CHAPTER

13

Analysis of Variance

W. Penny and R. Henson

INTRODUCTION

The mainstay of many scientific experiments is the factorial design. These comprise a number of experimental factors which are each expressed over a number of levels. Data are collected for each factor/level combination and then analysed using analysis of variance (ANOVA). The ANOVA uses F-tests to examine a pre-specified set of standard effects, e.g. 'main effects' and 'interactions', as described in Winer *et al*. (1991).

ANOVAs are commonly used in the analysis of position emission tomography (PET), electroencephalography (EEG), magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) data. For PET, this analysis usually takes place at the 'first' level. This involves direct modelling of PET scans. For EEG, MEG and fMRI, ANOVAs are usually implemented at the 'second level'. As described in the previous chapter, first level models are used to create contrast images for each subject. These are then used as data for a second level or 'random-effects' analysis.

Some different types of ANOVA are tabulated in Table 13-1. A *two-way* ANOVA, for example, is an ANOVA with 2 factors; a K_1-*by*-K_2 ANOVA is a two-way ANOVA with K_1 levels of one factor and K_2 levels of the other. A *repeated measures* ANOVA is one in which the levels of one or more factors are measured from the same unit (e.g. subjects). Repeated measures ANOVAs are also sometimes called *within-subject* ANOVAs, whereas designs in which each level is measured from a different group of subjects are called *between-subject* ANOVAs. Designs in which some factors are within-subject, and others between-subject, are sometimes called *mixed* designs.

This terminology arises because in a between-subject design the difference between levels of a factor is given by the difference between subject responses, e.g. the difference between levels 1 and 2 is given by the difference between those subjects assigned to level 1 and those assigned to level 2. In a within-subject design, the levels of a factor are expressed within each subject, e.g. the difference between levels 1 and 2 is given by the average difference of subject responses to levels 1 and 2. This is like the difference between two-sample t-tests and paired t-tests.

The benefit of repeated measures is that we can match the measurements better. However, we must allow for the possibility that the measurements are correlated (so-called 'non-sphericity' – see below).

The level of a factor is also sometimes referred to as a 'treatment' or a 'group' and each factor/level combination is referred to as a 'cell' or 'condition'. For each type of ANOVA, we describe the relevant statistical models and show how they can be implemented in a general linear model (GLM). We also give examples of how main effects and interactions can be tested for using F-contrasts.

The chapter is structured as follows: the first section describes one-way between-subject ANOVAs. The next section describes one-way within-subject ANOVAs and introduces the notion of non-sphericity. We then describe two-way within-subject ANOVAs and make a distinction

TABLE 13-1 Types of ANOVA

Factors	Levels	Simple	Repeated Measures
1	2	Two-sample t-test	Paired t-test
1	K	One-way ANOVA	One-way ANOVA within-subject
M	$K_1, K_2, .., K_M$	M-way ANOVA	M-way ANOVA within-subject

Statistical Parametric Mapping, by Karl Friston et al.
ISBN–13: 978-0-12-372560-8 ISBN–10: 0-12-372560-7

between models with pooled versus partitioned errors. The last section discusses issues particular to fMRI and we end with a discussion.

Notation

In the mathematical formulations below, $N(m, \Sigma)$ denotes a uni/multivariate Gaussian with mean m and variance/covariance Σ. I_K denotes the $K \times K$ identity matrix, X^T denotes transpose, X^{-T} the inverse transpose, X^- the generalized-inverse, 1_K is a $K \times 1$ vector of 1s, 0_K is a $K \times 1$ vector of zeros and 0_{KN} is a $K \times N$ matrix of zeros. We consider factorial designs with $n = 1..N$ subjects and $m = 1..M$ factors where the mth factor has $k = 1..K_m$ levels.

ONE-WAY BETWEEN-SUBJECT ANOVA

In a between-subject ANOVA, differences between levels of a factor are given by the differences between subject responses. We have one measurement per subject and different subjects are assigned to different levels/treatments/groups. The response from the nth subject (y_n) is modelled as:

$$y_n = \tau_k + \mu + e_n \tag{13.1}$$

where τ_k are the treatment effects, $k = 1..K$, $k = g(n)$ and $g(n)$ is an indicator function whereby $g(n) = k$ means the nth subject is assigned to the kth group, e.g. $g(13) = 2$ indicates the 13th subject being assigned to group 2. This is the single experimental factor that is expressed over K levels. The variable μ is sometimes called the *grand mean* or *intercept* or *constant term*. The random variable e_n is the residual error, assumed to be drawn from a zero mean Gaussian distribution.

If the factor is significant, then the above model is a significantly better model of the data than the simpler model:

$$y_n = \mu + e_n \tag{13.2}$$

where we just view all of the measurements as random variation about the grand mean. Figure 13.2 compares these two models on some simulated data.

In order to test whether one model is better than another, we can use an F-test based on the *extra sum of squares* principle (see Chapter 8). We refer to Eqn. **13.1** as the 'full' model and Eqn. **13.2** as the 'reduced' model. If RSS denotes the residual sum of squares (i.e. the sum of squares left after fitting a model) then:

$$F = \frac{(RSS_{reduced} - RSS_{full})/(K-1)}{RSS_{full}/(N-K)} \tag{13.3}$$

has an F-distribution with $K-1, N-K$ degrees of freedom. If F is significantly non-zero then the full model has a significantly smaller error variance than the reduced model. That is to say, the full model is a significantly better model, or the *main effect* of the factor is significant.

The above expression is also sometimes expressed in terms of sums of squares (SS) due to treatment and due to error:

$$F = \frac{SS_{treat}/DF_{treat}}{SS_{error}/DF_{error}} \tag{13.4}$$

where

$$\begin{aligned} SS_{treat} &= RSS_{reduced} - RSS_{full} \\ DF_{treat} &= K - 1 \\ SS_{error} &= RSS_{full} \\ DF_{error} &= N - K \\ DF_{total} &= DF_{treat} + DF_{error} = N - 1 \end{aligned} \tag{13.5}$$

Eqns **13.3** and **13.4** are therefore equivalent.

Numerical example

This subsection shows how an ANOVA can be implemented in a GLM. Consider a one-way ANOVA with $K = 4$ groups each having $n = 12$ subjects (i.e. $N = Kn = 48$ subjects/observations in total). The GLM for the full model in Eqn. **13.1** is:

$$y = X\beta + e \tag{13.6}$$

where the design matrix $X = [I_K \otimes 1_n, 1_N]$ is shown in Figure 13.1, where $\otimes$ denotes the Kronecker product (see Appendix 13.1). The vector of parameters is $\beta = [\tau_1, \tau_2, \tau_3, \tau_4, \mu]^T$.

Eqn. **13.3** can then be implemented using the *effects of interest* F-contrast, as introduced in Chapter 9:

$$C^T = \begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 & 0 \\ -1/3 & 1 & -1/3 & -1/3 & 0 \\ -1/3 & -1/3 & 1 & -1/3 & 0 \\ -1/3 & -1/3 & -1/3 & 1 & 0 \end{bmatrix} \tag{13.7}$$

or equivalently:

$$C^T = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{bmatrix} \tag{13.8}$$

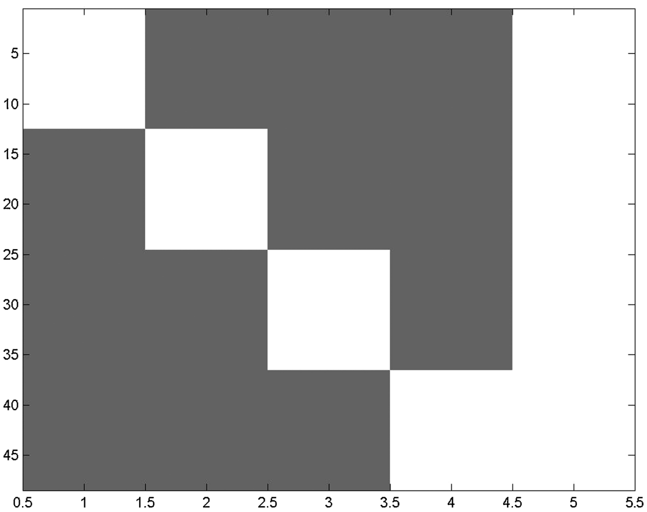

FIGURE 13.1 Design matrix for one-way (1×4) between-subjects ANOVA. White and grey represent 1 and 0. There are 48 rows, one for each subject ordered by condition, and 5 columns, the first 4 for group effects and the 5th for the grand mean.

These contrasts can be thought of as testing the null hypothesis $\mathcal{H}_0$:

$$\mathcal{H}_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 \qquad \textbf{13.9}$$

Note that a significant departure from $\mathcal{H}_0$ can arise from any pattern of these treatment means (parameter estimates) – they need not be monotonic across the four groups for example.

The correspondence between this F-contrast and the classical formulation in Eqn. **13.3** is detailed in Chapter 10. We now analyse the example data set shown in Figure 13.2. The results of a one-way between-subjects ANOVA are shown in Table 13-2. This shows that there is a significant main effect of treatment ($p < 0.02$).

Note that the design matrix in Figure 13.2 is rank-deficient (see Chapter 8) and the alternative design matrix $X = [I_K \otimes 1_n]$ could be used with appropriate F-contrasts (though the parameter estimates themselves would include a contribution of the grand mean, equivalent to the contrast $[1, 1, 1, 1]^T$). If β_1 is a vector of parameter estimates after the first four columns of X are mean-corrected (orthogonalized with respect to the fifth column), and β_0 is the parameter estimate for the corresponding fifth column, then:

$$\begin{aligned} SS_{treatment} &= n\beta_1^T\beta_1 = 51.6 \qquad \textbf{13.10}\\ SS_{mean} &= nK\beta_0^2 = 224.1\\ SS_{error} &= r^T r = 208.9\\ SS_{total} &= y^T y = SS_{treatment} + SS_{mean} + SS_{error} = 484.5 \end{aligned}$$

where the residual errors are $r = y - XX^{-}y$.

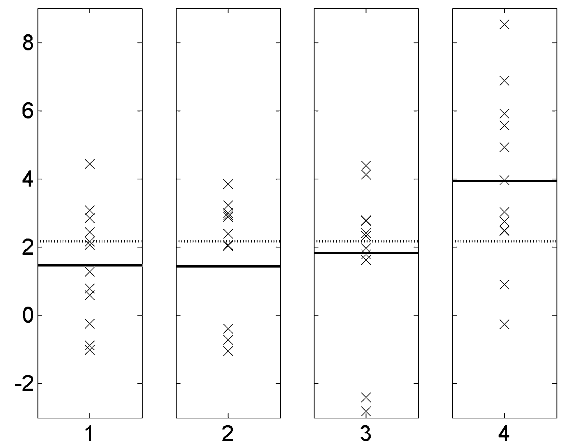

FIGURE 13.2 **One-way between-subject ANOVA.** 48 subjects are assigned to one of four groups. The plot shows the data points for each of the four conditions (crosses), the predictions from the 'one-way between-subjects model' or the 'full model' (solid lines) and the predicitons from the 'reduced model' (dotted lines). In the reduced model (Eqn. **13.2**), we view the data as random variation about a grand mean. In the full model (Eqn. **13.1**), we view the data as random variation about condition means. Is the full model significantly better than the reduced model? That responses are much higher in condition 4 suggests that this is indeed the case and this is confirmed by the results in Table 13-2.

TABLE 13-2 Results of one-way (1×4) between-subjects ANOVA

Main effect of treatment	$F = 3.62$	$DF = [3, 44]$	$p = 0.02$

ONE-WAY WITHIN-SUBJECT ANOVA

In this model we have K measurements *per* subject. The treatment effects for subject $n = 1 \ldots N$ are measured relative to the average response made by subject n on all treatments. The kth response from the nth subject is modelled as:

$$y_{nk} = \tau_k + \pi_n + e_{nk} \qquad \textbf{13.11}$$

where τ_k are the treatment effects (or *within-subject effects*), π_n are the *subject effects* and e_{nk} are the residual errors. We are not normally interested in π_n, but its explicit modelling allows us to remove variability due to differences in average responsiveness of each subject. See, for example, the data set in Figure 13.3. It is also possible to express the full model in terms of differences between treatments (see e.g. Eqn. **13.15** for the two-way case).

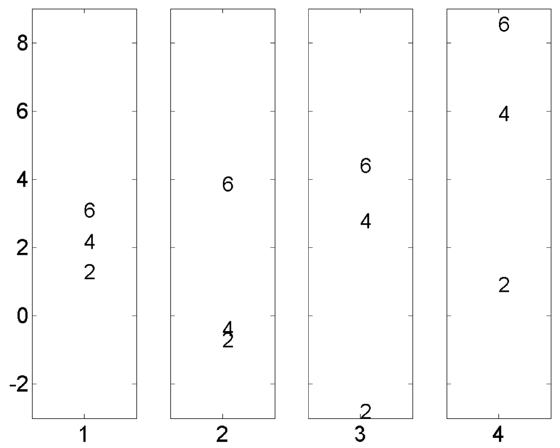

FIGURE 13.3 Portion of example data for one-way within-subject ANOVA. The plot shows the data points for 3 subjects in each of 4 conditions (in the whole data set there are 12 subjects). Notice how subject 6's responses are always high, and subject 2's are always low. This motivates modelling subject effects as in Eqns. ***13.11*** and ***13.12***.

To test whether the experimental factor is significant, we compare the full model in Eqn. **13.11** with the reduced model:

$$y_{nk} = \pi_n + e_{nk} \qquad \textbf{13.12}$$

An example of comparing these full and reduced models is shown in Figure 13.4. The equations for computing the relevant F-statistic and degrees of freedom are given, for example, in Chapter 14 of Howell (1992).

Numerical example

The design matrix $X = [I_K \otimes 1_N, 1_K \otimes I_N]$ for Eqn. **13.11**, with $K = 4$ and $N = 12$, is shown in Figure 13.5. The first 4 columns are treatment effects and the next 12 are subject effects. The main effect of the factor can be assessed using the same *effects of interest* F-contrast as in Eqn. **13.7**, but with additional zeros for the columns corresponding to the subject effects.

We now analyse another example data set, a portion of which is shown in Figure 13.3. Measurements have been obtained from 12 subjects under each of $K = 4$ conditions.

Assuming sphericity (see below), we obtain the ANOVA results in Table 13-3. In fact this dataset contains exactly the same numerical values as the between-subjects example data. We have just relabelled the data as being measured from 12 subjects with 4 responses each instead of from 48 subjects with 1 response each. The reason that the p-value is less than in the between-subjects example (it has reduced from 0.02 to 0.001) is that the data were created to include subject effects. Thus,

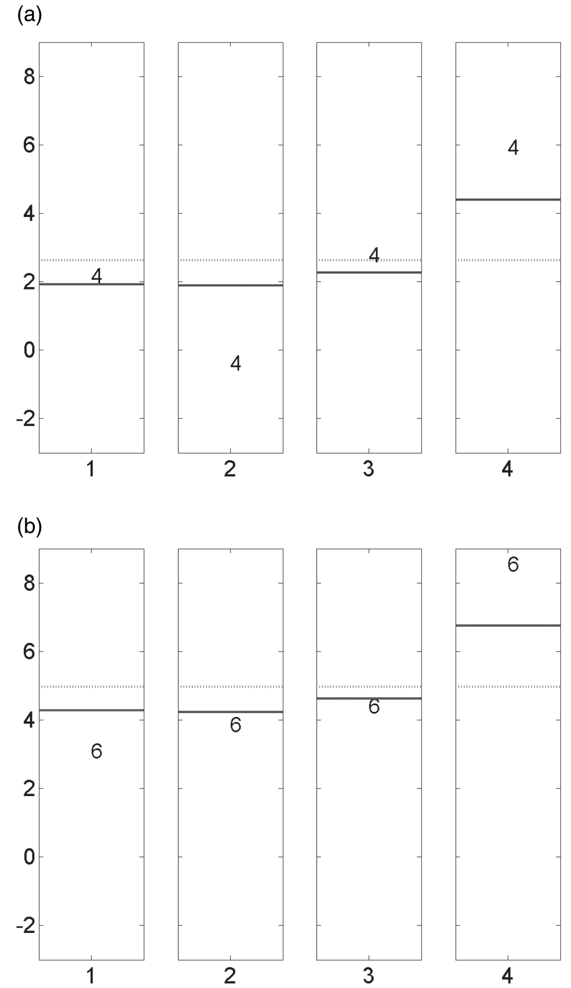

FIGURE 13.4 **One-way within-subjects ANOVA.** The plot shows the data points for each of the four conditions for subjects (a) 4 and (b) 6, the predictions from the one-way within-subjects model (solid lines) and the reduced model (dotted lines).

in repeated measures designs, the modelling of subject effects normally increases the sensitivity of the inference.

Non-sphericity

Due to the nature of the levels in an experiment, it may be the case that if a subject responds strongly to level i, he may respond strongly to level j. In other words, there may be a correlation between responses. In Figure 13.6 we plot subject responses for level i against level j for the example data set. These show that for some pairs of conditions there does indeed seem to be a correlation. This correlation can be characterized graphically by fitting a Gaussian to each 2D data cloud and then plotting probability contours. If these contours form a sphere (a circle in two dimensions) then the data are Independent and identically distributed (IID), i.e. same

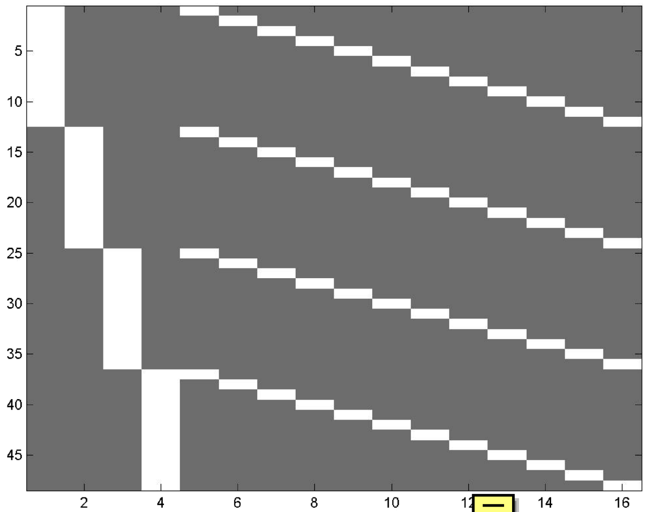

FIGURE 13.5 Design matrix for one-way (1 × 4) within-subjects ANOVA. The first 4 columns are treatment effects and the last 12 are subject effects.

TABLE 13-3 Results of one-way (1 × 4) within-subjects ANOVA

Main effect of treatment	F = 6.89	DF = [3, 33]	$p = 0.001$

variance in all dimensions and there is no correlation. The more these contours look like ellipses, the more 'non-sphericity' there is in the data.

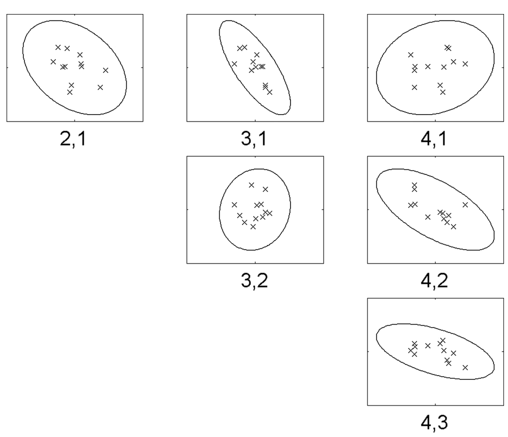

FIGURE 13.6 One-way within-subjects ANOVA: Non-sphericity. Each subgraph plots each subject's response to condition i versus condition j as a cross. There are twelve crosses, one from each subject. We also plot probability contours from the corresponding Gaussian densities. Subject responses, for example, to conditions 1 and 3 seem correlated – the sample correlation coefficient is −0.75. Overall, the more non-spherical the contours the greater the non-sphericity.

The possible non-sphericity can be taken into account in the analysis using a correction to the degrees of freedom (DFs). In the above example, a Greenhouse-Geisser (GG) correction (see Appendix 13.1 and Chapter 10) estimates $\epsilon = .7$, giving DFs of $[2.1, 23.0]$ and a p-value (with GG we use the same F-statistic, i.e. $F = 6.89$) of $p = 0.004$. Assuming sphericity, as before, we computed $p = 0.001$. Thus the presence of non-sphericity in the data makes us less confident of the significance of the effect.

An alternative representation of the within-subjects model is given in Appendix 13.2. This shows how one can take into account non-sphericity. Various other relevant terminology is also defined in Appendices 13.1 and 13.2.

TWO-WAY WITHIN-SUBJECT ANOVAS

The full model for a two-way, K_1-*by*-K_2 repeated measures ANOVA, with $P = K_1K_2$ measurements taken from each of N subjects, can be written as:

$$y_{nkl} = \tau_{kl} + \pi_n + e_{nkl} \qquad \mathbf{13.13}$$

where $k = 1 \ldots K_1$ and $l = 1 \ldots K_2$ index the levels of factor A and factor B respectively. Here we can think of indicator functions $k = g_k(i)$, $l = g_l(i)$ and $n = g_n(i)$ that return the levels of both factors and subject identity for the ith scan. Again, π_n are subject effects and e_{nkl} are residual errors. This equation can be written in matrix form:

$$y = X\beta + e \qquad \mathbf{13.14}$$

where $X = [I_P \otimes 1_N, 1_N \otimes I_P]$ is the design matrix and $\beta = [\tau_{kl}, \pi_n]^T$ are the regression coefficients. This is identical to the one-way within-subject design but with P instead of K treatment effects.

However, rather than considering each factor/level combination separately, the key concept of ANOVA is to model the data in terms of a standard set of *experimental effects*. These consist of *main effects* and *interactions*. Each factor has an associated main effect, which is the difference between the levels of that factor, averaging over the levels of all other factors. Each pair of factors (and higher-order tuples; see below) has an associated interaction. Interactions represent the degree to which the effect of one factor depends on the levels of the other factor(s). A two-way ANOVA thus has two main effects and one interaction.

Eqn. **13.13** can be rewritten as:

$$\begin{aligned} y &= X\beta + e \qquad \mathbf{13.15} \\ &= XC^{-T}C^T\beta + e \\ &= X_r\tilde{\beta} + e \end{aligned}$$

where $X_r = XC^{-T}$ is a rotated design matrix, the regression coefficients are $\tilde{\beta} = C^T\beta$, and C is a 'contrast matrix'. This equation is important as it says that the effects $\tilde{\beta}$ can be estimated by either (i) fitting the data using a GLM with design matrix X_r or (ii) fitting the original GLM, with design matrix X, and applying the contrast matrix $\tilde{\beta} = C^T\beta$.

For our two-way within-subjects ANOVA we choose C such that:

$$\tilde{\beta} = [\tau_q^A, \tau_r^B, \tau_{qr}^{AB}, m, \pi_n]^T \qquad \textbf{13.16}$$

Here, τ_q^A represents the differences between each *successive* level $q = 1 \ldots (K_1 - 1)$ of factor A (e.g. the differences between levels 1 and 2, 2 and 3, 3 and 4 etc.), averaging over the levels of factor B. In other words, the main effect of A is modelled as $K_1 - 1$ differences among K_1 levels. The quantity τ_r^B represents the differences between each *successive* level $r = 1 \ldots (K_2 - 1)$ of factor B, averaging over the levels of factor A; and τ_{qr}^{AB} represents the differences between the differences of each level $q = 1 \ldots (K_1 - 1)$ of factor A across each level $r = 1 \ldots (K_2 - 1)$ of factor B. The quantity m is the mean treatment effect. Examples of contrast matrices and rotated design matrices are given below.

Pooled versus partitioned errors

In the above model, e is sometimes called a *pooled error*, since it does not distinguish between different sources of error for each experimental effect. This is in contrast to an alternative model in which the original residual error e is split into three terms e_{nq}^A, e_{nr}^B and e_{nqr}^{AB}, each specific to a main effect or interaction. This is a different form of *variance partitioning*. Each error term is a random variable and is equivalent to the interaction between that effect and the subject variable.

The F-test for, say, the main effect of factor A is then:

$$F = \frac{SS_k/DF_k}{SS_{nk}/DF_{nk}} \qquad \textbf{13.17}$$

where SS_k is the sum of squares for the effect, SS_{nk} is the sum of squares for the interaction of that effect with subjects, $DF_k = K_1 - 1$ and $DF_{nk} = N(K_1 - 1)$.

Note that, if there are no more than two levels of every factor in an M-way repeated measures ANOVA (i.e, $K_m = 2$ for all $m = 1 \ldots M$), then the covariance of the errors Σ_e for each effect is a 2-by-2 matrix which necessarily has compound symmetry, and so there is no need for a nonsphericity correction.[1] A heuristic for this is that there is only one difference $q = 1$ between two levels $K_m = 2$. This is not necessarily the case if a pooled error is used, as in Eqn. **13.15**.

[1] Although one could model inhomogeneity of variance.

Models and null hypotheses

The difference between pooled and partitioned error models can be expressed by specifying the relevant models and null hypotheses.

Pooled errors

The pooled error model is given by Eqn. **13.15**. For the main effect of A we test the null hypothesis $\mathcal{H}_0 : \tau_q^A = 0$ for all q. Similarly, for the main effect of B. For an interaction we test the null hypothesis $\mathcal{H}_0 : \tau_{qr}^{AB} = 0$ for all q, r.

For example, for the 3-by-3 design shown in Figure 13.7 there are $q = 1..2$ differential effects for factor A and

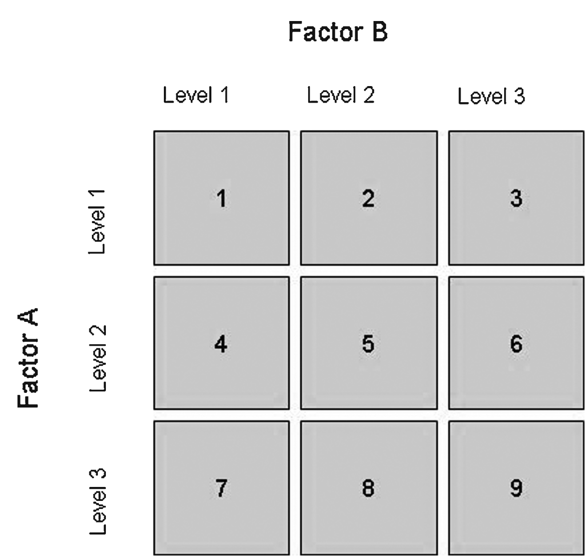

FIGURE 13.7 In a 3×3 ANOVA there are 9 cells or conditions. The numbers in the cells correspond to the ordering of the measurements when rearranged as a column vector y for a single-subject general linear model. For a repeated measures ANOVA there are 9 measurements per subject. The variable y_{nkl} is the measurement at the kth level of factor A, the lth level of factor B and for the nth subject. To implement the partitioned error models we use these original measurements to create differential effects for each subject. The differential effect τ_1^A is given by row 1 minus row 2 (or cells 1, 2, 3 minus cells 4,5,6 – this is reflected in the first row of the contrast matrix in Eqn. 13.52). The differential effect τ_2^A is given by row 2 minus row 3. These are used to assess the main effect of A. Similarly, to assess the main effect of B we use the differential effects τ_1^B (column 1 minus column 2) and τ_2^B (column 2 minus column 3). To assess the interaction between A and B, we compute the four 'simple interaction' effects τ_{11}^{AB} (cells (1–4)-(2–5)), τ_{12}^{AB} (cells (2–5)-(3–6)), τ_{21}^{AB} (cells (4–7)-(5–8)) and τ_{22}^{AB} (cells (5–8)-(6–9)). These correspond to the rows of the interaction contrast matrix in Eqn. **13.30**.

$r = 1..2$ for factor B. The pooled error model therefore has regression coefficients:

$$\tilde{\beta} = [\tau_1^A, \tau_2^A, \tau_1^B, \tau_2^B, \tau_{11}^{AB}, \tau_{12}^{AB}, \tau_{21}^{AB}, \tau_{22}^{AB}, m, \pi_n]^T \qquad \mathbf{13.18}$$

For the main effect of A we test the null hypothesis $\mathcal{H}_0 : \tau_1^A = \tau_2^A = 0$. For the interaction we test the null hypothesis $\mathcal{H}_0 : \tau_{11}^{AB} = \tau_{12}^{AB} = \tau_{21}^{AB} = \tau_{22}^{AB} = 0$.

Partitioned errors

For partitioned errors, we first transform our data set y_{nkl} into a set of differential effects for each subject and then model these in a GLM. This set of differential effects for each subject is created using appropriate contrasts at the 'first-level'. The models that we describe below then correspond to a 'second-level' analysis. The difference between first and second level analyses are described in the previous chapter on random effects analysis.

To test for the main effect of A, we first create the new data points ρ_{nq} which are the differential effects between the levels in A for each subject n. We then compare the full model:

$$\rho_{nq} = \tau_q^A + e_{nq}$$

to the reduced model $\rho_{nq} = e_{nq}$. We are therefore testing the null hypothesis, $\mathcal{H}_0 : \tau_q^A = 0$ for all q.

Similarly for the main effect of B. To test for an interaction, we first create the new data points ρ_{nqr} which are the differences of differential effects for each subject. For a K_1 by K_2 ANOVA there will be $(K_1 - 1)(K_2 - 1)$ of these. We then compare the full model:

$$\rho_{nqr} = \tau_{qr}^{AB} + e_{nqr}$$

to the reduced model $\rho_{nqr} = e_{nqr}$. We are therefore testing the null hypothesis, $\mathcal{H}_0 : \tau_{qr}^{AB} = 0$ for all q, r.

For example, for a 3-by-3 design, there are $q = 1..2$ differential effects for factor A and $r = 1..2$ for factor B. We first create the differential effects ρ_{nq}. To test for the main effect of A we compare the full model:

$$\rho_{nq} = \tau_1^A + \tau_2^A + e_{nq}$$

to the reduced model $\rho_{nq} = e_{nq}$. We are therefore testing the null hypothesis, $\mathcal{H}_0 : \tau_1^A = \tau_2^A = 0$. Similarly for the main effect of B.

To test for an interaction we first create the differences of differential effects for each subject. There are $2 \times 2 = 4$ of these. We then compare the full model:

$$\rho_{nqr} = \tau_{11}^{AB} + \tau_{12}^{AB} + \tau_{21}^{AB} + \tau_{22}^{AB} + e_{nqr}$$

to the reduced model $\rho_{nqr} = e_{nqr}$. We are therefore testing the null hypothesis, $\mathcal{H}_0 : \tau_{11}^{AB} = \tau_{12}^{AB} = \tau_{21}^{AB} = \tau_{22}^{AB} = 0$ i.e. that all the 'simple' interactions are zero. See Figure 13.7 for an example with a 3-by-3 design.

Numerical example

Pooled error

Consider a 2×2 ANOVA of the same data used in the previous examples, with $K_1 = K_2 = 2$, $P = K_1 K_2 = 4$, $N = 12$, $J = PN = 48$. The design matrix for Eqn. **13.15** with a pooled error term is the same as that in Figure 13.5, assuming that the four columns/conditions are ordered:

$$\begin{matrix} 1 & 2 & 3 & 4 \\ A_1B_1 & A_1B_2 & A_2B_1 & A_2B_2 \end{matrix} \qquad \mathbf{13.19}$$

where A_1 represents the first level of factor A, B_2 represents the second level of factor B etc, and the rows are ordered; all subjects data for cell A_1B_1; all for A_1B_2 etc. The basic contrasts for the three experimental effects are shown in Table 13-4 with the contrast weights for the subject-effects in the remaining columns 5–16 set to 0.

Assuming sphericity, the resulting F-tests give the ANOVA results in Table 13-5. With a Greenhouse-Geisser correction for non-sphericity, on the other hand, ϵ is estimated as 0.7, giving the ANOVA results in Table 13-6.

Main effects are not really meaningful in the presence of a significant interaction. In the presence of an interaction, one does not normally report the main effects, but proceeds by testing the differences between the levels of one factor for each of the levels of the other factor in the interaction (so-called *simple effects*). In this case, the presence of a significant interaction could be used to justify further simple effect contrasts (see above), e.g. the effect of B at the first and second levels of A are given by the contrasts $c = [1, -1, 0, 0]^T$ and $c = [0, 0, 1, -1]^T$.

Equivalent results would be obtained if the design matrix were *rotated* so that the first three columns reflect the experimental effects plus a constant term in the fourth column (only the first four columns would be rotated). This is perhaps a better conception of the ANOVA approach, since it is closer to Eqn. **13.15**, reflecting the

TABLE 13-4 Contrasts for experimental effects in a two-way ANOVA

Main effect of A	[1	1	−1	−1]
Main effect of B	[1	−1	1	−1]
Interaction, A × B	[1	−1	−1	1]

TABLE 13-5 Results of 2 × 2 within-subjects ANOVA with pooled error assuming sphericity

Main effect of A	F = 9.83	DF = [1, 33]	$p = 0.004$
Main effect of B	F = 5.21	DF = [1, 33]	$p = 0.029$
Interaction, A × B	F = 5.64	DF = [1, 33]	$p = 0.024$

TABLE 13-6 Results of 2 × 2 within-subjects ANOVA with pooled error using Greenhouse-Geisser correction

Main effect of A	F = 9.83	DF = [0.7, 23.0]	$p = 0.009$
Main effect of B	F = 5.21	DF = [0.7, 23.0]	$p = 0.043$
Interaction, A × B	F = 5.64	DF = [0.7, 23.0]	$p = 0.036$

conception of factorial designs in terms of the experimental effects rather than the individual conditions. This rotation is achieved by setting the new design matrix:

$$X_r = X \begin{bmatrix} C^T & 0_{4,12} \\ 0_{12,4} & I_{12} \end{bmatrix} \quad \mathbf{13.20}$$

where

$$C^T = \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \mathbf{13.21}$$

Notice that the rows of C^T are identical to the contrasts for the main effects and interactions plus a constant term (cf. Table 13-4). This rotated design matrix is shown in Figure 13.8. The three experimental effects can now be tested by the contrasts weight $[1, 0, 0, 0]^T$, $[0, 1, 0, 0]^T$, $[0, 0, 1, 0]^T$ (again, padded with zeros).

In this example, each factor only has two levels which results in one-dimensional contrasts for testing main effects and interactions. The contrast weights form a *vector*. But factors with more than two levels require multidimensional contrasts. Main effects, for example, can be expressed as a linear combination of differences between succesive levels (e.g. between levels 1 and 2, and 2 and 3). The contrast weights therefore form a *matrix*. An example using a 3-by-3 design is given later on.

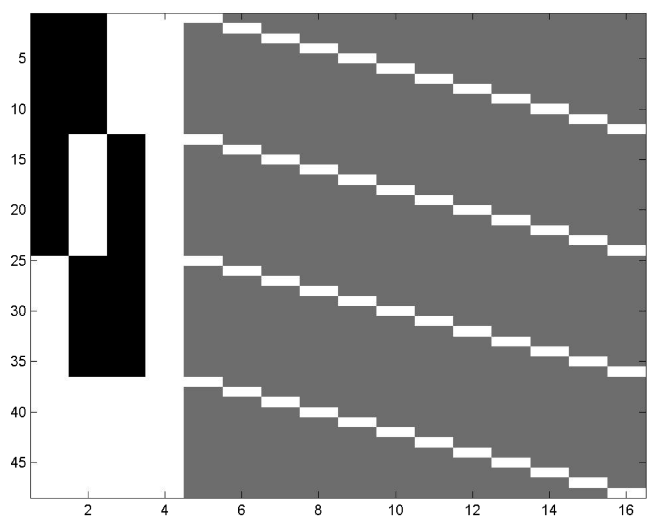

FIGURE 13.8 Design matrix for 2 × 2 within-subjects ANOVA. This design is the same as in Figure 13.5 except that the first four columns are rotated. The rows are ordered all subjects for cell A_1B_1, all for A_1B_2 etc. White, grey and black represent 1, 0 and −1. The first four columns model the main effect of A, the main effect of B, the interaction between A and B and a constant term. The last 12 columns model subject effects. This model is a GLM instantiation of Eqn. **13.15**.

TABLE 13-7 Results of ANOVA using partitioned errors

Main effect of A	F = 12.17	DF = [1, 11]	$p = 0.005$
Main effect of B	F = 11.35	DF = [1, 11]	$p = 0.006$
Interaction, A × B	F = 3.25	DF = [1, 11]	$p = 0.099$

Partitioned errors

Partitioned error models can be implemented by applying contrasts to the data, and then creating a separate model (i.e. separate GLM analysis) to test each effect. In other words, a two-stage approach can be taken, as described in the previous chapter on random effects analysis. The first stage is to create contrasts of the conditions for each subject, and the second stage is to put these contrasts or 'summary statistics' into a model with a block-diagonal design matrix.

Using the example dataset, and analogous contrasts for the main effect of B and for the interaction, we get the results in Table 13-7. Note how (1) the degrees of freedom have been reduced relative to Table 13-5, being split equally among the three effects; (2) there is no need for a non-sphericity correction in this case (since $K_1 = K_2 = 2$, see above); and (3) the p-values for some of the effects have decreased relative to Tables 13-5 and 13-6, while those for the other effects have increased. Whether p-values increase or decrease depends on the nature of the data (particularly correlations between conditions across subjects), but in many real datasets partitioned error comparisons yield more sensitive inferences. This is why, for repeated-measures analyses, the partitioning of the error into effect-specific terms is normally preferred over using a pooled error (Howell, 1992). But the partitioned error approach requires a new model to be specified for every effect we want to test.

GENERALIZATION TO M-WAY ANOVAS

The above examples can be generalized to M-way ANOVAs. For a K_1-by-K_2-..-by-K_M design, there are

$$P = \prod_{m=1}^{M} K_m \quad \mathbf{13.22}$$

conditions. An M-way ANOVA has $2^M - 1$ experimental effects in total, consisting of M main effects plus $M!/(M-r)!r!$ interactions of order $r = 2 \dots M$. A 3-way ANOVA for example has three main effects (A, B, C), three second-order interactions ($A \times B$, $B \times C$, $A \times C$) and one third-order interaction ($A \times B \times C$). Or more generally, an M-way ANOVA has $2^M - 1$ interactions of order $r = 0 \dots M$, where a 0th-order interaction is equivalent to a main effect.

We consider models where every cell has its own coefficient (like Eqn. **13.13**). We will assume these conditions are ordered in a GLM so that the first factor *rotates* slowest, the second factor next slowest, etc, so that for a 3-way ANOVA with factors A, B, C:

$$\begin{matrix} 1 & 2 & \dots & K_3 & \dots & P \\ A_1B_1C_1 & A_1B_1C_2 & \dots & A_1B_1C_{K_3} & \dots & A_{K_1}B_{K_2}C_{K_3} \end{matrix} \qquad \mathbf{13.23}$$

The data are ordered all subjects for cell $A_1B_1C_1$, all subjects for cell $A_1B_1C_2$ etc.

The F-contrasts for testing main effects and interactions can be constructed in an iterative fashion as follows. We define initial component contrasts.[2]

$$C_m = 1_{K_m} \quad D_m = -\mathrm{diff}(I_{K_m})^T \qquad \mathbf{13.24}$$

where $diff(A)$ is a matrix of column differences of A (as in the Matlab function *diff*). So for a 2-by-2 ANOVA:

$$C_1 = C_2 = [1, 1]^T \quad D_1 = D_2 = [1, -1]^T \qquad \mathbf{13.25}$$

The term C_m can be thought of as the *common effect* for the mth factor and D_m as the *differential effect*. Then contrasts for each experimental effect can be obtained by the Kronecker products of C_ms and D_ms for each factor $m = 1 \dots M$. For a 2-by-2 ANOVA, for example, the two main effects and interaction are respectively:

$$\begin{aligned} D_1 \otimes C_2 &= [1 \quad 1 \quad -1 \quad -1]^T \\ C_1 \otimes D_2 &= [1 \quad -1 \quad 1 \quad -1]^T \\ D_1 \otimes D_2 &= [1 \quad -1 \quad -1 \quad 1]^T \end{aligned} \qquad \mathbf{13.26}$$

This also illustrates why an interaction can be thought of as a *difference of differences*. The product $C_1 \otimes C_2$ represents the constant term.

For a 3-by-3 ANOVA:

$$C_1 = C_2 = [1, 1, 1]^T \quad D_1 = D_2 = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}^T \qquad \mathbf{13.27}$$

and the two main effects and interaction are respectively:

$$D_1 \otimes C_2 = \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix}^T \qquad \mathbf{13.28}$$

$$C_1 \otimes D_2 = \begin{bmatrix} 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 \end{bmatrix}^T \qquad \mathbf{13.29}$$

$$D_1 \otimes D_2 = \begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix}^T \qquad \mathbf{13.30}$$

The four rows of this interaction contrast correspond to the four 'simple interactions' τ_{11}^{AB}, τ_{12}^{AB}, τ_{21}^{AB}, and τ_{22}^{AB} depicted in Figure 13.7. This reflects the fact that an interaction can arise from the presence of one or more simple interactions.

Two-stage procedure for partitioned errors

Repeated measures M-way ANOVAs with partitioned errors can be implemented using the following summary-statistic approach.

1 Set up first level design matrices where each cell is modelled separately as indicated in Eqn. **13.23**.
2 Fit first level models.
3 For the effect you wish to test, use the Kronecker product rules outlined in the previous section to see what F-contrast you'd need to use to test the effect at the first level. For example, to test for an interaction in a 3×3 ANOVA you'd use the F-contrast in Eqn. **13.30** (application of this contrast to subject n's data tells you how significant that effect is in that subject).
4 If the F-contrast in the previous step has R_c rows then, for each subject, create the corresponding R_c contrast images. For N subjects this then gives a total of $N\ R_c$ contrast images that will be modelled at the second-level.
5 Set up a second-level design matrix, $X_2 = I_{R_c} \otimes 1_N$. The number of conditions is R_c. For example, in a 3×3 ANOVA, $X_2 = I_4 \otimes 1_N$ as shown in Figure 13.9.
6 Fit the second-level model.
7 Test for the effect using the F-contrast $C_2 = I_{R_c}$.

For each effect we wish to test we must get the appropriate contrast images from the first level (step 3) and

[2] In fact, the contrasts presented here are incorrect. But we present them in this format for didactic reasons, because the rows of the resulting contrast matrices, which test for main effects and interactions, are then readily interpretable. The correct contrasts, which normalize row lengths, are given in Appendix 13.2. We also note that the minus sign is unnecessary. It makes no difference to the results but we have included it so that the contrast weights have the canonical form $[1, -1, \dots]$ etc. instead of $[-1, 1, \dots]$.

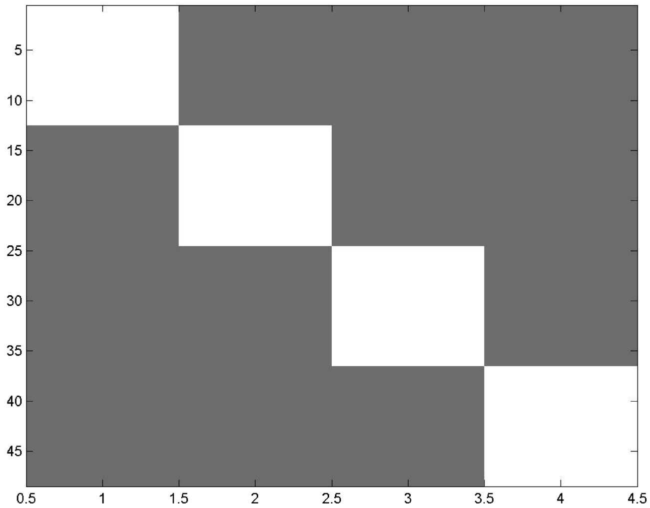

FIGURE 13.9 Second-stage design matrix for interaction in 3 × 3 ANOVA (partitioned errors).

implement a new second-level analysis (steps 4 to 7). Because we are taking *differential* effects to the second level we don't need to include subject effects at the second level.

fMRI BASIS FUNCTIONS

There are situations where one uses an 'ANOVA-type' model, but does not want to test a conventional main effect or interaction. One example is when one factor represents the basis functions used in an event-related fMRI analysis. So if one used three basis functions, such as a canonical haemodynamic response function (HRF) and two partial derivatives (see Chapter 14), to model a single event-type (versus baseline), one might want to test the reliability of this response over subjects. In this case, one would create for each subject the first-level contrasts: $[1, 0, 0]^T$, $[0, 1, 0]^T$ and $[0, 0, 1]^T$, and enter these as the data for a second-level 1-by-3 ANOVA, *without* a constant term.

In this model, we do not want to test for differences between the means of each basis function. For example, it is not meaningful to ask whether the parameter estimate for the canonical HRF differs from that for the temporal derivative. In other words, we do not want to test the null hypothesis for a conventional main effect, as described in Eqn. **13.9**. Rather, we want to test whether the sum of squares of the mean of each basis function explains significant variability relative to the total variability over subjects. This corresponds to the F-contrast:

$$c_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \textbf{13.31}$$

This is quite different from the F-contrast:

$$c_2 = \begin{bmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{bmatrix} \quad \textbf{13.32}$$

which is the default 'effects of interest' contrast given for a model that includes a constant term (or subject effects) in statistical parametric mapping (SPM), and would be appropriate instead for testing the main effect of such a 3-level factor.

DISCUSSION

The mainstay of many neuroimaging experiments is the factorial design and data from these experiments can be analysed using an analysis of variance. This chapter has described ANOVAs in terms of model comparison. To test, e.g. for a main effect of a factor, one compares two models, a 'full model' in which all levels of the factor are modelled separately, versus a 'reduced model', in which they are modelled together. If the full model explains the data significantly better than the reduced model then there is a significant main effect. We have shown how these model comparisons can be implemented using F-tests and general linear models.

This chapter has also revisited the notion of non-sphericity, within the context of within-subject ANOVAs. Informally, if a subject's response to levels i and j of a factorial manipulation is correlated, then a plot of the bivariate responses will appear non-spherical. This can be handled at the inferential stage by making an adjustment to the degrees of freedom. In current implementations of SPM this is generally unnecessary, as global non-sphericty estimates are used which have very high precision. This non-sphericity is then implicitly removed during the formation of maximum-likelihood parameter estimates (see Chapter 10).

We have also described inference in multiway within-subject ANOVAs and made a distinction between models with pooled versus partitioned errors and noted that partitioning is normally the preferred approach. One can implement partitioning using the multistage summary-statistic procedure until, at the last level, there is only one contrast per subject. This is a simple way to implement inference based on partitioned errors using the pooled-errors machinery of SPM.

APPENDIX 13.1 THE KRONECKER PRODUCT

If A is an $m_1 \times m_2$ matrix and B is an $n_1 \times n_2$ matrix, then the Kronecker product of A and B is the $(m_1 n_1) \times (m_2 n_2)$ matrix:

$$A \otimes B = \begin{bmatrix} a_{11}B & \ldots & a_{1m_2}B \\ \ldots & & \\ a_{m_1 1}B & & a_{m_1 m_2}B \end{bmatrix} \tag{13.33}$$

Circularity

A covariance matrix Σ is circular if:

$$\Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij} = 2\lambda \tag{13.34}$$

for all i, j.

Compound symmetry

If all the variances are equal to λ_1 and all the covariances are equal to λ_2 then we have *compound symmetry*.

Non-sphericity

If Σ is a $K \times K$ covariance matrix and the first $K-1$ eigenvalues are identically equal to:

$$\lambda = 0.5(\Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij}) \tag{13.35}$$

then Σ is *spherical*. Every other matrix is non-spherical or has *non-sphericity*.

Greenhouse-Geisser correction

For a 1-way ANOVA between subjects with N subjects and K levels the overall F statistic is approximately distributed as:

$$F\left[(K-1)\epsilon, (N-1)(K-1)\epsilon\right] \tag{13.36}$$

where

$$\epsilon = \frac{(\sum_{i=1}^{K-1} \lambda_i)^2}{(K-1)\sum_{i=1}^{K-1} \lambda_i^2} \tag{13.37}$$

and λ_i are the eigenvalues of the normalised matrix Σ_z where

$$\Sigma_z = M^T \Sigma_y M \tag{13.38}$$

and M is a K by $K-1$ matrix with orthogonal columns (e.g. the columns are the first $K-1$ eigenvectors of Σ_y).

APPENDIX 13.2 WITHIN-SUBJECT MODELS

The model in Eqn. **13.11** can also be written as:

$$y_n = 1_K \pi_n + \tau + e_n \tag{13.39}$$

where y_n is now the $K \times 1$ vector of measurements from the nth subject, 1_K is a $K \times 1$ vector of 1s, and τ is a $K \times 1$ vector with kth entry τ_k and e_n is a $K \times 1$ vector with kth entry e_{nk} where:

$$p(e_n) = \mathrm{N}(0, \Sigma_e) \tag{13.40}$$

We have a choice as to whether to treat the subject effects π_n as fixed-effects or random-effects. If we choose random-effects then:

$$p(\pi_n) = \mathrm{N}(\mu, \sigma_\pi^2) \tag{13.41}$$

and overall we have a mixed-effects model as the typical response for subject n, π_n, is viewed as a random variable whereas the typical response to treatment k, τ_k, is not a random variable. The reduced model is:

$$y_n = 1_K \pi_n + e_n \tag{13.42}$$

For the full model we can write:

$$p(y) = \prod_{n=1}^{N} p(y_n) \tag{13.43}$$

$$p(y_n) = \mathrm{N}(m_y, \Sigma_y)$$

and

$$m_y = 1_K \mu + \tau \tag{13.44}$$

$$\Sigma_y = 1_K \sigma_\pi^2 1_K^T + \Sigma_e$$

if the subject effects are random effects, and $\Sigma_y = \Sigma_e$ otherwise. If $\Sigma_e = \sigma_e^2 I_K$ then Σ_y has *compound symmetry*. It is also spherical (see Appendix 13.1). For $K = 4$ for example:

$$\Sigma_y = \begin{bmatrix} \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 \end{bmatrix} \tag{13.45}$$

If we let $\Sigma_y = (\sigma_\pi^2 + \sigma_e^2) R_y$ then:

$$R_y = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix} \tag{13.46}$$

where

$$\rho = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_e^2} \qquad \mathbf{13.47}$$

For a more general Σ_e, however, Σ_y will be non-spherical. In this case, we can attempt to correct for the non-sphericity. One approach is to reduce the degrees of freedom by a factor $\frac{1}{K-1} \leq \epsilon \leq 1$, which is an estimate of the degree of non-sphericity of Σ_y (the *Greenhouse-Geisser* correction; see Appendix 13.1). Various improvements of this correction (e.g. *Huhn-Feldt*) have also been suggested (Howell, 1992). Another approach is to parameterize explicitly the error covariance matrix Σ_e using a linear expansion and estimate the parameters using ReML, as described in Chapter 22.

Contrasts for M-way ANOVAs

The contrasts presented in the section 'Generalization to M-way ANOVAs' are actually incorrect. They were presented in a format that allowed the rows of the resulting contrast matrices, which test for main effects and interactions, to be readily interpretable. We now give the correct contrasts, which derive from speciying the initial differential component contrast as:

$$D_m = -\mathrm{orth}(\mathrm{diff}(I_{K_m})^T) \qquad \mathbf{13.48}$$

where orth(A) is the orthonormal basis of A (as in the Matlab function *orth*). This is identical to the expression in the main text but with the addition of an orth function which is necessary to ensure that the length of the contrast vector is unity.

This results in the following contrasts for the 2-by-2 ANOVA:

$$C_1 = C_2 = [1, 1]^T \quad D_1 = D_2 = [0.71, -0.71]^T \qquad \mathbf{13.49}$$

$$\begin{aligned} D_1 \otimes C_2 &= [0.71 \quad 0.71 \quad -0.71 \quad -0.71]^T \\ C_1 \otimes D_2 &= [0.71 \quad -0.71 \quad 0.71 \quad -0.71]^T \\ D_1 \otimes D_2 &= [0.71 \quad -0.71 \quad -0.71 \quad 0.71]^T \end{aligned} \qquad \mathbf{13.50}$$

For the 3-by-3 ANOVA:

$$C_1 = C_2 = [1, 1, 1]^T D_1 = D_2 = \begin{bmatrix} 0.41 & -0.82 & 0.41 \\ 0.71 & 0.00 & -0.71 \end{bmatrix}^T \qquad \mathbf{13.51}$$

and the two main effects and interaction are respectively:

$$D_1 \otimes C_2 = \begin{bmatrix} 0.41 & 0.41 & 0.41 & -0.82 & -0.82 & -0.82 & 0.41 & 0.41 & 0.41 \\ 0.71 & 0.71 & 0.71 & 0 & 0 & 0 & -0.71 & -0.71 & -0.71 \end{bmatrix}^T \qquad \mathbf{13.52}$$

$$C_1 \otimes D_2 = \begin{bmatrix} 0.41 & -0.82 & 0.41 & 0.41 & -0.82 & 0.41 & 0.41 & -0.82 & 0.41 \\ 0.71 & 0 & -0.71 & 0.71 & 0 & -0.71 & 0.71 & 0 & -0.71 \end{bmatrix}^T \qquad \mathbf{13.53}$$

$$D_1 \otimes D_2 = \begin{bmatrix} 0.17 & -0.33 & 0.17 & -0.33 & 0.67 & -0.33 & 0.17 & -0.33 & 0.17 \\ 0.29 & 0 & -0.29 & -0.58 & 0 & 0.58 & 0.29 & 0 & -0.29 \\ 0.29 & -0.58 & 0.29 & 0 & 0 & 0 & -0.29 & 0.58 & -0.29 \\ 0.5 & 0 & -0.5 & 0 & 0 & 0 & -0.5 & 0 & 0.5 \end{bmatrix}^T \qquad \mathbf{13.54}$$

REFERENCES

Howell DC (1992) *Statistical methods for psychology*. Duxbury Press, 1992.

Winer BJ, Brown DR, Michels KM (1991) *Statistical principles in experimental design*. McGraw-Hill,