alpha\beta)}_{i.}$.

The statement

```
LSMEANS B / PDIFF = ALL CL ADJUST = TUKEY ALPHA = 0.01;
```

of the first GLM procedure causes generation of Tukey's simultaneous 99% confidence intervals, comparing pairwise the main effects of the three levels of B, each averaged over the levels of A. The option PDIFF = ALL requests p-values for all pairwise comparisons, the option CL asks for the comparisons to be displayed as confidence intervals, and the option ADJUST = TUKEY when coupled with PDIFF = ALL requests Tukey's method for the pairwise comparisons. The output for the reaction time experiment, shown in Fig. 6.9, includes not only the confidence intervals for pairwise comparisons, but also p-values for simultaneous hypothesis tests using the Tukey method. Also given are individual 99% confidence intervals for the B means $\mu + \overline{\alpha}_. + \beta_j + \overline{(\alpha\beta)}_{.j}$. Because sample sizes are unequal, these least squares means are not simply the corresponding treatment sample means. If CL is omitted, then only the simultaneous tests and intervals for means are printed. The request TUKEY can be replaced by BON or SCHEFFE as appropriate.

In the second GLM procedure in Table 6.13, the cell-means model is used, with a treatment effect τ_{ij} associated with each treatment combination ij. The corresponding LSMEANS statements illustrate multiple comparisons of the effects τ_{ij} of the six treatment combinations, the first LSMEANS statement generating Tukey's method for all pairwise comparisons, and the remaining LSMEANS statement generating Dunnett's method for comparing all treatments with a control. To generate Dunnett's method, the option PDIFF = ALL is replaced by the option PDIFF = CONTROL for two-sided confidence intervals, and by the option PDIFF = CONTROLL or PDIFF = CONTROLU for upper or lower confidence bounds on the treatment-versus-control differences, as follows:

```
LSMEANS TC / PDIFF = CONTROL  CL ADJUST = DUNNETT ALPHA = 0.01;
LSMEANS TC / PDIFF = CONTROLL CL ADJUST = DUNNETT ALPHA = 0.01;
LSMEANS TC / PDIFF = CONTROLU CL ADJUST = DUNNETT ALPHA = 0.01;
```

Fig. 6.9 LSMEANS output for a two-way complete model with unequal sample sizes (reaction time experiment)

Results Viewer - SAS Output

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

B	Y LSMEAN	LSMEAN Number
1	0.22750000	1
2	0.21900000	2
3	0.23350000	3

Least Squares Means for effect B Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: Y			
i/j	1	2	3
1		0.7512	0.9010
2	0.7512		0.5166
3	0.9010	0.5166	

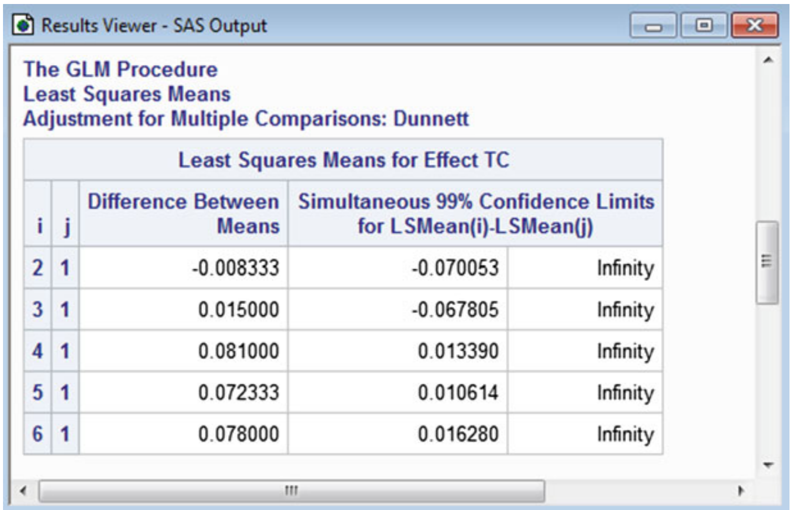
B	Y LSMEAN	99% Confidence Limits	
1	0.227500	0.197399	0.257601
2	0.219000	0.194422	0.243578
3	0.233500	0.198742	0.268258

Least Squares Means for Effect B				
i	j	Difference Between Means	Simultaneous 99% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	0.008500	-0.037650	0.054650
1	3	-0.006000	-0.060605	0.048605
2	3	-0.014500	-0.065055	0.036055

Figure 6.10 contains the output for the third set of simultaneous confidence intervals—namely, corresponding to the `CONTROLU` option. This set gives lower bounds for the treatment-minus-control comparisons, corresponding to upper-tailed inferences. The treatments are renumbered by SAS in numerical order. In our program, in Table 6.13, we have requested the treatment-versus-control contrasts be done for the treatment combinations 11, 12, 13, 21, 22, 23. SAS recodes these as 1–6, and treatment 1 (our treatment combination 11), as the lowest treatment combination, is taken as the control by default. One could specify treatment combination 23 as the control, for example, via the option `PDIFF = CONTROL('23')`. We have shown only the simultaneous confidence intervals, but simultaneous tests are also given by SAS software.

We remind the reader that for unequal sample sizes, it has not yet been proved that the overall confidence levels achieved by the Tukey and Dunnett methods are at least as great as those stated, except in some special cases such as the one-way layout.

An alternative method of obtaining simultaneous confidence intervals for pairwise comparisons can be obtained from the output of the `ESTIMATE` statement for each contrast. The corresponding


Results Viewer - SAS Output

The GLM Procedure
Least Squares Means
Adjustment for Multiple Comparisons: Dunnett

Least Squares Means for Effect TC				
i	j	Difference Between Means	Simultaneous 99% Confidence Limits for LSMean(i)-LSMean(j)	
2	1	-0.008333	-0.070053	Infinity
3	1	0.015000	-0.067805	Infinity
4	1	0.081000	0.013390	Infinity
5	1	0.072333	0.010614	Infinity
6	1	0.078000	0.016280	Infinity

Fig. 6.10 Dunnett's lower bound output for a two-way complete model with unequal sample sizes (reaction time experiment)

confidence intervals are of the form

$$\texttt{Estimate} \pm w\,(\texttt{Std Error of Estimate})\,,$$

where w is the critical coefficient given in (6.4.18) for the complete model and in (6.5.39) for the main-effects model.

6.8.3 Plots

Residual plots for checking the error assumptions on the model are generated in the same way as shown in Chap. 5. If the two-way main-effects model (6.2.3) is used, the assumption of additivity should also be checked. For this purpose it is useful to plot the standardized residuals against the level of one factor, using the levels of the other factor for plotting labels (see, for example, Fig. 6.3, p. 144, for the temperature experiment). A plot of the standardized residuals z against the levels of factor A using the labels of factor B can be generated using the following SAS program lines:

```
PROC SGPLOT;
  SCATTER X = A Y = Z / GROUP = B;
```

An interaction plot can be obtained by adding the following statements to the end of the program in Table 6.13:

```
PROC SORT DATA = RTIME; BY A B;
PROC MEANS DATA = RTIME NOPRINT MEAN VAR; BY A B;
  VAR Y;
  OUTPUT OUT = RTIME2 MEAN = AV_Y VAR = VAR_Y;
PROC PRINT;
  VAR A B AV_Y VAR_Y;
PROC SGPLOT;
  SERIES X = A Y = AV_Y / GROUP = B;
```

The `PROC PRINT` statement following `PROC MEANS` also gives the information about the variances that would be needed to check the rule of thumb that $s^2_{max}/s^2_{min} \le 3$.

In order to check for equal error variances, the residuals or the observations may be plotted against the treatment combinations using the following SAS code:

```
PROC SGPLOT;
  SCATTER X = TC Y = Z;   *or Y = Y for observations;
```

If the treatment combination codes were created as `TC = 10*A + B`, they will not be equally spaced along the axis, since the codes 11, 12, 13, 21, 22, 23 when regarded as 2-digit numbers are not equally spaced. A simple solution to this problem, as shown in Table 6.13, is to convert the variable `TC` from numeric to character via the statement

```
TC = PUT(10*A + B, 2.);
```

A plot of the residuals or the observations against the character variable `TC` will show the character variable codes evenly spaced along the axis.

When there are not sufficient observations to be able to check equality of error variances for all the cells, the standardized residuals should be plotted against the levels of each factor. The rule of thumb may be checked for the levels of each factor by comparing the maximum and minimum variances of the (nonstandardized) residuals. This is done for factor A by the following lines after creation of `RTIME` data set as in Table 6.13.

```
PROC GLM;
  CLASS TC;
  MODEL Y = TC;
  OUTPUT OUT = RESIDS RESIDUAL = E;
PROC SORT DATA = RESIDS; BY A;
PROC MEANS DATA = RESIDS NOPRINT VAR; BY A;
  VAR E;
  OUTPUT OUT = RTIME2 VAR = VAR_E;
PROC PRINT;
  VAR A VAR_E;
```

6.8.4 One Observation Per Cell

In order to split the interaction sum of squares into parts corresponding to negligible and nonnegligible orthogonal contrasts, we can enter the data in the usual manner and obtain the sums of squares for all of the contrasts via `CONTRAST` statements in the procedure `PROC GLM`. The analysis of variance table can then be constructed with the error sum of squares being the sum of the contrast sums of squares for the negligible contrasts. It is possible, however, to achieve this in a more direct way, as follows.

First, enter the contrast coefficients as part of the input data as shown in Table 6.14 for the air velocity experiment. In the air velocity experiment, factor A had $a = 3$ levels and factor B had $b = 6$ levels. The main-effect trend contrast coefficients are entered via the `INPUT` statement line by line directly from Table 6.1, p. 148, and the interaction trend contrast coefficients are obtained by multiplication following the `INPUT` statement. In the `PROC GLM` statement, the `CLASS` designation is omitted. If it were included, then `Aln`, for example, would be interpreted as one factor with three coded levels -1, 0, 1, and `Aqd` as a second factor with two coded levels 1, -2, and so on. The model is fitted using those contrasts that have not been declared to be negligible. The error sum of squares will be based on the three omitted contrasts `AlnBqn`, `AqdBqr`, and `AqdBqn`, and the resulting analysis of variance table will be equivalent to that in Table 6.12, p. 176.

It is not necessary to input the levels of A and B separately as we have done in columns 2 and 3 of the data, but these would be needed if plots of the data were required.

Table 6.14 Fitting a model in terms of contrasts (air velocity experiment)

```
DATA AIR;
  INPUT Y A B Aln Aqd Bln Bqd Bcb Bqr Bqn;
  AlnBln = Aln*Bln;
  AlnBqd = Aln*Bqd;
  AlnBcb = Aln*Bcb;
  AlnBqr = Aln*Bqr;
  AqdBln = Aqd*Bln;
  AqdBqd = Aqd*Bqd;
  AqdBcb = Aqd*Bcb;
  LINES;
  -24      1   1      -1    1      -5   5  -5    1  -1
  -23      1   2      -1    1      -3  -1   7   -3   5
    1      1   3      -1    1      -1  -4   4    2 -10
    8      1   4      -1    1       1  -4  -4    2  10
   29      1   5      -1    1       3  -1  -7   -3  -5
   23      1   6      -1    1       5   5   5    1   1
   33      2   1       0   -2      -5   5  -5    1  -1
   28      2   2       0   -2      -3  -1   7   -3   5
   45      2   3       0   -2      -1  -4   4    2 -10
   57      2   4       0   -2       1  -4  -4    2  10
   74      2   5       0   -2       3  -1  -7   -3  -5
   80      2   6       0   -2       5   5   5    1   1
   37      3   1       1    1      -5   5  -5    1  -1
   79      3   2       1    1      -3  -1   7   -3   5
   79      3   3       1    1      -1  -4   4    2 -10
   95      3   4       1    1       1  -4  -4    2  10
  101      3   5       1    1       3  -1  -7   -3  -5
  111      3   6       1    1       5   5   5    1   1
;
PROC PRINT;
PROC GLM; * omit the class statement;
  MODEL Y = Aln Aqd Bln Bqd Bcb Bqr Bqn AlnBln AlnBqd
                AlnBcb AlnBqr AqdBln AqdBqd  AqdBcb;
```

6.9 Using R Software

Table 6.15 contains a sample R program for analysis of the two-way complete model (6.2.2). For illustration, we use the data of the reaction time experiment shown in Table 4.4, p. 101, but with the last four observations missing, so that $r_{11} = r_{21} = 2$, $r_{12} = r_{22} = r_{23} = 3$, $r_{13} = 1$. The data file includes the levels of each of the two treatment factors A and B, as well as the response, the order in which the observations were collected, and treatment factor levels 1–6. A more descriptive two-digit code for each treatment combination `TC = 10*A + B` is easily generated and added to the data set `react.data`, along with factor variables `fTC`, `fA`, and `fB`, by the statement

```
react.data = within(react.data,
  {TC = 10*A + B; fTC = factor(TC); fA = factor(A); fB = factor(B)})
```

Table 6.15 R program to illustrate aspects of analysis of a two-way complete model (reaction time experiment)

```
react.data = read.table("data/reaction.time.txt", header=T)
react.data = head(react.data, 14) # Keep first 14 observations
head(react.data, 3)

     Order Trtmt A B     y
  1      1     6 2 3 0.256
  2      2     6 2 3 0.281
  3      3     2 1 2 0.167

# Create trtmt combo vbl TC and factors fTC, fA, and fB within data set
react.data = within(react.data,
  {TC = 10*A + B; fTC = factor(TC); fA = factor(A); fB = factor(B)})
summary(react.data[,c("fA","fB","fTC","y")])

# ANOVA
options(contrasts = c("contr.sum", "contr.poly"))
modelAB = aov(y ~ fA + fB + fA:fB, data = react.data)
anova(modelAB) # Type I ANOVA
drop1(modelAB, ~., test = "F") # Type III ANOVA
modelTC = aov(y ~ fTC, data = react.data)
anova(modelTC) # Model F-test

# Contrasts: estimates, CIs, tests
library(lsmeans)
# Main-effect-of-B contrast: B1-B2
lsmB = lsmeans(modelAB, ~ fB)
summary(contrast(lsmB, list(B12=c( 1,-1, 0))), infer=c(T,T))
# AB-interaction contrast: AB11-AB13-AB21+AB23
lsmAB = lsmeans(modelAB, ~ fB:fA) # Using "fB:fA" yields AB lex order
lsmAB # Display to see order of AB combos for contrast coefficients
summary(contrast(lsmAB, list(AB=c( 1 ,0,-1,-1, 0, 1))), infer=c(T,T))

# Multiple comparisons: B
confint(lsmB, level=0.99) # lsmeans for B and 99% CIs
# Tukey's method
summary(contrast(lsmB, method="pairwise", adjust="tukey"),
        infer=c(T,T), level=0.99)
# Dunnett's method
summary(contrast(lsmB, method="trt.vs.ctrl", adj="mvt", ref=1),
        infer=c(T,T), level=0.99)
```

This way of coding the treatment combinations works well for all applications except for drawing plots with TC on one axis. Such a plot would not show codes 11, 12, and 21 as equally spaced. An alternative way of creating the treatment combinations axis for plots will be given in Sect. 6.9.3.

Table 6.16 Analysis of variance output for the R program for a two-way complete model with unequal sample sizes (reaction time experiment)

```
> # ANOVA
> options(contrasts = c("contr.sum", "contr.poly"))
> modelAB = aov(y ~ fA + fB + fA:fB, data = react.data)
> anova(modelAB) # Type I ANOVA

Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value   Pr(>F)
fA         1 0.02102 0.02102   65.28 0.000041
fB         2 0.00033 0.00017    0.52     0.61
fA:fB      2 0.00018 0.00009    0.28     0.76
Residuals  8 0.00258 0.00032

> drop1(modelAB, ~., test = "F") # Type III ANOVA

Single term deletions

Model:
y ~ fA + fB + fA:fB
       Df Sum of Sq     RSS    AIC F value  Pr(>F)
<none>              0.00258 -108.4
fA      1   0.01683 0.01940  -82.1   52.27 0.00009
fB      2   0.00046 0.00303 -110.1    0.71    0.52
fA:fB   2   0.00018 0.00276 -111.5    0.28    0.76

> modelTC = aov(y ~ fTC, data = react.data)
> anova(modelTC) # Model F-test

Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value Pr(>F)
fTC        5 0.02153 0.00431    13.4  0.001
Residuals  8 0.00258 0.00032
```

6.9.1 Analysis of Variance

Referring to the R program in Table 6.15, the block of code under the comment "ANOVA" generates the analysis of variance output as shown in Table 6.16. As in the one-way analysis of variance, the treatment factors must be factor variables to be modeled as qualitative variables. The statements

```
modelAB = aov(y ~ fA + fB + fA:fB, data = react.data)
modelTC = aov(y ~ fTC, data = react.data)
```

fit the two-way complete model (6.2.2) and the cell-means model (6.2.1), respectively, saving the results as `modelAB` and `modelTC`. In the first model, the main effects may be listed in either order, but before the interaction effects `fA:fB`. Equivalently, the model could be specified as `y ~ fA*fB`, since `fA*fB` indicates inclusion of all main effects and interactions involving the factors `fA` and `fB`. The two-way main-effects model (6.2.3) would be represented as `y ~ fA + fB`.

Using the saved information from the fitted models, the three statements

```
anova(modelAB)
drop1(modelAB, ~., test ="F")
anova(modelTC)
```

respectively generate the "`Type I ANOVA`" shown at the top of Table 6.16, the "`Type III ANOVA`" shown in the middle of the table, and the "`Model F-test`" shown at the bottom of the table. In the first and last portion of the table, "`Residuals`" is synonymous with error. Totals for the degrees of freedom and sum of squares are not provided. For technical reasons, the statement

```
options(contrasts = c("contr.sum","contr.poly"))
```

must be executed prior to fitting the two-way complete model for all Type III sum of squares to be correct. (This option imposes common sum-to-zero constraints on the least squares estimates of the treatment factors effects, so each $\hat{\alpha}_i$ estimates a main effect of A contrast, each $\hat{\beta}_j$ estimates a main effect of B contrast, and each $\widehat{(\alpha_i\beta_j)}$ estimates an AB-interaction contrast.)

In the "`Type III ANOVA`" given in the middle of Table 6.16, the (Type III) sums of squares are the values *ssA*, *ssB*, and *ssAB* and are used for hypothesis testing whether or not the sample sizes are equal. Each is calculated by comparing the sum of squares for error in the full and reduced models as in Sect. 6.4.4, where each reduced model is obtained by dropping the corresponding term from the full model—the two-way complete model in this case. The hypothesis of no interaction should be tested first, even though its sum of squares is listed last in the output. For each effect, its listed `F value` may be computed from its listed sum of squares and degrees of freedom, and from the mean square for residuals listed in the analysis of variance table at either the top and bottom of Table 6.16. The corresponding p-value is listed under "`Pr(>F)`". (The reader may ignore the `RSS` and `AIC` columns.)

The sums of squares listed in the "`Type I ANOVA`" shown at the top of Table 6.16 are *Type I sum of squares*, or sequential sum of squares. For each effect, this is the additional variation in the data that is explained by adding that effect to a model containing the previously listed sources of variation. *Type I sum of squares* are discussed in more detail in Sect. 6.8.1. They are used for model building, not for hypothesis testing under an assumed model. Consequently, we will use only the Type III sums of squares. That said, the Type I and Type III sums of squares are identical when the sample sizes are equal, since the factorial effects are then estimated independently of one another. When the Type I and III analyses are the same, it seems preferable to use the cleaner Type I analysis of variance table as shown at the top of Table 6.16.

Finally, under "`Model F-test`" at the bottom of Table 6.16, an analysis of variance table is provided for testing model significance.

6.9.2 Contrasts and Multiple Comparisons

Information on individual contrasts is generated by coupling the `summary` and `contrast` functions of the `lsmeans` package, as was illustrated in Chap. 4. After loading the `lsmeans` package, the statements

```
lsmB = lsmeans(modelAB, ~ fB)
summary(contrast(lsmB, list(B12=c( 1,-1, 0))), infer=c(T,T))
```

Table 6.17 Contrasts output for the R program for a two-way complete model with unequal sample sizes (reaction time experiment)

```
> # Main-effect-of-B contrast: B1-B2
> lsmB = lsmeans(modelAB, ~ fB)

> summary(contrast(lsmB, list(B12=c( 1,-1, 0))), infer=c(T,T))

 contrast estimate       SE df  lower.CL upper.CL t.ratio p.value
 B12        0.0085 0.011582  8 -0.018207 0.035207   0.734  0.4839

Results are averaged over the levels of: fA
Confidence level used: 0.95

> # AB-interaction contrast: AB11-AB13-AB21+AB23
> lsmAB = lsmeans(modelAB, ~ fB:fA) # Using "fB:fA" yields AB lex order
> lsmAB # Display to see order of AB combos for contrast coefficients

 fB fA  lsmean       SE df lower.CL upper.CL
 1  1  0.18700 0.012687  8  0.15774  0.21626
 2  1  0.17867 0.010359  8  0.15478  0.20255
 3  1  0.20200 0.017942  8  0.16063  0.24337
 1  2  0.26800 0.012687  8  0.23874  0.29726
 2  2  0.25933 0.010359  8  0.23545  0.28322
 3  2  0.26500 0.010359  8  0.24111  0.28889

Confidence level used: 0.95

> summary(contrast(lsmAB, list(AB=c( 1 ,0,-1,-1, 0, 1))), infer=c(T,T))

 contrast estimate       SE df lower.CL upper.CL t.ratio p.value
 AB         -0.018 0.027407  8  -0.0812   0.0452  -0.657  0.5298

Confidence level used: 0.95
```

in Table 6.15 generate information for the main effect of B contrast $\beta_1^* - \beta_2^*$, comparing the effects of B averaged over the levels of A, using the two-way complete model previously fit and saved as `modelAB`. The following statements generate analogous information for the interaction contrast $(\alpha\beta)_{11} - (\alpha\beta)_{13} - (\alpha\beta)_{21} + (\alpha\beta)_{23}$.

```
lsmAB = lsmeans(modelAB, ~ fB:fA)
summary(contrast(lsmAB, list(AB=c( 1 ,0,-1,-1, 0, 1))), infer=c(T,T))
```

Using `fB:fA` rather than `fA:fB` yields least squares means in the standard lexicographical order, as can be seen by displaying `lsmAB`, and the contrast coefficients must be in the corresponding order. These statements and their output are shown in Table 6.17.

Multiple comparisons procedures are also implemented using functions of the `lsmeans` package, as illustrated for levels of B by sample code at the bottom of Table 6.15. The statements

```
lsmB = lsmeans(modelAB, ~ fB)
confint(lsmB, level = 0.99)
```

Table 6.18 Multiple comparisons output for a two-way complete model with unequal sample sizes (reaction time experiment)

```
> # Multiple comparisons: B
> confint(lsmB, level=0.99) # lsmeans for B and 99% CIs

 fB lsmean        SE df lower.CL upper.CL
 1  0.2275 0.0089710  8  0.19740  0.25760
 2  0.2190 0.0073248  8  0.19442  0.24358
 3  0.2335 0.0103588  8  0.19874  0.26826

Results are averaged over the levels of: fA
Confidence level used: 0.99

> # Tukey's method
> summary(contrast(lsmB, method="pairwise", adjust="tukey"),
+         infer=c(T,T), level=0.99)

 contrast estimate        SE df  lower.CL upper.CL t.ratio p.value
 1 - 2      0.0085 0.011582  8 -0.037650 0.054650   0.734  0.7512
 1 - 3     -0.0060 0.013703  8 -0.060606 0.048606  -0.438  0.9010
 2 - 3     -0.0145 0.012687  8 -0.065055 0.036055  -1.143  0.5166

Results are averaged over the levels of: fA
Confidence level used: 0.99
Confidence-level adjustment: tukey method for a family of 3 estimates
P value adjustment: tukey method for a family of 3 tests
```

in turn (i) compute least squares estimates of the means $\mu + \beta_j + \overline{\alpha}_. + \overline{(\alpha\beta)}_{.j}$ and related information, saving this information as `lsmB`; and (ii) display the least squares estimates, standard errors, degrees of freedom, and individual 99% confidence intervals shown at the top of Table 6.18. Because sample sizes are unequal, these least squares means are not simply the corresponding treatment sample means.

The statement

```
summary(contrast(lsmB, method = "pairwise", adjust = "tukey"),
        infer = c(T, T), level = 0.99)
```

applies Tukey's method, comparing pairwise the main effects of the three levels of B, each averaged over the levels of A. The `contrast` function coupled with the options `method="pairwise"` and `adjust="tukey"` requests tests including p-values for all pairwise comparisons via Tukey's method. Other adjustment options for pairwise comparisons include `"Scheffe"` for Scheffé's method, `"Bonf"` for the Bonferroni method, and `"none"` for individual inferences. The `summary` function with its `infer=c(T,T)` and `level=0.99` options requests Tukey's 99% confidence intervals. Specifically, the option `infer=c(T,T)` indicates "true" for confidence intervals and tests, respectively. The above statement and the corresponding output are shown in the bottom of Table 6.18.

Implementation of Dunnett's method for all treatment-versus-control comparisons is similar and illustrated by the following statement

```
summary(contrast(lsmB, method="trt.vs.ctrl", adj="mvt", ref=1),
        infer=c(T,T), level=0.99)
```

Table 6.19 Dunnett's lower band for a two-way complete model with unequal sample sizes (reaction time experiment)

```
> # Dunnett's method comparing cell means
> lsmTC = lsmeans(modelTC, ~ fTC)
> summary(contrast(lsmTC, method="trt.vs.ctrl", adj="mvt"),
+         infer=c(T,T), level=0.99, side=">")

 contrast   estimate        SE df  lower.CL upper.CL t.ratio p.value
 12 - 11  -0.0083333 0.016379  8 -0.070333      Inf  -0.509  0.9351
 13 - 11   0.0150000 0.021974  8 -0.068181      Inf   0.683  0.5629
 21 - 11   0.0810000 0.017942  8  0.013083      Inf   4.515  0.0037
 22 - 11   0.0723333 0.016379  8  0.010334      Inf   4.416  0.0042
 23 - 11   0.0780000 0.016379  8  0.016000      Inf   4.762  0.0025

Confidence level used: 0.99
Confidence-level adjustment: mvt method for 5 estimates
P value adjustment: mvt method for 5 tests
P values are right-tailed
```

from the bottom of Table 6.15. The option `method="trt.vs.ctrl"` yields all treatment-versus-control comparisons. Here, other levels of B are compared to the first level, which happens to be `"1"`, averaging over levels of A. Available options include the following, as discussed in Sect. 4.7.2. Dunnett's method uses (simulation based) critical values from the multivariate t-distribution, corresponding to `adjust="mvt"`, but the default option `adjust="dunnettx"` provides an approximation of Dunnett's method for two-sided confidence intervals that runs faster and dependably (but is appropriate only applicable when the contrast estimates have pairwise correlations of 0.5). The first level of a factor is the control by default, corresponding to reference level 1 (`ref=1`), but one could, for example, specify the second level as the control by the syntax `ref=2`. Also, `"two-sided"` is the default for confidence intervals and tests, but one can specify `side="<"` for the one-sided alternative $H_A : \tau_i < \tau_1$ and the corresponding upper confidence bound for $\tau_i - \tau_1$, or `side=">"` for the alternative $H_A : \tau_i > \tau_1$ and the corresponding lower confidence bound for $\tau_i - \tau_1$.

Multiple comparisons of all treatments may be obtained using the cell-means model, as illustrated for Dunnett's method by the following code, reproduced with corresponding output in Table 6.19.

```
lsmTC = lsmeans(modelTC, ~ fTC)
summary(contrast(lsmTC, method="trt.vs.ctrl", adj="mvt"),
        infer=c(T,T), level=0.99, side=">")
```

The same could be accomplished using `lsmAB` instead of `lsmTC`. Note that the default control here is `"11"`, which is the first level of `fTC`.

We remind the reader that for unequal sample sizes, it has not yet been proved that the overall confidence levels achieved by the Tukey and Dunnett methods are at least as great as those stated, except in some special cases such as the one-way layout.

6.9.3 Plots

Residual plots for checking the error assumptions on the model are generated in the same way as shown in Chap. 5. If the two-way main-effects model (6.2.3) is used, the assumption of additivity should also be checked. For this purpose it is useful to plot the standardized residuals against the level of one factor, using the levels of the other factor for plotting labels (see, for example, Fig. 6.3, p. 144, for the temperature experiment). A plot of the standardized residuals z against the levels of factor A using the labels of factor B can be generated using the following R program lines:

```
plot(z ~ A, data=react.data, xaxt="n", type="n") # Suppress x-axis, pts
  axis(1, at=seq(1,2,1)) # Add x-axis with tick marks from 1 to 2 by 1
  text(z ~ A, B, cex=0.75, data=react.data) # Plot z vs A using B label
  mtext("B=1,2,3", side=3, adj=1, line=1) # Margin text, top-rt, line 1
  abline(h=0) # Horizontal line at zero
```

An interaction plot of treatment means $\bar{y}_{ij.}$ versus levels of factor A using the labels of factor B can be generated by adding the following statement to the end of the program in Table 6.15.

```
interaction.plot(x.factor = react.data$fA, trace.factor = react.data$fB,
                 response = react.data$y, type ="b",
                 xlab ="A", trace.label ="B", ylab ="Mean of y")
```

The option `type="b"` plots both points and lines.

In order to check for equal error variances, the residuals or observations could be plotted against the treatment combinations using the following R code:

```
plot(modelAB$res ~ react.data$TC, xlab ="AB", ylab ="Residual")
plot(react.data$y ~ react.data$TC, xlab ="AB", ylab ="y")
```

However, since the treatment combination codes `TC = 10*A + B` are numeric as originally created in Table 6.13, they will not be equally spaced along the axis, since the codes 11, 12, 13, 21, 22, 23 when regarded as 2-digit numbers are not equally spaced. One solution to this problem is to plot the residuals or the observations against the treatment variable `Trtmt`, since its levels 1–6 are equally spaced, but replace each of the labels 1–6 with the corresponding treatment combination label. This is accomplished by the following code.

```
plot(modelAB$res ~ react.data$Trtmt, xaxt="n", xlab="AB", ylab="Residual")
  axis(1, at = react.data$Trtmt, labels = react.data$fTC)
plot(react.data$y ~ react.data$Trtmt, xaxt="n", xlab="AB", ylab="y")
  axis(1, at = react.data$Trtmt, labels = react.data$fTC)
```

In the first `plot` statement, for example, the residuals are plotted against `Trtmt`, which has equally spaced levels 1–6, but the x-axis is suppressed by the option `xaxt="n"`. The `axis` statement is then used to create an x-axis, still with tick marks at the equally-spaced `Trtmt` levels 1–6, but using the treatment combination labels 11, 12,…, 23 of `fTC`.

When there are not sufficient observations to be able to check equality of error variances for all the cells, the standardized residuals should be plotted against the levels of each factor. The rule of thumb may be checked for the levels of each factor by comparing the maximum and minimum variances of the (nonstandardized) residuals. The sample variance of the residuals may be computed by level of A, for example, by augmenting the statements in Table 6.15 with the following `by` command.

```
by(modelAB$res, react.data$A, var)
```

Table 6.20 Fitting a model in terms of contrasts (air velocity experiment)

```
air.data = read.table("data/air.velocity.contrasts.txt", header=T)
air.data

      y A B  Aln Aqd Bln Bqd Bcb Bqr Bqn
1   -24 1 1  -1   1  -5   5  -5   1  -1
2   -23 1 2  -1   1  -3  -1   7  -3   5
3     1 1 3  -1   1  -1  -4   4   2 -10
4     8 1 4  -1   1   1  -4  -4   2  10
5    29 1 5  -1   1   3  -1  -7  -3  -5
6    23 1 6  -1   1   5   5   5   1   1
7    33 2 1   0  -2  -5   5  -5   1  -1
8    28 2 2   0  -2  -3  -1   7  -3   5
9    45 2 3   0  -2  -1  -4   4   2 -10
10   57 2 4   0  -2   1  -4  -4   2  10
11   74 2 5   0  -2   3  -1  -7  -3  -5
12   80 2 6   0  -2   5   5   5   1   1
13   37 3 1   1   1  -5   5  -5   1  -1
14   79 3 2   1   1  -3  -1   7  -3   5
15   79 3 3   1   1  -1  -4   4   2 -10
16   95 3 4   1   1   1  -4  -4   2  10
17  101 3 5   1   1   3  -1  -7  -3  -5
18  111 3 6   1   1   5   5   5   1   1

# Fit linear regression model, save as model1
model1 = lm(y ~ Aln + Aqd + Bln + Bqd + Bcb + Bqr + Bqn
               + Aln:Bln + Aln:Bqd + Aln:Bcb + Aln:Bqr
               + Aqd:Bln + Aqd:Bqd + Aqd:Bcb, data=air.data)
# ANOVA
anova(model1)
```

6.9.4 One Observation Per Cell

In this section we present a direct way to split the interaction sum of squares into parts corresponding to negligible and nonnegligible orthogonal contrasts. The R program in Table 6.20 illustrates the method, using the data of the air velocity experiment.

First, the main effect contrast coefficients from Table 6.1, p. 148, are entered as part of the data, as is evident from the displayed data in Table 6.20. In the air velocity experiment, factor A had $a = 3$ levels and factor B had $b = 6$ levels.

The `lm` statement in Table 6.20 fits a linear model, including as predictors those contrasts that have not been declared to be negligible. For example, `Aln` and `Bln` terms contain the coefficients of the A-linear and B-linear contrasts, respectively, and the term `Aln:Bln` contains the coefficients of the A-linear-by-B-linear interaction trend contrast, which R obtains by multiplication of the corresponding `Aln` and `Bln` contrast coefficients. The results are saved as `model1`.

Intentionally, none of the predictor variables in the model are factor variables. If they were factor variables, then `Aln` would be interpreted as one factor with three coded levels -1, 0, 1, and `Aqd` as a second factor with two coded levels 1, -2, and so on. The error sum of squares will be based on the

Table 6.21 Data (beats per 15 seconds) for the weight lifting experiment

A	2	1	1	1	2	2	1	2	1	2	1
B	2	1	3	1	2	3	3	2	2	2	1
Rate	31	27	37	28	32	32	35	30	32	31	27
A	2	1	1	1	2	1	1	2	2	2	2
B	2	3	3	2	1	1	3	1	3	3	3
Rate	34	33	34	31	26	25	35	24	33	31	36
A	1	1	1	1	1	1	1	1	2	1	2
B	3	1	1	2	1	2	2	3	3	2	1
Rate	36	27	30	33	29	32	34	37	32	34	27
A	2	1	1	2	2	1	1	2	2	1	2
B	2	1	3	1	2	1	2	1	2	1	3
Rate	31	27	38	27	30	29	34	25	34	28	34
A	2	1	2	1	1	2	1	1	2	2	2
B	2	1	3	2	2	1	3	2	1	1	1
Rate	31	30	34	35	34	24	35	31	27	26	25
A	2	2	2	1	2	2	2	2	2	1	1
B	2	3	1	2	1	2	3	3	3	3	3
Rate	32	35	24	33	23	30	34	32	33	37	38

three omitted contrasts `Aln:Bqn`, `Aqd:Bqr`, and `Aqd:Bqn`, and the resulting analysis of variance table generated by the `anova(model1)` statement will be equivalent to that in Table 6.12, p. 176.

It is not necessary to input the levels of A and B separately as we have done in columns 2 and 3 of the data, but these would be needed if plots of the data were required.

Exercises

1. Under what circumstances should the two-way main effects model (6.2.3) be used rather than the two-way complete model (6.2.2)? Discuss the interpretation of main effects in each model.

2. Verify that $(\tau_{ij} - \overline{\tau}_{i.} - \overline{\tau}_{.j} + \overline{\tau}_{..})$ is an interaction contrast for the two-way complete model. Write down the list of contrast coefficients in terms of the τ_{ij}'s when factor A has $a = 3$ levels and factor B has $b = 4$ levels.

3. Consider the functions $\{\alpha_1^* - \alpha_2^*\}$ and $\{(\alpha\beta)_{11} - (\alpha\beta)_{21} - (\alpha\beta)_{12} + (\alpha\beta)_{22}\}$ under the two-way complete model (6.2.2).

   (a) Verify that the functions are estimable contrasts.
   (b) Discuss the meaning of each of these contrasts for plot (d) of Fig. 6.1, p. 140, and for plot (g) of Fig. 6.2, p. 141.
   (c) If $a = b = 3$, give the list of contrast coefficients for each contrast, first for the parameters involved in the contrast, and then in terms of the parameters τ_{ij} of the equivalent cell-means model.

4. Show that when the parentheses is expanded in formula (6.4.15) for ssE on p. 151, the computational formula (6.4.16) is obtained.

5. **Weight Lifting Experiment** (Gary Mirka 1986)
The experimenter was interested in the effect on pulse rate (heart rate) of lifting different weights with legs either straight or bent (factor A, coded 1, 2). The selected weights were 50 lb, 75 lb, 100 lb (factor B, coded 1, 2, 3). He expected to see a higher pulse rate when heavier weights were lifted. He also expected that lifting with legs bent would result in a higher pulse rate than lifting with legs straight.

(a) Write out a detailed checklist for running an experiment similar to this. In the calculation of the number of observations needed, either run your own pilot experiment or use the information that for a single subject in the above study, the error sum of squares was $ssE = 130.909$ bpfs^2 based on $df = 60$ error degrees of freedom (where bpfs is beats per 15 seconds).
(b) The data collected for a single subject by the above experimenter are shown in Table 6.21 in the order collected. The experimenter wanted to use a two-way complete model. Check the assumptions on this model, paying particular attention to the facts that (i) these are count data and may not be approximately normally distributed, and (ii) the measurements were made in groups of ten at a time in order to reduce the fatigue of the subject.
(c) Taking account of your answer to part (a), analyze the experiment, especially noting any trends in the response.

6. **Battery experiment, continued**

Consider the battery experiment introduced in Sect. 2.5.2, p. 24, for which $a = b = 2$ and $r = 4$. Suppose it is of interest to calculate confidence intervals for the four simple effects $\tau_{11} - \tau_{12}$, $\tau_{21} - \tau_{22}$, $\tau_{11} - \tau_{21}$, $\tau_{12} - \tau_{22}$, with an overall confidence level of 95%.

(a) Determine whether the Tukey or Bonferroni method of multiple comparisons would provide shorter confidence intervals.
(b) Apply the better method from part (a) and comment on the results. (The data give $\overline{y}_{11.} = 570.75$, $\overline{y}_{12.} = 860.50$, $\overline{y}_{21.} = 433.00$, and $\overline{y}_{22.} = 496.25$ minutes per unit cost and $msE = 2,367.71$.)
(c) Discuss the practical meaning of the contrasts estimated in (b) and explain what you have learned from the confidence intervals.

7. **Weld strength experiment**

The data shown in Table 6.22 are a subset of the data given by Anderson and McLean (1974) and show the strength of a weld in a steel bar. Two factors of interest were gage bar setting (the distance the weld die travels during the automatic weld cycle) and time of welding (total time of the automatic weld cycle). Assume that the levels of both factors were selected to be equally spaced.

(a) Using the cell-means model (6.2.1) for these data, test the hypothesis that there is no difference in the effects of the treatment combinations on weld strength against the alternative hypothesis that at least two treatment combinations have different effects.
(b) Suppose the experimenters had planned to calculate confidence intervals for all pairwise comparisons between the treatment combinations, and also to look at the confidence interval for the difference between gage bar setting 3 and the average of the other two. Write down the contrasts

Table 6.22 Strength of weld

	i	Time of welding (*j*) 1	2	3	4	5
Gage	1	10, 12	13, 17	21, 30	18, 16	17, 21
bar	2	15, 19	14, 12	30, 38	15, 11	14, 12
setting	3	10, 8	12, 9	10, 5	14, 15	19, 11

Source Reprinted from Anderson and McLean (1974), pp. 62–63, by courtesy of Marcel Dekker, Inc

in terms of the parameters τ_{ij} of the cell-means model, and suggest a strategy for calculating all intervals at overall level "at least 98%."

(c) Consider the intervals in part (b). Give the formulae and calculate the actual interval for $\tau_{13} - \tau_{15}$ (the difference in the true mean strengths at the 3rd and 5th times of welding for the first gage bar setting), and explain what this interval tells you. Also calculate the actual interval for the difference between gage bar setting 3 and the average of the other two, and explain what this interval tells you.

(d) Calculate an upper 90% confidence limit for σ^2.

(e) If the experimenters were to repeat this experiment and needed the pairwise comparison intervals in (b) to be of width at most 8, how many observations should they take on each treatment combination? How many observations is this in total?

8. **Weld strength experiment, continued**

For the experiment described in Exercise 7, use the two-way complete model instead of the equivalent cell means model.

(a) Test the hypothesis of no interaction between gage bar setting and time of weld and state your conclusion.

(b) Draw an interaction plot for the two factors Gage bar setting and Time of welding. Does your interaction plot support the conclusion of your hypothesis test? Explain.

(c) In view of your answer to part (b), is it sensible to investigate the differences between the effects of gage bar setting? Why or why not? Indicate on your plot what would be compared.

(d) Regardless of your answer to (c), suppose the experimenters had decided to look at the linear trend in the effect of gage bar settings. Test the hypothesis that the linear trend in gage setting is negligible (against the alternative hypothesis that it is not negligible).

9. **Sample size calculation**

An experiment is to be run to examine three levels of factor *A* and four levels of factor *B*, using the two-way complete model (6.2.2). Determine the required sample size if the error variance σ^2 is expected to be less than 15 and simultaneous 99% confidence intervals for pairwise comparisons between treatment combinations should have length at most 10 to be useful.

10. **Bleach experiment, continued**

Use the data of the bleach experiment of Example 6.4.4, on p. 154.

(a) Evaluate the effectiveness of a variance-equalizing transformation.

(b) Apply Satterthwaite's approximation to obtain 99% confidence intervals for the pairwise comparisons of the main effects of factor *A* using Tukey's method of multiple comparisons.

11. **Bleach experiment, continued**

The experimenter calculated that she needed $r = 5$ observations per treatment combination in order to be able to detect a difference in the effect of the levels of either treatment factor of 5 minutes (300 seconds) with probability 0.9 at significance level 0.05. Verify that her calculations were correct. She obtained a mean squared error of 43220.8 in her pilot experiment.

12. **Memory experiment** (James Bost 1987)

The memory experiment was planned in order to examine the effects of external distractions on short-term memory and also to examine whether some types of words were easier to memorize than others. Consequently, the experiment involved two treatment factors, "word type" and "type of distraction." The experimenter selected three levels for each factor. The levels of "word type" were

Level 1 (fruit): words representing fruits and vegetables commonly consumed;
Level 2 (nouns): words selected at random from Webster's pocket dictionary, representing tangible (i.e., visualizable) items;
Level 3 (mixed): words of any description selected at random from Webster's pocket dictionary.

A list of 30 words was prepared for each level of the treatment factor, and the list was not altered throughout the experiment.
The levels of "type of distraction" were

Level 1 : No distraction other than usual background noise;
Level 2 : Constant distraction, supplied by a regular banging of a metal spoon on a metal pan;
Level 3 : Changing distraction, which included vocal, music, banging and motor noise, and varying lighting.

The response variable was the number of words remembered (by a randomly selected subject) for a given treatment combination. The response variable is likely to have approximately a binomial distribution, with variance $30_p(1-p)$ where p is the probability that a subject remembers a given word and 30 is the number of words on the list. It is unlikely that p is constant for all treatment combinations or for all subjects. However, since $np(1-p)$ is less than $30(0.5)(0.5) = 7.5$, a reasonable guess for the variance σ^2 is that it is less than 7.5.
The experimenter wanted to reject each of the main-effect hypotheses H_0^A:{the memorization rate for the three word lists is the same} and H_0^B:{the three types of distraction have the same effect on memorization} with probability 0.9 if there was a difference of four words in memorization rates between any two word lists or any two distractions (that is $\Delta_A = \Delta_B = 4$), using a significance level of $\alpha = 0.05$. Calculate the number of subjects that are needed if each subject is to be assigned to just one treatment combination and measured just once.

13. **Memory experiment, continued**

(a) Write out a checklist for the memory experiment of Exercise 12. Discuss how you would obtain the subjects and how applicable the experiment would be to the general population.
(b) Consider the possibility of using each subject more than once (i.e., consider the use of a blocking factor). Discuss whether or not an assumption of independent observations is likely to be valid.

14. **Memory experiment, continued**

The data for the memory experiment of Exercise 12 are shown in Table 6.23 with three observations per treatment combination.

Table 6.23 Data and standardized residuals for the memory experiment

	Distraction								
Word list	None			Constant			Changing		
Fruit	20	14	24	15	22	17	17	13	12
	0.27	−2.16	1.89	−1.21	1.62	−0.40	1.21	−0.40	−0.81
Nouns	19	14	19	12	11	14	12	15	8
	0.67	−1.35	0.67	−0.13	−0.54	0.67	0.13	1.35	−1.48
Mixed	11	12	15	8	8	9	12	7	10
	−0.67	−0.27	0.94	−0.13	−0.13	0.27	0.94	−1.08	0.13

(a) The experimenter intended to use the two-way complete model. Check the assumptions on the model for the given data, especially the equal-variance assumption.
(b) Analyze the experiment. A transformation of the data or use of the Satterthwaite approximation may be necessary.

15. **Ink experiment**

Teaching associates who give classes in computer labs at the Ohio State University are required to write on white boards with "dry markers" rather than on chalk boards with chalk. The ink from these dry markers can stain rather badly, and an experiment was planned by M. Chambers, Y.-W. Chen, E. Kurali and R. Vengurlekar in 1996 to determine which type of cloth (factor A, 1 = cotton/polyester, 2 = rayon, 3 = polyester) was most difficult to clean, and whether a detergent plus stain remover was better than a detergent without stain remover (factor B, levels 1, 2) for washing out such a stain.
Pieces of cloth were to be stained with 0.1 ml of dry marker ink and allowed to air dry for 24 hours. The cloth pieces were then to be washed in a random order in the detergent to which they were allocated. The stain remaining on a piece of cloth after washing and drying was to be compared with a 19 point scale and scored accordingly, where 1 = black and 19 = white.

(a) Make a list of the difficulties that might be encountered in running and analyzing an experiment of this type. Give suggestions on how these difficulties might be overcome or their effects reduced.
(b) Why should each piece of cloth be washed separately? (Hint: think about the error variability.)
(c) The results of a small pilot study run by the four experimenters are shown in Table 6.24. Plot the data against the levels of the two treatment factors. Can you learn anything from this plot? Which model would you select for the main experiment? Why?
(d) Calculate the number of observations that you would need to take on each treatment combination in order to try to ensure that the lengths of confidence intervals for pairwise differences in the effects of the levels of each of the factors were no more than 2 points (on the 19-point scale).

Table 6.24 Data for the ink experiment in the order of collection

Cloth type	3	1	3	1	2	1	2	2	2	3	3	1
Stain remover	2	2	2	2	1	1	1	2	2	1	1	1
Stain score	1	6	1	5	11	9	9	8	6	3	4	8

Table 6.25 Data for the survival experiment (units of 10 hours)

	Treatment			
Poison	1	2	3	4
I	0.31	0.82	0.43	0.45
	0.45	1.10	0.45	0.71
	0.46	0.88	0.63	0.66
	0.43	0.72	0.76	0.62
II	0.36	0.92	0.44	0.56
	0.29	0.61	0.35	1.02
	0.40	0.49	0.31	0.71
	0.23	1.24	0.40	0.38
III	0.22	0.30	0.23	0.30
	0.21	0.37	0.25	0.36
	0.18	0.38	0.24	0.31
	0.23	0.29	0.22	0.33

Source Box and Cox (1964). Copyright 1964 Blackwell Publishers. Reprinted with permission

16. **Survival experiment** (G.E.P. Box and D.R. Cox, 1964)

The data in Table 6.25 show survival times of animals to whom a poison and a treatment have been administered. The data were presented by G. E. P. Box and D. R. Cox in an article in the *Journal of the Royal Statistical Society* in 1964. There were three poisons (factor A at $a = 3$ levels), four treatments (factor B at $b = 4$ levels), and $r = 4$ animals (experimental units) assigned at random to each treatment combination.

(a) Check the assumptions on a two-way complete model for these data. If the assumptions are satisfied, then analyze the data and discuss your conclusions.
(b) Take a reciprocal transformation (y^{-1}) of the data. The transformed data values then represent "rates of dying." Check the assumptions on the model again. If the assumptions are satisfied, then analyze the data and discuss your conclusions.
(c) Draw an interaction plot for both the original and the transformed data. Discuss the interaction between the two factors in each of the measurement scales.

17. Use the two-way main-effects model (6.2.3) with $a = b = 3$.

(a) Which of the following are estimable?

(i) $\mu + \alpha_1 + \beta_2$.
(ii) $\mu + \alpha_1 + \frac{1}{2}(\beta_1 + \beta_2)$.
(iii) $\beta_1 - \frac{1}{3}(\beta_2 + \beta_3)$.

(b) Show that $\overline{Y}_{i..} + \overline{Y}_{.j.} - \overline{Y}_{...}$ is an unbiased estimator of $\mu + \alpha_i + \beta_j$ with variance, $\sigma^2(a+b-1)/(abr)$.
(c) Show that $\sum_i c_i \overline{Y}_{i..}$ is an unbiased estimator of the contrast $\sum_i c_i \alpha_i$.

18. **Meat cooking experiment, continued**

The meat cooking experiment was introduced in Exercise 14 of Chap. 3, with the data given in Table 3.14, p. 68.

Table 6.26 Data for the water boiling experiment, in minutes. (Order of observation is in parentheses.)

	Salt (teaspoons)			
Burner	0	2	4	6
Right back	7(7)	4(13)	7(24)	5(15)
	8(21)	7(25)	7(34)	7(33)
	7(30)	7(26)	7(41)	7(37)
Right front	4(6)	4(36)	4(1)	4(28)
	4(20)	5(44)	4(14)	4(31)
	4(27)	4(45)	5(18)	4(38)
Left back	6(9)	6(46)	7(8)	5(35)
	7(16)	6(47)	6(12)	6(39)
	6(22)	5(48)	7(43)	6(40)
Left front	9(29)	8(5)	8(3)	8(2)
	9(32)	8(10)	9(19)	8(4)
	9(42)	8(11)	10(23)	7(17)

(a) Using the two-way complete model, conduct an analysis of variance, testing each hypothesis using a 1% significance level. State your conclusions.
(b) Draw an interaction plot for the two treatment factors. Does your interaction plot support the conclusion of your hypothesis test concerning interactions? Explain.
(c) Compare the effects of the three levels of fat content pairwise, averaging over cooking methods, using Scheffé's method for all treatment contrasts with a 95% confidence level. Interpret the results.
(d) Give a confidence interval for the average difference in weight after cooking between frying and grilling 110 g hamburgers, using Scheffé's method for all treatment contrasts with a 95% confidence level. Interpret the results.
(e) Obtain a 95% confidence interval for comparing the effect on post-cooked weight of the low fat content versus the average of the two higher fat contents (averaged over cooking method), using Scheffé's method.
(f) What is the overall confidence level of the intervals in parts (c), (d) and (e) taken together?
(g) If the contrast in part (e) had been the *only* contrast of interest, would your answer to part (e) have been different? If so, show the new calculation. If not, explain why not.

19. For the two-way main-effects model (6.2.3) with equal sample sizes,

(a) verify the computational formulae for ssE given in (6.5.38),
(b) and, if SSE is the corresponding random variable, show that $E[SSE]$ is $(n - a - b + 1)\sigma^2$. [Hint: $\text{E}[X^2] = \text{Var}(X) + \text{E}[X]^2$.]

20. An experiment is to be run to compare the two levels of factor A and to examine the pairwise differences between the four levels of factor B, with a simultaneous confidence level of 90%. The experimenter is confident that the two factors will not interact. Find the required sample size if the error variance will be at most 25 and the confidence intervals should have length at most 10 to be useful.

21. **Water boiling experiment** (Kate Ellis 1986)

The experiment was run in order to examine the amount of time taken to boil a given amount of water on the four different burners of her stove, and with 0, 2, 4, or 6 teaspoons of salt added to the water. Thus the experiment had two treatment factors with four levels each. The experimenter ran the experiment as a completely randomized design by taking $r = 3$ observations on each of the 16 treatment combinations in a random order. The data are shown in Table 6.26. The experimenter believed that there would be no interaction between the two factors.

(a) Check the assumptions on the two-way main-effects model.
(b) Calculate a 99% set of Tukey confidence intervals for pairwise differences between the levels of salt, and calculate separately a 99% set of intervals for pairwise differences between the levels of burner.
(c) Test a hypothesis that there is no linear trend in the time to boil water due to the level of salt. Do a similar test for a quadratic trend.
(d) The experimenter believed that observation number 13 was an outlier, since it has a large standardized residual and it was an observation taken late on a Friday evening. Using statistical software, repeat the analysis in (b) removing this observation. (Tukey's method is approximate for nearly balanced data.) Also repeat the test in part (c) but for the linear contrast only. (The formula for the linear contrast coefficients is given in (4.2.4) on p. 73.) Do you prefer the analysis that uses all the data, or that which removes observation 13? Explain your choice.

22. For $v = 5$ and $r = 4$, show that the first three "orthogonal polynomial contrasts" listed in Table A.2 are mutually orthogonal. (In fact all four are.) Find a pair of orthogonal contrasts that are not orthogonal polynomial contrasts. Can you find a third contrast that is orthogonal to each of these? How about a fourth? (This gets progressively harder!)

23. **Air velocity experiment, continued**

(a) For the air velocity experiment introduced in Sect. 6.7.4 (p. 176), calculate the sum of squares for each of the three interaction contrasts assumed to be negligible, and verify that these add to the value $ssE = 175.739$, as in Table 6.12.
(b) Check the assumptions on the model by plotting the standardized residuals against the predicted responses, the treatment factor levels, and the normal scores. State your conclusions.