

# 8

## Analysis of Designed Experiments

---

### 8.1 Introduction

In this chapter we discuss experiments whose main aim is to study and compare the effects of *treatments* (diets, varieties, doses) by measuring *response* (yield, weight gain) on *plots or units* (points, subjects, patients). In general the units are often grouped into *blocks* (groups, sets, strata) of ‘similar’ plots or units.

We assume we have a response  $y$  which is continuous (for example, a normal random variable). We also have possible explanatory variables which are discrete or qualitative, called *factors*. Here we consider treatment factors whose values are assigned by the experimenter. The values or categories of a factor are called *levels*, so that the levels of a treatment factor are labels for the different treatments. For example, in comparing engineering components from several different manufacturers, we could define indicator variables to denote the manufacturer. If there are  $k$  different manufacturers, one way to do it is to define

$$x_{ij} = \begin{cases} 1 & \text{if component } i \text{ comes from manufacturer } j, \\ 0 & \text{otherwise,} \end{cases} \quad (8.1)$$

for  $j = 1, 2, \dots, k - 1$ . We do not need to define  $x_{ik}$  because if  $x_{i1} = \dots = x_{i(k-1)} = 0$  then it automatically follows that component  $i$  is from manufacturer  $k$ . It is possible to use the regressor variables defined by 8.1, either on their own or in conjunction with other continuous regressors, to fit a linear regression model using standard software. If that were all there was to it, we would not need a separate theory for the analysis of variance. However, the specific issues raised by these models are such as to justify and require a detailed treatment in their own right.

### 8.2 Completely Randomized Design

We allocate  $r$  treatments randomly to the  $n$  sample units such that the  $i$ th treatment is allocated to  $n_i$  units ( $i = 1, 2, \dots, r$ ) (i.e. each treatment has  $n_i$  *replications*). Note there is no *blocking* here: very often in experimental

design, the units are blocked into groups of similar units before being allocated to treatments, but we are not assuming that kind of experiment yet (see Section 7.3) — all  $n = \sum_i n_i$  units are regarded as ‘similar’ and there is one treatment factor with  $r$  levels. If the  $n_i$  are all equal, then the experiment is called balanced. Note that a completely randomized design with  $r = 2$  gives an experiment which is the same set-up as for a two-sample  $t$ -test (Appendix A). An experiment with two treatments arranged in blocks, so that each block contains one member assigned to the first treatment and one member assigned to the second, corresponds to a matched pairs  $t$ -test.

For the completely randomized design we use the one-way model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

where  $j = 1, 2, \dots, n_i$ ,  $i = 1, 2, \dots, r$  and  $\sum_{i=1}^r n_i = n$ . In this model  $y_{ij}$  is the response of the  $j$ th unit receiving the  $i$ th treatment,  $\mu$  is an overall mean effect,  $\alpha_i$  is the effect due to the  $i$ th treatment and  $\epsilon_{ij}$  is random error. Another natural way to write this model is

$$y_{ij} = \mu_i + \epsilon_{ij}$$

where  $j = 1, 2, \dots, n_i$ ,  $i = 1, 2, \dots, r$  and  $\sum_{i=1}^r n_i = n$ . The only difference between these two models is that  $\mu_i$  is used to denote the total response to treatment  $i$ . The first specification has the advantage that  $\alpha_i$  has a specific interpretation as a “treatment effect” associated with treatment  $i$ .

We have the usual least squares assumptions that the  $\epsilon_{ij}$ s have zero mean and constant variance  $\sigma^2$  and are uncorrelated. In addition we usually make the normality assumption that the  $\epsilon_{ij}$ s are independent normal random variables. Together these mean that

$$y_{ij} \sim N(\mu + \alpha_i, \sigma^2)$$

and are independent.

We may note the following:

1. If treatment acts multiplicatively on the response, where  $y_{ij} = \mu\alpha_i\epsilon_{ij}$  (where in this kind of model  $\epsilon_{ij} > 0$  for all  $i$  and  $j$ ), then the variance of  $y_{ij}$  is non-constant, but by taking logarithms of the response, the one way model may then be valid. (The model assumes the variance is the same for each treatment group.)
2. Units, even in the same treatment group, are assumed to be uncorrelated.
3. The original response may need to be transformed to achieve normality.

Within this framework, some of the questions that we might want to answer include

- (a) estimating the  $\mu_i$ s, or equivalently  $\mu$  and the individual  $\alpha_i$ s,

- (b) estimating linear combinations of the  $\mu_i$ s or  $\alpha_i$ s, especially *contrasts*: linear combinations of the form  $\sum_i c_i \alpha_i$  where  $\sum_i c_i = 0$ ,
- (c) testing equality of all the  $\mu_i$ s, or equivalently the hypothesis  $H_0 : \alpha_1 = \dots = \alpha_r = 0$ .

Theoretical treatments of the analysis of variance often focus unduly on objective (c), though it is important to remember that this is rarely an interesting objective in its own right. It may well be a necessary preliminary to some more meaningful question such as determining which of several drugs or agricultural varieties is the best, or whether any of a number of engineering manufacturers is differing unacceptably from a predefined standard.

We can write the model in matrix form  $y = X\beta + \epsilon$ ,

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{rn_r} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_r \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1n_1} \\ \epsilon_{21} \\ \epsilon_{22} \\ \vdots \\ \epsilon_{rn_r} \end{pmatrix}.$$

We can estimate the unknown parameters  $\beta = (\mu, \alpha_1, \alpha_2, \dots, \alpha_r)$  by least squares or equivalently by maximum likelihood under the normality assumption by solving the normal equations

$$(X^T X)\hat{\beta} = X^T y.$$

These are  $(r + 1)$  equations in  $(r + 1)$  unknowns. We have

$$X^T X = \begin{pmatrix} n & n_1 & n_2 & \dots & n_r \\ n_1 & n_1 & 0 & \dots & 0 \\ n_2 & 0 & n_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_r & 0 & 0 & \dots & n_r \end{pmatrix},$$

$$X^T y = \begin{pmatrix} \sum_i \sum_j y_{ij} \\ \sum_j y_{1j} \\ \sum_j y_{2j} \\ \vdots \\ \sum_j y_{rj} \end{pmatrix} = \begin{pmatrix} G \\ T_1 \\ T_2 \\ \vdots \\ T_r \end{pmatrix},$$

where  $G$  is the grand total and  $T_i$  is the total for the  $i$ th treatment.

The normal equations are therefore

$$n\hat{\mu} + \sum_{i=1}^r n_i \hat{\alpha}_i = \sum_i \sum_j y_{ij}, \quad (8.2)$$

$$n_i \hat{\mu} + n_i \hat{\alpha}_i = \sum_j y_{ij} \quad i = 1, 2, \dots, r. \quad (8.3)$$

The equations are not independent as the sum over  $i$  in (8.3) equals the total in (8.2), so there are only  $r$  equations for  $r + 1$  unknowns and there are infinitely many solutions. However, this does not prevent the model being estimated, because the fitted values are always the same whichever solution is chosen and hence the residuals and residual sum of squares are the same.

In order to solve the equations we have to add another equation or constraint on the estimates. Consider three possible constraints:

1. Applying the constraint  $\sum_i n_i \alpha_i = 0$  implies that  $\mu$  equals the overall mean,

$$\hat{\mu} = \bar{y} = \frac{G}{n} = \frac{\sum_i \sum_j y_{ij}}{n}.$$

Hence  $\alpha_i$  is estimated by

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$$

where  $\bar{y}_{i.} = \sum_j y_{ij}/n_i$  and  $\bar{y}_{..} = \sum_i \sum_j y_{ij}/n$ .

2. A second method is to set  $\mu = 0$ , so that

$$\hat{\alpha}_i = T_i/n_i = \bar{y}_{i.}$$

3. Take  $\hat{\alpha}_1 = 0$ . This is the solution adopted by some statistical packages e.g. GLIM. It follows that  $\hat{\mu} = \bar{y}_{1.}$  and hence that

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{1.} \quad \text{for } i = 2, 3, \dots, n.$$

For each of the three solutions the fitted values (estimated means) are

$$\hat{\mu} + \hat{\alpha}_i = \bar{y}_{i.}$$

for a unit in the  $i$ th treatment group. So the fitted values are identical, as are the residuals,

$$e_{ij} = y_{ij} - \bar{y}_{i.},$$

the residual sum of squares (*SSE* or deviance) and hence  $s^2$ . The residual sum of squares is given by

$$SSE = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2.$$

### 8.2.1 The ANOVA Table

As in Chapters 2 and 3, a convenient way to represent the results of an analysis of variance is through a table known as the *ANOVA table*. Although this is mathematically just a special case of the general development in Section 3.5, it is useful to re-derive the results directly for this model.

The total (corrected) sum of squares is

$$SSTO = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2,$$

which is also known as the *deviance* after fitting the null model  $y_{ij} = \mu + \epsilon_{ij}$ , i.e. a model with no treatment effects,  $\alpha_1 = \alpha_2 = \dots = \alpha_r = 0$ . The total sum of squares  $SSTO$  can be decomposed:

$$\begin{aligned} & \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 \\ &= \sum_i \sum_j [(y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..})]^2 \\ &= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i n_i (\bar{y}_{i.} - \bar{y}_{..})^2 + 2 \sum_i (\bar{y}_{i.} - \bar{y}_{..}) \sum_j (y_{ij} - \bar{y}_{i.}) \\ &= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i n_i (\bar{y}_{i.} - \bar{y}_{..})^2 \end{aligned}$$

since

$$\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.}) = 0.$$

Hence

$$SSTO = SSE + SSTR$$

where  $SSE$  is the deviance under the full model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}.$$

The  $SSTR$  is the “sum of squares due to treatments”, analogous to  $SSR$  in an ANOVA table for regression. It is the difference in deviances between the null model and the full model, in other words, the extra sum of squares due to the treatment effects or the increase in deviance when we assume  $H_0 : \alpha_1 = \dots = \alpha_r = 0$  to be true. This is the extra sum of squares due to  $H_0$ . In order to test  $H_0$ , note that

$$E\{SSE\} = E\left\{\sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2\right\}$$

$$\begin{aligned}
&= \sum_i E\left\{\sum_j (y_{ij} - \bar{y}_{i\cdot})^2\right\} \\
&= \sum_i (n_i - 1)\sigma^2 \\
&= (n - r)\sigma^2
\end{aligned}$$

and hence

$$s^2 = \frac{SSE}{n - r}$$

is an unbiased estimate of  $\sigma^2$ . Note that the divisor is total number of observations minus number of independent parameters, which is also the total number of degrees of freedom when the treatment means are unconstrained. Under the null model  $y_{ij} = \mu + \epsilon_{ij}$ , all treatments have the same effect. If  $H_0$  is true

$$E\{SSTO\} = (n - 1)\sigma^2$$

and

$$\begin{aligned}
E\{\text{extra SS due to treatments}\} &= (n - 1)\sigma^2 - (n - r)\sigma^2 \\
&= (r - 1)\sigma^2
\end{aligned}$$

Hence if  $H_0$  is true

$$\frac{SSTR}{r - 1}$$

is an unbiased estimate of  $\sigma^2$ . Under the normality assumption it can be shown that

$$\frac{SSTR}{\sigma^2} \sim \chi_{r-1}^2$$

and

$$\frac{SSE}{\sigma^2} \sim \chi_{n-r}^2$$

and that they are independent. Hence

$$F = \frac{SSTR/(r - 1)}{SSE/(n - r)} = \frac{\text{Treatment Mean Square}}{s^2} \sim F_{r-1, n-r} \quad (8.4)$$

under  $H_0$ . So we reject  $H_0$  (and hence conclude that there are significant differences among the treatments) if  $F > F_{r-1, n-r; 1-\alpha}$ .

The whole information may be summarized in the form of an *analysis of variance* (ANOVA) table as in Table 8.1.

Source	DF	SS	MS	F
Treatments	$r - 1$	$\sum_i \frac{T_i^2}{n_i} - \frac{G^2}{n}$	$\frac{SS}{r-1}$	$\frac{SS/(r-1)}{s^2}$
Residual	$n - r$	by subtraction	$s^2$	
Total	$n - 1$	$SSTO = \sum_i \sum_j y_{ij}^2 - \frac{G^2}{n}$		

Table 8.1 ANOVA table for one-way analysis of variance.

### 8.2.2 Testing equality of variances

The preceding analysis has assumed that the variances are the same in all samples. If the variances are allowed to be completely arbitrary, then there is no exact procedure to test equality of the means — in the case  $r = 2$ , this is just the Behrens-Fisher problem discussed in Appendix A. However, it is advisable to test for the equality of variances. We describe here how to construct a likelihood ratio test for this.

Define  $S_i = \sum_j (y_{ij} - \bar{y}_{i.})^2$  for  $i = 1, \dots, r$ . Assume that  $y_{ij} \sim N(\mu_i, \sigma_i^2)$  independently for all  $i$  and  $j$ , and consider testing  $H_0 : \sigma_1^2 = \dots = \sigma_r^2$  against the alternative  $H_1$  that imposes no constraints on the  $\{\sigma_i^2\}$ . The likelihood  $L$ , maximized with respect to  $\{\mu_i\}$ , is given by

$$L \propto \prod_{i=1}^r \left\{ \left( \frac{1}{\sigma_i^2} \right)^{n_i/2} \exp \left( -\frac{S_i}{2\sigma_i^2} \right) \right\}.$$

This is maximized with respect to  $\sigma_i^2$  by setting  $\hat{\sigma}_i^2 = S_i/n_i$ . Hence the maximized likelihood under  $H_1$  is given, up to a constant of proportionality, by

$$L_1 = \prod_{i=1}^r \left\{ \left( \frac{1}{\hat{\sigma}_i^2} \right)^{n_i/2} e^{-n_i/2} \right\}.$$

Under  $H_0$ , we replace  $\sigma_i^2$  by a common  $\sigma^2$  in  $L$ , and maximize with respect to  $\sigma^2$ . This leads to  $\hat{\sigma}^2 = \sum_i S_i/N$ , and the resulting likelihood given by

$$L_0 = \left( \frac{1}{\hat{\sigma}^2} \right)^{N/2} e^{-N/2}.$$

Consequently the log likelihood ratio statistic is

$$T = 2 \log \left( \frac{L_1}{L_0} \right) = \sum_{i=1}^r n_i \log \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_i^2} \right). \quad (8.5)$$

Under  $H_0$ , the approximate distribution of this statistic is

$$T \sim \chi_{r-1}^2.$$

### 8.2.3 Bartlett's modification

Bartlett (1937), in the earliest example of what is now known as the Bartlett correction, proposed a modification of this test to improve the  $\chi^2$  approximation.

The steps in this modification are:

1. Replace  $n_i$  by  $n_i - 1$  everywhere, so that  $\hat{\sigma}_i^2 = S_i/(n_i - 1)$ ,  $\hat{\sigma}^2 = \sum_i S_i/(N - r)$  and replace  $n_i$  by  $n_i - 1$  in (8.5),
2. With the definition of  $T$  thus modified, the distributional approximation is

$$\left\{ 1 + \frac{1}{3(r-1)} \sum_{i=1}^r \left( \frac{1}{n_i - 1} - \frac{1}{N - r} \right) \right\}^{-1} T \sim \chi_{r-1}^2.$$

Hence reject  $H_0$  at level  $\alpha$  if  $T > \chi_{r-1; 1-\alpha}^2$ , the upper- $\alpha$  quantile of the  $\chi_{r-1}^2$  distribution.

### 8.2.4 Two examples.

(i) *The PEMA data.*

Our first example is based on Streete *et al.* (1986).

Table 8.2 gives the serum levels of 2-ethyl-2-phenylmalonamide (PEMA) in patients receiving anticonvulsant medication. The aim is to measure PEMA levels in patients who are prescribed primidone either singly or in combination with other anticonvulsants. In this example we shall only consider patients who had similar dose levels of each of the four drug combinations listed below. The first group were given primidone alone, the second primidone and phenobarbitone, the third primidone and phenytoin and the last a combination of all three, primidone, phenobarbitone and phenytoin. We shall fit a one-way model and look for evidence to indicate real differences among the four different drug combinations.

DRUG			
1	2	3	4
9.9	8.8	4.9	11.0
8.6	11.8	6.7	8.3
6.0	6.6	5.6	12.6
1.2	26.6	7.2	7.8
4.2	27.0	10.1	5.9
10.5	13.4	11.1	8.8
4.1	37.2	2.8	8.6

**Table 8.2** PEMA data.



Quick inspection of the data suggests a highly skewed distribution in some of the groups, so a logarithmic transformation was taken to improve the fit to a normal distribution.

The ANOVA table for these transformed data is in Table 8.3.

Source	DF	SS	MS	$F$
Treatments	3	4.883	1.628	5.142
Residual	24	7.598	0.317	
Total	27	12.481		

**Table 8.3** ANOVA table for the logged PEMA data.

From tables we find  $F_{24,3;0.99} = 4.718$  and therefore there is very strong evidence for differences between the treatments which in this case are the drugs (the  $p$ -value is .007).

Table 8.4 gives the fitted values and residuals.

DRUGS			
1	2	3	4
1.659+0.634	2.760-0.585	1.848-0.259	2.172+0.226
1.659+0.493	2.760-0.292	1.848+0.054	2.172-0.056
1.659+0.134	2.760-0.873	1.848-0.125	2.172+0.362
1.659-1.477	2.760+0.521	1.848+0.126	2.172-0.118
1.659-0.224	2.760+0.536	1.848+0.465	2.172-0.397
1.659+0.692	2.760-0.165	1.848+0.559	2.172+0.003
1.659-0.248	2.760+0.856	1.848-0.818	2.172-0.020

**Table 8.4** Fitted values and residuals for the PEMA data.

For example, for drug 1 the average of log PEMA values is 1.659 and the residuals are  $\log(9.9)-1.659=0.634$ ,  $\log(8.6)-1.659=0.493$ , etc. Note that from the ANOVA table  $s^2 = 0.317$  so that  $s = 0.563$ . The largest crude residual in modulus is 1.477, for a  $t$  statistic of 2.623. Based on the nominal  $t_{24}$  distribution, the two-sided  $p$ -value associated with that (in other words,  $2\Pr\{T > 2.623\}$  when  $T \sim t_{24}$ ) is .0149, not especially surprising considering that this is the largest of 28 values.

We could plot the residuals (or the standardized residuals) against both the fitted values and the treatment group number to check for homogeneity of variance. Using Bartlett's modified test produces the values in Table 8.5. Here  $n = 28$  and  $t = 4$ . We find that  $T = 6.66$ . With a null distribution of  $\chi_3^2$ , this has a  $p$ -value of .084, so there is some slight evidence against  $H_0$ .

$\bar{y}_i$	1.659	2.760	1.848	2.172
$s_i^2$	0.572	0.418	0.217	0.059
$n_i$	7	7	7	7

Table 8.5 Bartlett's test for the PEMA data

## (ii) Round-robin data.

Round-robin tests are tests performed under supposedly the same conditions in a series of laboratories. They play an important part in the standardization of measurement procedures in physics and engineering. However, the tests are often expensive to perform, so the results of such studies are often small unbalanced sets of data.

Table 8.6 is based on an actual example of such a study into the measurement of the rate of creep rupture. Samples of a material were sent to 11 laboratories which were asked to perform repeat measurements. Tabulated are  $n_i$ , the number of tests performed at laboratory  $i$ , and the mean and standard deviation (S.D.) of the measurements made at laboratory  $i$ . Also computed in column 5 is  $S_i = (n_i - 1)(\text{S.D.})^2$ . The last two columns will be explained a little further on.

Laboratory $i$	$n_i$	Mean	S.D.	$S_i$	$\hat{\alpha}_i$	S.E.
1	5	102.1	48.1	9254.44		
2	9	92.8	8.3	551.12	3.62	2.06
3	4	97.2	8.6	221.88	8.02	3.28
4	5	79.9	9.2	338.56	-9.28	2.90
5	5	87.0	4.8	92.16	-2.18	2.90
6	5	93.1	5.5	121.00	3.92	2.90
7	5	82.2	4.4	77.44	-6.98	2.90
8	6	54.9	1.9	18.05		
9	5	94.0	8.3	275.56	4.82	2.90
10	5	90.4	2.2	19.36	1.22	2.90
11	5	84.7	5.7	129.96	-4.48	2.90

Table 8.6 Round-robin test data

It can be seen immediately that the S.D. for laboratory 1 is much larger than the others, and an application of (8.5) immediately confirms that this is significant: the likelihood ratio statistic is 103.3 with a nominal  $\chi_{10}^2$  distribution. The 0.0001% point of  $\chi_{10}^2$  is 46.9. In fact, the raw data for laboratory 1 showed one extreme outlier which was not deleted from the

reported data, almost certainly due to a recording error. If we ignore the apparent discrepancy in variances and compute the  $F$  statistics (8.4) for equality of means, the  $F$  value is 3.86, which is highly significant (with a nominal distribution of  $F_{10,48}$ , the  $p$ -value is .0007). However, in this case the discrepancy in variances is to some extent masking differences in the means, and to avoid such a bias in other comparisons, we delete laboratory 1 for the rest of the discussion.

To proceed with the analysis, deleting laboratory 1 and repeating the tests for the other 10 laboratories, (8.5) yields the test statistic 22.6, nominally  $\chi_9^2$ , which is still high (significant at 1% — actual  $p$ -value 0.007) but not nearly as bad as before. In this case, the  $F$  statistic in (8.4) is 20.2, very highly significant against the nominal  $F_{9,44}$  distribution (theoretical  $p$ -value about  $10^{-13}$ ). Further inspection of the data indicates that laboratory 8 has a much lower mean than the others, and also the lowest standard deviation. Omission of laboratory 8 (as well as laboratory 1) leads to the test statistics 13.2 in (8.5), 3.67 in (8.4). The value 13.2 is not significant against  $\chi_8^2$  at the 10% level, so seems acceptable for the equality of variances.

If the likelihood ratio statistics (8.5) are modified by the Bartlett correction, they become 77.9 for the full model with all 11 laboratories, 16.9 for the model with laboratory 1 omitted, and 11.4 for the model with laboratories 1 and 8 both omitted. The modification in the first case is substantial but irrelevant. The value 16.9 is almost exactly at the 5% point for  $\chi_9^2$ , so the effect of the modification here is to conclude that the assumption of equality of variances might just be acceptable in this case (with laboratory 1 omitted), though since it is clear that laboratory 8 is discrepant in its mean value, this does not affect our overall conclusions. The correction from 13.2 to 11.4 in the third model strengthens the conclusion that the variances can be assumed equal in this case. Thus the Bartlett correction does not change any of our conclusions, but it does turn out to be numerically not negligible. In general it can be recommended that, in experiments of this nature when the individual  $n_i$  are small, it is worthwhile to apply the Bartlett correction.

For the final model, with laboratories 1 and 8 both omitted, an ANOVA table is shown in Table 8.7. The final  $F$  statistic for the hypothesis of no treatment effect is 3.66, for a  $p$ -value of .003. Thus, we conclude that even when laboratories 1 and 8 are omitted, there is still a significant difference among the remaining 9 laboratories.

The last two columns of Table 8.6 show the estimates  $\hat{\alpha}_i$  and their standard errors for all laboratories except 1 and 8. To compute the standard errors, note that

Source	DF	SS	MS	F
Treatments	8	1374.0	171.75	3.66
Residual	39	1827.0	46.85	
Total	47	3201.0		

**Table 8.7** One way ANOVA table for the round-robin test data with laboratories 1 and 8 omitted.

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..} = \left( \frac{1}{n_i} - \frac{1}{N} \right) \sum_j y_{ij} - \frac{1}{N} \sum_{i' \neq i} \sum_j y_{i'j}$$

and consequently

$$\begin{aligned} \text{Var}(\hat{\alpha}_i) &= \left( \frac{1}{n_i} - \frac{1}{N} \right)^2 \sigma^2 n_i + \left( \frac{1}{N} \right)^2 \sigma^2 (N - n_i) \\ &= \sigma^2 \frac{N - n_i}{N n_i} \end{aligned}$$

so the standard error of  $\hat{\alpha}_i$  may be computed as  $s\sqrt{(N - n_i)/(N n_i)}$  where  $s^2$  is the usual unbiased estimate of  $\sigma^2$ . From Table 8.7 this is calculated as  $s^2 = 46.85$  and hence  $s = 6.84$ , and this is used to calculate the standard errors in Table 8.6.

In the light of these calculations we can now see that the treatment effects are significantly negative in laboratories 4 and 7, and significantly positive in laboratory 3, while the others are not significantly different from 0.

The overall conclusion therefore is that laboratories 1 and 8 are gross outliers, while laboratories 3, 4 and 7 are also producing results significantly different from the remainder. Follow-up studies would be expected to identify the reasons for this — for example, it is possible that some of the laboratories were using a different measurement technique from the others, and the conclusion in that case would be that the different techniques could not be regarded as interchangeable.

### 8.2.5 Multiple Comparisons

If we have not pre-planned any specific comparisons between treatments and there is evidence for differences we may still wish to investigate where these differences lie. We shall describe three methods

(i) *Least significant differences (LSD)*

Suppose we have a balanced experiment so that  $n = rt$ . A general procedure for paired comparisons is to compute a “least significant difference”

$$LSD = t_{\nu;0.025} s \sqrt{\frac{2}{t}},$$

where  $\nu = n - r$  and  $s\sqrt{\frac{2}{t}}$  is the estimated standard error of the difference between two treatment means.

This gives a lower bound for a significant difference between pairs, so that if a preplanned pair of means differs by more than  $LSD$  then there is evidence to contradict the hypothesis that the two treatment effects are the same. Note that there are  $r(r-1)/2$  pairs of treatments, so even if there are no differences in treatment effects, approximately 5% of differences between pairs will exceed  $LSD$ . We should use  $LSD$  sparingly!

(ii) *Scheffé's method*

Suppose we have an arbitrary set of treatments for which the treatment sum of squares is significant (i.e. we reject a null hypothesis that the treatment means are all equal), but there are no specific comparisons (or contrasts) for which we planned in advance to test. In this context, it is natural to try to find *simultaneous*  $100(1 - \alpha)\%$  confidence intervals for *all* possible contrasts. This raises issues similar to the discussion of simultaneous confidence intervals in Sections 2.8 and 3.6. A contrast  $\sum_{i=1}^r a_i \alpha_i = a^T \alpha$  is estimated by  $a^T \hat{\alpha}$  or  $\sum_{i=1}^r a_i \bar{y}_i$  with variance  $\sum_{i=1}^r a_i^2 \frac{\sigma^2}{n_i}$ . Hence in the case of a preplanned comparison a  $100(1 - \alpha)\%$  confidence interval for this contrast would be

$$\sum_{i=1}^r a_i \bar{y}_i \pm t_{n-r; \frac{1}{2}\alpha} s \sqrt{\sum_{i=1}^r \frac{a_i^2}{n_i}}$$

and if this confidence interval includes the value zero, then this contrast is 'not significant', i.e. there is no evidence against the hypothesis

$$H_0 : \sum_{i=1}^r a_i \alpha_i = 0.$$

An adaptation of Scheffé's method (Section 3.8) to this scenario shows that the set of all confidence intervals of the form

$$\sum_{i=1}^r a_i \bar{y}_i \pm s \sqrt{\sum_{i=1}^r \frac{a_i^2}{n_i} \sqrt{(r-1)F_{r-1, n-r; \alpha}}}$$

contain the true values ( $\sum_{i=1}^r a_i \alpha_i$ ) with joint probability  $1 - \alpha$  (so are 'simultaneous' confidence intervals).

A useful property of Scheffé's method is that at least one contrast will be 'significant' (in other words, the confidence interval will exclude zero) if

the treatment sum for squares of the standard  $F$  test is significant at the same level of significance  $\alpha$ .

(iii) *Tukey's method*

This method gives simultaneous  $100(1 - \alpha)\%$  confidence intervals for all contrasts

$$\sum_{i=1}^r a_i \bar{y}_i \pm s \sum_{i=1}^r \frac{|a_i|}{2\sqrt{n_i}} q(\alpha, r, \nu)$$

where  $q(\alpha, r, \nu)$  represents the percentage points of the distribution of the studentized' range statistic

$$\frac{1}{s^2} \left\{ \max_i (\bar{y}_i) - \min_i (\bar{y}_i) \right\}$$

for which several published tables are available (e.g. Harter 1960, Rohlf and Sokal 1995). Here  $\nu = n - r$  the degrees of freedom in this case.

For paired comparisons with balanced design the three methods give us the following intervals:

$$\text{Scheffé : } \bar{y}_i - \bar{y}_j \pm s \sqrt{\frac{2}{t} (r - 1) F_{r-1, n-r; \alpha}},$$

$$\text{Tukey : } \bar{y}_i - \bar{y}_j \pm \frac{s}{\sqrt{t}} q(\alpha, r, n - r),$$

$$\text{LSD : } \bar{y}_i - \bar{y}_j \pm s \sqrt{\frac{2}{t} t_{n-r, \alpha/2}}.$$

A method of illustrating the comparisons is to write the means in order on a scaled line and underline pairs not significantly different from each other. For the PEMA data, the Scheffé interval width is  $\pm 0.904$ , Tukey is  $\pm 0.970$  and LSD is  $\pm 0.621$ . The four drug means are 1.659, 2.760, 1.848 and 2.172 for drugs 1 to 4 respectively. The only two drugs which have a difference larger than these values is obtained when comparing drugs 1 and 2 where the difference is 1.101. Hence using the methods of multiple comparisons suggests that only treatments 1 and 2 are significantly different.

### 8.3 The Two-Way Layout

In the one-way layout, there is only one factor that varies between the groups. A more complicated and typical situation, however, is that there is more than one factor. For example, in the round-robin example discussed in section 8.2.4, it is possible that the different laboratories, instead of simply repeating the same experiment a number of times, were in fact asked to perform a fixed sequence of experiments, the objective being to determine whether there was any overall difference among the laboratories. Another

example is in clinical trials. Suppose a trial is set up to compare two drugs. If the patients are simply assigned at random to the two drugs, it is likely that the final distribution of ages, sexes and other prognostic variables such as disease condition, will be quite different between the two groups. This could seriously bias the results. A better way is first to stratify, i.e. divide the patients into subgroups of roughly comparable individuals, and then assign the two drugs randomly within each subgroup (or stratum).

In general, there could be many different factors that have to be taken into account, and a complex arrangement could be necessary to do that. The two-way layout refers to a set-up in which there are just two kinds of effect, a *treatment effect* which is the main variable of interest, and a *block effect* which is not of direct interest but which would bias the results if it were not taken into account. In the round-robin example, assuming that it is still the differences among laboratories that are of interest, the treatment effect is the laboratory and the block effect is an individual experiment within a laboratory. For a clinical trial, the treatment effect is usually the treatment itself, e.g. the type of drug, while the block effect refers to the factors such as the age, sex and clinical condition of the patients, which determine the subdivision into strata.

We consider only a balanced experiment, in which the same number  $t$  of observations are collected for each treatment-block combination, known as a randomized block design. Initially, we only consider the case  $t = 1$ . Assume there are  $r$  treatments and  $c$  blocks, and let  $y_{ij}$  denote the observation for treatment  $i$  and block  $j$ . A plausible model for this situation is

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad 1 \leq i \leq r, \quad 1 \leq j \leq c, \quad (8.6)$$

in which  $\mu$  is an overall mean and  $\alpha_i$  and  $\beta_j$  are respectively a treatment effect and a block effect. The constraints on these are

$$\sum_i \alpha_i = 0, \quad \sum_j \beta_j = 0, \quad (8.7)$$

and we make the usual assumption that the  $\{\epsilon_{ij}\}$  are independent  $N(0, \sigma^2)$ . Typically the  $\{\beta_j\}$  are thought of as nuisance parameters and the real interest is still in the  $\{\alpha_i\}$ . The hypothesis of no treatment effect is still formulated as

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_r = 0 \quad (8.8)$$

but now we want to test this without making any implicit or explicit assumption that the  $\{\beta_j\}$  are also 0. We can also test the hypothesis of no block effect which is formulated as

$$H'_0 : \beta_1 = \beta_2 = \dots = \beta_c = 0. \quad (8.9)$$

### 8.3.1 Estimates of the treatment and block effects

If we define  $\bar{y}_i = \sum_j y_{ij}/c$  to be the  $i$ 'th treatment mean,  $\bar{y}_{.j} = \sum_i y_{ij}/r$  to be the  $j$ 'th block mean, and  $\bar{y}_{..} = \sum_i \sum_j y_{ij}/(rc)$  the overall mean, then the appropriate estimates are obtained by minimizing

$$S = \sum_{i=1}^r \sum_{j=1}^c (y_{ij} - \mu - \alpha_i - \beta_j)^2,$$

and this is achieved by differentiating with respect to  $\alpha_i$ ,  $i = 1, \dots, r$  and  $\beta_j$ ,  $j = 1, \dots, c$ . There are  $r+c+1$  equations, but only  $r+c-1$  independent equations. So there are an infinite number of solutions.

One possible solution is to assume that  $\sum_{i=1}^r \hat{\alpha}_i = 0$  and  $\sum_{j=1}^c \hat{\beta}_j = 0$  and hence

$$\hat{\mu} = \bar{y}_{..}, \quad \hat{\alpha}_i = \bar{y}_i - \bar{y}_{..}, \quad \hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$$

The fitted values (estimated means) are

$$\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{y}_i + \bar{y}_{.j} - \bar{y}_{..}$$

For all solutions the residual sum of squares

$$\begin{aligned} S_{\min} &= \sum_{i=1}^r \sum_{j=1}^c (y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2 \\ &= \sum_{i=1}^r \sum_{j=1}^c (y_{ij} - \bar{y}_i - \bar{y}_{.j} + \bar{y}_{..})^2 \end{aligned}$$

As in the one-way layout (with  $N = rc$ ,  $n_i = c$ ), we may calculate the variance of  $\hat{\alpha}_i$  to be  $(r-1)\sigma^2/(rc)$ , and similarly variance of  $\hat{\beta}_j$  is  $(c-1)\sigma^2/(rc)$ . For a contrast of the form  $\sum_i c_i \alpha_i$  ( $\sum_i c_i = 0$ ) we have the estimator  $\sum_i c_i \hat{\alpha}_i = \sum_i c_i \bar{y}_i$ , which has variance  $\sigma^2 \sum_i c_i^2/c$ . Orthogonal contrasts can be used as before for preplanned comparisons of both treatments and blocks when either of these two tests show evidence for differences. Otherwise the methods of multiple comparisons (e.g. Scheffé) can be used.

### 8.3.2 The ANOVA table

Consider the decomposition

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$$



$$\begin{aligned}
&= \sum_i \sum_j \{(y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..}) + (\bar{y}_i - \bar{y}_{..}) + (\bar{y}_j - \bar{y}_{..})\}^2 \\
&= \sum_i \sum_j (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..})^2 + c \sum_i (\bar{y}_i - \bar{y}_{..})^2 + r \sum_j (\bar{y}_j - \bar{y}_{..})^2 \\
&= \sum_i \sum_j (y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2 + c \sum_i \hat{\alpha}_i^2 + r \sum_j \hat{\beta}_j^2
\end{aligned}$$

where as in earlier ANOVA calculations, the cross-product terms all turn out to be 0. The final result may also be written in the form

$$SSTO = SSE + SSTR + SSB$$

as a decomposition of the total sum of squares  $SSTO$  into a sum of squares due to error (or residual)  $SSE$ , a sum of squares due to treatment  $SSTR$  and a sum of squares due to blocks  $SSB$ . The corresponding partition of the degrees of freedom is

$$rc - 1 = (r - 1)(c - 1) + (r - 1) + (c - 1).$$

Hence we derive the ANOVA table in Table 8.8.

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE
Treatments	$SSTR$	$r - 1$	$SSTR/(r - 1)$
Blocks	$SSB$	$c - 1$	$SSB/(c - 1)$
Residual	$SSE$	$(r - 1)(c - 1)$	$SSE/\{(r - 1)(c - 1)\}$
Total	$SSTO$	$rc - 1$	

**Table 8.8** Two-way ANOVA table.

To test the hypothesis (8.8) we use the  $F$  statistic

$$\frac{SSTR/(r - 1)}{SSE/\{(r - 1)(c - 1)\}} = \frac{(c - 1)SSTR}{SSE} \sim F_{r-1, (r-1)(c-1)} \text{ under } H_0.$$

Similarly to test the hypothesis (8.9) we use the  $F$  statistic

$$\frac{SSB/(c - 1)}{SSE/\{(r - 1)(c - 1)\}} = \frac{(r - 1)SSB}{SSE} \sim F_{c-1, (r-1)(c-1)} \text{ under } H'_0.$$

*Example - Tensile properties of an alloy.*

A quantity of an alloy was prepared under careful processing conditions to attain precision and homogeneity of its properties. The alloy was rolled

Bar	Laboratories						Mean
	1	2	3	4	5	6	
1	3.06	3.01	2.84	3.48	3.24	2.42	3.01
2	2.59	2.75	3.13	3.58	2.35	2.58	2.83
3	3.32	2.93	3.27	3.80	3.55	2.77	3.27
4	3.08	2.87	2.76	2.94	2.24	2.58	2.74
Mean	3.01	2.89	3.00	3.45	2.84	2.59	2.96

**Table 8.9** Tensile properties of four alloy bars measured at six laboratories

into bars and lengths cut from four of these bars were sent to six laboratories. The analysis concerned a particular tensile property. The data given in Crowder (1992), form a balanced layout in which each laboratory conducts a test on only one length from each of the four bars, and are given in Table 8.3.2. To calculate the ANOVA table by hand, we need

$$\begin{aligned}
 SSTO &= \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 \\
 &= \sum_i \sum_j y_{ij}^2 - cr\bar{y}_{..}^2 \\
 &= 214.6906 - 24 \times 2.9642^2 = 3.819, \\
 SST &= c \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \\
 &= c \sum_i \bar{y}_{i.}^2 - cr\bar{y}_{..}^2 \\
 &= 212.4756 - 24 \times 2.9642^2 = 1.605, \\
 SSB &= r \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 \\
 &= r \sum_j \bar{y}_{.j}^2 - cr\bar{y}_{..}^2 \\
 &= 211.8522 - 24 \times 2.9642^2 = 0.981, \\
 SSE &= SSTO - SST - SSB = 1.233.
 \end{aligned}$$

The complete ANOVA table is therefore given in Table 8.10. The respective  $F$ -tests that both the laboratories and bars are significant with resultant  $p$ -values of 0.018 and 0.029 respectively. Residual plots can be done to check the model assumptions.

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE	F
Laboratories	1.605	5	0.321	3.90
Bars	0.981	3	0.327	3.98
Residual	1.233	15	1.233	
Total	3.819	23		

Table 8.10 ANOVA table for tensile property data.

### 8.4 The two-way layout with interaction

The analysis in section 8.3 relies on the assumption that (8.6) is an appropriate model. This is known as an *additivity assumption*, since it implies that one can deduce the combined effect of treatment and block simply by adding up the separate effects for the two. One consequence of this assumption, for example, is that it implies that differences between treatments are the same for all blocks: if drug 1 is better than drug 2 on one group of patients, then it is also better (by the same amount) on any other group of patients. Such an assumption is often good enough as a first approximation but may not be valid in general. Indeed one may even argue that detecting *interactions* is sometimes the main purpose of the analysis. An example of an interaction would be a statement that a certain drug is particularly effective in comparison with another drug with one group of patients, but not necessarily with other groups.

This implies that we might want to consider a more general model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad 1 \leq i \leq r, 1 \leq j \leq c, 1 \leq k \leq t, \quad (8.10)$$

subject to the constraints (8.7) plus

$$\begin{aligned} \sum_j \gamma_{ij} &= 0 \text{ for each } i, \\ \sum_i \gamma_{ij} &= 0 \text{ for each } j. \end{aligned}$$

In this model we are assuming that there are  $t > 1$  observations for each treatment-block combination. The total number of parameters for the treatment-block means is 1 for  $\mu$ ,  $r - 1$  for the treatment effects (i.e.  $r$  treatment effects subject to 1 constraint),  $c - 1$  for the block effects and  $(r - 1)(c - 1)$  for the interactions, a total of  $rc$ . This means that any combination of the  $rc$  treatment-block means may be handled within the framework of (8.10). However, if  $t = 1$  then there are no degrees of freedom left over to estimate  $\sigma^2$ , hence our need to assume  $t > 1$ .

As in previous notation we use a dot to indicate averaging over the dotted parameter - thus  $\bar{y}_{i..}$  is the mean of all observations with treatment

$\bar{y}_{ij}$  is the mean of all observations with the  $i$ 'th treatment and the  $j$ 'th block, and so on. In this notation we have

$$\begin{aligned}\hat{\mu} &= \bar{y}_{...}, \\ \hat{\alpha}_i &= \bar{y}_{i..} - \bar{y}_{...}, \\ \hat{\beta}_j &= \bar{y}_{.j.} - \bar{y}_{...}, \\ \hat{\gamma}_{ij} &= \bar{y}_{ij.} - \bar{y}_{.j.} - \bar{y}_{i..} + \bar{y}_{...}.\end{aligned}$$

The ANOVA decomposition becomes

$$\begin{aligned}& \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2 \\ &= \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2 + t \sum_i \sum_j \hat{\gamma}_{ij}^2 + ct \sum_i \hat{\alpha}_i^2 + rt \sum_j \hat{\beta}_j^2\end{aligned}$$

which may also be written as

$$SSTO = SSE + SSI + SSTR + SSB$$

with degrees of freedom

$$rct - 1 = rc(t - 1) + (r - 1)(c - 1) + (r - 1) + (c - 1).$$

Here *SSI* stands for sum of squares due to interaction. The ANOVA table is Table 8.11.

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE
Treatments	<i>SSTR</i>	$r - 1$	$SSTR/(r - 1)$
Blocks	<i>SSB</i>	$c - 1$	$SSB/(c - 1)$
Interaction	<i>SSI</i>	$(r - 1)(c - 1)$	$SSI/\{(r - 1)(c - 1)\}$
Residual	<i>SSE</i>	$rc(t - 1)$	$SSE/\{rc(t - 1)\}$
Total	<i>SSTO</i>	$rct - 1$	

**Table 8.11** Two-way ANOVA table with interactions.

For example, one could test the hypothesis of no treatment effect with

$$\frac{SSTR/(r - 1)}{SSE/\{rc(t - 1)\}} \sim F_{r-1, rc(t-1)} \text{ under } H_0 : \alpha_i = 0 \text{ for all } i.$$

However, it might also be of interest to test whether there is any interaction, in which case the appropriate *F*-test is based on

$$\frac{SSI/\{(r-1)(c-1)\}}{SSE/\{rc(t-1)\}} \sim F_{(r-1)(c-1), rc(t-1)} \text{ under } H_0 : \gamma_{ij} = 0 \text{ for all } i, j.$$

#### 8.4.1 Tukey's 1-DF test for additivity

Suppose we only have one observation for each treatment-block combination, but we still want to test for interactions. As previously explained, we cannot fit the full model in which all the interactions are unconstrained. Tukey (1949) proposed a solution to this problem. The idea is to assume the model

$$y_{ij} = \mu + \alpha_i + \beta_j + \theta\alpha_i\beta_j + \epsilon_{ij}, \quad 1 \leq i \leq r, \quad 1 \leq j \leq c, \quad (8.11)$$

in which the interaction  $\gamma_{ij}$  takes the specific form  $\theta\alpha_i\beta_j$ . This is unlikely to be the correct model but a test of  $\theta = 0$  should be a useful indication of whether the model is additive or not.

To define a test for this, let us first define

$$z_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..},$$

and consider the model

$$z_{ij} = \theta a_i b_j + e_{ij} \quad (8.12)$$

in which the constants  $a_i$  and  $b_j$  are known and subject to  $\sum_i a_i = \sum_j b_j = 0$ , and the  $e_{ij}$  are errors.

In (8.12), the usual least squares estimation of  $\theta$  (treating the errors as independent) yields

$$\hat{\theta} = \frac{\sum_i \sum_j z_{ij} a_i b_j}{\sum_i a_i^2 \sum_j b_j^2} = \frac{\sum_i \sum_j y_{ij} a_i b_j}{\sum_i a_i^2 \sum_j b_j^2}, \quad (8.13)$$

the two expressions being identical because  $\sum_i a_i = \sum_j b_j = 0$ . From the second expression, it can be seen that the variance of  $\hat{\theta}$  is

$$\frac{\sigma^2}{\sum_i a_i^2 \sum_j b_j^2}.$$

Hence under  $H_0 : \theta = 0$  we have

$$\frac{\hat{\theta}^2 \sum_i a_i^2 \sum_j b_j^2}{\sigma^2} = \frac{(\sum_i \sum_j y_{ij} a_i b_j)^2}{\sigma^2 \sum_i a_i^2 \sum_j b_j^2} \sim \chi_1^2.$$

Consider the ANOVA decomposition

$$\sum_i \sum_j z_{ij}^2 = \sum_i \sum_j (z_{ij} - \hat{\theta} a_i b_j)^2 + \hat{\theta}^2 \sum_i a_i^2 \sum_j b_j^2$$

which we also write as

$$SSI = SSIE + SSG$$

with degrees of freedom

$$(r-1)(c-1) = (rc - r - c) + 1.$$

Let us calculate  $\text{Cov}(z_{ij}, \hat{\theta})$  for given  $i, j$ . From (8.13) we have that

$$\text{Cov}(z_{ij}, \hat{\theta}) = \frac{1}{\sum a_{i'}^2 \sum b_{j'}^2} \sum_{i'} \sum_{j'} a_{i'} b_{j'} \text{Cov}(z_{ij}, y_{i'j'}).$$

However

$$\begin{aligned} \text{Cov}(z_{ij}, y_{i'j'}) &= \text{Cov}(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}, y_{i'j'}) \\ &= \sigma^2 \left( \delta_{ii'} \delta_{jj'} - \frac{1}{c} \delta_{ii'} - \frac{1}{r} \delta_{jj'} + \frac{1}{rc} \right) \end{aligned}$$

where  $\delta$  is the Kronecker delta ( $\delta_{ij} = 1$  if  $i = j$ , 0 otherwise). Hence

$$\begin{aligned} \text{Cov}(z_{ij}, \hat{\theta}) &= \frac{\sigma^2}{\sum a_{i'}^2 \sum b_{j'}^2} \sum_{i'} \sum_{j'} a_{i'} b_{j'} \left( \delta_{ii'} \delta_{jj'} - \frac{1}{c} \delta_{ii'} - \frac{1}{r} \delta_{jj'} + \frac{1}{rc} \right) \\ &= \frac{a_i b_j \sigma^2}{\sum a_{i'}^2 \sum b_{j'}^2} \\ &= a_i b_j \text{Var}(\hat{\theta}). \end{aligned}$$

Thus  $z_{ij} - a_i b_j \hat{\theta}$  is uncorrelated with  $\hat{\theta}$  for each  $i, j$ . Since everything is jointly normal, uncorrelated implies independent and hence the two sums of squares,  $SSIE$  and  $SSG$ , are themselves independent. From this we deduce the  $F$  test, that under  $H_0 : \theta = 0$  we have

$$\frac{SSG}{SSIE/(rc - r - c)} \sim F_{1, rc - r - c}. \quad (8.14)$$

Now comes the key step: we can do all this with the constants  $a_i, b_j$  replaced by  $\hat{\alpha}_i, \hat{\beta}_j$  respectively, and the final result (8.14) is still valid. The reason is simply that all the  $\{z_{ij}\}$  are independent of the  $\{\hat{\alpha}_i\}$  and  $\{\hat{\beta}_j\}$  so that none of the distributional arguments leading to (8.14) are affected by the fact that the  $\{\alpha_i\}$  and  $\{\beta_j\}$  are estimated rather than known.

To summarize, the procedure is as follows. First calculate the  $\{\hat{\alpha}_i\}, \{\hat{\beta}_j\}$

and  $\{z_{ij}\}$ , then compute

$$SSI = \sum_i \sum_j z_{ij}^2, \quad SSG = \frac{(\sum_i \sum_j z_{ij} \hat{\alpha}_i \hat{\beta}_j)^2}{\sum_i \hat{\alpha}_i^2 \sum_j \hat{\beta}_j^2} = \frac{(\sum_i \sum_j y_{ij} \hat{\alpha}_i \hat{\beta}_j)^2}{\sum_i \hat{\alpha}_i^2 \sum_j \hat{\beta}_j^2},$$

and let  $SSIE = SSI - SSG$ . Then 8.14 gives an exact  $F$  test of the additivity of the model.

#### 8.4.2 Example - Fisher data.

The data in Table 8.12 are taken from Fisher (1971), page 68, and give the yields of five varieties of barley in six locations in each of two years, 1931 and 1932. The interest here is presumably in the comparison among different varieties of barley (so “treatment” is variety), and one possible strategy is to treat each place-year combination as a separate block. This suffers from the disadvantage that it may not be possible to detect interactions. A second strategy is to treat the places as blocks and the two years’ of data as separate replications of the whole experiment, producing a  $5 \times 6$  table with two observations per cell. The disadvantage of this is that the residuals may be correlated with year. We shall try both strategies, starting with the second one. (A third strategy might be to consider variety, place and year all as separate factors and to do a three-way analysis, but we shall not consider that.)

Place	year	Manchuria	Svansota	Velvet	Trebi	Peatland	Mean
1	1931	81.0	105.4	119.7	109.7	98.3	102.82
1	1932	80.7	82.3	80.4	87.2	84.2	82.96
2	1931	146.6	142.0	150.7	191.5	145.7	155.30
2	1932	100.4	115.5	112.2	147.7	108.1	116.78
3	1931	82.3	77.3	78.4	131.3	89.6	91.78
3	1932	103.1	105.1	116.5	139.9	129.6	118.84
4	1931	119.8	121.4	124.0	140.8	124.8	126.16
4	1932	98.9	61.9	96.2	125.5	75.7	91.64
5	1931	98.9	89.0	69.1	89.3	104.1	90.08
5	1932	66.4	49.9	96.7	61.9	80.3	71.04
6	1931	86.9	77.1	78.9	101.8	96.0	88.14
6	1932	67.7	66.7	67.4	91.8	94.1	77.54
Column Mean		94.392	91.133	99.183	118.200	102.542	101.09

**Table 8.12** Fisher’s data on barley varieties

Table 8.13 shows the analysis of variance table for this model. The  $F$  statistics are  $1327.5/458.9=2.89$  for varieties,  $4244.2/458.9=9.25$  for places,

SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE
<i>SST</i>	5309.97	4	1327.5
<i>SSB</i>	21220.90	5	4244.2
<i>SSI</i>	4433.02	20	221.7
<i>SSE</i>	13768.46	30	458.9
Total	44732.35	59	

Table 8.13 ANOVA Table for 2-way model with interactions

.15	11.55	19.65	11.25	7.05
-.15	-11.55	-19.65	-11.25	-7.05
23.10	13.25	19.25	21.90	18.80
-23.10	-13.25	-19.25	-21.90	-18.80
-10.40	-13.90	-19.05	-4.30	-20.00
10.40	13.90	19.05	4.30	20.00
10.45	29.75	13.90	7.65	24.55
-10.45	-29.75	-13.90	-7.65	-24.55
16.25	19.55	-13.80	13.70	11.90
-16.25	-19.55	13.80	-13.70	-11.90
9.60	5.20	5.75	5.00	.95
-9.60	-5.20	-5.75	-5.00	-.95

Table 8.14 Residuals from 2-way model

$221.7/458.9=0.48$  for interactions. These are barely significant for a variety effect (the 5% point of  $F_{4,30}$  is 2.69), highly significant for a place effect, and not at all significant for interactions.

However, Table 8.14, which gives the residuals from this model, shows that there is indeed a correlation with year. In nearly every case the yield in 1931 was higher than 1932, except for place 3 where the pattern is curiously reversed. The (linear) correlation coefficient between residual and year is  $-0.525$ , obviously a significant effect.

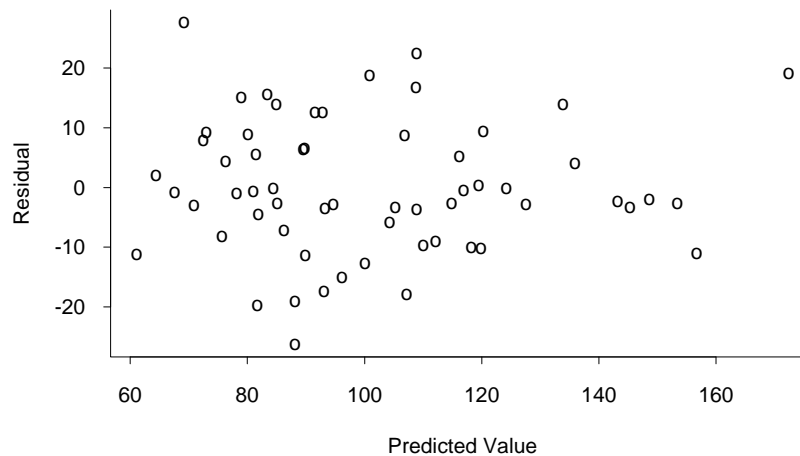
In view of this we try a second analysis, in which each place-year combination is treated as a separate block (12 blocks in all), and Tukey's test is used to test for interactions. The analysis of variance table is given in Table 8.4.2.

The  $F$  ratios are 8.18 for variety, 17.9 for the block effects, and 3.27 for the interaction effect. However the latter is not significant, the 5% point for the  $F_{1,43}$  distribution being about 4.1.



SOURCE	SUM OF SQUARES	D.F.	MEAN SQUARE
<i>SST</i>	5309.97	4	1327.5
<i>SSB</i>	31913.32	11	2901.2
<i>SSG</i>	531.09	1	531.1
<i>SSIE</i>	6977.97	43	162.3
Total	44732.35	59	

**Table 8.15** ANOVA Table for Tukey's 1-DF test



**Figure 8.1** Plot of residuals vs. fitted values for barley data, 2-way model without interactions

Finally, the residuals from this model have been examined with nothing suspicious being observed. As an example, Figure 8.1 shows a plot of the residuals against fitted values.

From this it appears that the second analysis is satisfactory. The problem with the first analysis was that, by failing to treat year as a separate block effect, the estimated variance was inflated because of what was evidently a significant difference between the two years.

## 8.5 Implementation in SAS and S-PLUS

### 8.5.1 ANOVA in SAS

The SAS procedure ANOVA may be used to fit any version of one-way or two-way analysis of variance, and many more different types of models.

Here is a simple example, applied to the PEMA data from Section 8.2.4.

First, we create a data file 'pema.dat', as follows:

```

1  9.9
1  8.6
1  6.0
1  1.2
1  4.2
1 10.5
1  4.1
2  8.8
2 11.8
...
4  8.8
4  8.6

```

Here, the data are arranged in two columns with the drug in the first column and the serum level in the second.

A SAS program to analyze these data is as follows:

```

options ls=64 ps=58;
data pema;
infile 'pema.dat'
input drug serum;
lserum=log(serum);
run;
;
proc anova;
class drug;
model lserum=drug;
means drug /lsd sheffe tukey;
run;
;

```

In the call to 'proc anova', the statement 'class drug' means that 'drug' is being treated as a classified variable (or factor), rather than as a numerical variable. The 'model' statement then fits the one-way ANOVA model, with log serum as the response variable. The 'means' statement asks for comparisons among the mean responses for each drug, with differences evaluated according to each of the three criteria discussed in Section 8.2.7.

The output from this procedure is as follows:

#### The ANOVA Procedure

##### Class Level Information

414 Chapter 8. Analysis of Designed Experiments

Class	Levels	Values
drug	4	1 2 3 4
Number of observations		28

Dependent Variable: lserum

Source	DF	Sum of Squares	Mean Square
Model	3	4.88313231	1.62771077
Error	24	7.59836533	0.31659856
Corrected Total	27	12.48149764	

Source	F Value	Pr > F
Model	5.14	0.0069
Error		
Corrected Total		

R-Square	Coeff Var	Root MSE	lserum Mean
0.391230	26.66922	0.562671	2.109814

Source	DF	Anova SS	Mean Square
drug	3	4.88313231	1.62771077

Source	F Value	Pr > F
drug	5.14	0.0069

t Tests (LSD) for lserum

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	24



NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	24
Error Mean Square	0.316599
Critical Value of F	3.00879
Minimum Significant Difference	0.9036

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	drug
A	2.7597	7	2
A			
B A	2.1719	7	4
B			
B	1.8482	7	3
B			
B	1.6594	7	1

The first part of the output reproduces the analysis of variance table given earlier. The three sets of results in response to the ‘means’ statement all come to the same conclusion: the order of the drugs in terms of decreasing serum levels is 2, 4, 3, 1; drugs 2 and 4 form a common group in the sense that they are no significantly different when judged by any of the three tests; similarly, drugs 4, 3 and 1 form a common group. The actual critical values of  $t$  or  $F$  which determine these groupings are, however, different.

Now let us consider a two-way ANOVA without interaction, using the alloys data set of Section 8.3.2. The data are again arranged in column format, in a file ‘alloy.dat’:

```

1  1  3.06
1  2  3.01
1  3  2.84
...
2  1  2.59
2  2  2.75
2  3  3.13
...
4  6  2.58

```

A sample SAS code is now as follows:

```

options ls=64 ps=58;
data alloy;

```

```

infile 'alloy.dat';
input bar lab strength;
run;
;
proc anova;
class bar lab;
model strength=bar lab;
means bar /sheffe;
means lab / sheffe;
run;
;

```

The only difference from the previous example is that there are now two 'class' variables, 'bar' and 'lab', and the model includes both of them. The 'means' statement asks for comparisons of both, using the Scheffé test criterion.

Here is part of the output (edited):

Dependent Variable: strength

Source	DF	Sum of Squares	Mean Square
Model	8	2.58635000	0.32329375
Error	15	1.23343333	0.08222889
Corrected Total	23	3.81978333	

...

Source	DF	Anova SS	Mean Square
bar	3	0.98141667	0.32713889
lab	5	1.60493333	0.32098667

Source	F Value	Pr > F
bar	3.98	0.0286
lab	3.90	0.0182

Scheffe's Test for strength

...

Scheffe	Grouping	Mean	N	bar
	A	3.2733	6	3
	A			
B	A	3.0083	6	1
B	A			
B	A	2.8300	6	2
B				
B		2.7450	6	4
...				
Scheffe	Grouping	Mean	N	lab
	A	3.4500	4	4
	A			
B	A	3.0125	4	1
B	A			
B	A	3.0000	4	3
B	A			
B	A	2.8900	4	2
B	A			
B	A	2.8450	4	5
B				
B		2.5875	4	6

The results of the ANOVA table are the same as given earlier, resulting in the conclusion that both 'bar' and 'lab' are significant factors. The 'means' comparison shows that bars 3, 1 and 2 form a common group (not separated by the Scheffé test); likewise, bars 1, 2 and 4 form a common group, but there are significant differences among all four groups. The corresponding comparisons for laboratories show that laboratories 4, 1, 3, 2, 5 form a common group, as do 1, 3, 2, 5, 6, but again, we reject the hypothesis that all six laboratories are the same.

Now let us consider the Fisher barley yields example of Section 8.4. The data set is again arranged in columns, with year in column 1, place in column 2, variety (coded as 1–5) in columns 3, and yield in column 4:

1931	1	1	81.0
1931	1	2	105.4
1931	1	3	119.7
1931	1	4	109.7
1931	1	5	98.3
1932	1	1	80.7
...			

1932 6 5 94.1

A possible SAS program is as follows:

```
options ls=64 ps=58;
data fisher;
infile 'fisher.dat';
input year place variety yield;
run;
;
proc anova;
class year place variety;
model yield=place variety place*variety;
run;
;
proc anova;
class year place variety;
model yield=year place variety;
run;
;
```

The first analysis ignores 'year', but treats the experiment as a two-way ANOVA with interaction — the variable 'place\*variety' creates this interaction. The second analysis is slightly different from the one given earlier — in effect a three-way analysis of variance without interactions, treating each of 'year', 'place' and 'variety' as a categorical variable.

The ANOVA table from the first analysis gives:

Source	DF	Sum of Squares	Mean Square
Model	29	30963.04083	1067.69106
Error	30	13767.46500	458.91550
Corrected Total	59	44730.50583	

Source	DF	Anova SS	Mean Square
place	5	21217.25483	4243.45097
variety	4	5313.39500	1328.34875
place*variety	20	4432.39100	221.61955

Source	F Value	Pr > F
place	9.25	<.0001



variety	2.89	0.0387
place*variety	0.48	0.9535

This analysis confirms that, analyzed as a two-way ANOVA, the interaction term is nowhere near significant. However, the three-way analysis including 'year' yields the following:

Source	DF	Sum of Squares	Mean Square
Model	10	30327.57133	3032.75713
Error	49	14402.93450	293.93744
Corrected Total	59	44730.50583	

Source	DF	Anova SS	Mean Square
year	1	3796.92150	3796.92150
place	5	21217.25483	4243.45097
variety	4	5313.39500	1328.34875

Source	F Value	Pr > F
year	12.92	0.0008
place	14.44	<.0001
variety	4.52	0.0035

This confirms that the 'year' variable is highly significant, with a  $p$ -value of .0008, and therefore speaks strongly towards including it in the model.

### 8.5.2 ANOVA in R

In R, it is possible to do an analysis of variance using the `lm` command that we usually use for linear regression, after first using `factor` to redefine the treatment or block variable as a factor variable. For example, to fit the PEMA data using `lm`, we could try

```
dat1<-matrix(scan('D:/r/b/s105/dat1/pema.dat'),ncol=2,byrow=T)
drug<-factor(dat1[,1])
lpema<-log(dat1[,2])
lm1<-lm(lpema~drug)
```

The command `summary(lm1)` then produces the output

```
Call:
lm(formula = lpema ~ drug)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.477082	-0.251050	-0.008663	0.471353	0.856561

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.6594	0.2127	7.803	4.9e-08 ***
drug2	1.1003	0.3008	3.659	0.00124 **
drug3	0.1888	0.3008	0.628	0.53614
drug4	0.5125	0.3008	1.704	0.10128

---

Signif. codes: 0 \*\*\* 0.001 \*\* 0.01 \* 0.05 . 0.1 1

Residual standard error: 0.5627 on 24 degrees of freedom  
 Multiple R-Squared: 0.3912, Adjusted R-squared: 0.3151  
 F-statistic: 5.141 on 3 and 24 DF, p-value: 0.006902

From this we see, in particular, that the  $F$  statistic for no treatment effect has a p-value of about .007, implying we should reject that null hypothesis. In this analysis, the treatment effect for drug 1 is by default assumed to be 0, and the analysis shows that the estimated treatment effects for drugs 2, 3, 4 are respectively 1.10, 0.19 and 0.51 about that for treatment 1.

Alternatively may replace the `lm` command with

```
aov1<-aov(lpema~drug)
```

The command `summary(aov1)` then yields

```
> summary(aov1)
          Df Sum Sq Mean Sq F value  Pr(>F)
drug       3  4.8831  1.6277  5.1412 0.006902 **
Residuals 24  7.5984  0.3166
---
```

Signif. codes: 0 \*\*\* 0.001 \*\* 0.01 \* 0.05 . 0.1 1

At first sight, this is not more informative. However, the `aov` object offers more. For example,

```
> TukeyHSD(aov1)
Tukey multiple comparisons of means
 95% family-wise confidence level
```

```
Fit: aov(formula = lpema ~ drug)
```

```
$drug
      diff      lwr      upr      p adj
2-1  1.1003440  0.2706641  1.93002384  0.0063672
3-1  0.1887808 -0.6408991  1.01846067  0.9221370
4-1  0.5125161 -0.3171638  1.34219600  0.3435628
```

```
3-2 -0.9115632 -1.7412430 -0.08188329 0.0275509
4-2 -0.5878278 -1.4175077 0.24185204 0.2329535
4-3 0.3237353 -0.5059445 1.15341521 0.7068677
```

This computes the mean and a 95% confidence limit for all the pairwise treatment differences, using the Tukey pairwise comparisons procedure to correct for multiple comparisons. The result shows that drug 2 yields a higher mean response than either drug 1 or drug 3; however, none of the other pairwise differences is statistically significant.

Another useful command is `plot(aov1)`. This produces a number of diagnostic plots, illustrated in Figure 8.2.

The next analysis performs a similar job for the alloys data set.

```
x<-matrix(scan(file='alloy.dat'),ncol=3,byrow=T)
strength<-x[,3]
bar<-factor(x[,1])
lab<-factor(x[,2])
aov2<-aov(strength~bar+lab)
```

This performs a two-way analysis without interactions, treating both “bar” and “lab” as factor variables. The command `summary(aov2)` produces

```
          Df Sum Sq Mean Sq F value Pr(>F)
bar           3  0.98142  0.32714   3.9784 0.02862 *
lab           5  1.60493  0.32099   3.9036 0.01820 *
Residuals    15  1.23343  0.08223
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

confirming that both the lab effect and the bar effect are statistically significant. If we wish to know which pairwise differences are significant, we can use `TukeyHSD(aov2)` yielding

```
Tukey multiple comparisons of means
 95% family-wise confidence level
```

```
Fit: aov(formula = strength ~ bar + lab)
```

```
$bar
      diff      lwr      upr    p adj
2-1 -0.1783333 -0.65549765 0.29883099 0.7081269
3-1 0.2650000 -0.21216432 0.74216432 0.4075721
4-1 -0.2633333 -0.74049765 0.21383099 0.4128340
3-2 0.4433333 -0.03383099 0.92049765 0.0730144
4-2 -0.0850000 -0.56216432 0.39216432 0.9546021
4-3 -0.5283333 -1.00549765 -0.05116901 0.0277449
```

```

$lab
      diff      lwr      upr      p adj
2-1 -0.1225 -0.78128343  0.53628343  0.9890835
....
6-4 -0.8625 -1.52128343 -0.20371657  0.0074210

```

For the bar variable, the result shows that the bar 3 has a higher mean than bar 4 but the other pairwise differences are not statistically significant. For the lab variable, most of the rows have been edited out, but only for labs 4 and 6 is the difference statistically significant.

For the third example, we consider the Fisher barley yield data. In this case, our R program is as follows:

```

x<-matrix(scan(file='fisher.dat'),ncol=4,byrow=T)
yield<-x[,4]
place<-factor(x[,2])
year<-factor(x[,1])
var<-factor(x[,3])
aov3<-aov(yield~var*place)
aov4<-aov(yield~year+var+place)

```

The first aov call is to a two-way interaction on var and place, including interaction (that is the meaning of writing `var*place` instead of `var+place`). The ANOVA table

```

> summary(aov3)
      Df Sum Sq Mean Sq F value    Pr(>F)
var      4  5313.4   1328.3   2.8945  0.03872 *
place    5 21217.3   4243.5   9.2467 2.077e-05 ***
var:place 20  4432.4    221.6   0.4829  0.95347
Residuals 30 13767.5    458.9
---

```

```

Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

```

shows significant effects for variety and place, but no interaction.

The second aov call is for a three-way analysis of variance without interactions. In this case the ANOVA table is

```

> summary(aov4)
      Df Sum Sq Mean Sq F value    Pr(>F)
year    1  3796.9   3796.9  12.9174 0.0007541 ***
var      4  5313.4   1328.3   4.5192 0.0034655 **
place    5 21217.3   4243.5  14.4366 1.091e-08 ***
Residuals 49 14402.9    293.9
---

```

```

Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

```

This shows all three main effects as statistically significant, but also a much smaller mean squared error (293.9 against 458.9). This shows the

importance of including the year variable in the model. The corresponding diagnostic plots are shown in Figure 8.3.

Finally, if we call the `TukeyHSD(aov4)` command, we get the following (most differences that are not statistically significant have been edited out):

```
Tukey multiple comparisons of means
 95% family-wise confidence level

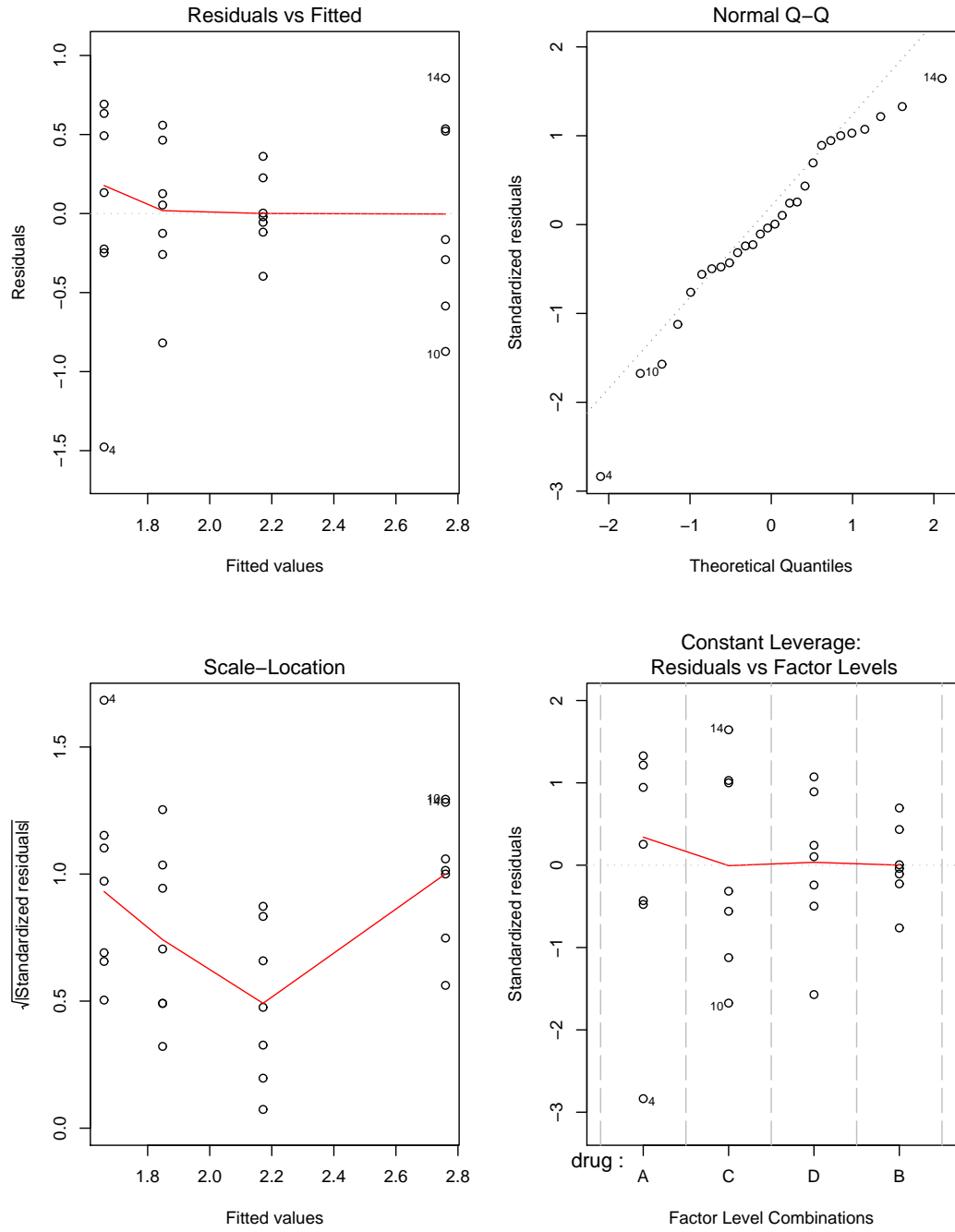
Fit: aov(formula = yield ~ year + var + place)

$year
      diff      lwr      upr    p adj
1932-1931 -15.91 -24.80582 -7.014178 0.0007541

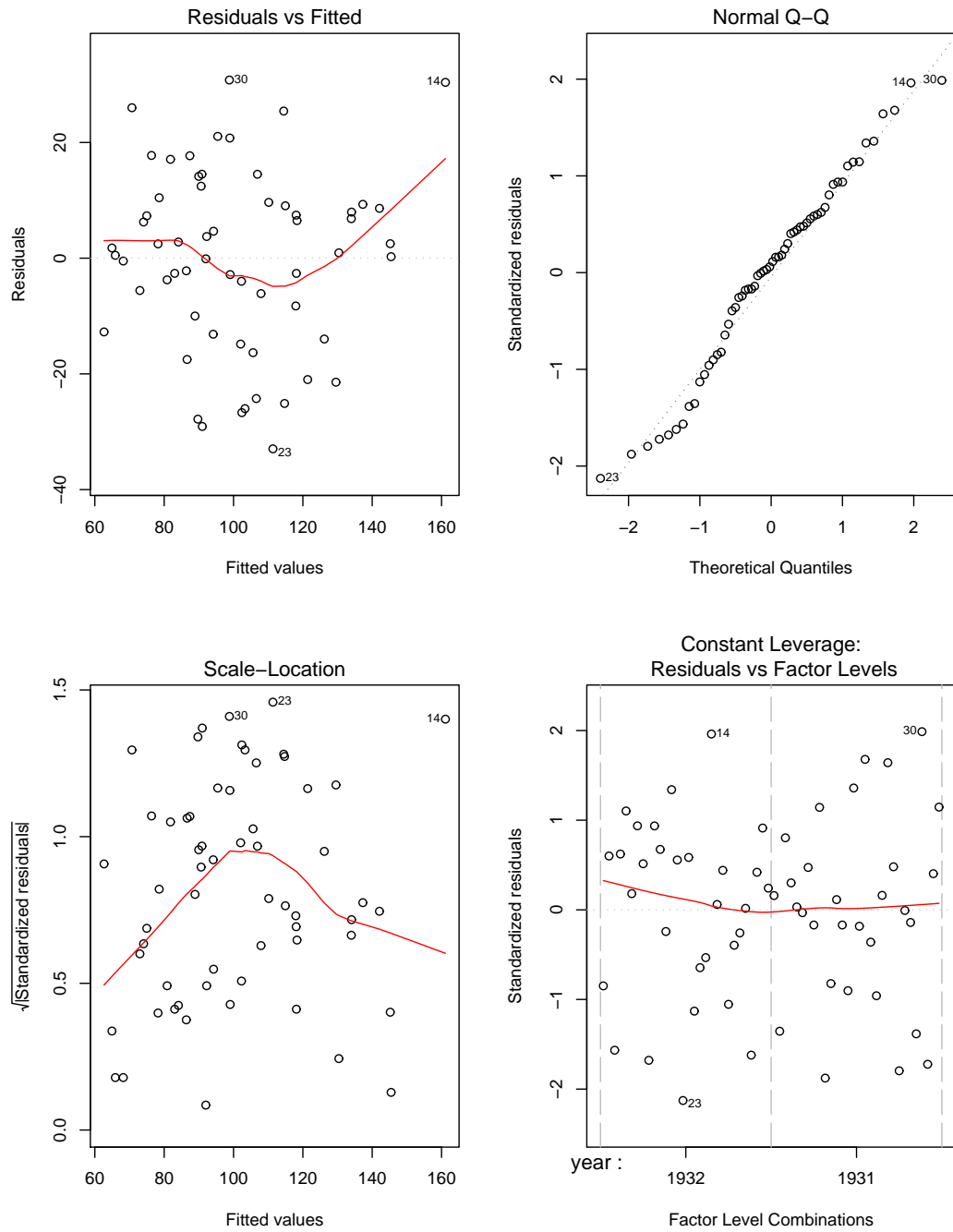
$var
      diff      lwr      upr    p adj
4-1  23.816667  3.9951888 43.638145 0.0111752
4-2  27.075000  7.2535221 46.896478 0.0028645
4-3  19.025000 -0.7964779 38.846478 0.0655987
5-4 -15.666667 -35.4881445  4.154811 0.1830523

$place
      diff      lwr      upr    p adj
2-1  43.15  20.413194 65.8868057 0.0000125
3-2 -30.73 -53.466806 -7.9931943 0.0027118
4-2 -27.14 -49.876806 -4.4031943 0.0108434
5-2 -55.48 -78.216806 -32.7431943 0.0000000
6-2 -53.19 -75.926806 -30.4531943 0.0000001
5-3 -24.75 -47.486806 -2.0131943 0.0255350
6-3 -22.46 -45.196806  0.2768057 0.0546248
5-4 -28.34 -51.076806 -5.6031943 0.0069063
6-4 -26.05 -48.786806 -3.3131943 0.0161431
```

This analysis confirms that both year and place are statistically significant effects, but if our main interest were in the effect of variety, the conclusion is that variety 4 is clearly better than varieties 1 and 2; the difference between variety 4 and either of 3 or 5, although in both cases favoring variety 4, is not statistically significant according to this analysis.



**Figure 8.2.** Analysis of variance plots for PEMA data (from `plot(aov1)` command in R).



**Figure 8.3.** Analysis of variance plots for Fisher data, second analysis (from `plot(aov4)` command in R).

## 8.6 Exercises

1. In a round-robin experiment, the individual sample numbers  $n_i$ , sample means  $\bar{y}_i$  and sample standard deviations  $s_i$  are as in Table 1.

$i$	$n_i$	$\bar{y}_i$	$s_i$
1	5	37.4	5.1
2	5	42.3	3.1
3	5	49.3	2.7
4	8	36.1	1.7

**Table 8.16** Data set for problem 8.1

It looks as though sample 3 has higher mean than the others. Do the data in fact support that this is a significant difference? (Assume the individual observations are all normally distributed with common variance  $\sigma^2$ ).

2. Ozonation as a secondary treatment for effluent following absorption by ferrous chloride was studied for three reaction times and three PH levels. The data in Table 8.17 were obtained for effluent decline.

		PH	Level	
		7	9	10.5
Reaction	20	23	16	14
		21	18	13
		22	15	16
Time	40	20	14	12
		22	13	11
		19	12	10
(Minutes)	60	21	13	11
		20	12	13
		19	12	12

**Table 8.17** Data for Problem 8.2.

- (a) Obtain the ANOVA table for a two-way analysis with interactions.

Based on the table just computed, answer the following questions:

- (b) Are the effects due to PH and reaction time significant?
- (c) Are the interactions significant?
- (d) In future operation, it is expected to maintain the PH level as 7, while the reaction time will be 20 minutes for half the total time of operation, and 40 minutes for the other half. Obtain an estimate,



with its standard error, for the mean effluent decline, assuming (i) interactions are present, and (ii) that they are not, and comment on any noticeable differences between the two results.

3. This question is about an analysis of variance experiment, reinterpreted as a linear regression. It does not assume detailed knowledge about analysis of variance.

A recent paper (Lee, J. and Wrolstad, R.E. (2004), Extraction of anthocyanins and polyphenolics from blueberry-processing waste, *Journal of Food Science* **69**, No. 7, C564–C573) discussed the following experiment related to the extraction of juice from blueberries. Three control variables were considered: temperature, level of sulfur dioxide (SO<sub>2</sub>) and citric acid (coded as 0 or 1). Two response variables were measured: ACY (anthocyanin) and TP (total phenolics), both of which are considered to have beneficial health effects. The data were as follows:

Number	Temp (deg C)	SO <sub>2</sub> (ppm)	Citric Acid	ACY	TP
1	50	0	0	27.5	55.9
2	50	0	1	42.6	62.6
3	80	0	0	50.2	71.4
4	80	0	1	62.4	88.8
5	50	50	0	92.2	307.3
6	50	50	1	96.5	316.4
7	80	50	0	97.5	420.6
8	80	50	1	102.2	413.8
9	50	100	0	90.6	386.0
10	50	100	1	82.2	337.5
11	80	100	0	92.1	641.0
12	80	100	1	91.4	684.3

Consider the model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_{ij} + \eta_{ik} + \zeta_{jk} + \epsilon_{ijk}, \quad (8.15)$$

where  $\alpha_i$ ,  $i = 1, 2$ ,  $\beta_j$ ,  $j = 1, 2, 3$ ,  $\gamma_k$ ,  $k = 1, 2$  are main effects due to temperature, SO<sub>2</sub> and citric acid respectively,  $\delta_{ij}$ ,  $\eta_{ik}$ ,  $\zeta_{jk}$  are interaction terms, and  $\epsilon_{ijk}$  are independent  $N(0, \sigma^2)$  errors. To make the model identifiable, assume any of  $\alpha_i, \beta_j, \gamma_k, \delta_{ij}, \eta_{ik}, \zeta_{jk}$  is 0 when any of  $i, j, k$  is 1 (note that this is a different identifiability condition from the ones assumed in most of the examples of this chapter).

- (a) Write the model (8.15) in the form  $Y = X\beta + \epsilon$ , where  $Y$  is the vector of responses (of dimension 12), the vector  $\beta$  consists of all

the non-zero unknown parameters, and  $X$  is a design matrix of zeros and ones. (You should find that  $X$  is  $12 \times 10$ .)

- (b) Using SAS's PROC REG or the "lm" command in R or S-PLUS, fit the model (8.15) to the data, where temperature,  $\text{SO}_2$  and citric acid are the three factor variables and ACY is the response. Also consider possible transformations of the response and indicate which you prefer. (For example, you should consider both the square root and the log transformation, and others in the Box-Cox family if you have time. It is not necessary to give detailed tables of parameter values, but state the value of the residual sum of squares or the estimated  $s$ , and any other statistics that are directly relevant to the question.)
- (c) Now using whatever transformation you selected in (b), decide which of the main effects and interactions is significant. (Again, I don't want very detailed regression output, but indicate the main steps of your analysis and how you did them.)
- (d) Repeat the steps of (b) and (c) for the TP response variable. (It's not necessary that the transformation of TP be the same as that for ACY.)
- (e) Write a short report on your conclusions for the company. Recall that the company's objective is to choose *one* setting of the three control variables so that both ACY and TP are high. Your report should indicate which settings you recommend, but should also make clear to what extent the differences among different possible settings are statistically significant, and whether you would recommend further experimentation.

*Note:* The proposed analysis in this question differs substantially from that in the paper from which the data are derived.