CHAPTER

# 9

# Contrasts and Classical Inference

*J. Poline, F. Kherif, C. Pallier and W. Penny*

## INTRODUCTION

The general linear model (GLM) characterizes the relationship between our experimental manipulations and observed data. It allows us to ask questions like: does frontal lobe activity in a memory task depend on age? Is the activity greater for normal subjects than for patients? While many questions concern only one effect (e.g. age, group), often our questions speak to multiple effects. In 1926, John Russel wrote 'An experiment is simply a question put to nature... Even in the best planned experiment the answer can simply be yes or no... The chief requirement is simplicity: only one question should be asked at a time',[**9.1] but R.A. Fisher's answer in his 1935 *Design of experiments* was: 'I am convinced that this view is wholly mistaken. If we ask Nature a single question, she will often refuse to answer until some other topic has been discussed'[**9.2]. In other words, we model several effects that may or may not influence our measures and ask several questions by comparing the relative importance of and interactions among those effects. This chapter explains how one models and tests for effects through the use of 'contrasts'. These enable us to focus on specific questions that are put to the data.

There is no unique model of an experimental paradigm. For example, in a functional imaging experiment with three conditions 'A', 'B' and 'C', the 'C' condition (say a 'baseline'[1] or low level condition) can be modelled explicitly or implicitly. This issue generalizes to more complex designs. Contrast specification and the interpretation of the ensuing results depend on model specification, which, in turn, depends on the design of the experiment. The most important step is the specification of the experimental paradigm: if a design is clearly thought through, the questions asked of the data are generally formulated easily and contrasts are straightforward to interpret.

In general, it is not very useful simply to show that the measured signal in a specific brain area is higher under one condition relative to another. Rather, we want to know whether this difference is statistically significant. We will therefore review the aspects of hypothesis testing that relate directly to the specification of contrasts.

This chapter is organized as follows. First, we review the theoretical background behind the construction of contrasts. In the next section, we describe the rules for constructing contrasts that specify *t*-tests. We then discuss *F*-contrasts and the important issue of correlations between predictors and their impact on the interpretation of *t*- or *F*-tests. We conclude with some general remarks and a summary.

## CONSTRUCTING MODELS

### What should be included in the model?

Put simply, the model should include all factors (continuous or discrete) that might have an impact on the measurements. Deciding what should or should not be included is crucial (for instance, in a functional magnetic resonance imaging (fMRI) model, should the subjects' movement estimates be included?). The question; 'should this factor be included in the model?' can be resolved with model selection, but *a-priori* knowledge is essential to limit the exploration of model space. With limited information about which factors influence the measured signal, the model will be larger and more complex.

[1] There is no absolute baseline condition. In fact, we generally only interpret the difference between two conditions, and therefore an activation pattern in neuroimaging is almost universally associated with at least two experimental conditions.

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

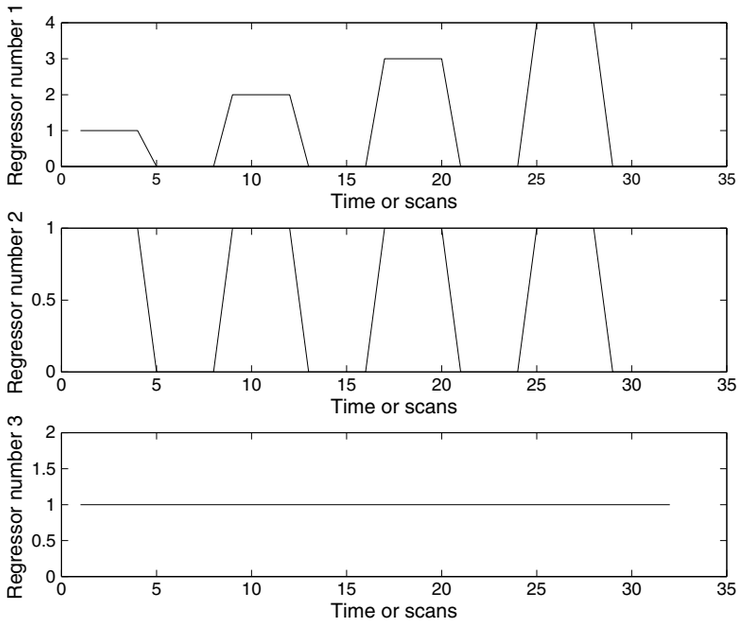


**FIGURE 9.1** Model-1: design with simple linear increase. The regressors, from top to bottom, model (i) the effects of a linear increase in force, (ii) the effect of force itself and (iii) the baseline response.

To make this point clear, consider an fMRI experiment looking at motor cortex responses when a subject presses a device with four different force levels: 'Press' conditions are interleaved with 'rest' periods. The conditions are ordered 'press force 1', 'rest', 'press force 2', 'rest', . . . , 'press force 4', etc.[2]

The first issue is how one models the 'press' and 'rest' conditions. One may have very specific prior assumptions, for example, that the response should be a *linear* function of the force. In this case, we construct a vector (a so-called *regressor, covariate*, or *predictor*) that represents this linear relationship. In the present example, this predictor could comprise 1s for all scans obtained during the first (lowest) force level, of 2s for all scans acquired during the second force level, etc. If the 'rest' periods are represented by zeros, the model assumes that the difference between rest and the first force level is the same as the difference between the first and the second force level (or between any two neighbouring force levels). To relax this assumption and construct a more flexible model, the difference between any 'press' condition and the rest period must be modelled explicitly in another predictor that takes value 1 during 'press' conditions and 0 during 'rest'.

Our model is then:

$$y_i = x_i^1\beta_1 + x_i^2\beta_2 + \epsilon_i \qquad \textbf{9.1}$$

for which $y_i$ is the $i$th measurement (scan), $x^1$ represents the predictor of the linear increase with force, and $x^2$ the difference between 'press' ($x_i^2 = 1$) and 'rest' ($x_i^2 = 0$). The parameters $\beta_1$ and $\beta_2$, which we need to estimate, are the coefficients of the linear functions encoded in our model. The error $\epsilon_i$ is the difference between the model prediction and the data $y_i$. If the signal is not zero during the rest condition (and this is always the case in neuroimaging), this offset has to be modelled by a constant term (i.e. a regressor consisting entirely of 1s). With this additional regressor, our model is written as:

$$y_i = x_i^1\beta_1 + x_i^2\beta_2 + 1\beta_3 + \epsilon_i \qquad \textbf{9.2}$$

in which $\beta_3$ represents the absolute offset of the data. Figure 9.1 shows an example for the three regressors from this model[3] which, throughout this chapter, we refer to as a 'linear parametric model' or simply 'model 1'. Note that this model may or may not provide a good explanation for the measured data. It may lack important predictors, or the measured response may not be a linear function of force. Two things can be done with this model once its parameters have been estimated. One can

[2] This order would not be used in an actual experiment, where one would normally randomize the different force levels.

[3] For models of fMRI data, one needs to take into account the delay and dispersion of the haemodynamic signal. This is usually done by convolving the regressors with a haemodynamic response function (see Chapter 8). Here, we have omitted this convolution step to concentrate on the modelling aspect.

make statistical inferences about its parameters (the $\beta$s), i.e. specify a contrast, and one can compare it with an alternative model.

### *Modelling the 'baseline'*

Should we add a predictor for the 'rest' periods to our model? This predictor could consist of 1 for scans during 'rest' and 0 for scans during all other conditions. This is not necessary because the difference between 'press' and 'rest' represented by predictor 2 ($x^2$) already encodes the difference between 'rest' and 'press'.

Given the model in Eqn. **9.2**, the following questions can be asked:

1 Does the measured response increase linearly with force, i.e. is $\beta_1$ significantly greater than zero?
2 Is there an additive offset for the 'press' condition that is not accounted for by the first predictor, i.e. is $\beta_2$ significantly greater than zero?
3 Is the signal during 'rest' above zero, i.e. is $\beta_3$ significantly greater than zero?

Note that the model in this example could be constructed differently, i.e. reparameterized, *while encoding exactly the same information*. For example, we could remove the average value of the first and second predictors ($x^1$ and $x^2$) so that their mean is zero. This operation is called 'mean centring'. This would not change the parameter estimates or interpretation of the first two predictors but would change the interpretation of the third predictor in this model (see below).

## Extending the first model

The assumption that the response increases linearly with force is a rather strong one. There are at least two ways in which this assumption can be relaxed.

First, the first covariate can be expanded using a Taylor-like expansion, such that not only linear but also higher-order (quadratic, cubic, etc.) increases are modelled. In this example, we restrict this expansion to second order, including a new regressor that is the square of the linear regressor. This results in a 'quadratic-parametric model' (model 2) which is shown in Figure 9.2.

Alternatively, one can choose a non-parametric form, enabling the model to capture any differences between the four force levels. This is achieved by representing each force level as a separate predictor. This 'non-parametric' model (model 3) is shown in Figure 9.3. Note that we would like to model two separate aspects of the data-first, the average activation over all force levels (the main effect of pressing). In model 3, this average can be computed from the sum of the different force levels. Second, we would like to model the differences between all pairs of neighbouring force levels, i.e. $(1-2)+(2-3)+(3-4)$. Modelling differences between levels is similar to modelling interactions in factorial designs (see Chapter 9.3). We therefore have the alternative choice to model the main effect and the interaction directly. This alternative model, model 4, is shown in Figure 9.4 (main effect and interactions). The questions that can be put to model 3 and model 4 are exactly the same, they just have to be 'rephrased' using appropriate contrasts.

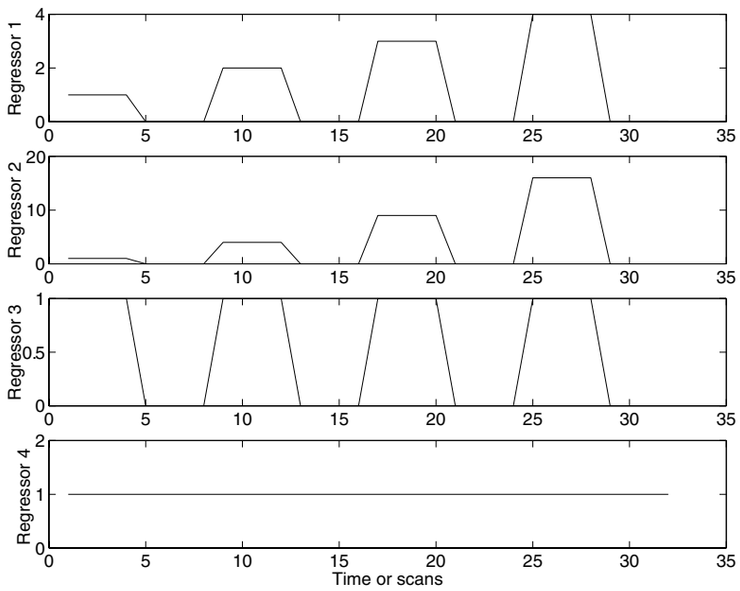


**FIGURE 9.2** Model-2: linear and quadratic increase covariates. Note the scale of the second covariate.

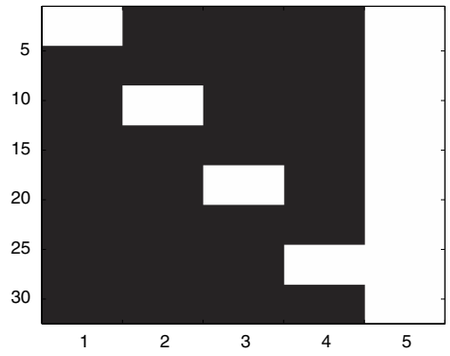


**FIGURE 9.3** Model-3: different force levels are modelled using separate covariates. Black is 0 and white is 1 on this panel.

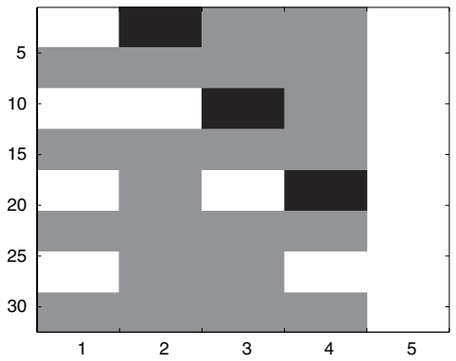


**FIGURE 9.4** Model-4: the main effect of force is modelled with the first regressor and the interactions are modelled with regressors 2 to 4.

The choice between parametric and non-parametric models often depends on the number of parameters that are required. If this number is large, then parametric models might be preferred. Relatively few parameters (compared to the number of data points) and limited prior information would speak to using non-parametric models that are more flexible.

For the parametric models, we might be interested in the following questions:

- Is there a linear increase or decrease in activation with force level (modelled by the first covariate)?
- Is there a quadratic change in activation with force level *additionally* to the linear variation (modelled by the second covariate)?
- Is there any linear or quadratic dependency of the response on force (a joint test on the first and second covariate)?

Note that in the parametric model, the linear and quadratic regressors are not uncorrelated and therefore influence each other's parameter estimates and statistical inference. Issues concerning correlated regressors or contrasts are reviewed later in this chapter.

For the non-parametric models, interesting questions might be:

- Is there an overall difference between force levels and the rest condition? This question can be addressed by means of the first four regressors in model 3 and the first regressor in model 4, respectively.
- Are there any differences between different force levels? This can be addressed by looking jointly at all differences in force levels versus rest in model 3 and at regressors 2 to 4 in model 4.
- Would it be possible to test for a linear dependency of the measured signal on force level? Because any differences between force levels have been modelled, it is possible (but not easy) to test for a *specific* linear increase.

These model specification questions are often framed in the following form: should conditions A and B be modelled separately, or should the common part of A and B (A + B) be modelled together with the difference (A − B)? Note that if there is no third condition (or implicit baseline) only (A − B) can be estimated from the data.

# CONSTRUCTING AND TESTING CONTRASTS

## Parameter estimation

We now turn to the issue of parameter estimation. As reviewed in depth in Chapter 8, the general linear model[4] rests on the equation:

$$Y = X\beta + \epsilon \qquad \textbf{9.3}$$

This equation models the data $Y$ (comprising $n$ measurements) as a linear combination of predictors which form the columns of the design matrix $X$. $X$ is of dimension $(n, p)$ and contains all effects $x^1, \ldots, x^p$ that are assumed to influence the measured data. The quantity $\epsilon$ is additive noise and has a normal distribution with zero mean and covariance $\sigma^2\Sigma_i$.

The model in Eqn. **9.3** states that the expectation of the data $Y$ is equal to $X\beta$. If the data cannot be modelled by a linear combination of the predictors in $X$ then the model is not appropriate and statistical results are difficult to interpret. This might occur if $X$ does not contain all effects

[4] Most of the notation used in this and Chapter 8 is identical but we also summarize notation in Appendix 9.1.

that influence the data, if it contains too many predictors that are unrelated to the data, or if the assumed linear relation between data and predictors does not hold.

A common method, used to solve the above equation, is called ordinary least squares (OLS).[5] OLS finds those parameter estimates $\hat{\beta}$ for which the sum of squared errors becomes minimal: $\| \epsilon \|^2 = \| Y - X\beta \|^2$.

This corresponds to finding a $\hat{\beta}$ such that $X\hat{\beta}$ is as close as possible to $Y$. This means that $X\hat{\beta}$ is the orthogonal projection of $Y$ onto $C(X)$, the vector space spanned by the columns of X (see Figure 9.15 for an illustration). Therefore, if $P_X$ is the orthogonal projection matrix (see Appendix 9.3) onto $C(X)$, $\hat{\beta}$ must satisfy:

$$P_X Y = X\hat{\beta}$$

This equation expresses the relationship between the parameters $\hat{\beta}$ and the data. For one-way analysis of variance ANOVA (Chapter 13), $P_X Y$ provides the means of the various groups, and the above equations describe the relationship between the $\hat{\beta}$ and these means (see below).

The matrix $P_X$ depends only on the space spanned by $X$s columns (i.e. C(X)). Therefore, two models with different design matrices $X_1$ and $X_2$ are equivalent if $C(X_1) = C(X_2)$: they explain the same aspects of the data ($X\beta$), have the same error components, and each contrast formulated for one model can be rephrased in the context of the other, such that it leads to the same statistical conclusions.

The parameters $\beta$ are estimated from the data using:

$$\hat{\beta} = (X^T X)^- X^T Y \quad \textbf{9.4}$$

where $X^-$ denotes the (Moore-Penrose) pseudoinverse of $X$. The fitted data $\hat{Y}$ are defined as:

$$\hat{Y} = X\hat{\beta} \quad \textbf{9.5}$$

and represent what is predicted by the model. The estimated noise (error) is:

$$Y - \hat{Y} = RY = \hat{\epsilon} \quad \textbf{9.6}$$

where

$$R = I_n - P_X \quad \textbf{9.7}$$

The noise variance is estimated with:

$$\hat{\sigma}^2 = Y^T R Y / \mathrm{tr}[R\Sigma_i] \quad \textbf{9.8}$$

Eqn. (**9.4**) has two important implications:

- Parameter estimates depend on the scaling of the regressors in $X$. This scaling is not important when a parameter estimate is compared to its standard deviation (see below). However, it is important if parameter estimates of different regressors are compared. When defined through statistical parametric mapping's (SPM) graphical user interface, regressors are appropriately scaled to ensure sensible comparisons.
- If X is not of full rank, there are infinitely many parameter vectors $\beta$ which solve the equation. In this case, estimation of $\hat{\beta}$ has a degree of arbitrariness and only some compounds will be meaningful. These are called *estimable* contrasts and are the subject of the next section.

## Estimability

One can appreciate that not all parameters may be estimable by looking at a model that contains the same regressor twice, say $x_1$ and $x_2 = x_1$ (with parameters $\beta_1$ and $\beta_2$). There is no information in the data on which to base the choice of $\hat{\beta}_1$ compared to $\hat{\beta}_2$. In this case, any solution of the form $\hat{\beta}_1 + \hat{\beta}_2 = constant$ will provide the same fitted data, the same residuals, but an infinity of solutions $\hat{\beta}_1$ and $\hat{\beta}_2$.

To generalize this argument, we can consider linear functions of the parameter estimates:

$$\lambda_1\hat{\beta}_1 + \cdots + \lambda_p\hat{\beta}_p = \lambda^T\hat{\beta} \quad \textbf{9.9}$$

The constants $\lambda_i$ are the coefficients of a function that 'contrasts' the parameter estimates. The vector $\lambda^T = [\lambda_1, \ldots, \lambda_p]$, where $p$ is the number of parameters in $X$, is referred to as the contrast vector. The word contrast is used for the result of the operation $\lambda^T\hat{\beta}$. A contrast is a random variable, because $\hat{\beta}$ is estimated from noisy data.

The matrix $X$ is said to be rank deficient or degenerate when (some of) the parameter estimates are not unique and therefore do not convey any meaning on their own. At first sight, this situation seems unlikely. However, many designs for position emission tomography (PET) data or population inference, are degenerate.

A contrast is estimable if (and only if) the contrast vector can be written as a linear combination of the rows of $X$. This is because the information about a contrast is obtained from combinations of the rows of $Y$. If no combination of rows of $X$ is equal to $\lambda^T$, then the contrast is not estimable.[6]

In more technical terms, the contrast $\lambda$ has to lie within the space of $X^T$, denoted by $\lambda \subset \mathcal{C}(X^T)$, or, equivalently, $\lambda$

[5] If the properties of the noise are known, the most efficient way to estimate the parameters is a maximum likelihood procedure. This entails whitening the noise.

[6] In Chapter 8, we define a contrast as an estimable function of the parameter estimates. If a linear combination of parameter estimates is not estimable then that linear combination is not a contrast. In this chapter, however, we often use the expression 'estimable contrast' for purposes of emphasis.

is unchanged when projected orthogonally onto the rows of $X$ (i.e. $P_{X^T}\lambda = \lambda$ with $P_{X^T}$ being the 'projector' onto $X^T$; see Appendix 9.3). The reason for this is as follows: if there is redundancy in $X$, for some linear combination $q$, we have $Xq = 0$. Therefore, $Y = X\beta + Xq + \epsilon = X(\beta + q) + \epsilon$. So, if we test $\lambda^T\beta$, we also test $\lambda^T(\beta + q)$, hence an estimable contrast $\lambda$ will satisfy $\lambda^T q = 0$. A necessary and sufficient condition for this is that $\lambda^T = vX$.

The SPM interface ensures that any specified contrast is estimable, hence offering protection against contrasts that would not make sense in degenerate designs. However, a contrast may be estimable but misinterpreted. In this chapter, we hope clarify the interpretation of contrasts.

## Three design matrices for a two sample t-test

The (unpaired) two sample $t$-test, comparing the mean of two groups, can be implemented in the linear model framework as follows. Consider an experiment with two groups of 2 (group 1) and 3 (group 2) subjects. In imaging experiments, these numbers will be larger (at least 10 or so). We have:

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

then

$$P_X Y = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \end{bmatrix} Y = X\beta = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_2 \\ \bar{y}_2 \end{bmatrix}$$

where $\bar{y}_i$ is the mean observation in group $i$. We will now describe two other parameterizations of the same model (such that the matrix $P_X$ is identical in all cases) and show how to specify meaningful contrasts.

Design matrix Parameters Contrasts

(1) $X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$ $\begin{cases} \hat{\beta}_1 = \bar{y}_1 \\ \hat{\beta}_2 = \bar{y}_2 \end{cases}$ $(1, 0)\hat{\beta} = \bar{y}_1$; $(0, 1)\hat{\beta} = \bar{y}_2$; $(1, -1)\hat{\beta} = \bar{y}_1 - \bar{y}_2$; $(.5, .5)\hat{\beta} = \text{mean}(\bar{y}_1, \bar{y}_2)$

(2) $X = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$ $\begin{cases} \hat{\beta}_1 + \hat{\beta}_2 = \bar{y}_1 \\ \hat{\beta}_2 = \bar{y}_2 \end{cases}$ $(1, 1)\hat{\beta} = \bar{y}_1$; $(0, 1)\hat{\beta} = \bar{y}_2$; $(1, 0)\hat{\beta} = \bar{y}_1 - \bar{y}_2$; $(.5, 1)\hat{\beta} = \text{mean}(\bar{y}_1, \bar{y}_2)$

(3) $X = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ $\begin{cases} \hat{\beta}_1 + \hat{\beta}_3 = \bar{y}_1 \\ \hat{\beta}_2 + \hat{\beta}_3 = \bar{y}_2 \end{cases}$ $(1, 0, 1)\hat{\beta} = \bar{y}_1$; $(0, 1, 1)\hat{\beta} = \bar{y}_2$; $(1, -1, 0)\hat{\beta} = \bar{y}_1 - \bar{y}_2$; $(.5, .5, 1)\hat{\beta} = \text{mean}(\bar{y}_1, \bar{y}_2)$

The only intuitive case is the first parameterization. In the two other cases, the interpretation of the parameter estimates is not obvious and the contrasts are not intuitive. In case 3, parameters are not estimable and not all contrasts are meaningful. Estimable contrasts are orthogonal to $[1\ 1\ -1]$, because column 1 plus column 2 equals column 3.

## Constructing and testing t-contrasts

If it is clear what the parameter estimates represent, then specification of contrasts is simple, especially in the case of $t$-contrasts. These contrasts are of the form described above, i.e. univariate linear combinations of parameter estimates. We return to our first model, which includes the four forces and 'rest' as regressors. For model 1, we can ask if there is a linear increase by testing $\beta_1$ using the combination $1\beta_1 + 0\beta_2 + 0\beta_3$ with the contrast vector $\lambda^T = [1\ 0\ 0]$. A linear decrease can be tested with $\lambda^T = [-1\ 0\ 0]$.

To test for the additive offset of the 'press' condition, not accounted for by the linear increase, we use $\lambda^T = [0\ 1\ 0]$. Note here that the linear increase is starting with a value of one for the first force level, and increases to 4 for the fourth level (see Figure 9.1).

When testing for the second regressor, *we are effectively removing that part of the signal that can be accounted for by the first regressor*. This means that the second parameter estimate is not the average of the difference between the 'press' conditions and the rest condition. To obtain the latter difference, we have to construct a re-parameterization of model 1 and replace the first regressor so that it models *only* differences of 'force levels' around an average difference between 'press' and 'rest'. This is achieved by orthogonalizing the first regressor with respect to the second. This new model, model 5, is shown in Figure 9.5. The parameter estimates of this new model are [10 30 100] as compared to [10 5 100] for model 1. This issue is detailed in Andrade *et al.* (1999) and an equivalent effect can be seen for $F$-tests. This emphasizes the principle that one should have in mind not only what is, but also what is *not*, tested by a contrast.

Another solution (useful in neuroimaging where estimating the parameters can be time consuming) is to

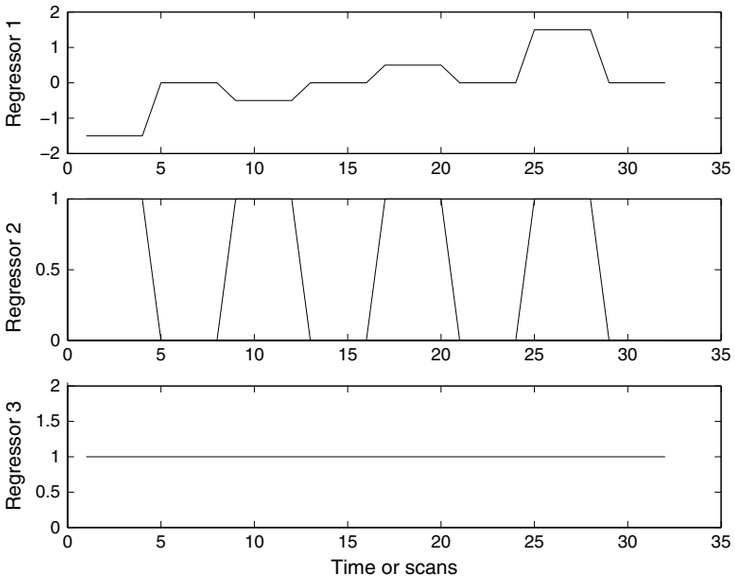


**FIGURE 9.5** Model-5 This is the same as Model-1 but the main effect of force has been removed from the first regressor. This changes the interpretation of the second regressor.

identify an equivalent contrast: the contrast vector $\lambda^T = [1\ 1\ 0]$ is valid but difficult to interpret. For example, the individual effects may be strong but, because they can have different signs, the overall effect may be weak.

For model 3 the average amplitude of the 'press' condition compared to 'rest' would be tested with $\lambda^T = [1\ 1\ 1\ 1\ 0]$. With model 4, the same effect can be tested with $\lambda^T = [1\ 0\ 0\ 0\ 0]$. The two contrasts give exactly the same *t*-maps. Note that in both cases the average over levels is tested, which could be significant just because of the effect of a single level.

An interesting question is whether we can test for the linearity of the response over the four levels. For model 3, the intuitive contrast to enter would be $\lambda^T = [1\ 2\ 3\ 4\ 0]$. This would indeed test for a linear increase with force level, but in a very unspecific manner; in the sense that the test might be significant in a situation where only the fourth condition has a greater signal than in rest condition. This is because we are testing for the weighted sum of the corresponding parameters. The test is valid, but does not ensure that the signal changes linearly with force. In other words, the model is flexible and we are testing a very restricted hypothesis, such that the shape of the predicted signal may be distinct from the shape of the component tested.

## Computing *t*-statistics

Whatever contrast is used, the contrast *t*-statistics are produced using (Friston *et al*., 1995; Worsley and Friston, 1995):

$$t_{df} = \lambda^T \hat{\beta} / \mathrm{SD}(\lambda^T \hat{\beta}) \qquad \textbf{9.10}$$

where SD($z$) denotes the standard deviation of $z$ and is computed as the square root of the variance:

$$\mathrm{var}[\lambda^T \hat{\beta}] = \hat{\sigma}^2 \lambda^T (X^T X)^- X^T \Sigma_i X (X^T X)^- \lambda \qquad \textbf{9.11}$$

For Gaussian errors, $t_{df}$ follows approximately a Student distribution with degrees of freedom given by $df = \mathrm{tr}[R\Sigma_i]^2 / \mathrm{tr}[R\Sigma_i R\Sigma_i]$. At the voxel level, the $p$-value of $t_{df}$ is computed using its null distribution.

The important point is that the standard deviation of the contrast depends on the matrix $X$. More specifically, when regressors are correlated, the variance of the corresponding parameter estimates increases. In other words, the precision of the estimation for one component is greater when other components in the model are not correlated. The dependence of the covariance of the estimated effects and the correlation within the model can be used, for instance, to optimize event-related designs.

The test of $t_{df}$ is one-tailed when testing exclusively for a positive (or negative) effect, and two-tailed when jointly testing for positive or negative effects.

# CONSTRUCTING AND TESTING *F*-CONTRASTS

In this section, we consider an experiment with two event-related conditions using the simple case of right and left motor responses. The subject is asked to press a button with the right or left hand with a visual instruction. The events arrive pseudo-randomly but with a long

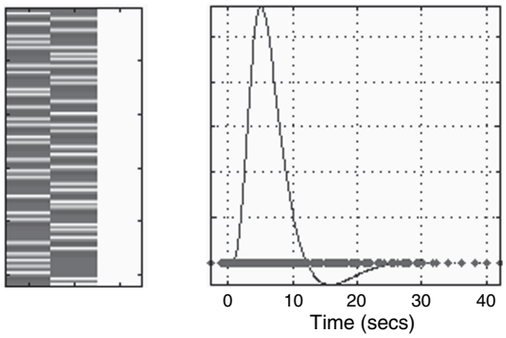


**FIGURE 9.6** The left panel shows the design matrix for analysing two event-related conditions (left or right motor responses). The shape of the HRF is assumed to be known, up to a scaling factor. The two first regressors have been constructed by convolution of a series of Dirac functions with the 'canonical' HRF (right panel).

inter-stimulus interval. We are interested in brain regions that are more activated for right versus left movements.

Our first model assumes that the shape of the haemodynamic response function (HRF) can be modelled by a 'canonical HRF' (see Chapter 14). This model is shown in Figure 9.6. To find brain regions that are more active for left versus right motor responses we can use $\lambda^T = [1 - 10]$. Using Eqn. **9.10** we can compute the *t*-map shown in Figure 9.7. This shows activation of contralateral motor cortex plus other typical regions, such as ipsilateral cerebellum.

Because there is an implicit baseline, the parameters are also interpretable individually, and when tested (*t*-maps not shown) they reveal the appropriate visual and motor regions.[7] Instead of having the two regressors encoding the left and right responses separately, an equivalent model could have the first regressor modelling the response common to right and left and the second modelling the difference between them.

The fact that the HRF varies across brain regions and subjects can be accommodated as follows. A simple extension of the model of Figure 9.6 is presented in Figure 9.8, for which each response is modelled with three basis functions. These functions can model small variations in the delay and dispersion of the HRF, as described in Chapter 14. They are mean centred, so the mean parameter will represent the overall average of the data.

In this new model, how do we test for the effects of, for instance, the right motor response? The most obvious approach is to test for all regressors modelling this response. This does not entail the sum (or average) of the parameter estimates because the sign of those parameter estimates is not interpretable, but rather the (weighted) sum of squares of those parameter estimates. The appropriate *F*-contrast is shown in Figure 9.9.

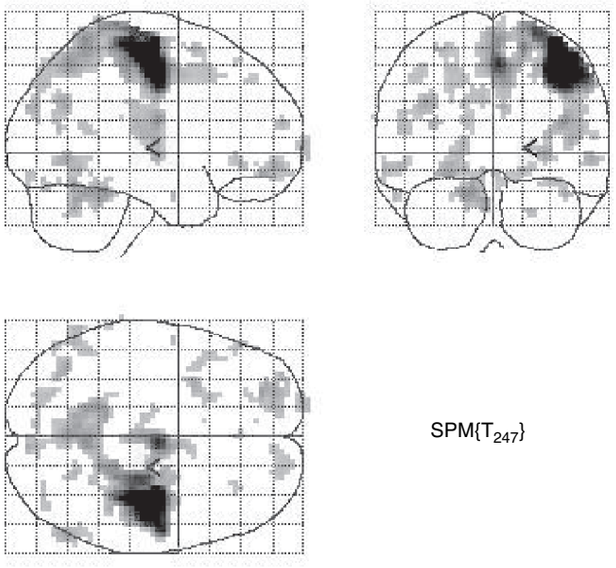


**FIGURE 9.7** SPM-t image corresponding to the overall difference between the left and right responses. This map was produced using the [1−10] contrast weights, using the model shown in Figure 9.6.

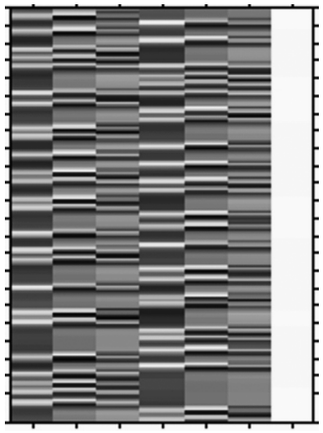

**FIGURE 9.8** The same model as in Figure 9.6, but we use three regressors to model each condition. The first three columns model the first condition (left motor response) while columns 4 to 6 model the second condition (right motor response). Each set of three regressors is the result of the convolution of the stimulus onsets with the canonical HRF and its derivatives with respect to time and dispersion.

[7] Interestingly, there is some ipsilateral activation in the motor cortex such that the 'left-right' contrast is slightly less significant in the motor regions than the 'left' [1 0 0] contrast.

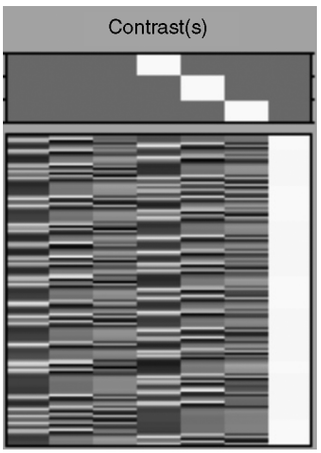


**FIGURE 9.9** An 'F-contrast' testing for the regressors modelling the right motor response. As described in the text, this corresponds to constructing the reduced model that does not contain the regressors that are 'marked' with the F-contrast.

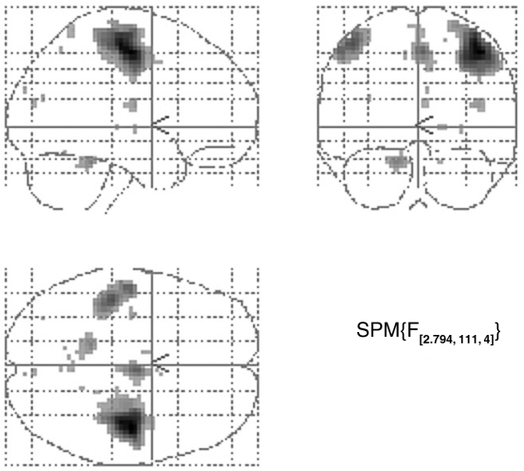


**FIGURE 9.11** SPM-F image corresponding to the overall difference between the left and right responses. This map was produced using the F-contrast in Figure 9.10 and the design matrix in Figure 9.8.

One interpretation of the *F*-contrast is that it is a series of one-dimensional contrasts, each testing the null hypothesis that the relevant parameter is zero. To test for the *overall* difference between right and the left responses we use the contrast shown in Figure 9.10. Note that multiplying the *F*-contrast coefficients by −1 does not change the statistic. The *F*-test image corresponding to this contrast is shown in Figure 9.11. This image is very similar to the corresponding image for the simpler model (Figure 9.12). Finally, Figure 9.13 shows that the more complex model provides a better fit to the data.

To conclude this section, we look at another example; a 2 by 3 factorial design. In this experiment, words are presented either visually (V) or aurally (A) and belong to three different categories (C1, C2, C3). In the design matrix, the six event-types are ordered as follows: V-C1 (presented visually and in category one), V-C2, V-C3, A-C1, A-C2, A-C3. We can then test for the interaction between the modality and category factors. We suppose that the experiment is a rapid event-related design with no implicit baseline, such that only comparisons between different event-types are meaningful. In the first instance, we model each event using a single basis function. A test for the main effect of modality is presented in Figure 9.14(a). Figure 9.14(b) shows the test for the main effect of category. Note that because there is no implicit baseline here, the main effects of factors are given by differences between levels. Finally, the interaction term would be tested for as in Figure 9.14(c).

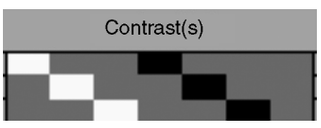


**FIGURE 9.10** F-contrast used to test the overall difference (across basis functions) between the left and right responses.

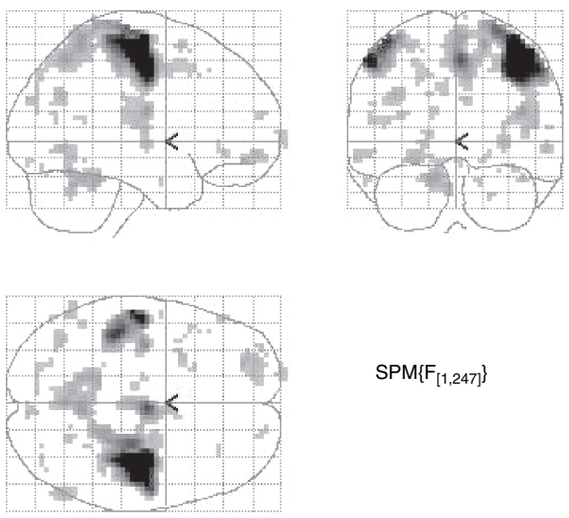


**FIGURE 9.12** SPM-F image corresponding to the overall difference (positive or negative) from the left and right responses. This map was produced with an F-contrast [1 0 0;0 1 0] using the model shown in Figure 9.6.

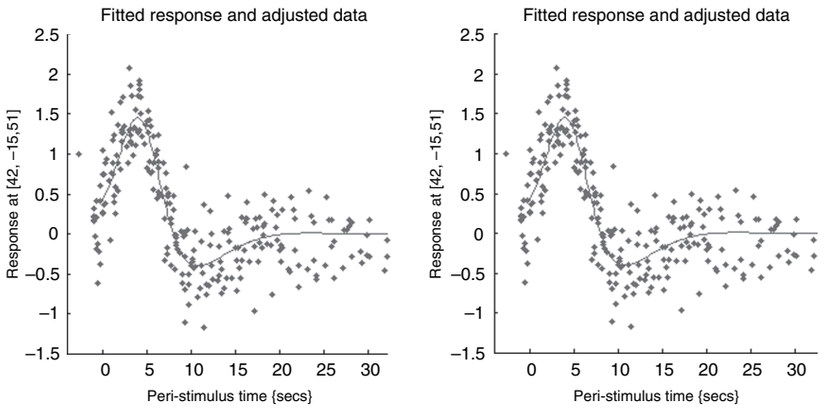


**FIGURE 9.13** Haemodynamic responses at a single voxel (the maxima of the SPM-F map in Figure 9.11). The left plot shows the HRF as estimated using the simple model (Figure 9.6) and demonstrates a certain lack of fit. The fit based on a more flexible model (Figure 9.8) is better (right panel).

The number of rows in an interaction contrast (without implicit baseline) is given by:

$$N_{rows} = \prod_{i=1}^{N}(l_i - 1) \qquad \textbf{9.12}$$

where $N$ is the number of factors and $l_i$ the number of levels of factor $i$.

## Interpretation of *F*-contrasts

There are two equivalent ways of thinking about h*F*-contrasts. For example, we can think about the *F*-contrast in Figure 9.9 as fitting a reduced model that does not contain the 'right motor response'. This reduced model would have a design matrix $X_0$ with zero entries in place of the 'right motor response' regressors of the 'full' design matrix $X$. The test then compares the variance of the residuals as compared to that of the full model $X$. The *F*-test simply computes the extra sum of squares that can be accounted for by inclusion of the three 'right hand' regressors. Following any statistical textbook (e.g., Christensen, 1996) and the work of Friston *et al*. (1995) and Worsley and Friston (1995), this is expressed by testing the following quantity:

$$F_{df_1,df_2} = \frac{(Y^T(I-P_{X_0})Y - Y^T(I-P_X)Y)/\nu_1}{Y^T(I-P_X)Y/\nu_2} \qquad \textbf{9.13}$$

with

$$\nu_1 = \mathrm{tr}((R_0 - R)\Sigma_i)$$
$$\nu_2 = \mathrm{tr}(R\Sigma_i) \qquad \textbf{9.14}$$

and

$$df_1 = \mathrm{tr}((R_0 - R)\Sigma_i(R_0 - R)\Sigma_i)/tr((R_0 - R)\Sigma_i)^2 \qquad \textbf{9.15}$$
$$df_2 = tr(R\Sigma_i R\Sigma_i)/\mathrm{tr}(R\Sigma_i)^2 \qquad \textbf{9.16}$$

where $R_0$ is the projector onto the residual space of $X_0$ and $P_X$ is the orthogonal projector onto $X$.

The second interpretation of the *F*-test is as a series of one-dimensional contrasts, each of them testing the null hypothesis that the respective contrast of parameters is zero.

We now show formally how these two interpretations are linked. The model in Eqn. **9.3**, $Y = X\beta + \epsilon$ is restricted by the test $c^T\beta = 0$ where $c$ is now a 'contrast matrix'. If $c$ yields an estimable function, then we can define a matrix $H$ such that $c = H^T X$. Therefore, $H^T X\beta = 0$ which, together with Eqn. **9.3**, is equivalent to $Y \subset \mathcal{C}(X)$ and $Y \subset \mathcal{C}(H^{\perp})$, the space orthogonal to $H$. It can be shown that the reduced model corresponding to this test is $X_0 =$

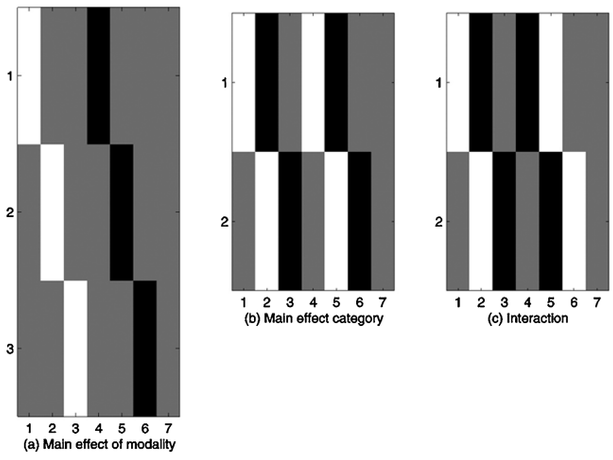


**FIGURE 9.14** F-contrasts testing respectively for (a) the main effect of modality, (b) the main effect of categories, and (c) the interaction modality × category.

$P_X - P_H$. This is valid if, and only if, the space spanned by $X_0$ is the space defined by $\mathcal{C}(H)^{\perp} \bigcap \mathcal{C}(X)$: it is easy to show that this is indeed the case.

If $\mathcal{C}(H) \subset \mathcal{C}(X)$, the numerator of Eqn. **9.13** can be rewritten as:

$$Y^T(R_0 - R)Y = Y^T(X_0 - R)Y = Y^T(P_X - X_0)Y = Y^T(P_H)Y \quad \textbf{9.17}$$

We choose $H$ such that it satisfies the condition above with $H = (X^T)^- c$, which yields:

$$\begin{aligned} Y^T(P_H)Y &= Y^T X(X^T X)^- X^T H(H^T H)^- H^T X(X^T X)^- X^T Y \\ &= \hat{\beta}^T c(H^T H)^- c^T \hat{\beta} \end{aligned} \quad \textbf{9.18}$$

This reformulation of the $F$-test is important for several reasons. First, it makes the specification and computation of $F$-tests feasible in the context of large data sets. Specifying a reduced model and computing the extra sum of squares using Eqn. **9.13** would be computationally too demanding. Second, it links the $t$-test and the test of a reduced model, and therefore makes it explicit that the 'extra' variability cannot be explained by the reduced model. Third, it makes the test of complex interactions using $F$-tests more intuitive.

The $F$-contrast that looks at the total contribution of all the 'right regressors' is, however, quite a non-specific test. One may have a specific hypothesis about the magnitude or the delay of the response and would like to test for this specifically. A reasonable test would be a $t$-test with contrast [0 0 0 1 0 0 0 0], testing for a positive value of the parameter that scales the standard HRF. This is perfectly valid, but it is not a test of the magnitude of the response. For instance, if the response has the shape implied by the standard model but is delayed significantly, the test might produce poor results, even if the delay is taken into account by the temporal derivative (Chapter 14). This may be important when comparing the magnitude of responses between two conditions: if the magnitudes are the same but the delays are different, across conditions, the test comparing the standard response regressors might be misinterpreted: a difference in delays might appear as a difference of magnitude *even if the basis functions are orthogonal to each other*.

Note that the simplest $F$-contrasts are unidimensional, in which case the $F$-statistic is simply the square of the corresponding $t$-statistic. To differentiate between unidimensional $F$-contrasts and $t$-contrasts in the SPM interface, the former are displayed in terms of images and the latter as bars.

An important point is that, generally, if we are confident about the shape of the expected response, $F$-tests are often less sensitive than $t$-tests. The reason is that, with increased model complexity, it becomes more likely that a signal of no interest could be captured by the $F$-contrast. The $F$-test implicitly corrects for this (Eqn. **9.13**), but this decreases sensitivity of the test, as compared to the more constrained $t$-test.

## CORRELATION BETWEEN REGRESSORS

Correlations among regressors can make the interpretation of tests difficult. Unfortunately, such correlation is often imposed by the brain's dynamics, experimental design or the method of measurement. The risks of misinterpretation have been extensively discussed in Sen and Srivastava (1990) and Andrade *et al.*, (1999). To summarize, one could miss activations when testing for a given contrast if there is a substantial correlation with the rest of the design. A frequently encountered example is when the response to a stimulus is highly correlated with a motor response

If one believes that a region's activity will not be influenced by the motor response, then it is advisable to test this specific region by first removing, from the motor response regressor, all that can be explained by the stimulus. This can be seen as a 'dangerous' procedure because if, in fact, the motor response does influence the signal in this region, then an 'activation' could be wrongly attributed to a stimulus-induced effect.

Because the issue of what is and what is not tested in a model is so important, we use two complementary perspectives that might shed light on it. First, from a geometrical perspective, the model is understood as some low-dimensional space; for purposes of visualization we choose a two-dimensonal space. The data lie in a greater 3D-space. The fitted data are an orthogonal projection of the data onto the model space (Figure 9.15). If the model space is spanned by two predictors $C1$ and $C2$, testing for $C2$ will, in effect, test for the part of $C2$ that is orthogonal to $C1$. If the two vectors are very similar (correlated), this

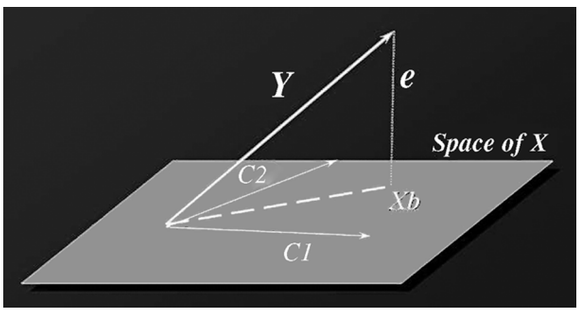


**FIGURE 9.15** Geometrical perspective: estimation. The data $Y$ are projected orthogonally onto the space of the design matrix ($X$) defined by two regressors $C1$ and $C2$. The error $e$ is the distance between the data and the smallest possible within the model space.

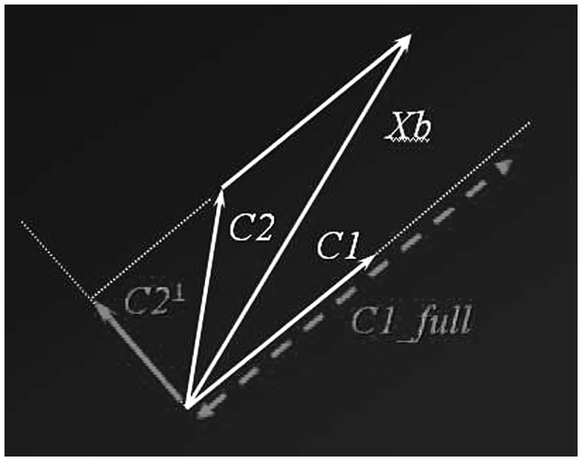


**FIGURE 9.16** Hypothesis testing: the geometrical perspective. With a model defined by the two regressors *C*1 and *C*2, testing for *C*2 in effect measures its part orthogonal to C1. If the model is explicitly orthogonalized, (i.e. *C*2 is replaced by $C2^{orth}$), the test of *C*2 is unchanged, but the test of *C*1 is, and will capture more variablity, as indicated by $C1_{full}$.

part can be very small. Explicit orthogonalization of *C*2 will make the effect tested by *C*1 appear much greater, while the effect tested by the $C2_{orth}$ is left unchanged (Figure 9.16).

A second perspective obtains from the following analogy. Let us consider a series of discs of different colours. Each disc represents a predictor, or more generally, a series of predictors in our model. Say we have two discs, a blue and a red one. The discs are placed on a table, where they might overlap. Testing for the effect of the first regressor would be analogous to measuring the surface of the blue disc that can be seen. If the two discs are non-overlapping (i.e. the regressors are not correlated), the two tests can be performed independently. But if the two discs do overlap (there is some correlation between the two regressors), testing for the blue disc corresponds to placing the red on top and measuring what remains of the blue. To put the blue on top amounts to orthogonalizing the red. Testing for the full surface of both discs corresponds to an *F*-test, and this does not depend on how the discs are placed on each other.

### Moving the variance across correlated regressors

If one decides that regressors, or a combination of regressors, should be orthogonalized with respect to some part of the design matrix, it is not necessary to reparameterize and fit the model again. Once the model has been fitted, all the information needed can be found in the fitted parameter estimates. For instance, instead of testing for the *additional* variance explained by a regressor, one may wish to test for all the variance that can be explained by this regressor. If $c$ is the contrast testing for the extra sum of squares, it is easy to show that the contrast matrix:

$$c_{Full\_space} = X^T X c \qquad \textbf{9.19}$$

tests for all the variance explained by the subspace of $X$ defined by $Xc$ since we then have $H = Xc$.

### Contrasts and reparameterized models

The above procedure can be generalized as follows: if the design matrix contains three subspaces say $(S_1, S_2, S_3)$, one may wish to test for what is in $S_1$, having removed what could be explained by $S_2$ (but not by $S_3$). Other examples are conjunction analyses, in which a series of contrasts can be modified such that the effects they test are orthogonal. This involves orthogonalizing the subsequent subspaces tested. The results may therefore differ depending on the order in which these contrasts are entered.

The principle for computing the same contrast in two different model parameterizations, which span the same space, is simple. If $X$ and $X_p$ are two differently parameterized versions of the same model then we can define a matrix $T$ such that $X_p = XT$. If $c_p$ is a test expressed in $X_p$ while the data have been fitted using $X$, the equivalent of $c_p$ using the parameter estimates of $X$ is

$$c = c_p (T^T X^T X T)^- T^T X^T X \qquad \textbf{9.20}$$

## DESIGN COMPLEXITY

Before acquiring neuroimaging data one should think about how to model them and which contrasts are of interest. Most of the problems concerning contrast specification derive from poor design specification. Poor designs may be unclear about the objectives pursued, include factors that are confounded, or may try to answer too many questions in a single experiment. This often leads to compromises and it can become difficult to provide clear answers to the questions of interest.

This does not preclude the use of a complex paradigm, in the sense that many conditions can and (often should be) included in the design. The process of recruiting subjects and acquiring data is long and costly, and it is only natural that one would like to answer as many questions as possible with the same data. However, this requires careful thought about which contrasts will be specified and whether they actually answer the question of interest.

## SUMMARY

In functional imaging experiments, one is often interested in many sorts of effects, e.g. the main effect of a factor and the possible interactions between factors. To analyse each of these effects one could fit several different GLMs and test hypotheses by looking at individual parameter estimates. However, this approach is impractical, because functional imaging data sets are very large. A more expedient approach is to fit larger models and test for effects using specific contrasts.

In this chapter, we have seen how the specification of the design matrix is intimately related to the specification of contrast weights. For example, it is often the case that main effects and interactions can be set up using parametric or non-parametric designs. These different designs lead to the use of different contrasts. Parametric approaches are favoured for factorial designs with many levels per factor. Contrasts must be estimable to be interpretable, and we have described the conditions for estimability.

In fMRI, one can model haemodynamic responses using the canonical HRF. This allows one to test for activations using *t*-contrasts. To account for the variability in the haemodynamic response, across subjects and brain regions, one can model the HRF using a canonical HRF plus its derivatives, with respect to time and dispersion. Inferences about differences in activation can then be made using *F*-contrasts. We have shown that there are two equivalent ways of interpreting *F*-contrasts, one employing the extra-sum-of-squares principle to compare the model and a reduced model, and one specifying a series of one-dimensional contrasts. Designs with correlations between regressors are less efficient and correlation can be removed by orthogonalizing one effect with respect to others. However, this may have a strong impact on the interpretation of subsequent tests. Finally, we have shown how such orthogonalization can be applied retrospectively, i.e. without having to refit the models.

In this chapter, we have focused on how to test for specific treatment effects encoded by the design matrix of the general linear model. However, the general linear model also entails assumptions about the random errors. In the next chapter, we examine these assumptions, in terms of covariance component and non-sphericity.

## APPENDIX 9.1 NOTATION

| | | |
|---|---|---|
| $Y$: | Data | The $(n, 1)$ time series, where $n$ is the number of time points or scans. $y_i$: one of those measures. |
| $c$ or $\lambda$: | Contrast weights | Linear combination of the parameter estimates used to form the (numerator) of the statistics |
| $X$: | Design matrix or design model | the $(n, p)$ matrix of regressors |
| $\beta$: | Model parameters | The true (unobservable) coefficients such that the weighted sum of the regressors is the expectation of our data (if $X$ is correct) |
| $\hat{\beta}$: | Parameter estimates | The computed estimation of the $\beta$ using the data $Y$: $\hat{\beta} = (X^T X)^- X^T Y$ |
| $C(X)$: | Vector space spanned by X | Given a model $X$, the vector space spanned by X are all vectors $v$ that can be written as $v = X\lambda$ |
| $P_X(X)$ or $M(X)$: | The orthogonal projector onto X | $P_X = X(X^T X)^- X^T$ |
| $R$: | Residual forming matrix | Given a model $X$, the residual forming matrix $R = I_n - XX^-$ transforms the data $Y$ into the residuals $r = RY$. |
| $\sigma^2\Sigma_i$: | scan (time) covariance | This $(n, n)$ matrix describes the (noise) covariance between scans |

## APPENDIX 9.2 SUBSPACES

Let us consider a set of $p$ vectors $x_i$ of dimension $(n, 1)$ (with $p < n$), e.g. regressors in fMRI. The space spanned by this set of vectors is formed from all possible vectors (say $u$) that can be expressed as a linear combination of the $x_i$: $u = \alpha_1 x_1 + \alpha_2 x_2 + \ldots \alpha_p x_p$. If the matrix $X$ is formed with the $x_i : X = [x_1 x_2 \ldots x_p]$, we denote this space as $\mathcal{C}(X)$.

Not all the $x_i$ may be necessary to form $\mathcal{C}(X)$. The minimal number needed is called the rank of the matrix $X$. If only a subset of the $x_i$ is selected, they form a smaller matrix $X_0$. The space spanned by $X_0$, $\mathcal{C}(X_0)$, is called a subspace of $X$. A contrast defines two subspaces of the design matrix $X$: one that is tested and one of 'no interest', corresponding to the reduced model.

## APPENDIX 9.3 ORTHOGONAL PROJECTION

The orthogonal projection of a vector $x$ onto the space of a matrix $A$ is the vector (e.g a time-series) that is closest in the space $C(A)$, where distance is measured as the sum of squared errors. The projector onto $A$, denoted $P_A$, is unique and can be computed with $P_A = AA^-$, with $A^-$ denoting the Moore-Penrose pseudoinverse[8] of $A$. For instance, the fitted data $\hat{Y}$ can be computed with

$$\hat{Y} = P_X Y = XX^- Y = X(X^T X)^- X^T Y = X\hat{\beta} \qquad \textbf{9.21}$$

Most of the operations needed when working with linear models only involve computations in parameter space, as is shown in Eqn. **9.18**. For a further gain in computational expediency, one can work with an orthonormal basis of the space of $X$, if the design is degenerate. This is how the SPM code is implemented.

## REFERENCES

Andrade A, Paradis A-L, Rouquette S *et al*. (1999) Ambiguous results in functional neuroimaging data analysis due to covariate correlation. *Neuroimage* **10**: 83–86

Christensen R (1996) *Plane answers to complex questions: the theory of linear models*. Springer, Berlin

Friston KJ, Holmes AP, Worsley KJ *et al*. (1995) Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* **2**: 189–210

Sen A, Srivastava M (1990) *Regression analysis – theory, methods, and applications*. Springer-Verlag, Berlin

Worsley KJ, Friston KJ (1995) Analysis of fMRI time-series revisited – again. *NeuroImage* **2**: 73–81

[8] Any generalized inverse could be used.