

ANOVAs and SPM

R. Henson⁽¹⁾ and W. Penny⁽²⁾,

(1) Institute of Cognitive Neuroscience,

(2) Wellcome Department of Imaging Neuroscience,
University College London.

July 12, 2005

Abstract

This note describes ANOVAs and how they can be implemented in SPM. For each type of ANOVA we show what are the relevant statistical models and how they can be implemented in a GLM. We give examples of how main effects and interactions can be tested for using F-contrasts in SPM. We also show how to implement a repeated-measures M-way ANOVA with partitioned errors using a two-level procedure in SPM.

1 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 (*main effects* and *interactions* - see below).

The definitive reference for ANOVAs is Winer et al. [6]. Some different types of ANOVA are tabulated below. 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 eg. 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 eg. 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 section 3.2).

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 show what are the relevant statistical models and how they can be implemented in a GLM. We also give examples of how main effects and interactions can be tested for using F-contrasts in SPM.

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, \dots, K_M	M-way ANOVA	M-way ANOVA within-subject

Table 1: *Types of ANOVA*

1.1 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, 1_K is a $K \times 1$ vector of 1’s, 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 m th factor has $k = 1..K_m$ levels.

2 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 n th subject (y_n) is modelled as

$$y_n = \tau_k + \mu + e_n \quad (1)$$

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 n th subject is assigned to the k th group eg. $g(13) = 2$ 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

$$y_n = \mu + e_n \quad (2)$$

where we just view all of the measurements as random variation about the grand mean. Figure 1 compares these two models on some 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 Appendix A). We refer

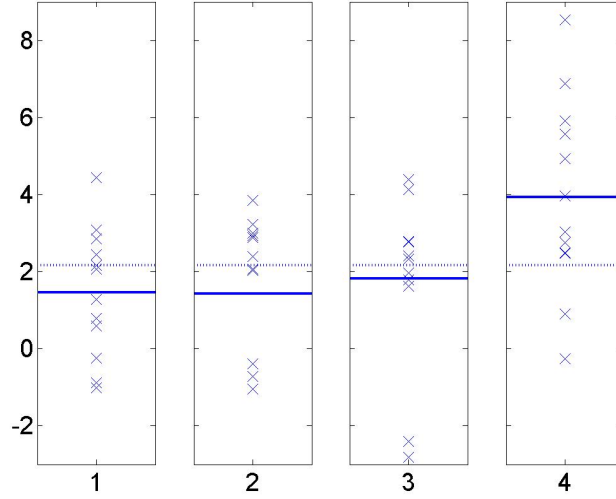


Figure 1: **One-way between-subject ANOVA.** 48 subjects are assigned to one of four groups. The plot shows the data points for each of the four conditions (crosses), the predictions from the ‘one-way between-subjects model’ or the ‘full model’ (solid lines) and the predictions from the ‘reduced model’ (dotted lines). In the reduced model (equation 2) we view the data as random variation about a grand mean. In the full model (equation 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 2.

to Equation 1 as the ‘full’ model and Equation 2 as the ‘reduced’ model. If RSS denotes the residual sum of squares (ie. the sum of squares left after fitting a model) then

$$F = \frac{(RSS_{reduced} - RSS_{full}) / (K - 1)}{RSS_{full} / (N - K)} \quad (3)$$

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

$$F = \frac{SS_{treat} / DF_{treat}}{SS_{error} / DF_{error}} \quad (4)$$

where

$$\begin{aligned} SS_{treat} &= RSS_{reduced} - RSS_{full} & (5) \\ DF_{treat} &= K - 1 \\ SS_{error} &= RSS_{full} \\ DF_{error} &= N - K \\ DF_{total} &= DF_{treat} + DF_{error} = N - 1 \end{aligned}$$

Expressions 3 and 4 are therefore equivalent.

2.1 Example Implementation in SPM

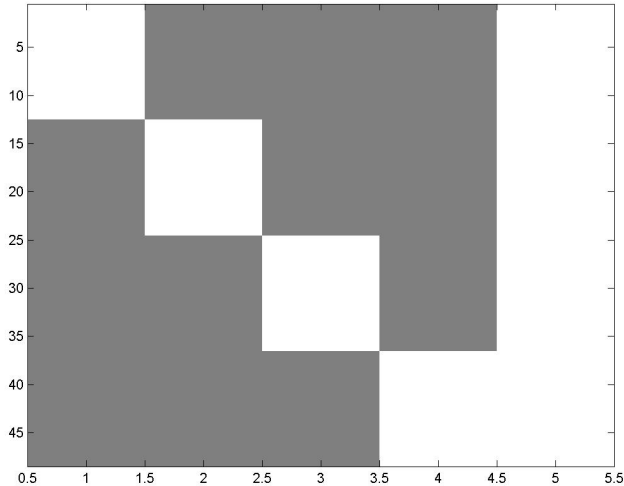


Figure 2: *Design matrix for one-way (1x4) between-subjects ANOVA. White and gray represent 1 and 0. There are 48 rows, one for each subject ordered by condition, and 5 columns, the first 4 for group effects and the 5th for the grand mean.*

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 equation 1 is

$$y = X\beta + e \quad (6)$$

where the design matrix $X = [I_K \otimes 1_n, 1_N]$ is shown in Figure 2, where \otimes denotes the Kronecker product (see Appendix B). The vector of parameters is $\beta = [\tau_1, \tau_2, \tau_3, \tau_4, \mu]^T$.

Equation 3 can then be implemented in SPM using the *effects of interest* F-contrast (see Appendix A.1)

$$C^T = \begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 & 0 \\ -1/3 & 1 & -1/3 & -1/3 & 0 \\ -1/3 & -1/3 & 1 & -1/3 & 0 \\ -1/3 & -1/3 & -1/3 & 1 & 0 \end{bmatrix} \quad (7)$$

or equivalently

$$C^T = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{bmatrix} \quad (8)$$

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

$$\mathcal{H}_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 \quad (9)$$

Main effect of treatment	F=3.62	DF=[3,44]	p=.02
--------------------------	--------	-----------	-------

Table 2: *Results of one-way (1x4) between-subjects ANOVA.*

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 in SPM and the classical formulation in equation 3 is detailed in Appendix A1.

We now analyse an example data set shown in Figure 1. The results of a one-way between-subjects ANOVA are in Table 2.

Note that the above design matrix is rank-deficient 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 (orthogonalised with respect to the fifth column), and β_0 is the parameter estimate for the corresponding fifth column, then

$$\begin{aligned}
 SS_{treatment} &= n\beta_1^T\beta_1 = 51.6 \\
 SS_{mean} &= nK\beta_0^2 = 224.1 \\
 SS_{error} &= r^T r = 208.9 \\
 SS_{total} &= y^T y = SS_{treatment} + SS_{mean} + SS_{error} = 484.5
 \end{aligned}
 \tag{10}$$

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

3 One-way within-subject ANOVA

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

$$y_{nk} = \tau_k + \pi_n + e_{nk} \tag{11}$$

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 3. It is also possible to express the full model in terms of differences between treatments (see eg. equation 14 for the two-way case).

To test whether the experimental factor is significant we compare the full model in equation 11 with the reduced model

$$y_{nk} = \pi_n + e_{nk} \tag{12}$$

An example of comparing these full and reduced models is shown in Figure 4. The equations for computing the relevant F-statistic and degrees of freedom are given, for example, in Chapter 14 of [3].

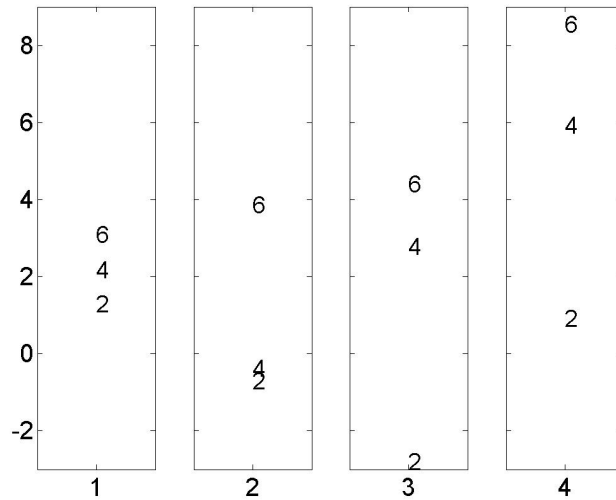


Figure 3: **Portion of example data for one-way within-subject ANOVA.** *The plot shows the data points for 3 subjects in each of 4 conditions (in the whole data set there are 12 subjects). Notice how subject 6's responses are always high, and subject 2's are always low. This motivates modelling subject effects as in equations 11 and 12.*

Main effect of treatment	F=6.89	DF=[3,33]	p=.001
--------------------------	--------	-----------	--------

Table 3: *Results of one-way (1x4) within-subjects ANOVA.*

3.1 Example Implementation in SPM

The design matrix $X = [I_K \otimes 1_N, 1_K \otimes I_N]$ for equation 11, with $K = 4$ and $N = 12$, is shown in Figure 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 equation 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 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 3. In fact this data set 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, in repeated measures designs, the modelling of subject effects normally increases the sensitivity of the inference.

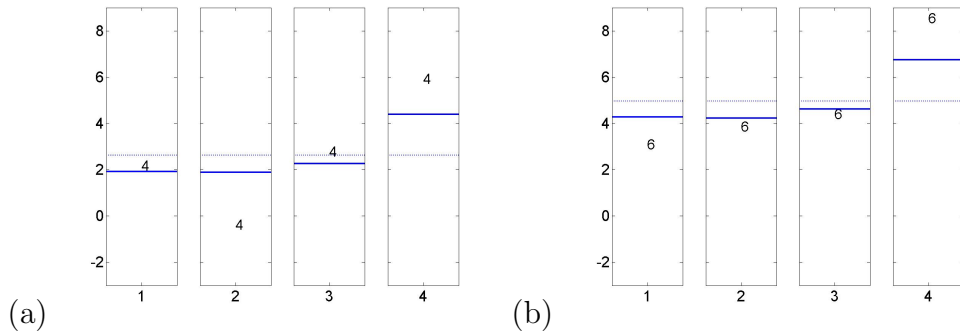


Figure 4: **One-way within-subjects ANOVA.** The plot shows the data points for each of the four conditions for subjects (a) 4 and (b) 6, the predictions from the one-way within-subjects model (solid lines) and the reduced model (dotted lines).

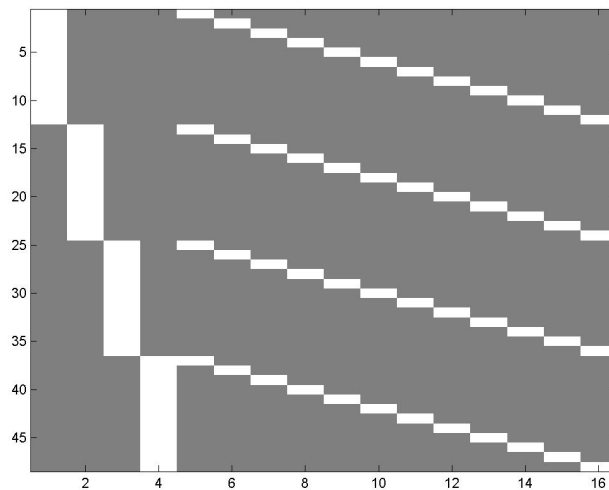


Figure 5: *Design matrix for one-way (1x4) within-subjects ANOVA.* The first 4 columns are treatment effects and the last 12 are subject effects.

3.2 Nonsphericity

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 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 characterised 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 is Independent and Identically Distributed (IID) (ie. same variance in all dimensions and there is no correlation). The more these contours look like ellipses, the more ‘nonsphericity’ there is in the data.

The possible nonsphericity can be taken into account in the analysis using a correction to the degrees of freedom. In the above example, a Greenhouse-Geisser (GG) correction estimates $\epsilon = .7$, giving DFs of [2.1, 23.0] and a p-value

(with GG we use the same F-statistic ie. $F = 6.89$) of $p = 0.004$. Assuming sphericity as in section 3.1 we computed $p = 0.001$. Thus the presence of nonsphericity 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 D. This shows how one can take into account nonsphericity. Various other relevant terminology is defined in Appendix C.

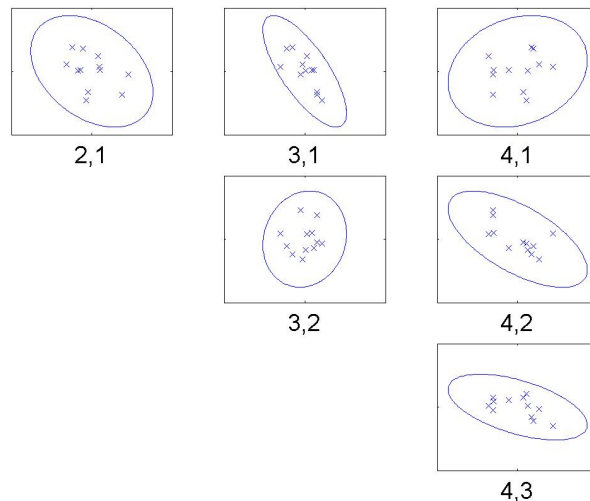


Figure 6: **One-way within-subjects ANOVA: Nonsphericity.** *Each subgraph plots each subjects response to condition i versus condition j as a cross. There are twelve crosses from 12 subjects. We also plot probability contours from the corresponding Gaussian densities. Subject responses, for example, to conditions 1 and 3 seem correlated - the sample correlation coefficient is -0.75 . Overall, the more non-spherical the contours the greater nonsphericity there is.*

4 Two-way within-subject ANOVAs

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

$$y_{nkl} = \tau_{kl} + \pi_n + e_{nkl} \quad (13)$$

where $k = 1 \dots K_1$ and $l = 1 \dots K_2$ index the levels of factor A and factor B respectively. Again, π_n are subject effects and e_{nkl} are residual errors.

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 section 5) has an associated interaction. Interactions represent the degree to which the effect of one factor depends on the levels of the other factors that comprise the interaction. A two-way ANOVA thus has two main effects and one interaction.

Equation 13 can be rewritten as

$$y_{nkl} = \tau_q^A + \tau_r^B + \tau_{qr}^{AB} + \pi_n + e_{nkl} \quad (14)$$

where τ_q^A represents the differences between each *successive* level $q = 1 \dots (K_1 - 1)$ of factor A, averaging over the levels of factor B; τ_r^B represents the differences between each *successive* level $r = 1 \dots (K_2 - 1)$ of factor B, averaging over the levels of factor A; and τ_{qr}^{AB} represents the differences between the differences of each level $q = 1 \dots (K_1 - 1)$ of factor A across each level $r = 1 \dots (K_2 - 1)$ of factor B. This rewriting corresponds to a rotation of the design matrix (see section 4.3). Again, π_n are subject effects and e_{nkl} are residual errors.

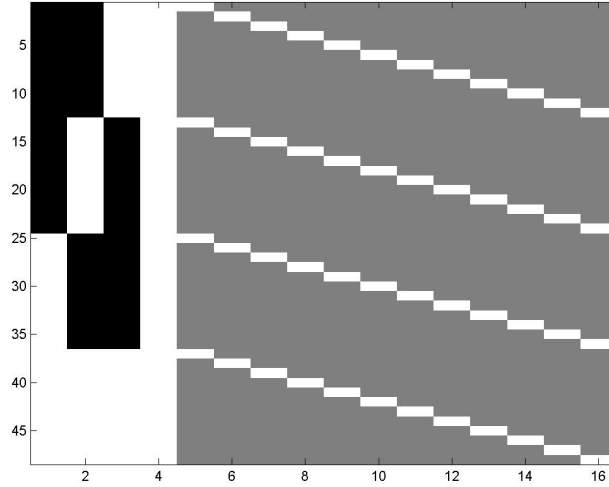


Figure 7: *Design matrix for 2x2 within-subjects ANOVA. This design is the same as in Figure 5 except that the first four columns are rotated. The rows are ordered all subjects for cell A_1B_1 , all for A_1B_2 etc. White, gray and black represent 1, 0 and -1. The first four columns model the main effect 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 equation 14.*

4.1 Pooled versus partitioned errors

In the above model, e_{nkl} 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

$$y_{nkl} = \tau_q^A + \tau_r^B + \tau_{qr}^{AB} + \pi_n + \rho_{nq}^A + \rho_{nr}^B + \rho_{nqr}^{AB} \quad (15)$$

in which the original residual error e_{nkl} has been split into three terms ρ_{nq}^A , ρ_{nr}^B and ρ_{nqr}^{AB} , each specific to a main effect or interaction. This is a different form of *variance partitioning*. Each such 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 1 is then

$$F = \frac{SS_k/DF_k}{SS_{nk}/DF_{nk}} \quad (16)$$

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 $m = 1 \dots M$), 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. This is not necessarily the case if a pooled error is used, as in equation 14.

4.2 Full and reduced models

The difference between pooled and partitioned error models is perhaps more conventionally expressed by specifying the relevant full and reduced models.

4.2.1 Pooled errors

The full model is given by equation 14. To test for the main effect of A we compare this to the reduced model

$$y_{nkl} = \tau_r^B + \tau_{qr}^{AB} + \pi_n + e_{nkl}$$

Similarly for the main effect of B. To test for an interaction we compare the full model to the reduced model

$$y_{nkl} = \tau_r^A + \tau_r^B + \pi_n + e_{nkl}$$

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. The full model is therefore

$$y_{nkl} = \tau_1^A + \tau_2^A + \tau_1^B + \tau_2^B + \tau_{11}^{AB} + \tau_{12}^{AB} + \tau_{21}^{AB} + \tau_{22}^{AB} + \pi_n + e_{nkl}$$

(there are 8 factor variables here, the 9th would be the grand mean). The reduced model used in testing for the main effect of A is

$$y_{nkl} = \tau_1^B + \tau_2^B + \tau_{11}^{AB} + \tau_{12}^{AB} + \tau_{21}^{AB} + \tau_{22}^{AB} + \pi_n + e_{nkl}$$

The reduced model used in testing for the interaction is

$$y_{nkl} = \tau_1^A + \tau_2^A + \tau_1^B + \tau_2^B + \pi_n + e_{nkl}$$

4.2.2 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 SPM. This set of differential effects for each subject is created using appropriate contrasts at the ‘first-level’. The models that we describe below are then at SPM’s ‘second-level’.

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 (see eg. section 5). 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 .

¹Although one could model inhomogeneity of variance.

		Factor B		
		Level 1	Level 2	Level 3
Factor A	Level 1	1	2	3
	Level 2	4	5	6
	Level 3	7	8	9

Figure 8: In a 3×3 ANOVA there are 9 cells or conditions. The numbers in the cells correspond to the ordering of the measurements when re-arranged 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 k th level of factor A, the l th level of factor B and for the n th 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 equation 32). 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 contrast matrix in equation 34.

Main effect of A	[1	1	-1	-1]
Main effect of B	[1	-1	1	-1]
Interaction, AxB	[1	-1	-1	1]

Table 4: *Contrasts for experimental effects in a two-way ANOVA.*

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, $\mathcal{H}_0 : \tau_{11}^{AB} = \tau_{12}^{AB} = \tau_{21}^{AB} = \tau_{22}^{AB} = 0$ ie. that all the ‘simple’ interactions are zero. See eg. figure 8.

4.3 Example Implementation in SPM

4.3.1 Pooled Error

Consider a 2x2 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 equation 14 with a pooled error term is the same as that in Figure 5, assuming that the four columns/conditions are ordered

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ A_1 B_1 & A_1 B_2 & A_2 B_1 & A_2 B_2 \end{array} \quad (17)$$

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 $A_1 B_1$, all for $A_1 B_2$ etc. The basic contrasts for the three experimental effects are shown in Table 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 5. With a Greenhouse-Geisser correction for nonsphericity, on the other

Main effect of A	F=9.83	DF=[1,33]	p=.004
Main effect of B	F=5.21	DF=[1,33]	p=.029
Interaction, AxB	F=5.64	DF=[1,33]	p=.024

Table 5: *Results of 2x2 within-subjects ANOVA with pooled error assuming sphericity.*

Main effect of A	F=9.83	DF=[0.7,23.0]	p=.009
Main effect of B	F=5.21	DF=[0.7,23.0]	p=.043
Interaction, AxB	F=5.64	DF=[0.7,23.0]	p=.036

Table 6: *Results of 2x2 within-subjects ANOVA with pooled error using Greenhouse-Geisser correction.*

hand, ϵ is estimated as .7, giving the ANOVA results in Table 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 equation 14, reflecting the 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

$$X_r = X \begin{bmatrix} C^T & 0_{4,12} \\ 0_{12,4} & I_{12} \end{bmatrix} \quad (18)$$

where

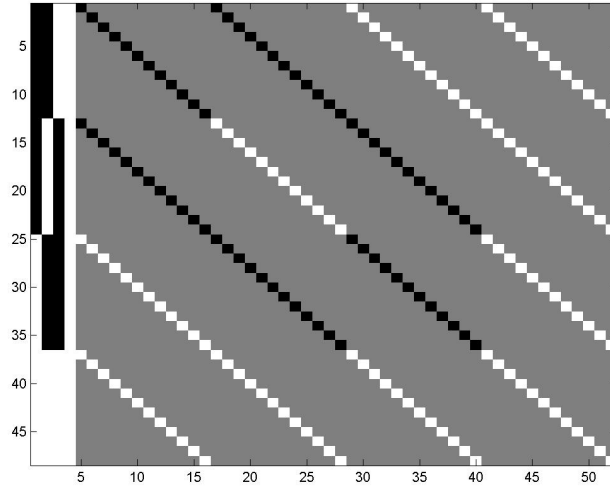
$$C^T = \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad (19)$$

Notice that the rows of C^T are identical to the contrasts for the main effects and interactions plus a constant term (cf. Table 4). This rotated design matrix is shown in Figure 7. The three experimental effects can now be tested by the contrasts $[1, 0, 0, 0]^T$, $[0, 1, 0, 0]^T$, $[0, 0, 1, 0]^T$ (again, padded with zeros).

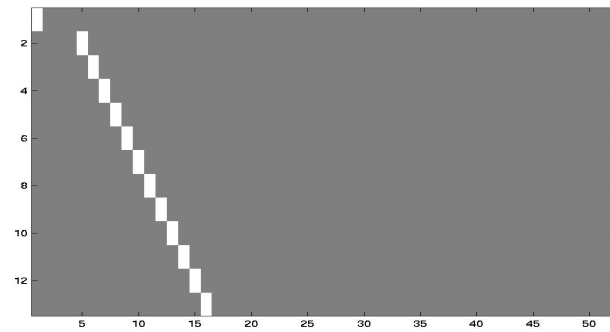
4.3.2 Partitioned errors

The design matrix for the model with partitioned errors in equation 15 is shown in Figure 9. This is created by $X_f = [C \otimes 1_N, C \otimes I_N]$ to give the model

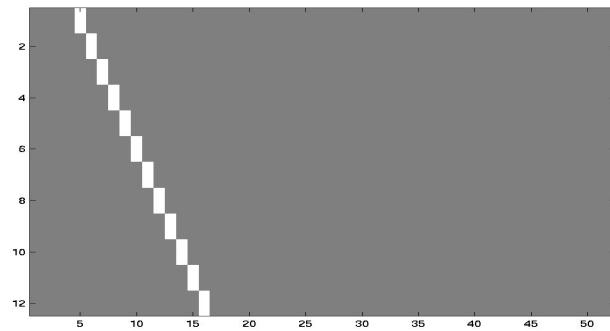
$$y = X_f \beta \quad (20)$$



(a)



(b)



(c)

Figure 9: 2×2 within-subjects ANOVA with partitioned errors (a) design matrix, (b) contrast matrix for main effect of A in the full model, C_1^T and (c) contrast matrix for main effect in the reduced model, C_0^T . White, gray and black correspond to 1, 0 and -1.

with no residual error.

The parameters corresponding to this design matrix could not actually be estimated in SPM, because there is no residual error (ie, $\text{rank}(X_f) = J$). However, we show in Appendix A.2 how the relevant effects could still be calculated in principle using two F-contrasts to specify subspaces of X that represent two embedded models to be compared with an F-ratio. We next show how the same effects could be tested in SPM in practice using a two-stage procedure.

4.3.3 Two-stage procedure

Even though equation 15 cannot be implemented in a single SPM analysis one can obtain identical results by applying contrasts to the data, and then creating a separate model (ie. separate SPM analysis) to test each effect. In other words, a two-stage approach can be taken [2], in which the first stage is to create contrasts of the conditions for each subject, and the second stage is to put these contrasts into a model with a block-diagonal design matrix (actually the *one-way ANOVA (without a constant term)* option in SPM).

The equivalence of this two-stage procedure to the multiple F-contrasts approach in Appendix A.2 is seen by projecting the data onto the subspace defined by the matrix

$$C = c \otimes I_N \quad (21)$$

where c is the contrast of interest (e.g, main effect of A, $[1, 1, -1, -1]^T$), such that

$$y_2 = C^- y \quad (22)$$

Continuing the 2-by-2 ANOVA example (eg. in Figure 7), y_2 is an N-by-1 vector. This vector comprises the data for the second-stage model

$$y_2 = X_2 \beta_2 + e_2 \quad (23)$$

in which the design matrix $X_2 = 1_N$, a column of 1's, is the simplest possible block-diagonal matrix (the *one-sample t-test* option in SPM). This is interrogated by the F-contrast $c_2 = [1]$.

The equivalence of this second-stage model to that specified by the appropriate subspace contrasts for the model $y = X_f \beta$ in equation 20 is because $C^- X_f = C_1$, where C_1 is identical to that in equation 40. This means

$$y_2 = C^- y = C^- X_f \beta = C_1 \beta \quad (24)$$

Given that

$$C_1^T = \begin{bmatrix} c & 0_{1,J} \\ 0_{N,P} & c \otimes I_N \end{bmatrix} \quad (25)$$

comparison of equations 23 and 24 reveals that β_2 is equal to the appropriate one of the first three elements of β (depending on the specific effect tested by contrast c) and that e_2 is comprised of the remaining nonzero elements of $C_1 \beta$, which are the appropriate interactions of that effect with subjects.

Using the example dataset, and analogous contrasts for the main effect of B and for the interaction, we get the results in Table 7. Note how (1) the degrees of freedom have been reduced relative to table 5, being split equally among the three effects, (2) there is no need for a nonsphericity correction in this case

Main effect of A	F=12.17	DF=[1,11]	p=.005
Main effect of B	F=11.35	DF=[1,11]	p=.006
Interaction, AxB	F=3.25	DF=[1,11]	p=.099

Table 7: *Results of ANOVA using partitioned errors.*

(since $K_1 = K_2 = 2$, see section 4.1, and (3) the p-values for some of the effects have decreased relative to tables 5 and 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[3].

5 Generalisation to M-way ANOVAs

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

$$P = \prod_{m=1}^M K_m \quad (26)$$

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 3 main effects (A, B, C), three second-order interactions (AxB, BxC, AxC) and one third-order interaction (AxBxC). 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 Equation 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

$$\begin{matrix} 1 & 2 & \dots & K_3 & \dots & P \\ A_1 B_1 C_1 & A_1 B_1 C_2 & \dots & A_1 B_1 C_{K_3} & \dots & A_{K_1} B_{K_2} C_{K_3} \end{matrix} \quad (27)$$

The data is ordered all subjects for cell $A_1 B_1 C_1$, all subjects for cell $A_1 B_1 C_2$ etc.

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

$$C_m = \mathbf{1}_{K_m} \quad D_m = \text{orth}(\text{diff}(I_{K_m})^T) \quad (28)$$

where $\text{diff}(A)$ is a matrix of column differences of A (as in the Matlab function *diff*) and $\text{orth}(A)$ is the orthonormal basis of A (as in the Matlab function *orth*). So for a 2-by-2 ANOVA

$$C_1 = C_2 = [1, 1]^T \quad D_1 = D_2 = [1, -1]^T \quad (29)$$

(ignoring overall scaling of these matrices). C_m can be thought of as the *common effect* for the m th factor and D_m as the *differential effect* for the m th

factor. Then contrasts for each experimental effect can be obtained by the Kronecker products of C_m 's and D_m 's for each factor $m = 1 \dots M$. For a 2-by-2 ANOVA for example the two main effects and interaction are respectively

$$\begin{aligned} D_1 \otimes C_2 &= [1 \quad 1 \quad -1 \quad -1]^T \\ C_1 \otimes D_2 &= [1 \quad -1 \quad 1 \quad -1]^T \\ D_1 \otimes D_2 &= [1 \quad -1 \quad -1 \quad 1]^T \end{aligned} \quad (30)$$

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

$$C_1 = C_2 = [1, 1, 1]^T \quad D_1 = D_2 = \begin{bmatrix} 0.82 & -0.41 & -0.41 \\ 0 & 0.71 & -0.71 \end{bmatrix}^T \quad (31)$$

and the two main effects and interaction are respectively

$$D_1 \otimes C_2 = \begin{bmatrix} 0.82 & 0.82 & 0.82 & -0.41 & -0.41 & -0.41 & -0.41 & -0.41 & -0.41 \\ 0 & 0 & 0 & 0.71 & 0.71 & 0.71 & -0.71 & -0.71 & -0.71 \end{bmatrix}^T \quad (32)$$

$$C_1 \otimes D_2 = \begin{bmatrix} 0.82 & -0.41 & -0.41 & 0.82 & -0.41 & -0.41 & 0.82 & -0.41 & -0.41 \\ 0 & 0.71 & -0.71 & 0 & 0.71 & -0.71 & 0 & 0.71 & -0.71 \end{bmatrix}^T \quad (33)$$

$$D_1 \otimes D_2 = \begin{bmatrix} 0.67 & -0.34 & -0.34 & -0.34 & 0.17 & 0.17 & -0.34 & 0.17 & 0.17 \\ 0 & 0.58 & -0.58 & 0 & -0.29 & 0.29 & 0 & -0.29 & 0.29 \\ 0 & 0 & 0 & 0.58 & -0.29 & -0.29 & -0.58 & 0.29 & 0.29 \\ 0 & 0 & 0 & 0 & 0.50 & -0.50 & 0 & -0.50 & 0.50 \end{bmatrix}^T \quad (34)$$

5.1 Two-stage procedure for partitioned errors

Repeated measures M-way ANOVAs with partitioned errors can be implemented in SPM using the following recipe.

1. Set up first level design matrices where each cell is modelled separately as indicated in equation 27.
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 equation 34 (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 NR_c contrast images that will be modelled at the second-level.
5. Set up a second-level design matrix, $X_2 = I_{R_c} \otimes 1_N$, obtained in SPM by selecting a *one-way ANOVA (without a constant term)*. The number of conditions is R_c . For example, in a 3x3 ANOVA, $X_2 = I_4 \otimes 1_N$ as shown in Figure 10.

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 implement a new 2nd level analysis (steps 4 to 7).

Note that, because we are taking *differential* effects to the second level we don't need include subject effects at the second level.

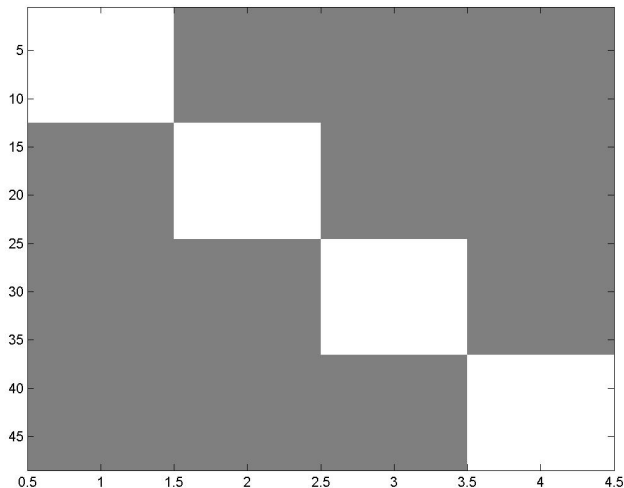


Figure 10: *Second-stage design matrix for interaction in 3x3 ANOVA (partitioned errors).*

6 Note on effects other than main effects and interactions

Finally, note that 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 SPM's canonical HRF and two partial derivatives, 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 equation 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:

$$c_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (35)$$

(which is the default "effects of interest" contrast given for this model in SPM). This is quite different from the F-contrast:

$$c_2 = \begin{bmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{bmatrix} \quad (36)$$

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

References

A The extra sum-of-squares principle

The statistical machinery behind analysis of variance is model comparison. In classical statistics this rests on the use of F-tests as described in the following material taken from [5].

The *extra sum-of-squares* principle provides a method of assessing general linear hypotheses, and for comparing models in a hierarchy, where inference is based on a F-statistic. Here, we will describe the classical F-test based on the assumption of an independent identically distributed error.

We first describe the classical F-test as found in nearly all introductory statistical texts. After that we will point at two critical limitations of this description and derive a more general and better suited implementation of the F-test for typical models in neuroimaging.

Suppose we have a model with parameter vector β that can be partitioned into two, $\beta = [\beta_1, \beta_2]^T$, and suppose we wish to test $\mathcal{H}_0 : \beta_1 = 0$. The corresponding partitioning of the design matrix X is $X = [X_1, X_2]$, and the *full model* is:

$$y = [X_1, X_2] \begin{bmatrix} \beta_1 \\ \cdots \\ \beta_2 \end{bmatrix} + \epsilon$$

which, when \mathcal{H}_0 is true, reduces to the *reduced model*: $Y = X_2\beta_2 + \epsilon$. Denote the residual sum-of-squares for the full and reduced models by $S(\beta)$ and $S(\beta_2)$ respectively. The *extra sum-of-squares* due to β_1 after β_2 is then defined as $S(\beta_1|\beta_2) = S(\beta_2) - S(\beta)$. Under \mathcal{H}_0 , $S(\beta_1|\beta_2) \sim \sigma^2\chi_p^2$ independent of $S(\beta)$, where the degrees of freedom are $\text{rank}(X) - \text{rank}(X_2)$. (If \mathcal{H}_0 is not true, then $S(\beta_1|\beta_2)$ has a non-central chi-squared distribution, still independent of $S(\beta)$.) Therefore, the following *F*-statistic expresses evidence against \mathcal{H}_0 :

$$F = \frac{\frac{S(\beta_2) - S(\beta)}{p - p_2}}{\frac{S(\beta)}{J - p}} \sim F_{p - p_2, J - p} \quad (37)$$

where $p = \text{rank}(X)$, $p_2 = \text{rank}(X_2)$ and J is the number of observations (length of y). The larger F gets, the more unlikely it is that F was sampled under the null hypothesis \mathcal{H}_0 . Significance can then be assessed by comparing this statistic with the appropriate F -distribution.

This formulation of the F -statistic has two limitations. The first is that two (nested) models, the full and the reduced model, have to be fitted to the data. In practice, this is implemented by a two-pass procedure on a, typically, large data set. The second limitation is that a partitioning of the design matrix into two blocks of regressors is not the only way one can partition the design matrix space. Essentially, one can partition X into two sets of linear combinations of the regressors. As an example, one might be interested in the difference between two effects. If each of these two effects is modelled by one regressor, a simple partitioning is not possible and one cannot use Eq. 37 to test for the difference. Rather, one has to re-parameterize the model such that the differential effect is explicitly modelled by a single regressor. As we will show in the following section this re-parameterization is unnecessary.

The key to implement an F-test that avoids these two limitations is the notion of contrast matrices. A contrast matrix is a generalisation of a contrast vector (see [4]). Each column of a contrast matrix consists of one contrast vector. Importantly, the contrast matrix controls the partitioning of the design matrix X .

A.1 Equivalence of contrast approach

A (user-specified) contrast matrix C is used to determine a subspace of the design matrix, i.e. $X_c = XC$. The orthogonal contrast to C is given by $C_0 = I_J - CC^-$. Then, let $X_0 = XC_0$ be the design matrix of the reduced model. We wish to compute what effects X_c explain, *after* first fitting the reduced model X_0 . The important point to note is that although C and C_0 are orthogonal to each other, X_c and X_0 are possibly not, because the relevant regressors in the design matrix X can be correlated. If the partitions X_0 and X_c are not orthogonal, the temporal sequence of the subsequent fitting procedure attributes their shared variance to X_0 . However, the subsequent fitting of two models is unnecessary, because one can construct a projection matrix from the data to the subspace of X_c , which is orthogonal to X_0 . We denote this subspace by X_1 .

The projection matrix M due to X_1 can be derived from the residual forming matrix of the reduced model X_0 . This matrix is given by $R_0 = I_J - X_0X_0^-$. The projection matrix is then $M = R_0 - R$, where R is the residual forming matrix of the full model, i.e. $R = I_J - XX^-$.

The F-statistic can then be written as

$$F = \frac{(My)^T My}{(Ry)^T Ry} \frac{J - p}{p_1} = \frac{y^T My}{y^T Ry} \frac{J - p}{p_1} \sim F_{p_1, J-p} \quad (38)$$

where p_1 is the rank of X_1 . Since M is a projector onto a subspace within X , we can also write

$$F = \frac{\hat{\beta}^T X^T M X \hat{\beta}}{y^T Ry} \frac{J - p}{p_1} \sim F_{p_1, J-p} \quad (39)$$

This equation means that we can conveniently compute a F-statistic for any user-specified contrast without any re-parameterization. In SPM, all F -statistics are based on the full model so that $y^T R y$ needs only to be estimated once and be stored for subsequent use.

In summary, the formulation of the F-statistic in Eq. 39 is a powerful tool, because by using a contrast matrix C we can test for a subspace spanned by contrasts of the design matrix X . Importantly, we do not need to reparameterise the model and estimate an additional parameter set, but can use estimated parameters of the full model. This is very important in neuroimaging because model parameters must be estimated at every voxel in the image, and typically there are 10^5 voxels.

A.2 F-contrasts to specify subspaces

To test, for example, the main effect of factor A, we can specify two contrasts that specify subspaces of X in Figure 9(a). The contrast for the *full* model of the main effect of A is

$$C_1^T = \begin{bmatrix} c & 0_{1,J} \\ 0_{N,P} & c \otimes I_N \end{bmatrix} \quad (40)$$

where $0_{N,P}$ is an $N \times P$ zero matrix and $c = [1, 0, 0, 0]$, while the contrast for the corresponding *reduced model* is

$$C_0^T = \begin{bmatrix} 0_{N,P} & c \otimes I_N \end{bmatrix} \quad (41)$$

These two contrasts are shown in Figures 9(b) and (c) respectively. We can then define

$$\begin{aligned} D_0 &= I_J - C_0 C_0^- \\ D_1 &= I_J - C_1 C_1^- \\ X_0 &= X D_0 \\ X_1 &= X D_1 \\ R_0 &= I_J - X_0 X_0^- \\ R_1 &= I_J - X_1 X_1^- \end{aligned} \quad (42)$$

with which (see Appendix A.1)

$$\begin{aligned} M &= R_0 - R_1 \\ DF &= [\text{rank}(X_1) - \text{rank}(X_0), J - \text{rank}(X_1)] \\ \hat{\beta} &= X_1^- y \\ F &= \frac{\hat{\beta}^T X_1^T M X_1 \hat{\beta}}{y^T R_1 y} \frac{(J - \text{rank}(X_1))}{\text{rank}(X_1) - \text{rank}(X_0)} \end{aligned} \quad (43)$$

B 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 $(m_1n_1) \times (m_2n_2)$ matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1m_2}B \\ \dots & & \\ a_{m_11}B & & a_{m_1m_2}B \end{bmatrix} \quad (44)$$

C Definitions

C.1 Circularity

A covariance matrix Σ is circular if

$$\Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij} = 2\lambda \quad (45)$$

for all i, j .

C.2 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*.

C.3 Nonsphericity

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

$$\lambda = 0.5(\Sigma_{ii} + \Sigma_{jj} - 2\Sigma_{ij}) \quad (46)$$

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

C.4 Greenhouse-Geisser correction

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

$$F[(K - 1)\epsilon, (N - 1)(K - 1)\epsilon] \quad (47)$$

where

$$\epsilon = \frac{(\sum_{i=1}^{K-1} \lambda_i)^2}{(K - 1) \sum_{i=1}^{K-1} \lambda_i^2} \quad (48)$$

and λ_i are the eigenvalues of the normalised matrix Σ_z where

$$\Sigma_z = M^T \Sigma_y M \quad (49)$$

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

D Within-subject models

The model in equation 11 can also be written as

$$y_n = 1_K \pi_n + \tau + e_n \quad (50)$$

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

$$p(e_n) = \mathbf{N}(0, \Sigma_e) \quad (51)$$

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

$$p(\pi_n) = \mathbf{N}(\mu, \sigma_\pi^2) \quad (52)$$

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

$$y_n = 1_K \pi_n + e_n \quad (53)$$

For the full model we can write

$$\begin{aligned} p(y) &= \prod_{n=1}^N p(y_n) \\ p(y_n) &= \mathbf{N}(m_y, \Sigma_y) \end{aligned} \quad (54)$$

and

$$\begin{aligned} m_y &= 1_K \mu + \tau \\ \Sigma_y &= 1_K \sigma_\pi^2 1_K^T + \Sigma_e \end{aligned} \quad (55)$$

of 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 C). For $K = 4$ for example

$$\Sigma_y = \begin{bmatrix} \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 & \sigma_\pi^2 \\ \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 & \sigma_\pi^2 + \sigma_e^2 \end{bmatrix} \quad (56)$$

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

$$R_y = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix} \quad (57)$$

where

$$\rho = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_e^2} \quad (58)$$

For a more general Σ_e , however, Σ_y will be non-spherical. In this case, we can attempt to correct for the nonsphericity. 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 nonsphericity of Σ_y (the *Greenhouse-Geisser* correction; see Appendix C.4). Various improvements of this correction (eg. *Huhn-Feldt*) have also been suggested (see [3]). Another approach is to explicitly parameterise the error covariance matrix Σ_e using a linear expansion and estimated using ReML (see [1] for a full treatment).