Chapter 12
Two-Way Analysis of Variance

In Chapter 6 we consider situations where a response variable is measured on groups of observations classified by a single factor and look at ways to compare the changes in the mean of the response variable attributable to the various levels of this factor. Here we extend this to situations where there are two factors. In Chapters 13 and 14 we will discuss instances where there are more than two factors.

12.1 Example—Display Panel Data

12.1.1 Study Objectives

An air traffic controller must be able to respond quickly to an emergency condition indicated on her display panel. It was desired to compare three types of display panel. Each panel was tested under four simulated emergency situations. Two well-trained controllers were assigned to each of the 12 combinations of emergency condition and display panel type; 24 controllers in all. The data in `data(display)` are from Bowerman and O'Connell (1990). It is clear that the type of display panel is a fixed factor, but unclear from this reference whether emergency situation is a fixed or random factor (review these concepts in Sections 6.2 and 6.4). That is, do these four situations represent the totality of incidents to which air traffic controllers might be exposed, or are they four of far more situations? In the former case, emergency situation is a fixed factor; in the latter case, emergency situation is a random factor.

© Springer Science+Business Media New York 2015

R.M. Heiberger, B. Holland, *Statistical Analysis and Data Display*,
Springer Texts in Statistics, DOI 10.1007/978-1-4939-2122-5_12

12.1.2 Data Description

The data in `data(display)` is structured as 24 rows with four variables.

time: the response variable, time in seconds

panel: factor with three levels indicating the panel being tested

emergenc: factor describing four simulated emergencies

panel.ordered: repeat of the panel factor with the levels reordered to match the order of the response means.

12.1.3 Analysis Goals

We seek to determine whether the three panels afford significantly different display times and whether such conclusions are consistent across different types of emergency.

Exhibited here are graphs and tables that will aid in answering these questions. Discussion of this output is deferred until Section 12.11.

Figure 12.1 shows plots for assessing interaction between `panel` and `emergenc` as well as boxplots for examining the main effects of these factors. The concept of *interaction* is introduced in Section 12.2. The structure of the interaction plot in Figure 12.1 is discussed in Section 12.4.

Table 12.1 shows the `aov` and `anova` statements assuming that `emergenc` is a fixed factor. Table 12.2 and Figures 12.2 and 12.3 show the `panel` means and the results of the multiple comparisons by the Tukey method. As will be explained in Section 12.11, the conclusion derived from this table is that there is a significant difference in response times for the three panels. Panel 3 affords a significantly longer response time than panels 1 or 2; response times for panels 1 and 2 do not differ significantly.

Table 12.3 shows the `aov` and `summary` statements assuming that `emergenc` is a random factor.

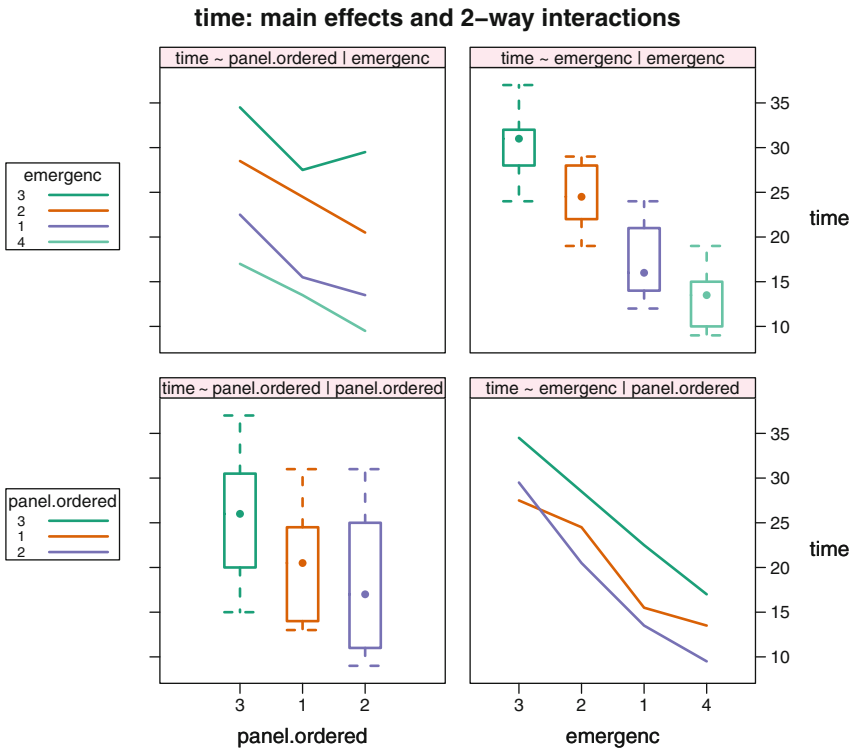


Fig. 12.1 Interaction plot for display panel experiment. The nearly parallel traces suggest the absence of interaction between `panel` and `emergenc`. Note that we reordered the emergencies and the panels by average time in order to simplify the appearance of the plot. The structure of the interaction plot is discussed in Section 12.4.

Table 12.1 Display panel data: ANOVA table with test of `panel` appropriate if `emergenc` is fixed. The test of `panel` is from the "both factors fixed" column of Table 12.8. That is, all sums of squares are compared to the `Residuals` line of the ANOVA table. The listing is continued in Table 12.2.

```
> displayf.aov <- aov(time ~ emergenc * panel, data=display)

> anova(displayf.aov)
Analysis of Variance Table

Response: time
               Df  Sum Sq Mean Sq F value    Pr(>F)
emergenc        3 1052.46  350.82 60.5731 1.612e-07 ***
panel           2  232.75  116.38 20.0935 0.0001478 ***
emergenc:panel  6   28.92    4.82  0.8321 0.5675015
Residuals      12   69.50    5.79
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 12.2 Display panel data: ANOVA table with test of `panel` appropriate if `emergenc` is fixed. Multiple comparisons of `panel` by Tukey method. The standard deviation for the comparison is based on the `Residuals` line of the ANOVA table in Table 12.1. We show plots of the multiple comparisons in Figures 12.3 and 12.2.

```
> displayf.mmc <- mmc(displayf.aov, focus="panel")

> displayf.mmc
Tukey contrasts
Fit: aov(formula = time ~ emergenc * panel, data = display)
Estimated Quantile = 2.668615
95% family-wise confidence level
$mca
    estimate   stderr     lower     upper  height
3-1    5.375 1.203294  2.163871  8.586129 22.9375
3-2    7.375 1.203294  4.163871 10.586129 21.9375
1-2    2.000 1.203294 -1.211129  5.211129 19.2500
$none
  estimate    stderr    lower    upper height
3   25.625 0.8508574 23.35439 27.89561 25.625
1   20.250 0.8508574 17.97939 22.52061 20.250
2   18.250 0.8508574 15.97939 20.52061 18.250
```

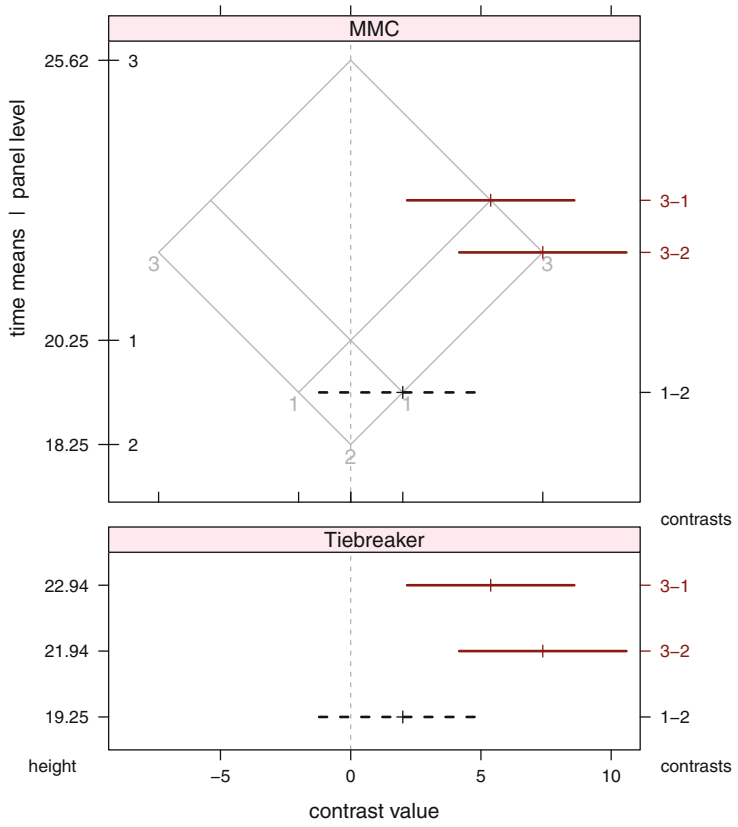


Fig. 12.2 MMC plot of pairwise comparisons of `panel` means by the Tukey method. The top panel shows the `panel` means along the *y*-axis and the confidence intervals for the differences along the *x* axis. The Tiebreaker plot in the bottom panel shows the contrasts equally spaced along the *y*-axis and in the same sequence as the top panel. The heights displayed as the *y*-axis tick labels in the Tiebreaker panel are the actual heights along the *y*-axis for the contrasts in the MMC panel. These heights are the weighted averages of the means being compared by the contrasts. The Tiebreaker panel is not needed in this example.

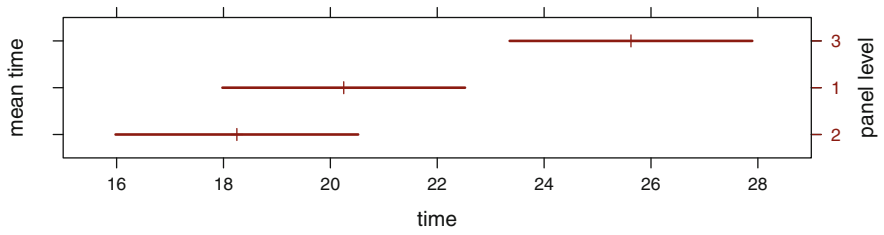


Fig. 12.3 Confidence intervals on each of the `panel` means.

Table 12.3 Display panel data: ANOVA table with test of `panel` appropriate if `emergenc` is random. In this example, the test is from the "A fixed, B random" column of Table 12.8 with `panel` taking the role of A. That is, the sum of squares for `panel` is compared to the `panel:emergenc` interaction line of the ANOVA table.

```
> displayr.aov <- aov(time ~ Error(emergenc/panel) + panel,
+                     data=display)

> summary(displayr.aov)

Error: emergenc
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  3   1052   350.8

Error: emergenc:panel
          Df Sum Sq Mean Sq F value  Pr(>F)
panel      2 232.75  116.38   24.15 0.00135 **
Residuals  6  28.92    4.82
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 12   69.5   5.792
```

12.2 Statistical Model

To model an experiment with two factors, we begin by calling the factors A and B, where A has a levels and B has b levels. We use n_{ij} to denote the number of observations taken from cell (i, j), i.e., the treatment combination corresponding to level i of A and level j of B, $i = 1, \ldots, a$ and $j = 1, \ldots, b$. Our discussion in this chapter is confined to the case where the n_{ij} are equal for all i, j, and sometimes $n_{ij} = 1$. We extend the notation of Equation (6.1) by replacing the singly indexed symbol α_i with a doubly indexed set of symbols $\alpha_i + \beta_j + (\alpha\beta)_{ij}$ and model the k^{th} observation at the i^{th} level of A, j^{th} level of B, as

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} = \mu_{ij} + \epsilon_{ijk} \tag{12.1}$$

for $1 \le i \le a$, $1 \le j \le b$, and $1 \le k \le n_{ij}$. The expectations for the cell means are denoted

$$E(Y_{ijk}) = \mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} \tag{12.2}$$

We assume the errors $\epsilon_{ijk} \sim \text{NID}(0, \sigma^2)$, that is they are assumed to be normally independently distributed with a common variance σ^2. The parameter μ represents the grand mean of all ab populations.

Each of the factors A and B can be either fixed or random. If A is fixed, then we assume that $\sum_i \alpha_i = 0$. If A is random, we assume that each $\alpha_i \sim N(0, \sigma_A^2)$. Similarly, if B is fixed, then we assume that $\sum_j \beta_j = 0$ and if B is random, we assume that each $\beta_j \sim N(0, \sigma_B^2)$.

The term $(\alpha\beta)_{ij}$ models the possibility of interaction between the two factors. If A and B are both fixed factors, then the sum of $(\alpha\beta)_{ij}$ over either i or j is zero. If both factors are random, then $(\alpha\beta)_{ij} \sim N(0, \sigma_{AB}^2)$. In the case of a mixed model, where for concreteness we have A fixed and B random, $(\alpha\beta)_{ij} \sim N(0, \frac{a-1}{a}\sigma_{AB}^2)$ subject to $\sum_i (\alpha\beta)_{ij} = 0$ for each $j = 1, \ldots, b$.

Factors A and B are said to *interact* if the difference in response between two levels of A differs according to the level of B. Equivalently, there is *interaction* between factors A and B if the difference in response between two levels of B differs according to the level of A. Graphically, the traces for each level of factor A across levels of B are parallel if there is no interaction, and are not parallel when there is interaction. Equivalently, the traces for each level of B across levels of A are parallel if there is no interaction. In Figure 12.1 we see essentially parallel traces, consistent with the non-significance of the test of the interaction in Table 12.1. In Figure 12.12 we will see nonparallel, actually crossing, traces consistent with the significance of the interaction in Table 12.12.

12.3 Main Effects and Interactions

As in one-way ANOVA, we are interested in comparing the means of observations in each *cell*, that is for each *treatment combination* (combination of factor levels), in the design, and for combinations of cells. We work with the *cell means*

$$\bar{Y}_{ij} = \sum_k Y_{ijk} \Big/ n_{ij} \tag{12.3}$$

and the *marginal means*. The marginal means for the rows are calculated by averaging the cell means in each row over the columns. The marginal means for the columns are calculated by averaging the cell means in each column over the rows:

$$\bar{Y}_{i.} = \sum_j \bar{Y}_{ij} \Big/ b \tag{12.4}$$

$$\bar{Y}_{.j} = \sum_i \bar{Y}_{ij} \Big/ a \tag{12.5}$$

Table 12.4 Table of means for the rhizobium clover experiment of Section 12.14. Means from Table 12.12 have been arranged in a two-way table to display the cell means in the body of the table, the marginal means on the margins of the table, and the grand mean as the margin of the marginal means. `clover` and `clover+alfalfa` are the two levels of the factor `comb`. The left side of the table shows the means symbolically using $\bar{Y}_{ij}$ notation. The right side show the numerical values from Table 12.12.

Strain	Clover	Clover+alfalfa	Mean		Strain	Clover	Clover+alfalfa	Mean
3DOk1	$\bar{Y}_{11}$	$\bar{Y}_{12}$	$\bar{Y}_{1.}$		3DOk1	29.04	28.41	28.72
3DOk5	$\bar{Y}_{21}$	$\bar{Y}_{22}$	$\bar{Y}_{2.}$		3DOk5	36.29	27.44	31.86
3DOk4	$\bar{Y}_{31}$	$\bar{Y}_{32}$	$\bar{Y}_{3.}$		3DOk4	21.35	23.98	22.66
3DOk7	$\bar{Y}_{41}$	$\bar{Y}_{42}$	$\bar{Y}_{4.}$	=	3DOk7	22.93	24.96	23.95
3DOk13	$\bar{Y}_{51}$	$\bar{Y}_{52}$	$\bar{Y}_{5.}$		3DOk13	22.49	24.30	23.39
k.composite	$\bar{Y}_{61}$	$\bar{Y}_{62}$	$\bar{Y}_{6.}$		k.composite	25.97	24.92	25.45
Mean	$\bar{Y}_{.1}$	$\bar{Y}_{.2}$	$\bar{Y}_{..}$		Mean	26.35	25.67	26.01

where $n_{i.} = \sum_j n_{ij}$, $n_{.j} = \sum_i n_{ij}$, and $n_{..} = \sum_{ij} n_{ij}$. Marginal means get their name because they are often displayed on the margins of a two-way table of cell means, as in Table 12.4. We also use the *grand mean*:

$$\bar{Y}_{..} = \sum_i n_{i.}\bar{Y}_{i.} \Big/ n_{..} = \sum_j n_{.j}\bar{Y}_{.j} \Big/ n_{..} = \sum_{ijk} Y_{ijk} \Big/ n_{..} \tag{12.6}$$

When more than one factor is present, there are three principal types of comparisons that we will investigate.

Main effects are comparisons of the marginal means for one of the factors, for example, $\bar{Y}_{1.} - \bar{Y}_{2.}$. It is usually valid to compare main effects only when there is no interaction.

Interactions (or interaction effects) are comparisons of the cell means across levels of both factors, for example, $(\bar{Y}_{13} - \bar{Y}_{23}) - (\bar{Y}_{14} - \bar{Y}_{24})$. When interaction is present, that is when differences in the cell means across rows depend on the column or equivalently, when comparisons of the form indicated here are significantly different from 0, we usually must use simple effects, not main effects, to discuss the factors.

Simple effects are separate comparisons of the cell means across levels of one factor for some or all levels of the other factor, for example, $\bar{Y}_{13} - \bar{Y}_{23}$. See Section 13.3.

The analyst should be alert to the possibility that interaction is present. The nature of the analysis when interaction exists is different from that when interaction is absent.

Without interaction, the analysis proceeds similarly to the procedures for one-way analysis. The marginal means are calculated and compared, perhaps by using one of the multiple comparisons techniques discussed in Sections 6.3, 7.1.3, or 7.1.4.1. The advantage of the two-way analysis in this case is in the efficiency, hence increased power, of the comparisons. Because we use the same residual sum of squares for the denominator of both F-tests (for the rows and for the columns), we can run the combined experiment to test the effect of both factors for less expense than if we were to run two separate experiments.

When interaction between two factors is present, it is not appropriate to compare the main effects, the levels of one of these factors averaged over the levels of the other factor. It is possible, for example, that the mean of Y increases over factor B for level 1 of factor A and decreases over factor B for level 2 of factor A. Averaging over the levels of factor A would mask that behavior of the response.

We explore main effects, interactions, and simple effects with the rhizobium data in Section 12.14.

12.4 Two-Way Interaction Plot

The two-way interaction plot, first shown in Figure 12.1 and used throughout the remainder of this book, shows all main effects and two-way interactions for designs with two or more factors. We construct it using the `interaction2wt` function in the **HH** package by analogy with the `splom` (scatterplot matrix) function in the **lattice** package. The rows and columns of the two-way interaction plot are defined by the Cartesian product of the factors.

1. Each main diagonal panel shows a boxplot for the marginal effect of a factor.
2. Each off-diagonal panel is a standard interaction plot of the factors defining its position in the array. Each point in the panel is the mean of the response variable conditional on the values of the two factors. Each line in the panel connects the cell means for a constant level of the *trace* factor. Each vertically aligned set of points in the panel shows the cell means for a constant value of the x-factor.
3. Panels in mirror-image positions interchange the trace- and x-factors. This duplication is helpful rather than redundant because one of the orientations is frequently much easier to interpret than the other.
4. The rows are labeled with a key that shows the line type and color for the trace factor by which the row is defined.
5. Each box in the boxplot panels has the same color, and optionally the same line type, as the corresponding traces in its row.
6. The columns are labeled by the x-factor.

12.5 Sums of Squares in the Two-Way ANOVA Table

Table 12.5 presents the structure of the analysis of variance table for a balanced two-way ANOVA with a levels of the A factor, b levels of the B factor, and n observations at each of the ab AB-treatment combinations, analogous to Table 6.2 for one-way ANOVA. If the test F_{AB} shows that AB interaction is present, the F-tests on A and B are not interpretable.

If the AB interaction is not significant, then the form of the tests for the main effects A and B depends on whether the factors A and B are fixed or random factors. See the discussion in Section 12.10 where Table 12.8 lists the expected mean squares and F-tests under various assumptions.

Table 12.5 Two-way ANOVA structure with both factors representing fixed effects.

Analysis of Variance of Dependent Variable y					
Source	Degrees of Freedom	Sum of Squares	Mean Square	F	p-value
Treatment A	df_A	SS_A	MS_A	F_A	p_A
Treatment B	df_B	SS_B	MS_B	F_B	p_B
AB Interaction	df_{AB}	SS_{AB}	MS_{AB}	F_{AB}	p_{AB}
Residual	df_{Res}	SS_{Res}	MS_{Res}		
Total	df_{Total}	SS_{Total}			

Terms of the table are defined by:

Treatment A	
df_A	$a-1$
SS_A	$bn\sum(\bar{Y}_{i.}-\bar{Y}_{..})^2$
MS_A	$\mathsf{SS}_A/\mathsf{df}_A$
F_A	$\mathsf{MS}_A/\mathsf{MS}_{Res}$
p_A	$1-\mathcal{F}_F(F_A \mid \mathsf{df}_A, \mathsf{df}_{Res})$

Treatment AB	
df_{AB}	$(a-1)(b-1)$
SS_{AB}	$n\sum(\bar{Y}_{ij}-\bar{Y}_{..})^2-\mathsf{SS}_A-\mathsf{SS}_B$
MS_{AB}	$\mathsf{SS}_{AB}/\mathsf{df}_{AB}$
F_{AB}	$\mathsf{MS}_{AB}/\mathsf{MS}_{Res}$
p_{AB}	$1-\mathcal{F}_F(F_{AB} \mid \mathsf{df}_{AB}, \mathsf{df}_{Res})$

Treatment B	
df_B	$b-1$
SS_B	$an\sum(\bar{Y}_{.j}-\bar{Y}_{..})^2$
MS_B	$\mathsf{SS}_B/\mathsf{df}_B$
F_B	$\mathsf{MS}_B/\mathsf{MS}_{Res}$
p_B	$1-\mathcal{F}_F(F_B \mid \mathsf{df}_B, \mathsf{df}_{Res})$

Residual	
df_{Res}	$ab(n-1)$
SS_{Res}	$\sum_i\sum_j(Y_{ijk}-\bar{Y}_{ij})^2$
MS_{Res}	$\mathsf{SS}_{Res}/\mathsf{df}_{Res}$

Total	
df_{Total}	$abn-1$
SS_{Total}	$\sum_i\sum_j\sum_k(Y_{ijk}-\bar{Y}_{..})^2$

Table 12.5 shows the F-statistics and their p-values for tests on the main effects A and B under the assumption that both factors represent fixed effects. Most ANOVA programs calculate these values by default whether or not they are appropriate.

12.6 Treatment and Blocking Factors

Treatment factors are those for which we wish to determine if there is an effect. *Blocking* factors are those for which we believe there is an effect. We wish to prevent a presumed blocking effect from interfering with our measurement of the treatment effect.

An experiment with two factors may have either two treatment factors or one treatment factor and one blocking factor. The primary objective of a factorial experiment is comparisons of the levels of treatment factors. By contrast, a blocking factor is set up in order to enhance one's ability to distinguish between the levels of treatment factors. The term *block* was chosen by analogy to two of the dictionary definitions: a rectangular section of land bounded on each side by consecutive streets; or a set of similar items sold or handled as a unit, such as shares of stock.

We are not interested in comparing the *blocks*, i.e., the levels of a blocking factor. In a well-designed experiment, we anticipate that the response differs across the levels of a blocking factor because if the levels of this factor cover a variety of experimental conditions, this broadens the scope of our inferences about treatment differences. Multiple comparisons across blocks are not meaningful because we know in advance that the blocks are different. In general, blocking is advisable and successful as an experimental and analytical technique if the experimental units can reasonably be grouped into blocks such that the units within every block are homogeneous, while the units in any given block are different from those in any other block. By homogeneous units, we mean that they will tend to respond alike if treated alike. Usually, there is no interaction between blocking and treatment factors; otherwise blocking will not have accomplished its objective and the analysis will be much less able to detect significant differences than if blocking were properly done.

Blocking is the natural extension to three or more treatments of the matched pairs design introduced in Section 5.5. The F-test of the treatment effect against the residual is the generalization of the paired t-test. It is exactly true that a blocked design with two levels of the treatment factor and with many blocks of size two is identical to the matched pairs design.

For example, in an experiment on tire wear, the location of the tire on the car (say, Right Front) is a treatment effect and the specific car (of the many used in the experiment) is a blocking effect.

12.7 Fixed and Random Effects

As mentioned in Sections 6.2 and 6.4, treatment factors may be regarded as either fixed or random. The levels of a fixed factor are the only levels of interest in the experiment, and we wish to see if the response is homogeneous across these levels. The levels of a random factor are a random sample from some large population of levels, and we are interested in assessing whether the variance of responses over this population of levels is essentially zero. Block factors are almost always regarded as random.

The levels of a treatment factor can be either categorical or quantitative. For example, in an experiment where the `fertilizer` treatment has four levels, the experimental levels of fertilizer could be four different fertilizer compounds, or four different applications per acre of one fertilizer compound. When the levels are quantitative, it is usually preferable to regard the factor as a single degree-of-freedom predictor variable.

12.8 Randomized Complete Block Designs

A randomized complete block design (RCBD) has one treatment factor involving t treatment levels and one blocking factor having b levels. The b blocks each contain experimental units arranged according to the principles discussed in Section 12.6. That is, experimental units in the same block are expected to respond alike if treated alike, while the blocks should reflect a variety of experimental conditions to broaden the scope of conclusions to be drawn from inferences about the treatments. It is assumed that blocks and treatments do not interact. This assumption permits us to compare the treatment levels when each block contains exactly t experimental units, i.e., there is no replication of treatments within any block. If there are $n > 1$ observations on each treatment within each block, then additional degrees of freedom are available for comparing treatments. We outline the effect of larger sample size, which usually means more degrees of freedom in the denominator of statistical tests, in Section 3.10. In summary, more degrees of freedom move us up the t-table or F-table or χ^2-table and the critical value gets smaller.

The model for the RCBD with one observation on each treatment in each block is

$$y_{ij} = \mu + \tau_i + \rho_j + \epsilon_{ij} \tag{12.7}$$

where μ represents the overall mean, τ_i is the differential effect of treatment level i, ρ_j is the differential effect of block j, and the ϵ's are random $N(0, \sigma^2)$ residuals. We further define

$$\bar{y}_i = \sum_i y_{ij}/b, \ \bar{y}_j = \sum_j y_{ij}/t, \text{ and } \bar{\bar{y}}_{..} = \sum_i \sum_j y_{ij}/bt \tag{12.8}$$

Table 12.6 ANOVA table structure for a randomized complete block design with no replication.

Analysis of Variance of Dependent Variable y					
Source	Degrees of Freedom	Sum of Squares	Mean Square	F	p-value
Blocks	df_{Blk}	SS_{Blk}	MS_{Blk}		
Treatments	df_{Tr}	SS_{Tr}	MS_{Tr}	F_{Tr}	p_{Tr}
Residuals	df_{Res}	SS_{Res}	MS_{Res}		
Total	$\mathsf{df}_{\text{Total}}$	$\mathsf{SS}_{\text{Total}}$			

The terms of the table are defined by:

Blocks	
df_{Blk}	$b-1$
SS_{Blk}	$\sum_i \sum_j (\bar{y}_j - \bar{\bar{y}}_{\cdot\cdot})^2$
MS_{Blk}	$\mathsf{SS}_{\text{Blk}}/\mathsf{df}_{\text{Blk}}$

Residual	
df_{Res}	$(b-1)(t-1)$
SS_{Res}	$\sum_i \sum_j (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{\bar{y}}_{\cdot\cdot})^2$
MS_{Res}	$\mathsf{SS}_{\text{Res}}/\mathsf{df}_{\text{Res}}$

Treatments	
df_{Tr}	$t-1$
SS_{Tr}	$\sum_i \sum_j (\bar{y}_i - \bar{\bar{y}}_{\cdot\cdot})^2$
MS_{Tr}	$\mathsf{SS}_{\text{Tr}}/\mathsf{df}_{\text{Tr}}$
F_{Tr}	$\mathsf{MS}_{\text{Tr}}/\mathsf{MS}_{\text{Res}}$
p_{Tr}	$1 - \mathcal{F}_F(F_{\text{Tr}} \mid \mathsf{df}_{\text{Tr}}, \mathsf{df}_{\text{Res}})$

Total	
$\mathsf{df}_{\text{Total}}$	$bt-1$
$\mathsf{SS}_{\text{Total}}$	$\sum_i \sum_j (y_{ij} - \bar{\bar{y}}_{\cdot\cdot})^2$

The setup of the ANOVA table for an RCBD with $n = 1$ is shown in Table 12.6. Some ANOVA programs also display an F-statistic and p-value for blocks, but it is inappropriate to interpret these since the experiment is designed in such a way that responses will differ across blocks and artificially force high F values for blocks. We could do efficiency of blocking calculations. See, for example, Cochran and Cox (1957) (Section 4.37).

12.9 Example—The Blood Plasma Data

12.9.1 Study Objectives

The dataset `data(plasma)` comes from Anderson et al. (1981) and is reproduced in Hand et al. (1994). The data are measurements on plasma citrate concentrations in micromols/liter obtained from 10 subjects at 8 am, 11 am, 2 pm, 5 pm, and 8 pm.

To what extent is there a normal profile for the level in the human body during the day?

This experiment is viewed as an RCBD with treatment factor `time` and blocking factor `id`. It is desirable here that the subjects (blocks) be as unlike as possible in order to broaden the scope of the conclusion about normal profiles as much as possible. The no-interaction assumption amounts to assuming that the daily response profile is constant across subjects.

12.9.2 Data Description

The data in `data(plasma)` is structured as 50 rows with three variables.

plasma: the response variable, plasma citrate concentrations in micromols/liter

time: factor with five values: 8 am, 11 am, 2 pm, 5 pm, and 8 pm

id: factor with 10 levels, one per subject

12.9.3 Analysis

We begin our analysis with the interaction plots in Figure 12.4. There seem to be anomalies for `id=3` at 8 pm and for `id=6` at 11 am, but otherwise both sets of traces look reasonably parallel.

We proceed with an additive model in Table 12.7 and discover that the ratio of the `id` stratum Residual Mean Square to the Within stratum Residual Mean Square (1177/147.5 = 7.98) is large (had this been a valid test, which it is not because `id` is a blocking factor, it would have been $F = 7.98$), confirming our decision to block on patients. This is not a hypothesis test, because we know at the beginning of our analysis that patients are different from each other.

The test of differences due to `time` rejects the null hypothesis that the response at all times is the same. Since there appears to be no interaction, we can act as if there is a single pattern that applies to everyone. We investigate the `time` pattern with the MMC plot in Figure 12.5. The only significant single contrast is between the low at 5PM and the high at 11AM. The low at 5PM is clearly visible in the `plasma ~ time | id` panel of Figure 12.4. The high at 11AM is hinted at in Figure 12.4.

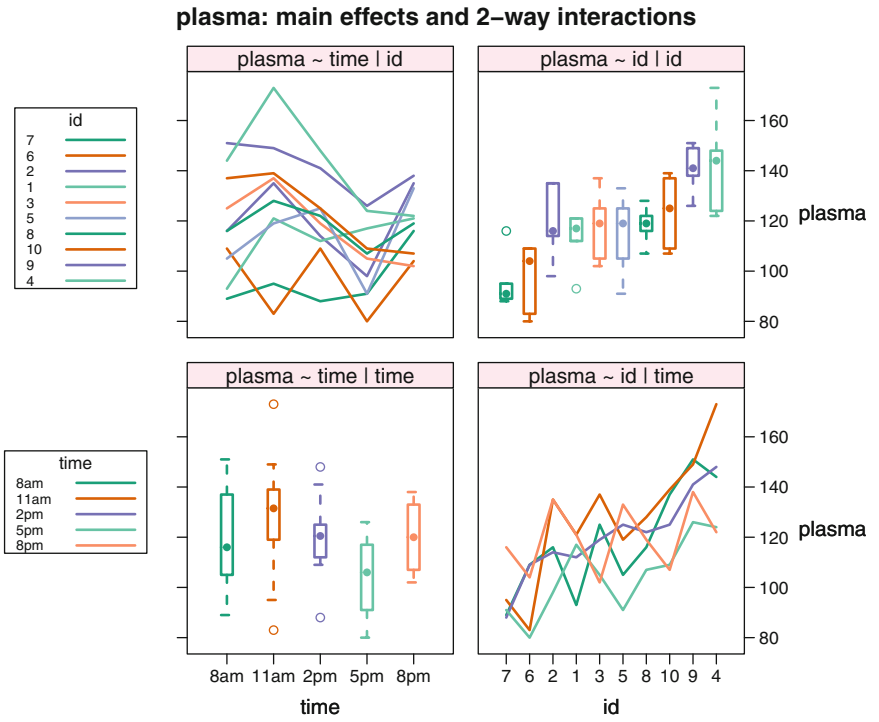


Fig. 12.4 Interaction Plot for Plasma Citrate. The `id` factor has been sorted by median `plasma` value. The `time` factor must be displayed in chronological order.

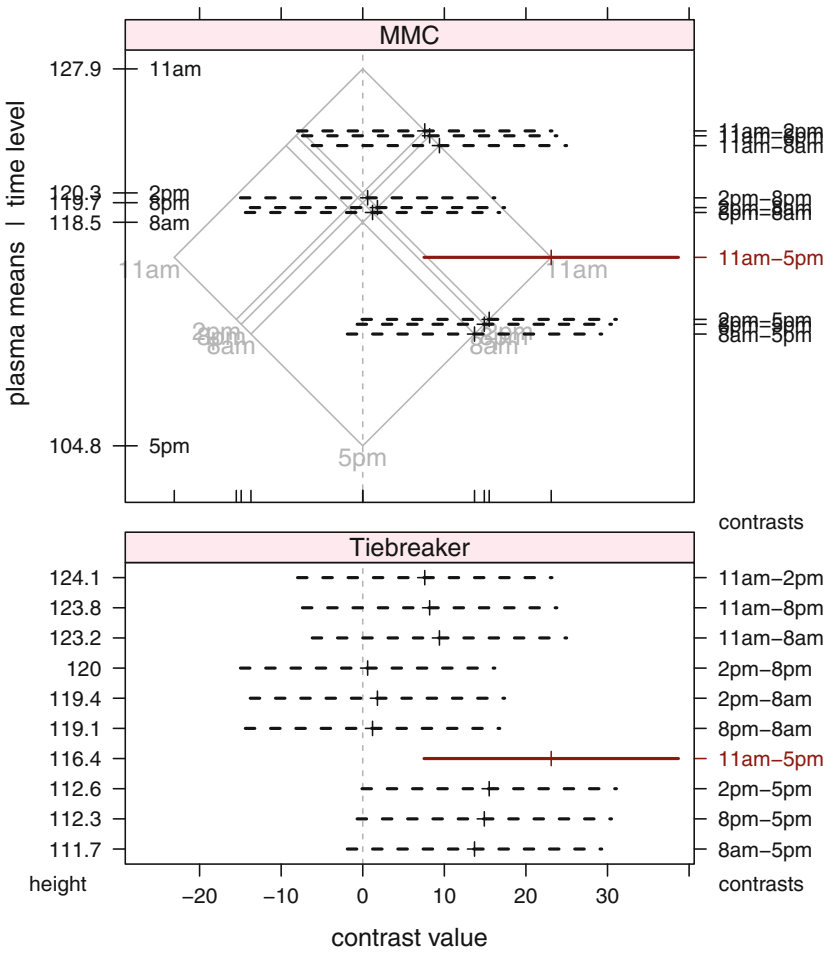
Table 12.7 ANOVA Table for Plasma Citrate Experiment

```
> plasma.aov <- aov(plasma ~ Error(id) + time, data=plasma)

> summary(plasma.aov)

Error: id
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  9  10593    1177

Error: Within
          Df Sum Sq Mean Sq F value  Pr(>F)
time       4   2804   701.0   4.754 0.00349 **
Residuals 36   5308   147.5
---
Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```



The MMC panel shows informative overprinting. Please see Tiebreaker panel and caption.

Fig. 12.5 MMC plot and Tiebreaker plot of `time` in the plasma data. The low at 5PM is clearly visible in the `plasma ~ time | id` panel of Figure 12.4. The high at 11AM is hinted at in Figure 12.4. The Tiebreaker plot in the bottom panel is imperative for this example. The means at many of the levels of `time` are very close and therefore their labels are overprinted. As a consequence, the heights of the contrasts are similar and their labels are also overprinted. The Tiebreaker plot shows the contrasts equally spaced along the *y*-axis and in the same sequence as the top panel. The heights displayed as the *y*-axis tick labels in the Tiebreaker panel are the actual heights along the *y*-axis for the contrasts in the MMC panel.

12.10 Random Effects Models and Mixed Models

In Section 6.4, we compare two analyses of the same data assuming the single factor is fixed or random. There we indicate that a table of expected mean squares may be used to formulate the correct mean square ratio to test the hypothesis of interest. We also show that in the single factor case, while the same ratio is used in both the fixed and random cases, the hypothesis tested about the factor differs in the two cases.

When we have two or more factors and interactions, the test statistics as well as the hypotheses depend on whether the factors are fixed or random. The formulas for standard errors for comparing the levels of fixed factors also depend on whether the other factor(s) are fixed or random.

Table 12.8 is an algebraically derived table of expected mean squares for an experiment with two possibly interacting factors A and B and equal sample sizes $n_{ij} = n \geq 2$ at each of the ab treatment combinations under each of three assumptions: the fixed model where both factors are fixed, the mixed model where one factor is fixed and the other factor is random, and the random model where both factors are random. Each entry in the table is derived by evaluating, for example (using the notation of Table 12.5), the statement

$$E(\mathrm{MS_A}) + E\left(bn \sum_{\cdot}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right)/(a-1)$$

where we model Y_{ijk} and $E(Y_{ijk})$ by Equations 12.1 and 12.2.

From the lineups of the expected mean squares, we see that for testing the A main effect, the appropriate denominator mean square is the Residual mean square when factor B is fixed (from the "Both factors fixed" column, EMS(A) = $\sigma^2 + nb\kappa_A^2$ and EMS(Residual) = σ^2).

Table 12.8 Expected mean squares in two-way analysis of variance. Compare to Tables 6.4, 12.5, and 13.11. See Section 12.10 for the discussion on when to use each of the columns.

Source	df	Both factors fixed	A fixed, B random	Both factors random
Treatment A	$a-1$	$\sigma^2 + nb\kappa_A^2$	$\sigma^2 + n\sigma_{AB}^2 + nb\kappa_A^2$	$\sigma^2 + n\sigma_{AB}^2 + nb\sigma_A^2$
Treatment B	$b-1$	$\sigma^2 + na\kappa_B^2$	$\sigma^2 + na\sigma_B^2$	$\sigma^2 + n\sigma_{AB}^2 + na\sigma_B^2$
AB Interaction	$(a-1)(b-1)$	$\sigma^2 + n\kappa_{AB}^2$	$\sigma^2 + n\sigma_{AB}^2$	$\sigma^2 + n\sigma_{AB}^2$
Residual	$ab(n-1)$	σ^2	σ^2	σ^2
Total	$abn-1$			

where $\kappa_A^2 = \dfrac{\sum_i \alpha_i^2}{a-1}$ $\kappa_B^2 = \dfrac{\sum_j \beta_j^2}{b-1}$ $\kappa_{AB}^2 = \dfrac{\sum_i \sum_j (\alpha\beta)_{ij}^2}{(a-1)(b-1)}$

The appropriate denominator mean square for testing the A main effect is the AB-interaction mean square when B is random (from the other two columns $\text{EMS(A)} = \sigma^2 + n\sigma^2_{AB} + nb\, f(A)$ and $\text{EMS(AB)} = \sigma^2 + n\sigma^2_{AB}$, where $f(A) = \kappa^2_A$ when A is fixed and $f(A) = \sigma^2_A$ when A is random). The ratio of these mean squares is appropriate for testing equality of the levels of factor A because the corresponding ratio of these expected mean squares exceeds one if and only if $\sigma^2_A > 0$ or $\kappa^2_A > 0$. Use of the Residual mean square as the denominator of the F-test would be inappropriate because such a ratio would exceed one if there is an AB interaction effect.

The conclusions for testing the B main effect follow from interchanging "A" and "B" in the previous sentence.

12.11 Example—Display Panel Data—Continued

In Section 12.1 we introduced the display panel example illustrating a two-way analysis of variance. We continue here with the analysis by discussing Figures 12.1–12.3 and Tables 12.1–12.3.

In Figure 12.1 we display two-way interaction plots and boxplots for the factors `panel` and `emergenc`. The two interaction plots in the off-diagonal panels contain equivalent information, but in general, one of them is more readily interpretable than the other. In this instance, the close-to-parallel traces suggest the absence of interaction between `panel` and `emergenc`. This is anticipated because `emergenc` is a block factor and confirmed by the large p-value for the interaction test in Table 12.1. One set of boxplots in Figure 12.1 evinces a greater response time with panel 3 than with either panel 1 or panel 2. The other set of boxplots shows substantial differences in the response times of the four emergencies; this is anticipated since `emergenc` is regarded as a blocking factor and differences in response across blocks are expected by design.

The simplest ANOVA specification in Table 12.1 assumes all factors are fixed. We see that when `emergenc` is a fixed factor, the F-statistic for `panel` is 20.09 on 2 and 12 degrees of freedom. The small corresponding p-value suggests that response `time` varies with the type of `panel`.

If `emergenc` is a random factor, as in Table 12.3, the pattern of expected mean squares in Table 12.8 indicates that the appropriate denominator mean square for testing `panel` is the interaction mean square. This test is specified by placing `emergenc/panel` inside the `Error()` function in the model formula. We see that `panel` is tested with $F = 24.15$ on 2 and 6 degrees of freedom.

The F-statistic for `panel` corresponds to a small p-value under either assumption on `emergenc`. Therefore, in this example, we reach the same conclusion under both assumptions: that response time differs across `panels`. However, since in general the F-statistic differs in the two cases, the ultimate conclusion concerning a fixed factor may depend crucially on our assumption concerning the other factor. If `emergenc` is a fixed factor, the conclusions regarding panels applies to these four emergencies

only. If emergenc is a random factor, the panel conclusions apply to the entire population of emergencies from which these four emergencies are assumed to be a random sample.

The F-test for interaction between panel and emergenc when emergenc is a random factor is the same test as when emergenc is a fixed factor.

Since panel is a fixed factor, an appropriate follow-up is a Tukey test to compare the response time for each display panel. This is shown in Table 12.2 for the case where emergenc is fixed. The means are in the estimate column of the $none section. We find that both display panel 1 and display panel 2 have significantly shorter response times than display panel 3, but panels 1 and 2 are not significantly different. Therefore, we conclude that display panel 3 can safely be eliminated from further consideration. The absence of interaction tells us that these conclusions are consistent over emergencies. If interaction had existed in this experiment, one would have concluded that the optimal panel differs according to the type of emergency. Then one would need to make separate panel recommendations for each emergency type. Since we will normally select just one panel type for the entire facility, and since we have no control over emergencies, the decision process would become more difficult.

The confidence intervals in the $mca section of Table 12.2 and in both panels Figure 12.2 display the differences between all pairs of panel means using the two-sided Tukey multiple comparisons procedure introduced in Section 6.3. The $mca (the term *mca* stands for *multiple comparisons analysis*) section of Table 12.2 shows the results of the $\binom{3}{2} = 3$ pairwise tests. The negative lower bound and positive upper bound for the 1-2 comparison indicates that the confidence interval for the difference between the corresponding population means ($\bar{y}_1 = 20.250$ and $\bar{y}_2 = 18.250$) includes zero, hence the difference is not significant. The comparisons between $\bar{y}_3 = 25.625$ and the other two panel means have positive lower and upper bounds, hence these confidence intervals exclude zero. This indicates that the population mean of panel 3 is significantly different from both other population means.

Figure 12.2 provides two confidence interval displays for pairwise comparisons of the population means of the three panels. Both contain the confidence intervals on each pairwise difference taken directly from the $mca section of Table 12.2. A pairwise difference of means is significantly different from zero; equivalently, the two means differ significantly if the confidence interval for the pairwise difference excludes zero. If this confidence interval includes zero, then conclude that the two population means do not significantly differ. Thus the "1–2" interval says that these two panel means are indistinguishable. The "1–3" and "2–3" intervals says that the mean of panel 3 differs from the means of the other two panels. The top panel is an MMC plot (see Chapter 7) with the contrasts displayed on the isomeans grid as a background that shows the individual panel means. The bottom panel uses equal vertical spacing between contrasts.

The `$none` (the term *none* indicating no contrasts) section of Table 12.2 shows the results for the individual group means. Figure 12.3 contains simultaneous confidence intervals for the three population means, where the confidence coefficient, here 95%, is the probability that each interval contains its respective population mean. If two of these confidence intervals overlap, then the corresponding population means are not significantly different. Since panels 1 and 2 have overlapping intervals, these two panel means are not distinguishable. If a pair of these confidence intervals does not overlap, then the corresponding population means are declared to differ significantly. Since the panel 3 interval does not overlap the other two, we conclude that the mean of panel 3 differs from the means of the other two panels.

12.12 Studentized Range Distribution

The tabled values of the Studentized Range Distribution (see Section J.1.10) of a set of a means are scaled for the random variable $Q = (\bar{y}_{(a)} - \bar{y}_{(1)})/s_{\bar{y}}$. The denominator is the standard error of a single $\bar{y}$. The estimated quantile (critical point) shown in Table 12.2 and used in Figure 12.2 is 2.668. This is not the Studentized range tabular value $q_{.05}$ but instead $q_{.05}/\sqrt{2}$. Details of the R calculation can be followed in file `HHscriptnames(12)`. The equivalent SAS code reports the Studentized range $q_{.05}$.

We use the tabled values in two places in the MMC display. In Table 12.2, $q_{.05} = 3.77278$, the *Estimated Quantile* is $\frac{q_{.05}}{\sqrt{2}} = 2.668$, $\mathsf{MS}_{\text{Res}} = 5.791667$ (from Table 12.1), and $m = 8$.

In the `$none` section of Table 12.2 we show the sample means $\bar{y}_i$ in the `estimate` column and the standard error $s_{\bar{y}}$ of an individual $\bar{y}$ in the `stderr` column. We must adjust the Q value by dividing by $\sqrt{2}$. The formula for the simultaneous 95% confidence intervals on individual means is

$$\mu_i: \qquad \bar{Y}_i \pm \frac{q_{.05}}{\sqrt{2}} \sqrt{\frac{\mathsf{MS}_{\text{Res}}}{m}} \tag{12.9}$$

where $q_{.05}$ is the 95$^{\text{th}}$ percentile of the Studentized range distribution and m is the common sample size used in calculating each sample mean. The "minimum significant difference" in this table is the "±" part of formula (12.9),

$$\frac{q_{.05}}{\sqrt{2}} \sqrt{\frac{\mathsf{MS}_{\text{Res}}}{m}} = 2.27 \tag{12.10}$$

In the `$mca` section of Table 12.2 we show the differences $\bar{y}_i - \bar{y}_j$ in the `estimate` column and the standard error $\sqrt{2}s_{\bar{y}}$ of the difference in the `stderr` column. Again we must adjust the Q value by dividing by $\sqrt{2}$. The formula for the simultaneous 95% confidence intervals on pairwise mean differences shown in Figure 12.2 is

$$\mu_i - \mu_j: \quad \bar{Y}_i - \bar{Y}_{i'} \pm \frac{q_{.05}}{\sqrt{2}}\sqrt{2\frac{\mathsf{MS}_{\text{Res}}}{m}} = \bar{Y}_i - \bar{Y}_{i'} \pm q_{.05}\sqrt{\frac{\mathsf{MS}_{\text{Res}}}{m}} \tag{12.11}$$

The "minimum significant difference" in this table is the "±" part of formula (12.11),

$$q_{.05}\sqrt{\frac{\mathsf{MS}_{\text{Res}}}{m}} = 3.21 \tag{12.12}$$

12.13 Introduction to Nesting

In the previous examples the two factors have a *crossed* relationship. Saying that factors A and B are crossed indicates that each level of A may be observed in a treatment combination with any level of B. Alternatively, two factors may have a *nested* or hierarchical relationship. When B is nested within A, the levels of B are similar but not identical for different levels of A.

12.13.1 Example—Workstation Data

A small electronics firm wishes to compare three methods for assembling an electronic device. For this purpose, the plant has available six different workstations. The study is conducted by randomly assigning $s = 2$ workstations to each of the $m = 3$ assembly methods. At each workstation–method combination $w = 5$ randomly selected production workers will assemble the device for one hour using the appropriate assembly method. The response is the number of devices produced in one hour. The data from Bowerman and O'Connell (1990) (p. 890) are accessible as `data(workstation)` and are displayed in Figure 12.6.

12.13.2 Data Description

The data in `data(plasma)` is structured as 30 rows with three variables.

method: factor with three levels describing the assembly methods

station: factor with two levels describing the workstations

devices: response variable, number of devices produced in one hour.

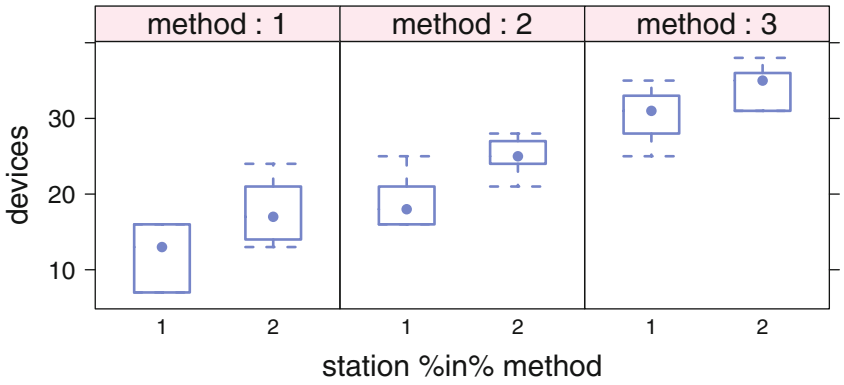


Fig. 12.6 Boxplot of workstation data. The significance of `method` and `station` within `method` are confirmed in Table 12.10.

12.13.3 Analysis Goals

Note that the workstations assigned to any assembly method are different from those assigned to any other method. As a consequence, the factors (which we call `station` and `method`) are not crossed with one another, and an analysis using a model we have previously studied would be incorrect. The factor `station` is said to be *nested* within the factor `method` because each workstation is associated with exactly one of the methods.

Our analysis assumes that `station` is a fixed factor. If instead `station` were assumed to be a random factor, the code would have to be modified to force `station` to be tested against the `station within method` mean square instead of against the `Residuals` mean square. The procedures for doing so are demonstrated in the data analysis in Section 13.4.

The basic structure of the ANOVA table is in Table 12.9. In R, we use the formula

```
devices ~ method / station
```

to indicate that `station` is nested within `method`. The analysis is in Table 12.10.

We conclude that when using at least one of the three methods, the two workstations for that method produced a significantly different number of devices. We also conclude that the three methods produced significantly different numbers of devices.

In this example there is balanced sampling. That is, each method has the same number of workstations and each workstation has the same number of workers. Without much additional difficulty, the above nested factorial analysis can be extended to situations with unbalanced sampling. (In contrast, when one has unbalanced sampling and crossed factors, the analysis is considerably more difficult than with balanced sampling.)

Table 12.9 Basic structure of the ANOVA table for a nested design with $m = 3$, $s = 2$, and $w = 5$.

Source	df			MS	F
	Algebra	Example			
Method	$m-1$	$= 3-1$	$= 2$	MS_m	$\frac{MS_m}{MS_w}$
Station within Method	$m(s-1)$	$= 3\times(2-1)$	$= 3$	MS_s	$\frac{MS_s}{MS_w}$
Worker within Station (Residual)	$ms(w-1)$	$= 3\times2\times(5-1)$	$= 24$	MS_w	
Total	$msw-1$	$= 3\times2\times5-1$	$= 29$		

Table 12.10 Workstation data. ANOVA table and means.

```
> workstation.aov <- aov(devices ~ method / station,
+                        data=workstation)

> summary(workstation.aov)
              Df Sum Sq Mean Sq F value   Pr(>F)
method         2 1545.3   772.6  51.452 2.09e-09 ***
method:station  3  210.2    70.1   4.666   0.0105 *
Residuals      24  360.4    15.0
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> model.tables(workstation.aov, "means", se=TRUE)
Tables of means
Grand mean

23.06667

 method
method
   1    2    3
14.8 22.1 32.3

 method:station
      station
method 1    2
     1 11.8 17.8
     2 19.2 25.0
     3 30.4 34.2

Standard errors for differences of means
        method method:station
         1.733          2.451
replic.     10              5
```

12.14 Example—The *Rhizobium* Data

12.14.1 Study Objectives

Erdman (1946) discusses experiments to determine if antibiosis occurs between *Rhizobium Meliloti* and *Rhizobium Trifolii.* Rhizobium is a bacteria, growing on the roots of clover and alfalfa, that fixes nitrogen from the atmosphere into a chemical form the plants can use. The research question for Erdman was whether there was an interaction between the two types of bacteria, one specialized for alfalfa plants and the other for clover plants. If there was an interaction, it would indicate that clover bacteria mixed with alfalfa bacteria changed the nitrogen fixing response of alfalfa to alfalfa bacteria or of clover to clover bacteria. The biology of the experiment says that interaction indicates antibiosis or antagonism of the two types of rhizobium. That is, the goal was to test whether the two types of rhizobium kill each other off. If they do, then there will be less functioning bacteria in the root nodules and consequently nitrogen fixation will be slower.

Erdman ran two sets of experiments in parallel. In one the response variable was the nitrogen content in clover plants, in the other the nitrogen content in alfalfa plants. The treatments were combinations of bacterial cultures in which the plants were grown. As a historical note, beginning with Steel and Torrie (1960), the one-way analysis of the clover plus alfalfa combination of the Clover experiment has been frequently used as an example to illustrate multiple comparisons procedures. Here we examine the complete data from two related two-way experiments.

12.14.2 Data Description

Both experiments are two-way factorial experiments with two treatment factors:

`strain:` one of six rhizobium cultures, five pure strains and one a mixture of all five strains. Five strains of alfalfa rhizobium were used for the alfalfa plants and five strains of clover rhizobium were used for the clover plants.

`comb:` a factor at two levels. At one level the rhizobium cultures consisted of only strains specialized for the host plant. At the other level each of the six cultures was combined with a mixture of rhizobium strains specialized for the other plant.

12.14.3 *First Rhizobium* Experiment: Alfalfa Plants

Five observations on the response variable, nitrogen content, were taken at each of the 12 `strain*comb` treatment combinations. Primary interest was in the differences in responses to the six rhizobium treatments. Erdman originally analyzed the response variable "milligrams of nitrogen per 20 plants". After studying his analysis and his discussion we choose to analyze a related response variable, "milligrams of nitrogen per gram of dry plant weight". We give the original analysis as Exercise 12.1.

12.14.4 *Second Rhizobium* Experiment: Clover Plants

Five observations on the response variable, nitrogen content, were taken at each of the 12 `strain*comb` treatment combinations. Primary interest was in the differences in responses to the six rhizobium treatments. Erdman originally analyzed the response variable "milligrams of nitrogen per 10 plants". After studying his analysis and his discussion, we choose to analyze a related response variable, "milligrams of nitrogen per gram of dry plant weight". We give the original analysis as Exercise 12.2.

12.14.5 Initial Plots

Datasets `data(rhiz.alfalfa)` and `data(rhiz.clover)` contain the complete data for both experiments. The alfalfa data is plotted in Figure 12.7. The clover data is plotted in Figure 12.8. Erdman's response variable is shown as `nitro` in both figures. Our response variable is shown as `Npg`. The single most evident feature from the clover boxplots is the large response to the pure culture `3DOk5`. This observation is the one that caused us to consider the alternate response variable. There were fewer plants, hence larger plants, for this strain. We posit that the reported values were scaled up, that is reported as grams per 10 plants. We hope that analyzing the ratio, milligrams of nitrogen per gram of plants, rather than the reported rate, milligrams per 10 plants, will adjust for the outliers. Nothing in the alfalfa plots is as clear.

As a graphical aside, we looked at four different layouts for these plots. In Figures 12.7 and 12.8 we show vertical boxplots by `strain` conditioned on `comb`. We also looked at vertical boxplots by `comb` conditioned on `strain` and horizontal boxplots with both conditionings. We chose this one because we have a preference for the response variable on the vertical axis and because we believe the patterns

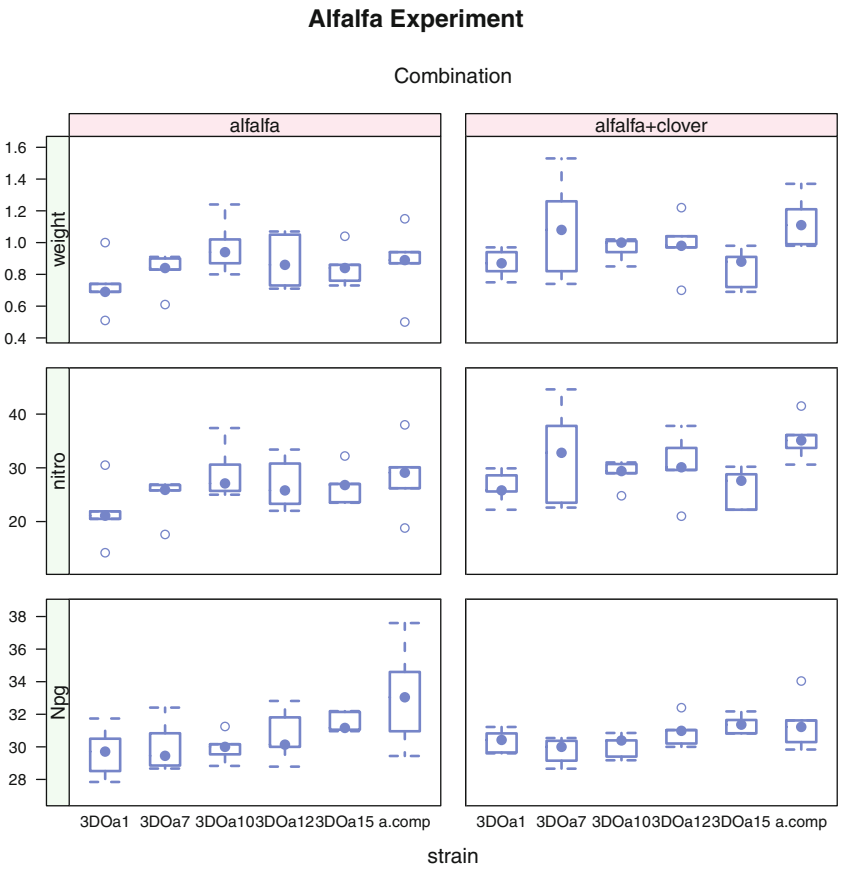


Fig. 12.7 Boxplot of alfalfa data. The `Npg` response variable has the least variability. We shall continue our analysis with `Npg`.

are easier to see when this example is conditioned on the factor `comb`. The other three layouts for the data in Figure 12.8 can be viewed by running the code in file `HHscriptnames(12)`. Also see the discussion in Section 13.A.

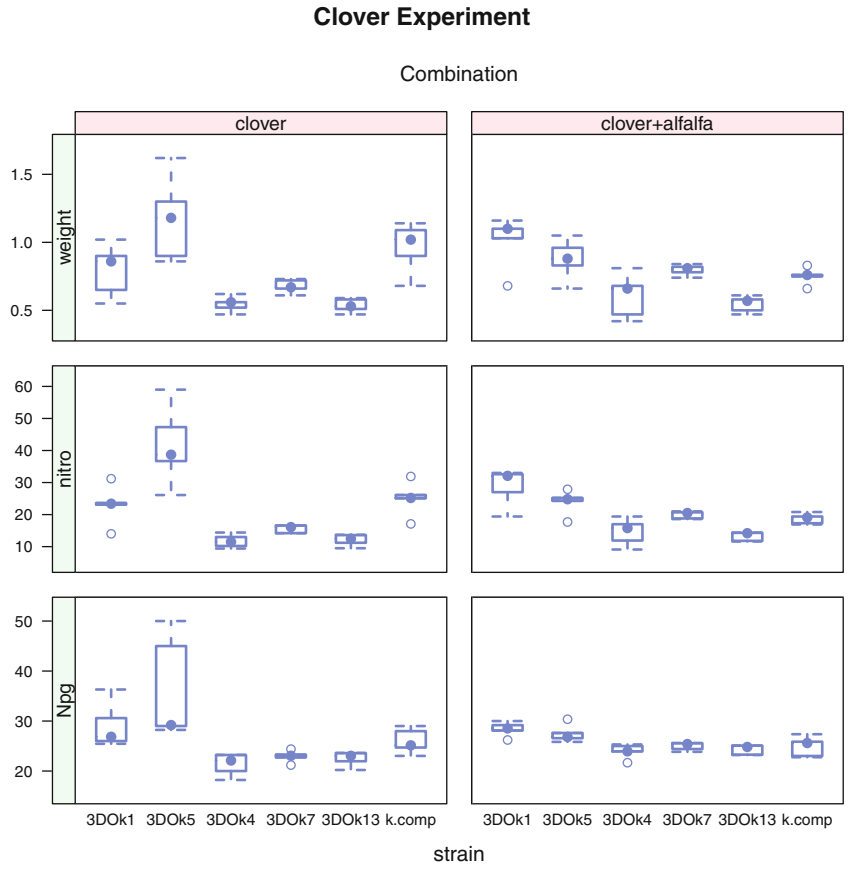


Fig. 12.8 Boxplot of clover data. The `Npg` response variable has the least variability. We shall continue our analysis with `Npg`. The large variability in the `3DOk5` strain is visible in all three response variables.

12.14.6 Alfalfa Analysis

The ANOVA table and table of means for the alfalfa experiment are in Table 12.11. Since there was no interaction with the combination of clover strains of bacteria (`strain:comb` interaction p-value = .53 in Table 12.11), there is no evidence of antibiosis or antagonism.

Since only the `strain` main effect is significant, we confine our investigation to differences among the means for `strain`. Figures 12.9, 12.10, and 12.11 display the results of the Tukey multiple comparison procedure for comparing `strain` mean differences. Since `strain` has 6 levels, we simultaneously examine all $\binom{6}{2} = 15$ pairwise mean differences. Any mean difference having a confidence interval in Figure 12.10 that doesn't include 0 is declared a significantly differing pair. There are three such confidence intervals, therefore we conclude that `a.composite` has a

Table 12.11 ANOVA table and table of means for alfalfa experiment. See Figure 12.7.

```
> ## unset position(rhiz.alfalfa$comb) for glht
> data(rhiz.alfalfa) ## fresh copy of the data.

> rhiz.alfalfa.aov <- aov(Npg ~ strain * comb, data=rhiz.alfalfa)

> summary(rhiz.alfalfa.aov)
            Df Sum Sq Mean Sq F value  Pr(>F)
strain       5  46.22   9.244   4.565 0.00174 **
comb         1   0.57   0.573   0.283 0.59714
strain:comb  5   8.44   1.687   0.833 0.53275
Residuals   48  97.21   2.025
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> alf.means <- model.tables(rhiz.alfalfa.aov, type="means",
+                           se=TRUE, cterms="strain")

> alf.means
Tables of means
Grand mean

30.73547

 strain
strain
 3DOa1  3DOa7 3DOa10 3DOa12 3DOa15 a.comp
 30.00  29.89  30.00  30.82  31.43  32.27

Standard errors for differences of means
        strain
        0.6364
replic.     10
```

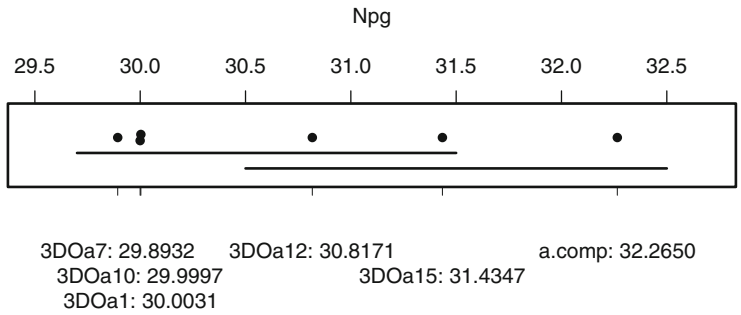
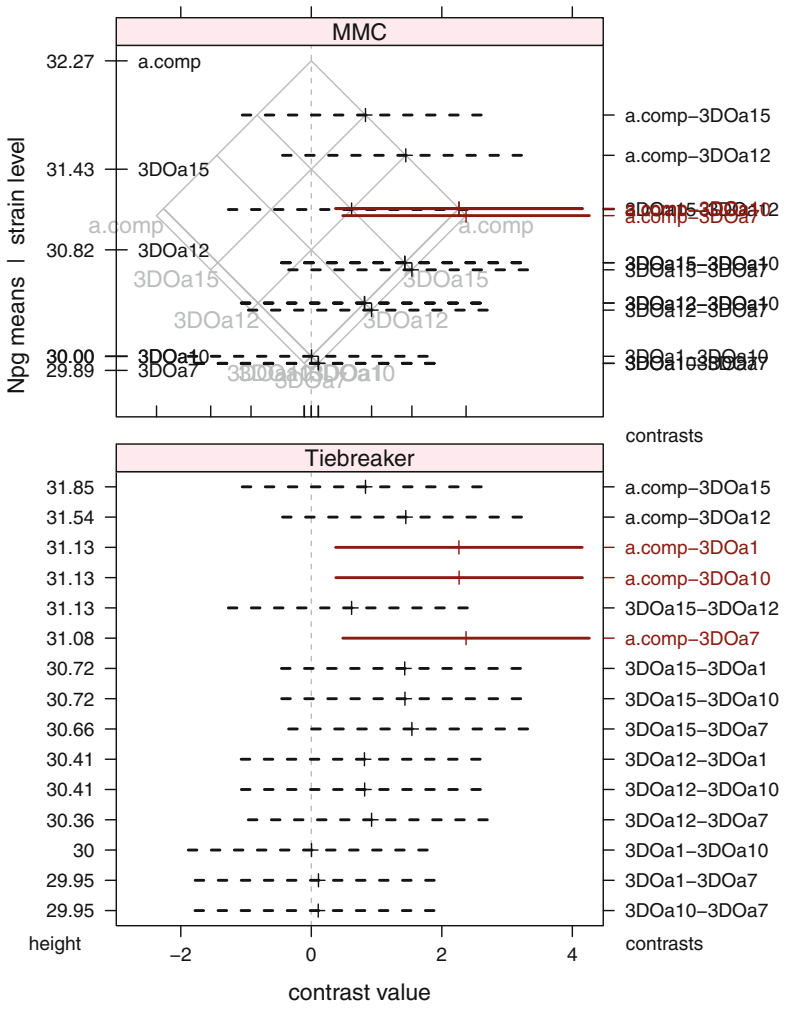


Fig. 12.9 Means for alfalfa experiment. Dots that appear over a common horizontal line correspond to population means that do not differ significantly according to the Tukey multiple comparisons procedure with simultaneous 95% confidence intervals. Compare this figure to Figure 12.10.



The MMC panel shows informative overprinting. Please see Tiebreaker panel and caption.

Fig. 12.10 Mean–mean multiple comparisons plot and Tiebreaker plot showing Tukey contrasts for alfalfa data. The MMC panel shows informative overprinting. Note that `3DOa1`, `3DOa10`, and `3DOa7` have almost identical means. Consequently (i) their means overprint on the left axis, (ii) their differences overprint on the right axis, and (iii) their contrasts are displayed at the same vertical position in the MMC panel. Most of the overprinting contrasts cross zero and are not significant. Their details are displayed in the Tiebreaker panel. The only significant contrasts (the red solid lines) are on the right corner of the isomeans grid. All three are contrasts of `a.comp` with the three almost identical means of the lower three strains. Again the details are clear in the Tiebreaker panel. The MMC panel displays the contrasts at heights constructed as the average of the two means being compared. The Tiebreaker panel shows the contrasts in the same data-dependent vertical order as the MMC panel. The Tiebreaker panel breaks the ties in the MMC panel by placing the confidence intervals at equally spaced vertical positions. See also Figure 12.11 where we have constructed a set of orthogonal contrasts to capture and illustrate the relationships among the levels.

significantly higher mean response than each of 3D0a1, 3D0a7, and 3D0a10; these were the only significant differences detected. The inference is that any of the three treatments with high response (3D0a12, 3D0a15, or a.composite) should be used.

Equivalent information is contained in Figure 12.9, where two population means are declared significantly different if their corresponding sample means are *not* underlined by the same line. (Such an underlining display may be used only when all samples have the same size, as is the case here.)

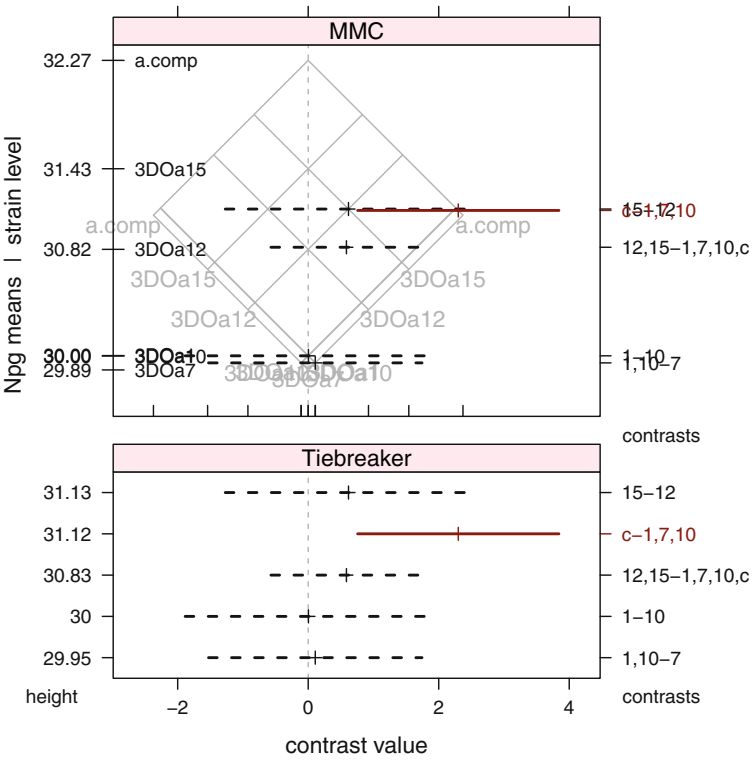
Figure 12.10 is very busy because it shows 15 pairwise contrasts for only 5 degrees of freedom. In Figure 12.11 we provide a graphical summary of our conclusions by constructing an orthogonal basis for the contrasts. We believe the orthogonal contrasts in Figure 12.11 are easier to use in expository settings. The detail of Figure 12.10 is needed to help us construct a useful and meaningful set of orthogonal contrasts. We see that the single comparison between a.composite and the average of the three strains with low means (3D0a1, 3D0a7, and 3D0a10) is the only significant effect.

Figure 12.10 and 12.11 each have two panels. The MMC (mean–mean multiple comparisons) panel shows the MMC plot discussed in Chapter 7. There is severe overprinting of the confidence intervals and their labeling because so many of the means and estimates of their differences have almost identical values. The overprinting is itself information on similarity of level means. Nonetheless we need a tiebreaker that will return legibility to the plot. We provide the tiebreaker in the Tiebreaker panel, an ordinary multiple comparisons plot of the individual contrasts placed at equally spaced vertical positions, sorted to be in the same data-dependent order that is used in the MMC panel. This sort order is based on the values of the level means. The standard sort order used by both R (see for example Table 6.3) and SAS is based on the names of the levels.

12.14.7 Clover Analysis

The ANOVA table and table of means for the clover experiment are shown in Table 12.12. In this experiment the strain:comb interaction effect is significant and the comb main effect is not significant.

The significance of the strain:comb in Table 12.12 (p-value $< .01$) implies that we can't immediately, if at all, interpret the main effect of strain. Main effect comparisons of the levels of comb and strain are inappropriate because the difference in response to two levels of strain will differ according to the level of comb. From the table of means in Table 12.12 and the interaction plots in Figure 12.12 we discover, again, that strain 3D0k5 is the anomaly. The interaction is made evident by the lack of parallel profiles in both interaction plot panels. The three points marked



The MMC panel shows informative overprinting. Please see Tiebreaker panel and caption.

Fig. 12.11 MMC plot and Tiebreaker plot of an orthogonal set of Tukey contrasts for alfalfa data. There are six strains, hence five independent comparisons. This orthogonal set has been chosen to summarize the information in Figure 12.10 and show that only one comparison is significantly different from 0. We see now that the three strains with low means are indistinguishable, that the two intermediate strains are indistinguishable, and that these two clusters are not significantly different from each other. The only significant comparison is from the mean of the composite to the mean of the cluster of three strains with low means. The MMC panel shows the same overprinting discussed in the caption to Figure 12.10. The Tiebreaker panel shows clearly the single significant comparison of the composite mean to the mean of the cluster of three strains with low means.

as outliers in both boxplot panels are the points that drive much of the remaining analysis. We show the simple effects in Figure 12.13.

Once we decide that main effects are not meaningful in the presence of strong interaction, we must look at the behavior separately for each level of the `comb` factor. We continue to do so in the context of a single analysis because we are still able to use the residual term constructed from all levels of `comb`. This residual term has 48 degrees of freedom. Had we been forced to run separate analyses each would have had a residual with much fewer degrees of freedom. Recall from Section 3.10 that tests are more powerful when the denominator has higher degrees of freedom.

Table 12.12 ANOVA table and table of means for clover experiment. See Figure 12.8.

```
> rhiz.clover.aov <- aov(Npg ~ strain * comb, data=rhiz.clover)

> summary(rhiz.clover.aov)
            Df Sum Sq Mean Sq F value   Pr(>F)
strain       5  642.3  128.45   9.916 1.47e-06 ***
comb         1    6.9    6.88   0.531  0.46955
strain:comb  5  228.2   45.65   3.524  0.00857 **
Residuals   48  621.8   12.95
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> model.tables(rhiz.clover.aov, type="means", se=TRUE)
Tables of means
Grand mean

26.00674

 strain
strain
 3DOk1  3DOk5  3DOk4  3DOk7 3DOk13 k.comp
 28.72  31.86  22.66  23.95  23.39  25.45

 comb
comb
        clover clover+alfalfa
        26.345         25.668

 strain:comb
        comb
strain   clover clover+alfalfa
  3DOk1  29.04  28.41
  3DOk5  36.29  27.44
  3DOk4  21.35  23.98
  3DOk7  22.93  24.96
  3DOk13 22.49  24.30
  k.comp 25.97  24.92

Standard errors for differences of means
        strain   comb strain:comb
        1.6096 0.9293      2.2763
replic.     10     30           5
```

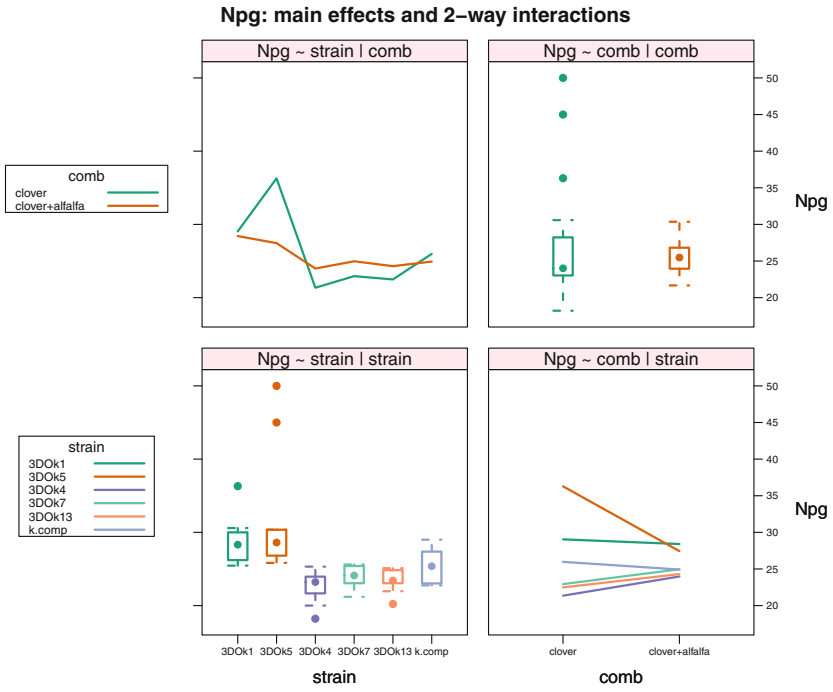


Fig. 12.12 Interaction plot for clover experiment. The three points marked as outliers in clover 3DOk5 are the points that drive much of the remaining analysis. We show the simple effects for this interaction in Figure 12.13.

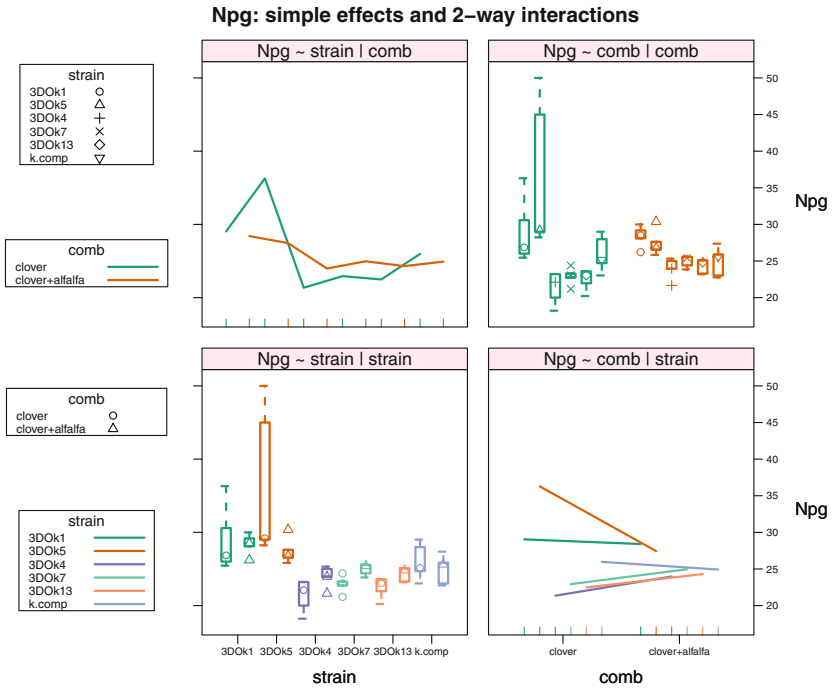


Fig. 12.13 Simple effects plot for the clover experiment interaction in Figure 12.12. It is now even more clear that the clover 3DOk5 points differ from the others.

We therefore repartition the sums of squares in Table 12.13 and look separately at the simple effect of `strain` within each of the levels of `comb`. The notation in Table 12.13 is the mechanics by which the 10 degrees of freedom are separated into two meaningful groups of 5 degrees of freedom. The differences in the clover strains of rhizobium alone are significant. The differences with the combination clover and alfalfa strains of rhizobium are not. Therefore, we examine only the *simple effects* within the clover strains. These simple effects are the differences between pairs of means of `strain` within the `clover` level of the factor `comb`. We examine and report on those such differences that are statistically significant. Since the simple effect for `strain` within the `clover+alfalfa` level of `comb` is not significant, we do not look further at those means.

Erdman's interpretation of the analysis shows that bacteria strain `3DOk5` showed antibiosis with the alfalfa bacteria strains. With `3DOk5` the response was strong alone and suppressed when combined with the alfalfa bacteria culture.

Table 12.13 ANOVA table showing simple effects for strain in clover experiment. We partitioned the sums of squares for the nesting with the `split` argument to the `summary` function. We needed to display the names of the individual regression coefficients in order to determine which belonged to each of the levels of `comb`. In this example the `comb` and `strain` effects are orthogonal, hence the partitioning is valid. The individual degrees of freedom are usually not interpretable.

```
> rhiz.clover.nest.aov <-
+     aov(Npg ~ comb/strain, data=rhiz.clover)

> summary(rhiz.clover.nest.aov)
            Df Sum Sq Mean Sq F value Pr(>F)
comb         1    6.9    6.88   0.531   0.47
comb:strain 10  870.5   87.05   6.720  2e-06 ***
Residuals   48  621.8   12.95
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> old.width <- options(width=35)

> names(coef(rhiz.clover.nest.aov))
 [1] "(Intercept)"
 [2] "combclover+alfalfa"
 [3] "combclover:strain3DOk5"
 [4] "combclover+alfalfa:strain3DOk5"
 [5] "combclover:strain3DOk4"
 [6] "combclover+alfalfa:strain3DOk4"
 [7] "combclover:strain3DOk7"
 [8] "combclover+alfalfa:strain3DOk7"
 [9] "combclover:strain3DOk13"
[10] "combclover+alfalfa:strain3DOk13"
[11] "combclover:straink.comp"
[12] "combclover+alfalfa:straink.comp"

> options(old.width)

> summary(rhiz.clover.nest.aov,
+         split=list("comb:strain"=
+           list(clover=c(1,3,5,7,9),
+                "clover+alf"=c(2,4,6,8,10))))
                          Df Sum Sq Mean Sq F value   Pr(>F)
comb                       1    6.9    6.88   0.531    0.470
comb:strain               10  870.5   87.05   6.720 2.00e-06 ***
  comb:strain: clover      5  788.4  157.68  12.172 1.22e-07 ***
  comb:strain: clover+alf  5   82.1   16.42   1.268    0.293
Residuals                 48  621.8   12.95
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

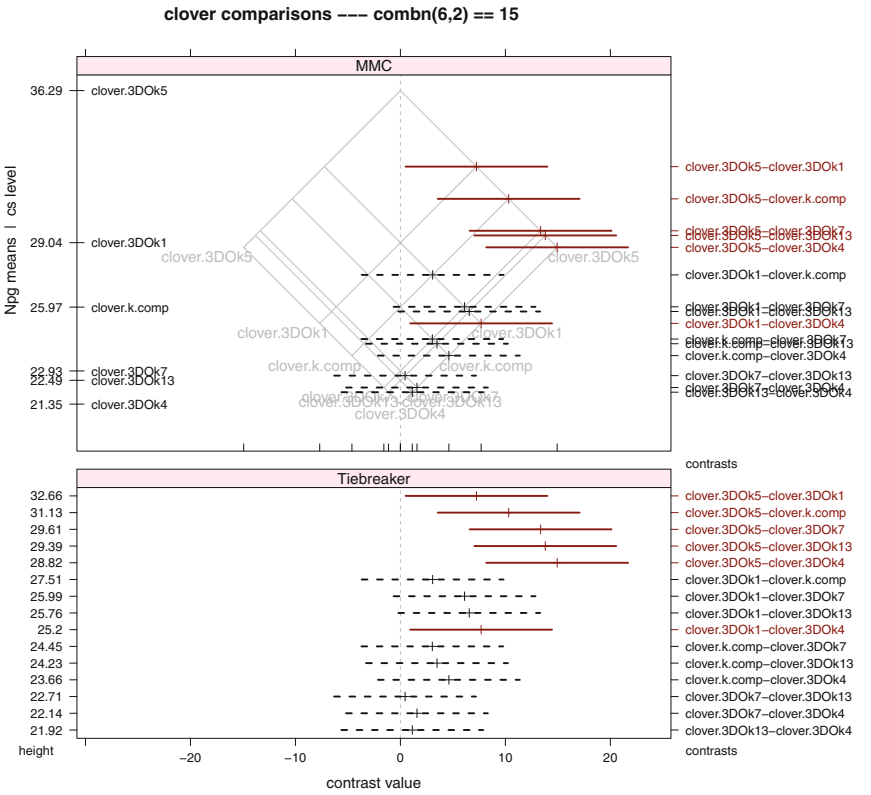


Fig. 12.14 MMC plot and Tiebreaker plot of Tukey simple effect contrasts for `comb="clover"` data. It is visually quite clear that the strain `3DOk5` differs from the rest (very strongly for the bottom three strains and less so for the middle two strains.) There is also one marginally significant contrast between the second largest mean and the smallest mean. We illustrate this observation in Figure 12.15 with an appropriately chosen set of orthogonal contrasts. The Tiebreaker plot in the bottom panel is imperative for this example. The means at many of the levels of `cs` are very close and therefore their labels are overprinted. The Tiebreaker plot shows the contrasts equally spaced along the *y*-axis and in the same sequence as the top panel.

Table 12.14 shows the dummy variables and Table 12.15 shows the regression coefficients for the simple effects of `strain` in the clover experiment displayed in Table 12.13. The names for the columns of the dummy variables generated by the program are excessively long and would force the matrix of dummy variables to occupy many pages just to accommodate the column names. Therefore, we abbreviated them. We see the nesting structure in the dummy variables as the `cmb`*n* columns for pure strains and the `cm+`*n* columns for combination strains are identical in structure. Only the `cmb`*n* regression coefficients are significant. The dummy variables are constructed from the default treatment contrasts.

Since there is interaction in the clover experiment, we must look at the multiple comparisons for the simple effects of `strain` at each value of `comb`.

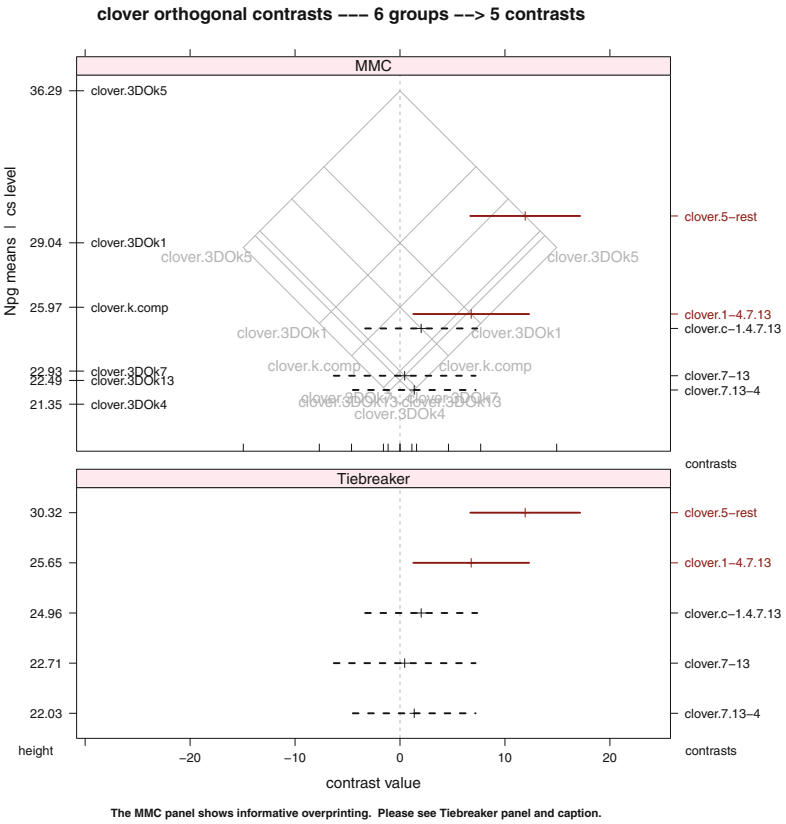


Fig. 12.15 MMC plot and Tiebreaker plot of an orthogonal basis set of Tukey simple effect contrasts for `comb="clover"` data. We summarize the conclusions from Figure 12.14. The strongest contrast compares `3DOk5` to the rest. There is one other marginally significant contrast. Two contrasts show that the three strains with the lowest means are indistinguishable. The Tiebreaker plot in the bottom panel is imperative for this example. The means at many of the levels of `cs` are very close and therefore their labels are overprinted. The Tiebreaker plot shows the contrasts equally spaced along the *y*-axis and in the same sequence as the top panel.

Figure 12.14 shows the simple effects for `comb="clover"`. The only strongly significant contrasts are the ones centered on the upper right isomeans grid line (`clover.3DOk5`) comparing `3DOk5` to the rest of the strains. There is one other borderline significant contrast. The Tiebreaker panel makes it slightly easier to identify the names of the contrasts. The set of orthogonal contrasts in Figure 12.15 shows that the single contrast comparing `3DOk5` to the others carries almost all the significance in Figure 12.15.

Figure 12.16 shows that there are no significant contrasts in the simple effects for `comb="clover+alfalfa"`. We forced Figure 12.16 to be on the same scale as Figure 12.14.

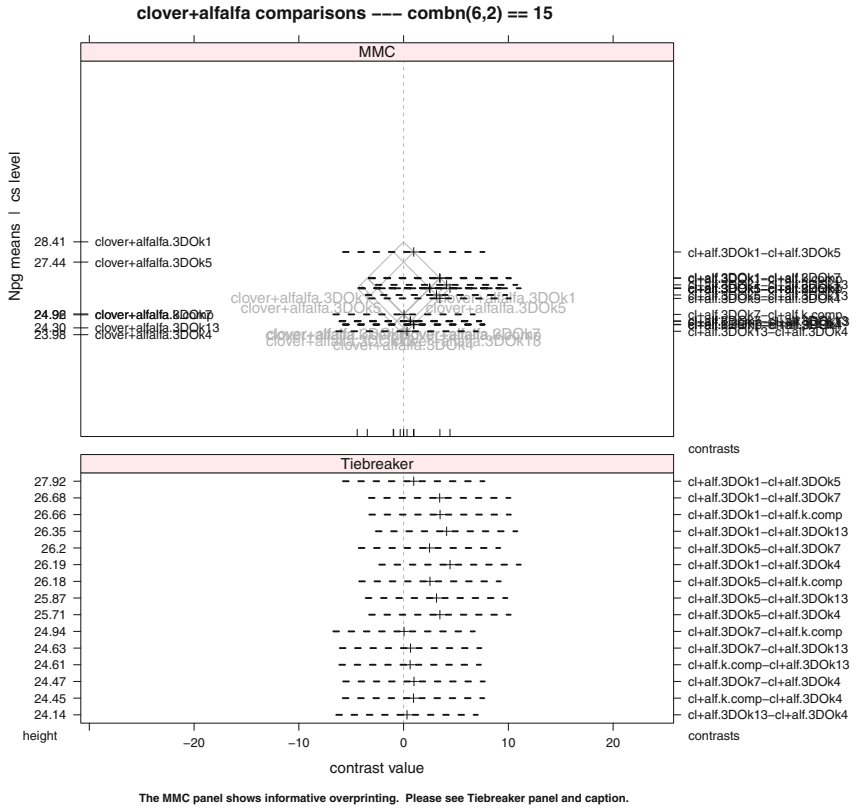


Fig. 12.16 MMC plot and Tiebreaker plot of Tukey simple effect contrasts for `comb="clover+alfalfa"` data. This plot is on the same scale as Figure 12.14. This common scale emphasizes the disparity between `3DOk5` in `comb="clover"` and any values of `strain` in `comb="clover+alfalfa"`. None of the simple effects for `strain` within the `clover+alfalfa` level of `comb` are significant. The Tiebreaker plot in the bottom panel is imperative for this example as all the means are almost identical and therefore their labels are overprinted.

Table 12.14 Dummy variables for simple effects of `strain` in clover experiment. These dummy variables are based on the treatment contrasts. The sums of squares from these dummy variables are displayed in Table 12.13. The regression coefficients are in Table 12.15. The dummy variables and regression coefficients have been reordered to place the within-`clover` values together and the within-`clover+alfalfa` values together.

```
> ## Look at the contrasts, their generated dummy variables,
> ## and their regression coefficients.
> ## Abbreviate their names for presentation.
> tmp <- abbreviate(names(coef(rhiz.clover.nest.aov)))

> ## tmp
>
> ## contrasts(rhiz.clover$comb)
> ## contrasts(rhiz.clover$strain)
>
> cnx <- aov(Npg ~ comb/strain, data=rhiz.clover, x=TRUE)$x

> dimnames(cnx)[[2]] <- tmp

> ## cnx
> cnx[seq(1,60,5), c(1,2,  3,5,7,9,11)]
   (In) cmb+ c:3D05 c:3D04 c:3D07 c:3D01 cm:.
1     1    0      0      0      0      0    0
6     1    0      1      0      0      0    0
11    1    0      0      1      0      0    0
16    1    0      0      0      1      0    0
21    1    0      0      0      0      1    0
26    1    0      0      0      0      0    1
31    1    1      0      0      0      0    0
36    1    1      0      0      0      0    0
41    1    1      0      0      0      0    0
46    1    1      0      0      0      0    0
51    1    1      0      0      0      0    0
56    1    1      0      0      0      0    0

> cnx[seq(1,60,5), c(4,6,8,10,12)]
   c+:3D05 c+:3D04 c+:3D07 c+:3D01 c+:.
1        0       0       0       0    0
6        0       0       0       0    0
11       0       0       0       0    0
16       0       0       0       0    0
21       0       0       0       0    0
26       0       0       0       0    0
31       0       0       0       0    0
36       1       0       0       0    0
41       0       1       0       0    0
46       0       0       1       0    0
51       0       0       0       1    0
56       0       0       0       0    1
```

Table 12.15 Regression coefficients for simple effects of `strain` in clover experiment. The contrasts and dummy variables are displayed in Table 12.14. The dummy variables and regression coefficients have been reordered to place the within-`clover` values together and the within-`clover+alfalfa` values together.

```
> cnxb <- round(coef(summary.lm(rhiz.clover.nest.aov)), 3)

> dimnames(cnxb)[[1]] <- tmp

> ## cnxb
> cnxb[c(1,2,  3,5,7,9,11, 4,6,8,10,12),]
        Estimate Std. Error t value Pr(>|t|)
(In)      29.042      1.610  18.043    0.000
cmb+      -0.637      2.276  -0.280    0.781
c:3D05     7.243      2.276   3.182    0.003
c:3D04    -7.688      2.276  -3.378    0.001
c:3D07    -6.110      2.276  -2.684    0.010
c:3D01    -6.556      2.276  -2.880    0.006
cm:.      -3.070      2.276  -1.349    0.184
c+:3D05   -0.966      2.276  -0.424    0.673
c+:3D04   -4.430      2.276  -1.946    0.057
c+:3D07   -3.441      2.276  -1.512    0.137
c+:3D01   -4.106      2.276  -1.804    0.078
c+:.      -3.482      2.276  -1.530    0.133
```

12.15 Models Without Interaction

Experiments with two factors are normally designed with a large enough sample size to investigate the possibility that the factors interact. When the analyst has previous experience with these factors, or subject area knowledge that the factors are unlikely to interact, it is possible to set up the model without an interaction term:

Algebra	$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$
R	`Y ~ A + B`
SAS	`Y = A B`

The residual portion of this no-interaction model includes the $(a-1)(b-1)$ degrees of freedom that would otherwise have been attributable to the AB interaction. If the no-interaction assumption is correct, the no-interaction model provides a more precise estimate of the residual than a model incorporating interaction and this in turn implies more power for tests involving the individual main effects or the means of their levels. With this model, comparisons among the levels of factor A or among the levels of factor B are undertaken in much the same way as in a one-way experiment, but using this model's residual sum of squares and degrees of freedom.

When we initially posit a model containing the two-factor interaction, it may happen that the analysis of variance test for interaction leads to non-rejection of the no-interaction hypothesis. If the evidence for no interaction is sufficiently strong (a large p-value for this test and/or no strong subject area feeling about the existence of interaction), the analyst may feel comfortable about reverting to the no-interaction model and proceeding with the analysis as above. This amounts to pooling a nonsignificant interaction sum of squares with the previous residual sum of squares (calculated under the now rejected assumption of an interaction) to produce a revised residual mean square (under the assumption of no interaction). This combined or pooled estimate is justified because in the absence of interaction, the interaction mean square estimates the same quantity, the residual variance, as does the residual mean square. The pooled estimate of the residual variance is an improvement over the individual estimates because it is constructed with additional degrees of freedom. Therefore, the pooled estimate provides more powerful inferences on the level means of the two factors than would a residual mean square in a model including interaction. See Section 5.4.2 for further discussion of pooling.

12.16 Example—Animal Feed Data

12.16.1 Study Objectives

A manufacturer of animal feed investigated the influence on the amount of vitamin A retained in feed. The manufacturer considered 15 treatment combinations formed from 5 amounts of feed supplement and 3 levels of temperature at which the supplements were added to the feed. Two samples were selected at each treatment combination. The data from Anderson and McLean (1974), accessible as `data(feed)`, are said to be on transformed scales that this reference does not specify. The response variable is `retained` and the two factors are `temp` and `supp`.

12.16.2 Analysis

The data is displayed in the interaction plot in Figure 12.17. The profiles in the interaction plot are sufficiently close to parallel to suggest that there is no interaction between `temp` and `supp`.

Table 12.16 Feed data: ANOVA table for model with interaction. The interaction is not significant.

```
> feed.int.aov <- aov(retained ~ temp * supp, data=feed)

> anova(feed.int.aov)
Analysis of Variance Table

Response: retained
          Df Sum Sq Mean Sq F value    Pr(>F)
temp       2 1479.2  739.60 26.0423 1.321e-05 ***
supp       4 3862.1  965.53 33.9977 2.334e-07 ***
temp:supp  8  243.5   30.43  1.0716    0.4313
Residuals 15  426.0   28.40
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Initially, in Table 12.16, we fit an interaction model leading to an interaction p-value of 0.43, confirming our impression from the interaction plot that `temp` and `supp` do not interact. It is not unreasonable to conclude that temperature affects each concentration of feed supplement in roughly the same way. Therefore, we abandon the assumption of interaction and move to a no-interaction model.

Table 12.17 Feed data: ANOVA with main effects and their polynomial contrasts.

```
> feed.aov <- aov(retained ~ temp + supp, data=feed)

> anova(feed.aov)
Analysis of Variance Table

Response: retained
          Df Sum Sq Mean Sq F value    Pr(>F)
temp       2 1479.2  739.60  25.410 1.499e-06 ***
supp       4 3862.1  965.53  33.172 3.037e-09 ***
Residuals 23  669.5   29.11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(feed.aov, split=
+         list(temp=list(linear=1, quadratic=2),
+              supp=list(linear=1, quadratic=2, rest=3:4)))
                Df Sum Sq Mean Sq F value   Pr(>F)
temp             2   1479   739.6  25.409 1.50e-06 ***
  temp: linear   1    370   369.8  12.705  0.00165 **
  temp: quadratic 1   1109  1109.4  38.114 2.68e-06 ***
supp             4   3862   965.5  33.172 3.04e-09 ***
  supp: linear   1   2912  2912.1 100.046 7.61e-10 ***
  supp: quadratic 1    947   946.7  32.525 8.30e-06 ***
  supp: rest     2      3     1.7   0.058  0.94418
Residuals       23    669    29.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> model.tables(feed.aov, type="means", se=TRUE)
Tables of means
Grand mean

68.8

 temp
temp
  40   80  120
60.2 77.4 68.8

 supp
supp
    2     4     6     8    10
48.33 64.67 76.00 79.00 76.00

Standard errors for differences of means
         temp  supp
        2.413 3.115
replic.    10     6
```

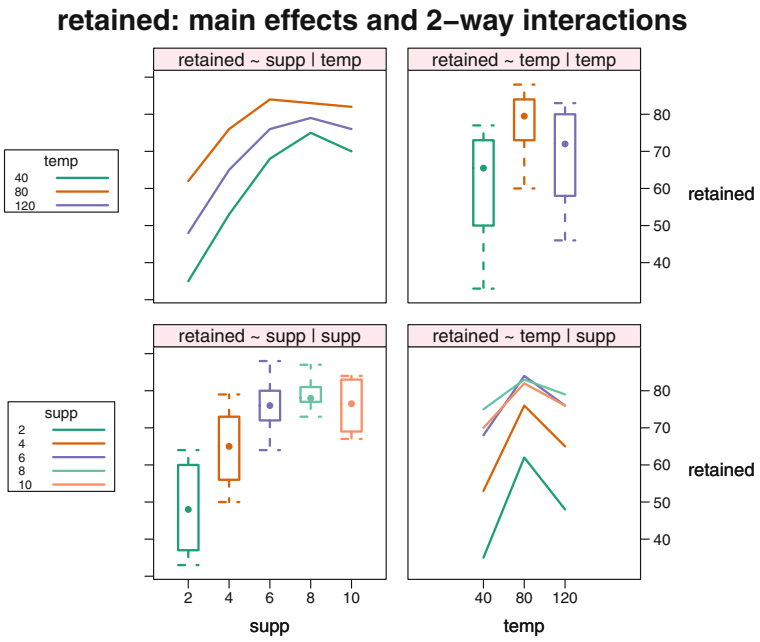


Fig. 12.17 Feed data interaction plots.

The fit of the no-interaction model in Table 12.17 suggests that both `temp` and `supp` impact significantly on `retained`. Since both of these factors have quantitative levels, our analysis of the nature of the mean differences involves modeling the response `retained` as polynomial functions of both temperature and the amount of supplement. The method for accomplishing such modeling was introduced in Section 10.4.

The interaction plot in Figure 12.17 suggests that the response to changes in the level of `supp` is quadratic in nature and that possibly the response to changes in the level of `temp` is quadratic as well. Therefore, for both of these factors we calculated the one degree-of-freedom tests on the linear and quadratic contrasts among the factor levels, and show the results in Table 12.17. Since the p-values for both quadratic contrasts are close to 0, there is strong evidence that the response of vitamin A retention is a quadratic function of both temperature and amount of feed supplement.

We show the MMC plot for supplement in Figure 12.18 for all pairwise contrasts and in Figure 12.19 for the orthogonal contrasts. The Tiebreaker panel is needed because two of the supplement means are identical. The MMC plot of the orthogonal polynomial contrasts shows the linear and quadratic effects are significant.

Our findings implies that for maximum vitamin A retention we should recommend intermediate amounts of `temp` and `supp`, perhaps in the vicinity of `temp`=80 and `supp`=6. An enlargement of this experiment could more accurately determine the optimal values.

If the analyst had been told, prior to the design of the experiment, that the primary goal was to determine the optimizing combination of the inputs `temp` and `supp`, the analyst would have considered using a *response surface design*, the most efficient design for this purpose. A brief introduction to such designs is contained in Montgomery (2001).

12.17 Exercises

12.1. Do the original Erdman alfalfa analysis of Section 12.14.3 with `nitro` as the response variable. Use the data accessible as `data(rhiz.alfalfa)`.

12.2. Do the original Erdman clover analysis of Section 12.14.4 with `nitro` as the response variable. Use the data accessible as `data(rhiz.clover)`.

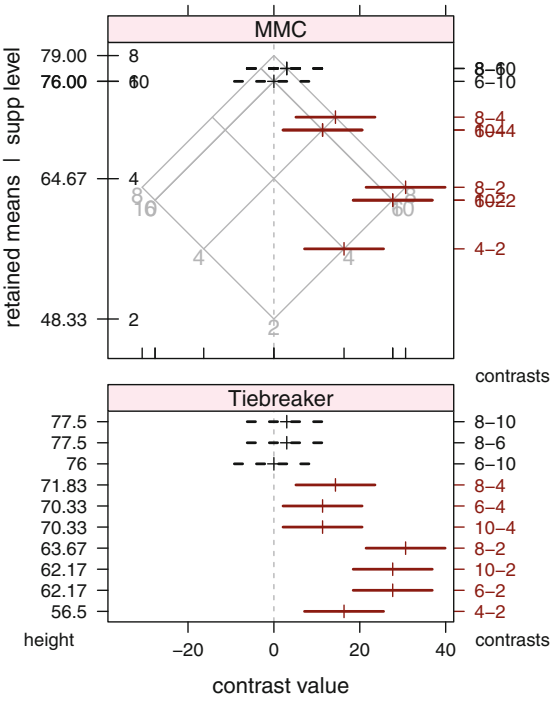
12.3. Analyze the two factor experiment with data accessible as `data(testing)`. This is a 3×3 design with 4 observations per treatment combination. The factors are breaker at levels 1 to 3 and Gauger at levels 1 to 3. The observations are strengths of cement. The cement is "gauged" or mixed with water and worked by three different gaugers before casting it into cubes. Three testers or "breakers" later tested the cubes for compressive strength, measured in pounds per square inch. Each gauger gauged 12 cubes, which were divided into 3 sets of 4, and each breaker tested one set of 4 cubes from each gauger. Breakers and gaugers are fixed in this experiment. Breakers and gaugers are people, not machines. Are there differences in the strength of the cement that depend on the handling by the breakers and gaugers?

We got the data from Hand et al. (1994). The data originally appeared in Davies and Goldsmith (1972). There the data were coded by $.1(X - 1000)$ before analysis.

The term *coded data* means that they have been centered and scaled to make the numerical values easier to work with by hand. The F-tests in the ANOVA table and the t-tests for regression coefficients with coded data are identical to the tests for the original data.

12.4. An agronomist compared five different sources of nitrogen fertilizer and a control (no fertilization) on the yield of barley. A randomized block design was used with four types of soil as the blocks. Each soil type was randomly divided into six plots, and the six treatments were randomly assigned to plots within each type. The treatments were, respectively, $(NH_4)SO_4$, NH_4NO_3, $CO(NH_2)_2$, $Ca(NO_3)_2$, $NaNO_3$, and control. The data, taken from Peterson (1985), are accessible as `data(barleyp)`.

a. Plot the data. Does it appear from the plot that yield is related to treatment? Does it appear from the plot that blocking was successful?



The MMC panel shows informative overprinting. Please see Tiebreaker panel and caption.

Fig. 12.18 Feed data MMC plot for pairwise contrasts of supplement. The means of `retained` at levels of feed 6 and 10 are identical. We need the Tiebreaker panel to distinguish them. Since Table 12.17 shows that the polynomial contrasts are very significant, we show the MMC plot for orthogonal polynomial contrasts in Figure 12.19.

b. Set up the two-way analysis of variance table for these data and explain what you conclude from it.

c. Use the Dunnett procedure, introduced in Section 7.1.3, to compare the five fertilizers with the control. Report your findings to the agronomist.

12.5. A chemist compared the abilities of three chemicals used on plywood panels to retard the spread of fire. Twelve panels were obtained, and each chemical was randomly sprayed on four of these twelve panels. Two pieces were cut from each panel and the time was measured for each piece to be completely consumed by a standard flame. (Thus Panel is nested within Chemical and Sample is nested within Panel.) The data, from Peterson (1985), are accessible as `data(retard)`. Carefully noting the relationship between the factors `chemical` and `panel`, and considering whether these factors are fixed or random, set up an analysis of variance and followup analysis of `chemical` means in order to make a recommendation of the best chemical for retarding the spread of plywood fires.

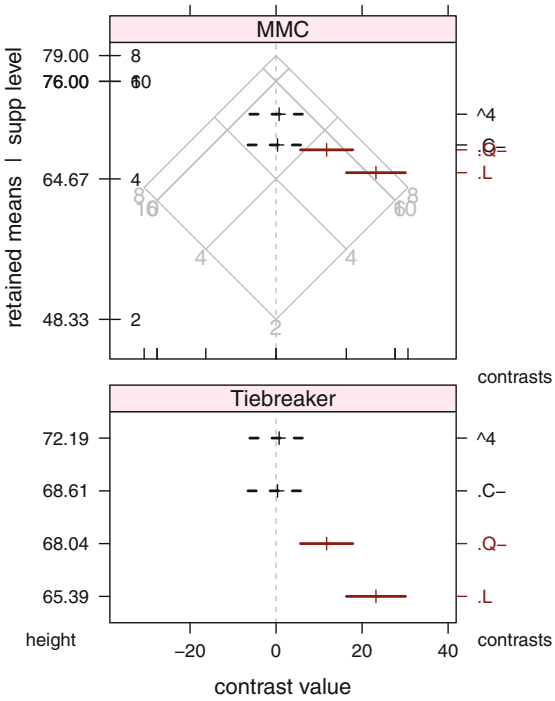


Fig. 12.19 Feed data MMC plot for orthogonal polynomial contrasts of supplement. As indicated in the ANOVA table, the linear and quadratic contrasts are significant and the cubic and quartic are not. We show the Tiebreaker panel even though it is not really needed in this example.

12.6. The judging of the ice skating events at the 2002 Winter Olympics in Salt Lake City was very controversial. The data, accessible as `data(skateslc)`, are taken from Olympic Committee (2001). The dataset contains the scores on both `technique` and `presentation` of the five leading `skaters`, assigned by each of nine `judges`. We have recoded the data with $10(X - 5)$. Perform a two-way analysis of variance where the response is the total of both scores. Do further analysis and comment on the consistency of the nine judges across skaters.

12.7. Box and Cox (1964), reprinted in Hand et al. (1994), present the results of a 3× 4 factorial experiment with four replications per treatment combination to illustrate the importance of investigating the normality assumption underlying analyses of variance. The original response variable is the survival time, `survtime` of each of four antidotes, `treatment` to each of three `poisons`. The data are accessible as `data(animal)`.

a. Perform a two-way analysis of variance using `survtime` as the response, taking care to save the calculated cell residuals.

b. Produce a normal probability plot (described in Chapter 5) of the cell residuals and use it to conclude that the residuals are not normally distributed.

c. Redo the two-way analysis of variance with a reciprocal transformation of the response variable `survtime`, and again save the cell residuals. From a normal probability plot of these cell residuals, conclude that these new residuals are normally distributed and hence the transformation was successful.

d. Report your findings to the antidote researchers.

12.8. An experiment was constructed to compare the `effects` on etchings of wafers of four etching `compounds` and heat treatment by each of three `furnaces`. The experiment was reported in Johnson and Leone (1967) and the data are accessible as `data(furnace)`. Viewing `furnace` as a random factor and allowing for the possibility of interaction, provide a thorough analyses of these data.

12.9. Anemia, caused by iron deficiency in the blood, is common in some countries. It was hypothesized that food cooked in iron pots would gain iron content from the pot and, hence when eaten, contribute to alleviation of iron deficiency. Research performed by Adish et al. (1999) compares the iron content (mg/100g) of three types of (traditional Ethiopian) food when cooked in pots of aluminum, clay, or iron. The data, accessible as `data(ironpot)`, give the `Iron` content in mg/100g of food, the type of `Pot`, and the type of `Food`. Perform a two-way analysis of variance and provide interaction plots. Based on your analysis, is the hypothesis supported? Does your conclusion apply to all `Foods` studied?

12.10. To check the consistency of new car fuel efficiency, the miles per gallon of gasoline consumed was recorded for each of 5 `cars` of the same year and brand, on each of 10 randomly selected `days`. The investigation was reported in Johnson and Leone (1967) and the data are accessible as `data(mpg)`. Viewing `car` as a random treatment factor and `day` as a random blocking factor, analyze the data and carefully state your conclusions. Suggest ways to elaborate on and improve this experiment.

12.11. Williams (1959), originally in Sulzberger (1953), examined the effects of temperature on the strength of wood, plywood, and glued joints. The data are accessible as `data(hooppine)`. The studied wood came from hoop pine trees. The response is compressive `strength` parallel to the grain, and the treatment factor is `temperature` in degrees C. An available covariate is the `moisture` content of the wood, and `tree` is a blocking factor.

a. Fit a full model where both `strength` and `moisture` are adjusted for the blocking factor `tree`, allowing for the possibility that `temp` interacts with `moisture`.

b. Conclude that the interaction term can be deleted from this model. Reanalyze without this term. Carefully state your conclusions.

c. Investigate the nature of the relationship between `strength` and `temp`. Conclude that a linear fit will suffice.

d. Provide plots illustrating the conclusion from part a and the final model in parts b and c.

12.A Appendix: Computation for the Analysis of Variance

When there is more than a single factor, the discussion in this chapter is usually limited to the case where the sample size n_{ij} is the same for each cell. The programs we use for the computation do not usually have this limitation. We will discuss more general cases in Chapters 13 and 14.

With R we will be using `aov` for the calculations and `anova` and related commands for the display of the results. `aov` can be used with equal or unequal cell sizes. Model (12.1)

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} = \mu_{ij} + \epsilon_{ijk} \tag{12.1}$$

is denoted in R either by the formula

```
Y ~ A + B + A:B
```

which uses the operator `+` to indicate the sum of two terms and the operator `:` to indicate the interaction of two factors, or by the formula

```
Y ~ A * B
```

which uses the operator `*` to denote the crossing of two factors. The operator `~` is read as "is modeled by". The second formula is syntactically expanded by the program to the first formula before it is evaluated. We usually prefer the more compact notation `Y ~ A * B` because it more closely captures the English statement, "Y is modeled by the crossing of factors A and B."

With SAS we use `PROC ANOVA` and `PROC GLM`. `PROC ANOVA` is limited to the equal sample size cases (actually, to balanced designs; see the SAS documentation for details). Where there are at least two factors and unequal cell sizes [that is, the n_{ij} are not constrained to be equal and some cells may be empty (with $n_{ij} = 0$)] `PROC GLM` should be used. `PROC ANOVA` may not give sensible answers in such cases. Model (12.1) is denoted in SAS either by the expression

```
Y = A   B   A*B
```

which uses a space to indicate the sum of two terms and the operator `*` to indicate the interaction term, or by the expression

```
Y = A | B
```

which uses the operator `|` to denote the crossing of two factors. The operator `=` is read as "is modeled by". The second expression is syntactically expanded by the program to the first expression before it is evaluated. We usually prefer the more compact notation `Y = A | B` because it more closely captures the English statement, "Y is modeled by the crossing of factors A and B."

The intercept term μ and the error term ϵ_{ijk} are assumed in both statistical languages. The existence of the subscripts is implied and the actual values are specified by the data values.

The formula language also includes a notation for nesting of factors. We introduce nesting in Section 12.13 and say more in Section 13.5, especially in Tables 13.18 and 13.21. In R use the formula

```
Y ~ A + A:B
```

or the formula (which will be expanded to the first formula)

```
Y ~ A / B
```

which uses the `/` to indicate that A is an outer factor and `B` is nested within `A`.

SAS doesn't have the equivalent of the second formula. In SAS, use either the equivalent of the first formula

```
Y = A   A*B
```

or an alternative notation

```
Y = A   B(A)
```

which uses the parenthesis notation to indicate that `B` is nested within `A`.

Note that the `A:B` notation (or `A*B` in SAS) tells about the relation of the levels of the factors, not the degrees of freedom. Degrees of freedom depend on the linear dependencies with earlier terms in the model formula.