Chapter 12: Random Effects Analysis

W.D. Penny and A.J. Holmes

May 8, 2006

Introduction

In this chapter we are concerned with making statistical inferences involving many subjects. One can envisage two main reasons for studying multiple subjects. The first is that one may be interested in individual differences, as in many areas of psychology. The second, which is the one that concerns us here, is that one is interested in what is common to the subjects. In other words, we are interested in the stereotypical effect in the population from which the subjects are drawn.

As every experimentalist knows, a subject's response will vary from trial to trial. Further, this response will vary from subject to subject. These two sources of variability, within-subject (also called between-scan) and between-subject, must both be taken into account when making inferences about the population.

In statistical terminology, if we wish to take the variability of an effect into account we must treat it as a 'random effect'. In a 12-subject fMRI study, for example, we can view those 12 subjects as being randomly drawn from the population at large. The subject variable is then a random effect and, in this way, we are able to take the sampling variability into account and make inferences about the population from which the subjects were drawn. Conversely, if we view the subject variable as a 'fixed effect' then our inferences will relate only to those 12 subjects chosen.

The majority of early studies in neuroimaging combined data from multiple subjects using a 'Fixed-Effects' (FFX) approach. This methodology only takes into account the within-subject variability. It is used to report results as case studies. It is not possible to make formal inferences about population effects using FFX. Random-Effects (RFX) analysis, however, takes into account both sources of variation and makes it possible to make formal inferences about the population from which the subjects are drawn.

In this chapter we describe FFX and RFX analyses of multiple-subject data. We first describe the mathematics behind RFX, for balanced designs, and show how RFX can be implemented using the computationally efficient 'summary-statistic' approach. We then describe the mathematics behind FFX and show that it only takes into account within-subject variance. The next section shows that RFX for unbalanced designs is optimally implemented using the PEB algorithm described in the previous chapter. This section includes a numerical example which shows that, although not optimal, the summary statistic approach performs well even for unbalanced designs.

Random effects analysis

Maximum likelihood

Underlying RFX analysis is a probability model defined as follows. We first envisage that the mean effect in the population (ie. averaged across subjects) is of size w_{pop} and that the variability of this effect between subjects is σ_b^2. The mean effect for the ith subject (ie. averaged across scans), w_i, is then assumed to be drawn from a Gaussian with mean w_{pop} and variance σ_b^2. This process reflects the fact that we are drawing subjects at random from a large population. We then take into account the within-subject (ie. across scan) variability by modelling the jth observed effect in subject i as being drawn from a Gaussian with mean w_i and variance σ_w^2. Note that σ_w^2 is assumed to be the same for all subjects. This is a requirement of a balanced design. This two-stage process is shown graphically in Figure 1.

Given a data set of effects from N subjects with n replications of that effect per subject, the population effect is modelled by a two level process

$$\begin{aligned} y_{ij} &= w_i + e_{ij} \\ w_i &= w_{pop} + z_i \end{aligned} \tag{1}$$

where w_i is the true mean effect for subject i and y_{ij} is the jth observed effect for subject i, and z_i is the between subject error for the ith subject. These Gaussian errors have the same variance, σ_b^2. For the PET data considered below this is a differential effect, the difference in activation between word generation and word shadowing. The first equation captures the within-subject variability and the second equation the between-subject variability.

The within-subject Gaussian error e_{ij} has zero mean and variance $\mathsf{Var}[e_{ij}] = \sigma_w^2$. This assumes that the errors are independent over subjects and over replications within subject. The between-subject Gaussian error z_i has zero mean and variance $\mathsf{Var}[z_i] = \sigma_b^2$. Collapsing the two levels into one gives

$$y_{ij} = w_{pop} + z_i + e_{ij} \tag{2}$$

The maximum-likelihood estimate of the population mean is

$$\hat{w}_{pop} = \frac{1}{Nn} \sum_{i=1}^{N} \sum_{j=1}^{n} y_{ij} \tag{3}$$

We now make use of a number of statistical relations defined in the appendix to show that this estimate has a mean $\mathsf{E}[\hat{w}_{pop}] = w_{pop}$ and a variance given by

$$\begin{aligned} \mathsf{Var}[\hat{w}_{pop}] &= \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{n} \frac{1}{n}(w_{pop} + z_i + e_{ij})\right] \\ &= \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} z_i\right] + \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{n} \frac{1}{n} e_{ij}\right] \\ &= \frac{\sigma_b^2}{N} + \frac{\sigma_w^2}{Nn} \end{aligned} \tag{4}$$

The variance of the population mean estimate contains contributions from both the within-subject and between-subject variance.

Summary statistics

Implicit in the summary-statistic RFX approach is the two-level model

$$
\begin{aligned}
\bar{w}_i &= w_i + e_i \qquad (5)\\
w_i &= w_{pop} + z_i
\end{aligned}
$$

where w_i is the true mean effect for subject i, $\bar{w}_i$ is the sample mean effect for subject i and w_{pop} is the true effect for the population.

The Summary-Statistic (SS) approach is of interest because it is computationally much simpler to implement than the full random effects model of equation 1. This is because it is based on the sample mean value, $\bar{w}_i$, rather than on all of the samples y_{ij}. This is important for neuroimaging as in a typical functional imaging group study there can be thousands of images, each containing tens of thousands of voxels.

In the first level we consider the variation of the sample mean for each subject around the true mean for each subject. The corresponding variance is $\mathsf{Var}[e_i] = \sigma_w^2/n$, where σ_w^2 is the within-subject variance. At the second level we consider the variation of the true subject means about the population mean where $\mathsf{Var}[z_i] = \sigma_b^2$, the between-subject variance. We also have $\mathsf{E}[e_i] = \mathsf{E}[z_i] = 0$. Consequently

$$\bar{w}_i = w_{pop} + z_i + e_i \qquad (6)$$

The population mean is then estimated as

$$\hat{w}_{pop} = \frac{1}{N}\sum_{i=1}^{N} \bar{w}_i \qquad (7)$$

This estimate has a mean $\mathsf{E}[\hat{w}_{pop}] = w_{pop}$ and a variance given by

$$
\begin{aligned}
\mathsf{Var}[\hat{w}_{pop}] &= \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N}\bar{w}_i\right] \qquad (8)\\
&= \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} z_i\right] + \mathsf{Var}\left[\sum_{i=1}^{N} \frac{1}{N} e_i\right]\\
&= \frac{\sigma_b^2}{N} + \frac{\sigma_w^2}{Nn}
\end{aligned}
$$

Thus, the variance of the estimate of the population mean contains contributions from both the within-subject and between-subject variances. Importantly, both $\mathsf{E}[\hat{w}_{pop}]$ and $\mathsf{Var}[\hat{w}_{pop}]$ are identical to the maximum-likelihood estimates derived earlier. This validates the summary-statistic approach. Informally, the validity of the summary-statistic approach lies in the fact that what is brought forward to the second-level is a *sample* mean. It contains an element of within-subject variability which when operated on at the second level produces just the right balance of within and between subject variance.

Fixed effects analysis

Implicit in FFX analysis is a single-level model

$$y_{ij} = w_i + e_{ij} \qquad (9)$$

The parameter estimates for each subject are

$$\hat{w}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij} \tag{10}$$

which have a variance given by

$$\begin{aligned} \mathsf{Var}[\hat{w}_i] &= \mathsf{Var}\left[\sum_{j=1}^{n}\frac{1}{n}y_{ij}\right] \\ &= \frac{\sigma_w^2}{n} \end{aligned} \tag{11}$$

The estimate of the group mean is then

$$\hat{w}_{pop} = \frac{1}{N}\sum_{i=1}^{N}\hat{w}_i \tag{12}$$

which has a variance

$$\begin{aligned} \mathsf{Var}[\hat{w}_{pop}] &= \mathsf{Var}\left[\sum_{i=1}^{N}\frac{1}{N}\hat{w}_i\right] \\ &= \frac{1}{N}\mathsf{Var}[\hat{d}_i] \\ &= \frac{\sigma_w^2}{Nn} \end{aligned} \tag{13}$$

The variance of the fixed-effects group mean estimate contains contributions from within-subject terms only. It is not sensitive to between-subject variance. We are not therefore able to make formal inferences about population effects using FFX. We are restricted to informal inferences based on separate case studies or summary images showing the average group effect. This will be demonstrated empirically in a later section.

Parametric Empirical Bayes

We now return to RFX analysis. We have previously shown how the SS approach can be used for the analysis of balanced designs ie. identical σ_w^2 for all subjects. This section starts by showing how PEB can also be used for balanced designs. It then shows how PEB can be used for unbalanced designs and provides a numerical comparison between PEB and SS on unbalanced data.

Before proceeding we note that an algorithm from classical statistics, known as Restricted Maximum Likelihood (ReML), can also be used for variance component estimation. Indeed, many of the papers on random effects analysis use ReML instead of PEB [Friston et al. 2002, Friston et al. 2005].

The model described in this section is identical to the separable model in the previous chapter but with $x_i = 1_n$ and $\beta_i = \beta$. Given a data set of contrasts from N subjects with n scans per subject, the population effect can be modelled by the two level process

$$\begin{aligned} y_{ij} &= w_i + e_{ij} \\ w_i &= w_{pop} + z_i \end{aligned} \tag{14}$$

where y_{ij} (a scalar) is the data from the ith subject and the jth scan at a particular voxel. These data points are accompanied by errors e_{ij} with w_i being the size of the effect for subject i, w_{pop} being the size of the effect in the population and z_i being the between subject error. This may be viewed as a Bayesian model where the first equation acts as a likelihood and the second equation acts as a prior. That is

$$\begin{aligned} p(y_{ij}|w_i) &= \mathsf{N}(w_i, \sigma_w^2) \\ p(w_i) &= \mathsf{N}(w_{pop}, \sigma_b^2) \end{aligned} \tag{15}$$

where σ_b^2 is the between subject variance and σ_w^2 is the within subject variance. We can make contact with the hierarchical formalism of the previous chapter by making the following identities. We place the y_{ij} in the column vector y in the order - all from subject 1, all from subject 2 etc (this is described mathematically by the *vec* operator and is implemented in MATLAB (Mathworks, Inc.) by the colon operator). We also let $X = I_N \otimes 1_n$ where $\otimes$ is the Kronecker product and let $w = [w_1, w_2, ..., w_N]^T$. With these values the first level in equation 2 of the previous chapter is then the matrix equivalent of the first level in equation 14 (ie. it holds for all i, j). For $y = Xw + e$ and eg. $N = 3$,$n = 2$ we then have

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \end{bmatrix} \tag{16}$$

We then note that $X^T X = nI_N$, $\hat{\Sigma} = \mathsf{diag}(\mathsf{Var}[w_1], \mathsf{Var}[w_2], ..., \mathsf{Var}[w_N])$ and the ith element of $X^T y$ is equal to $\sum_{j=1}^n y_{ij}$.

If we let $M = 1_N$ then the second level in equation 2 of the previous chapter is then the matrix equivalent of the second-level in equation 14 (ie. it holds for all i). Plugging in our values for M and X and letting $\beta = 1/\sigma_w^2$ and $\alpha = 1/\sigma_b^2$ gives

$$\mathsf{Var}[\hat{w}_{pop}] = \frac{1}{N} \frac{\alpha + \beta n}{\alpha \beta n} \tag{17}$$

and

$$\begin{aligned} \hat{w}_{pop} &= \frac{1}{N} \frac{\alpha + \beta n}{\alpha \beta n} \frac{\alpha \beta}{\alpha + \beta n} \sum_{i,j} y_{ij} \\ &= \frac{1}{Nn} \sum_{i,j} y_{ij} \end{aligned} \tag{18}$$

So the estimate of the population mean is simply the average value of y_{ij}. The variance can be re-written as

$$\mathsf{Var}[\hat{w}_{pop}] = \frac{\sigma_b^2}{N} + \frac{\sigma_w^2}{Nn} \tag{19}$$

This result is identical to the maximum-likelihood and summary-statistic results derived earlier. The equivalence between the Bayesian and ML results derives from the fact that there is no prior at the population level. Hence, $p(Y|\mu) = p(\mu|Y)$ as indicated in the previous chapter.

Unbalanced designs

The model described in this section is identical to the separable model in the previous chapter but with $x_i = 1_{n_i}$. If the error covariance matrix is non-isotropic ie. $C \neq \sigma_w^2 I$, then the population estimates will change. This can occur, for example, if the design matrices are different for different subjects (so-called 'unbalanced-designs'), or if the data from some of the subjects is particularly ill-fitting. In these cases, we consider the within subject variances $\sigma_w^2(i)$ and the number of events n_i to be subject-specific. This will be the case in experimental paradigms where the number of events is not under experimental control eg. in memory paradigms where n_i may refer to the number of remembered items.

If we let $M = 1_N$ then the second level in equation 2 in the previous chapter is then the matrix equivalent of the second-level in equation 14 (ie. it holds for all i). Plugging in our values for M and X gives

$$\mathsf{Var}[\hat{w}_{pop}] = \left(\sum_{i=1}^{N} \frac{\alpha \beta_i n_i}{\alpha + n_i \beta_i} \right)^{-1} \tag{20}$$

and

$$\hat{w}_{pop} \quad = \quad \left(\sum_{i=1}^{N} \frac{\alpha \beta_i n_i}{\alpha + \beta_i n_i} \right)^{-1} \sum_{i=1}^{N} \frac{\alpha \beta_i}{\alpha + \beta_i n_i} \sum_{j=1}^{n_i} y_{ij} \tag{21}$$

This reduces to the earlier result if $\beta_i = \beta$ and $n_i = n$. Both of these results are different to the summary statistic approach, which we note is therefore mathematically inexact for unbalanced designs. But as we shall see in the numerical example below, the summary statistic approach is remarkably robust to departures from assumptions about balanced designs.

Estimation

To implement the PEB estimation scheme for the unequal variance case we first compute the errors $\hat{e}_{ij} = y_{ij} - X\hat{w}_i$, $\hat{z}_i = \hat{w}_i - M\hat{w}_{pop}$. We then substitute $x_i = 1_{n_i}$ into the update rules derived in the PEB section of the previous chapter to obtain

$$\sigma_b^2 \equiv \frac{1}{\alpha} = \frac{1}{\gamma} \sum_{i=1}^{N} \hat{z}_i^2 \tag{22}$$

$$\sigma_w^2(i) \equiv \frac{1}{\beta_i} = \frac{1}{n_i - \gamma_i} \sum_{j=1}^{n_i} \hat{e}_{ij}^2 \tag{23}$$

where

$$\gamma = \sum_{i=1}^{N} \gamma_i \tag{24}$$

and

$$\gamma_i = \frac{n_i \beta_i}{\alpha + n_i \beta_i} \tag{25}$$

For balanced designs $\beta_i = \beta$ and $n_i = n$ we get

$$\sigma_b^2 \equiv \frac{1}{\alpha} = \frac{1}{\gamma} \sum_{i=1}^{N} \hat{z}_i^2 \tag{26}$$

$$\sigma_w^2 \equiv \frac{1}{\beta} = \frac{1}{Nn-\gamma} \sum_{i=1}^{N} \sum_{j=1}^{n} \hat{e}_{ij}^2 \tag{27}$$

where

$$\gamma = \frac{n\beta}{\alpha + n\beta} N \tag{28}$$

Effectively, the degrees of freedom in the data set (Nn) are partitioned into those that are used to estimate the between-subject variance, γ, and those that are used to estimate the within-subject variance, $Nn - \gamma$.

The posterior distribution of the first-level coefficients is

$$p(w_i|y_{ij}) \equiv p(\hat{w}_i) = \mathsf{N}(\bar{w}_i, \mathsf{Var}[\hat{w}_i]) \tag{29}$$

where

$$\mathsf{Var}[\hat{w}_i] = \frac{1}{\alpha + n_i\beta_i} \tag{30}$$

$$\hat{w}_i = \frac{\beta_i}{\alpha + n_i\beta_i} \sum_{j=1}^{n_i} y_{ij} + \frac{\alpha}{\alpha + n_i\beta_i} \hat{w}_{pop} \tag{31}$$

Overall, the PEB estimation scheme is implemented by first initialising $\hat{w}_i$, $\hat{w}_{pop}$ and α, β_i (for example to values given from the equal error-variance scheme). We then compute the errors $\hat{e}_{ij}$, $\hat{z}_i$ and re-estimate the α and β_i's using the above equations. The coefficients $\hat{w}_i$ and $\hat{w}_{pop}$ are then re-estimated and the last two steps are iterated until convergence. This algorithm is identical to the PEB algorithm for the separable model in the previous chapter but with $x_i = 1_{n_i}$.

Numerical example

We now give an example of random effects analysis on simulated data. The purpose is to compare the PEB and SS algorithms. We generated data from a three-subject, two-level model with population mean $\mu = 2$, subject effect sizes $w = [2.2, 1.8, 0.0]^T$ and within subject variances $\sigma_w^2(1) = 1$, $\sigma_w^2(2) = 1$. For the third subject $\sigma_w^2(3)$ was varied from 1 to 10. The second level design matrix was $M = [1, 1, 1]^T$ and the first-level design matrix was given by $X = \mathsf{blkdiag}(x_1, x_2, x_3)$ with x_i being a boxcar.

Figure 2 shows a realisation of the three time series for $\sigma_w^2(3) = 2$. The first two time series contain stimulus-related activity but the third does not. We then applied the PEB algorithm, described in the previous section, to obtain estimates of the population mean $\hat{\mu}$ and estimated variances, σ_μ^2. For comparison, we also obtained equivalent estimates using the SS approach. We then computed the accuracy with which the population mean was estimated using the criterion $(\hat{\mu} - \mu)^2$. This was repeated for 1000 different data sets generated using the above parameter values, and for 10 different values of $\sigma_w^2(3)$. The results are shown in figures 3 and 4.

Firstly we note that, as predicted by theory, both PEB and SS give identical results when the first level error variances are equal. When the variance on the 'rogue' time series approaches double that of the others we see different estimates of both $\hat{\mu}$ and σ_μ^2. With increasing rogue error variance the SS estimates get worse but the PEB estimates get better. There is an improvement with respect to the true values, as shown in Figure 3, and with respect to the variability of the estimate, as shown in Figure 4. This is because the third time series is more readily recognised by PEB as containing less reliable information about the population mean and is increasingly ignored. This gives better estimates $\hat{\mu}$ and a reduced uncertainty, σ_{μ}^2.

We created the above example to reiterate a key point of this chapter, that SS gives identical results to PEB for equal within subject error variances (homoscedasticity) and unbalanced designs, but not otherwise. In the numerical example, divergent behaviour is observed when the error variances differ by a factor of two. For studies with more subjects (12 being a typical number), however, this divergence requires a much greater disparity in error variances. In fact we initially found it difficult to generate data sets where PEB showed a consistent improvement over SS ! It is therefore our experience that the vanilla SS approach is particularly robust to departures from homoscedasticity. This conclusion is supported by what is known of the robustness of the t-test that is central to the SS approach. Lack of homoscedasticity only causes problems when the sample size (ie. number of subjects) is small. As sample size increases so does the robustness (see eg. [Yandell 1997]).

PET data example

We now illustrate the difference between FFX and RFX analysis using data from a PET study of verbal fluency. These data come from 5 subjects and were recorded under two alternating conditions. Subjects were asked to either repeat a heard letter or to respond with a word that began with that letter. These tasks are referred to as word shadowing and word generation and were performed in alternation over 12 scans and the order randomized over subjects. Both conditions were identically paced with one word being generated every two seconds. PET images were re-aligned, normalised and smoothed with a 16mm isotropic Gaussian kernel. [1]

Fixed-Effects Analysis

Analysis of multiple-subject data takes place within the machinery of the General Linear Model (GLM) as described in earlier chapters. However, instead of having data from a single-subject at each voxel we now have data from multiple subjects. This is entered into a GLM by concatenating data from all subjects into the single column vector Y. Commensurate with this augmented data vector is an augmented multi-subject design matrix [2], X, which is shown in Figure 5. Columns 1 and 2 indicate scans taken during the word shadowing and word generation conditions respectively, for the first subject. Columns 3 to 10 indicate these conditions for the other subjects. The time variables in columns 11 to 15 are used to probe habituation effects. These variables are not of interest to us in this chapter but we include them to improve the fit of the

[1] This data set and full details of the pre-processing are available from http : //www.fil.ion.ucl.ac.uk/spm/data.

[2] This design was created using the 'Multi-subject: condition by subject interaction and covariates' option in SPM-99.

model. The GLM can be written as

$$Y = X\beta + E \tag{32}$$

where β are regression coefficients and E is a vector of errors. The effects of interest can then be examined using an augmented contrast vector, c. For example, for the verbal fluency data the contrast vector

$$c = [-1, 1, -1, 1, -1, 1, -1, 1, -1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T \tag{33}$$

would be used to examine the differential effect of word generation versus word shadowing, averaged over the group of subjects. The corresponding t-statistic,

$$t = \frac{c^T \hat{\beta}}{\sqrt{\mathsf{Var}[c^T \hat{\beta}]}} \tag{34}$$

where $\mathsf{Var}[]$ denotes variance, highlights voxels with significantly non-zero differential activity. This shows the 'average effect in the group' and is a type of fixed-effects analysis. The resulting Statistical Parametric Map is shown in Figure 6(b).

It is also possible to look for differential effects in each subject separately using subject-specific contrasts. For example, to look at the activation from subject 2 one would use the contrast vector

$$c_2 = [0, 0, -1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T \tag{35}$$

The corresponding subject-specific SPMs are shown in Figure 6(a).

We note that we have been able to look at subject-specific effects because the design matrix specified a 'subject-separable model'. In these models the parameter estimates for each subject are unaffected by data from other subjects. This arises from the block-diagonal structure in the design matrix.

Random-Effects Analysis via Summary-Statistics

An RFX analysis can be implemented using the 'Summary-Statistic (SS)' approach as follows [Frison and Pocock 1992, Holmes and Friston 1998].

1. Fit the model for each subject using different GLMs for each subject or by using a multiple-subject subject-separable GLM, as described above. The latter approach may be procedurally more convenient whilst the former is less computationally demanding. The two approaches are equivalent for the purposes of RFX analysis.

2. Define the effect of interest for each subject with a contrast vector. Each produces a contrast image containing the contrast of the parameter estimates at each voxel.

3. Feed the contrast images into a GLM that implements a one-sample t-test.

Modelling in step 1 is referred to as the 'first-level' of analysis whereas modelling in step 3 is referred to as the 'second-level'. A balanced design is one in which all subjects have identical design matrices and error variances. Strictly, balanced designs are a requirement for the SS approach to be valid. But as we have seen with the numerical example, the SS approach is remarkably robust to violations of this assumption.

If there are, say, two populations of interest and one is interested in making inferences about differences between populations then a two-sample t-test is used at the second level. It is not necessary that the numbers of subjects in each population be the same, but it is necessary to have the same design matrices for subjects in the same population ie. balanced designs at the first-level.

In Step 3, we have specified that only one contrast per subject be taken to the second level. This constraint may be relaxed if one takes into account the possibility that the contrasts may be correlated or be of unequal variance. This can be implemented using within-subject ANOVAs at the second level, a topic which is is covered in chapter 13.

An SPM of the RFX analysis is shown in Figure 6(c). We note that, as compared to the SPM from the average effect in the group, there are far fewer voxels deemed significantly active. This is because RFX analysis takes into account the between-subject variability. If, for example, we were to ask the question 'Would a new subject drawn from this population show any significant posterior activity ?', the answer would be uncertain. This is because three of the subjects in our sample show such activity but two subjects do not. Thus, based on such a small sample, we would say that our data do not show sufficient evidence against the null hypothesis that there is no population effect in posterior cortex. In contrast, the average effect in the group (in Figure 6(b)) is significant over posterior cortex. But this inference is with respect to the group of five subjects, not the population.

We end this section with a disclaimer, which is that the results presented in this section, have been presented for tutorial purposes only. This is because between-scan variance is so high in PET that results on single subjects are unreliable. For this reason, we have used uncorrected thresholds for the SPMs and, given that we have no prior anatomical hypothesis, this is not the correct thing to do [Frackowiak et al. 1997] (see Chapter 14). But as our concern is merely to present a tutorial on the difference between RFX and FFX we have neglected these otherwise important points.

fMRI data example

This section compares RFX analysis as implemented using SS versus PEB. The dataset we chose to analyse comprised 1,200 images that were acquired in 10 contiguous sessions of 120 scans. These data have been described elsewhere [Friston et al. 1998].

The reason we chose these data was that each of the 10 sessions was slightly different in terms of design. The experimental design involved 30-second epochs of single word streams and a passive listening task. The words were concrete, monosyllabic nouns presented at a number of different rates. The word rate was varied pseudo-randomly over epochs within each session.

We modelled responses using an event-related model where the occurrence of each word was modelled with a delta function. The ensuing stimulus function was convolved with a canonical hemodynamic response function and its temporal derivative to give two regressors of interest for each of the 10 sessions. These effects were supplemented with confounding and nuisance effects comprising a mean and the first few components of a discrete cosine transform, removing drifts lower than 1/128 Hz. Further details of the paradigm and analysis details are given in [Friston et al. 2005].

The results of the SS and PEB analyses are presented in Figure 7 and have been thresholded at $p < 0.05$, corrected for the entire search volume. These results are taken from [Friston et al. 2005] where PEB was implemented using the ReML formulation. It is evident that the inferences from these two procedures are almost identical, with PEB being slightly more sensitive. The results remain relatively unchanged despite the fact that the first level designs were not balanced. This contributes to non-sphericity at the second level which is illustrated in Figure 8 for the SS and PEB approaches. This figure shows that heteroscedasticity can vary by up to a factor of 4.

Discussion

We have shown how neuroimaging data from multiple subjects can be analysed using fixed-effects (FFX) or random-effects (RFX) analysis. FFX analysis is used for reporting case studies and RFX is used to make inferences about the population from which subjects are drawn. For a comparison of these and other methods for combining data from multiple subjects see [Lazar et al. 2002].

In neuroimaging, RFX is implemented using the computationally efficient summary-statistic approach. We have shown that this is mathematically equivalent to the more computationally demanding maximum likelihood procedure. For unbalanced designs, however, the summary-statistic approach is no longer equivalent. But we have shown using a simulation study and fMRI data, that this lack of formal equivalence is not practically relevant.

For more advanced treatments of random effects analysis [3] see eg. [Yandell 1997]. These allow, for example, for subject-specific within-subject variances, unbalanced designs and for Bayesian inference [Carlin and Louis 2000]. For a recent application of these ideas to neuroimaging, readers are referred to Chapter 17 in which hierarhical models are applied to single and multiple subject fMRI studies. As groundwork for this more advanced material readers are encouraged to first read the tutorial in Chapter 11.

A general point to note, especially for fMRI, is that because the between-subject variance is typically larger than the within-subject variance your scanning time is best used to scan more subjects rather than to scan individual subjects for longer. In practice, this must be traded off against the time required to recruit and train subjects [Worsley et al. 2002].

Further points

We have so far described how to make inferences about univariate effects in a single population. This is achieved in the summary statistic approach by taking forward a single contrast image per subject to the second level and then using a one sample t-test.

[3] Strictly, what in neuroimaging is known as random effects analysis is known in statistics as mixed effects analysis as the statistical models contain both fixed and random effects.

This methodology carries over naturally to more complex scenarios where we may have multiple populations or multivariate effects. For two populations, for example, we perform two-sample t-tests at the second level. An extreme example of this approach is the comparison of a single case study with a control group. Whilst this may sound unfeasible, as one population has only a single member, a viable test can in fact be implemented by assuming that the two populations have the same variance.

For multivariate effects we take forward multiple contrast images per subject to the second level and perform an analysis of variance. This can be implemented in the usual way with a GLM but, importantly, we must take into account the fact that we have repeated measures for each subject and that each characteristic of interest may have a different variability. This topic is covered in the next chapter.

As well as testing for whether univariate population effects are significantly different from hypothesized values (typically zero) it is also possible to test whether they are correlated with other variables of interest. For example, one can test whether task-related activation in the motor system correlates with age [Ward and Frackowiak 2003]. It is also possible to look for conjunctions at the second level eg. to test for areas that are conjointly active for pleasant, unpleasant and neutral odour valences [Gottfried et al. 2002]. For a statistical test involving conjunctions of contrasts it is necessary that the contrast effects be uncorrelated. This can be ensured by taking into account the covariance structure at the second level. This is also described in the next chapter on analysis of variance.

The validity of all of the above approaches relies on the same criteria that underpin the univariate single population summary statistic approach. Namely, that the variance components and estimated parameter values are, on average, identical to those that would be obtained by the equivalent two-level maximum likelihood model.

References

[Carlin and Louis 2000] B.P. Carlin and T.A. Louis. *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman and Hall, 2000.

[Frackowiak et al. 1997] R.S.J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, and J.C. Mazziotta, editors. *Human Brain Function.* Academic Press USA, 1997.

[Frison and Pocock 1992] L. Frison and S.J. Pocock. Repeated measures in clinical trials: An analysis using mean summary statistics and its implications for design. *Statistics in medicine*, 11:1685–1704, 1992.

[Friston et al. 1998] K.J. Friston, O. Josephs, G. Rees, and R. Turner. Nonlinear event-related responses in fMRI. *Magnetic Resonance in Medicine*, 39:41–52, 1998.

[Friston et al. 2002] K.J. Friston, W.D. Penny, C. Phillips, S.J. Kiebel, G. Hinton, and J. Ashburner. Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage*, 16:465–483, 2002.

[Friston et al. 2005] K.J. Friston, K.E. Stephan, T.E. Lund, A. Morcom, and S.J. Kiebel. Mixed-effects and fMRI studies. *NeuroImage*, 24:244–252, 2005.

[Gottfried et al. 2002] J. A. Gottfried, R. Deichmann, J.S. Winston, and R.J. Dolan. Functional Heterogeneity in Human Olfactory Cortex: An Event-Related Functional Magnetic Resonance Imaging Study. *The Journal of Neuroscience*, 22(24):10819–10828, 2002.

[Holmes and Friston 1998] A.P. Holmes and K.J. Friston. Generalisability, random effects and population inference. In *NeuroImage*, volume 7, page S754, 1998.

[Lazar et al. 2002] N.A. Lazar, B. Luna, J.A. Sweeney, and W.F. Eddy. Combining brains: a survey of methods for statistical pooling of information. *Neuroimage*, 16(2):538–550, 2002.

[Wackerley et al. 1996] D.D. Wackerley, W. Mendenhall, and R.L. Scheaffer. *Mathematical statistics with applications*. Duxbury Press, 1996.

[Ward and Frackowiak 2003] N.S. Ward and R.S.J. Frackowiak. Age related changes in the neural correlates of motor performance. *Brain*, 126:873–888, 2003.

[Worsley et al. 2002] K. J. Worsley, C. H. Liao, J. Aston, V. Petre, G. H. Duncan, F. Morales, and A. C. Evans. A general statistical analysis for fMRI data. *NeuroImage*, 15(1), January 2002.

[Yandell 1997] B.S. Yandell. *Practical data analysis for designed experiments*. Chapman and Hall, 1997.

Expectations and transformations

We use $\mathsf{E}[]$ to denote the expectation operator and $\mathsf{Var}[]$ to denote variance and make use of the following results. Under a linear transform $y = ax + b$, the variance of x changes according to

$$\mathsf{Var}[ax + b] = a^2 \mathsf{Var}[x] \tag{36}$$

Secondly, if $\mathsf{Var}[x_i] = \mathsf{Var}[x]$ for all i then

$$\mathsf{Var}\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right] = \frac{1}{N}\mathsf{Var}[x] \tag{37}$$

For background reading on expectations, variance transformations and introductory mathematical statistics see [Wackerley et al. 1996].

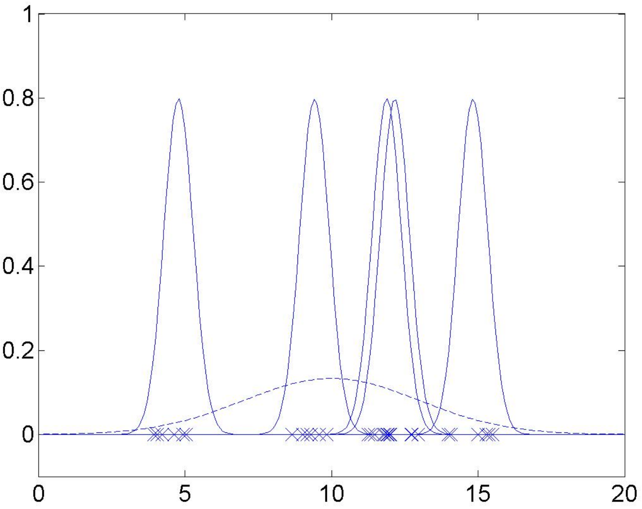

Figure 1: *Synthetic data illustrating the probability model underlying random effects analysis. The dotted line is the Gaussian distribution underlying the second level model with mean w_{pop}, the population effect, and variance σ_b^2, the between-subject variance. The mean subject effects, w_i, are drawn from this distribution. The solid lines are the Gaussians underlying the first level models with means w_i and variances σ_w^2. The crosses are the observed effects y_{ij} which are drawn from the solid Gaussians.*

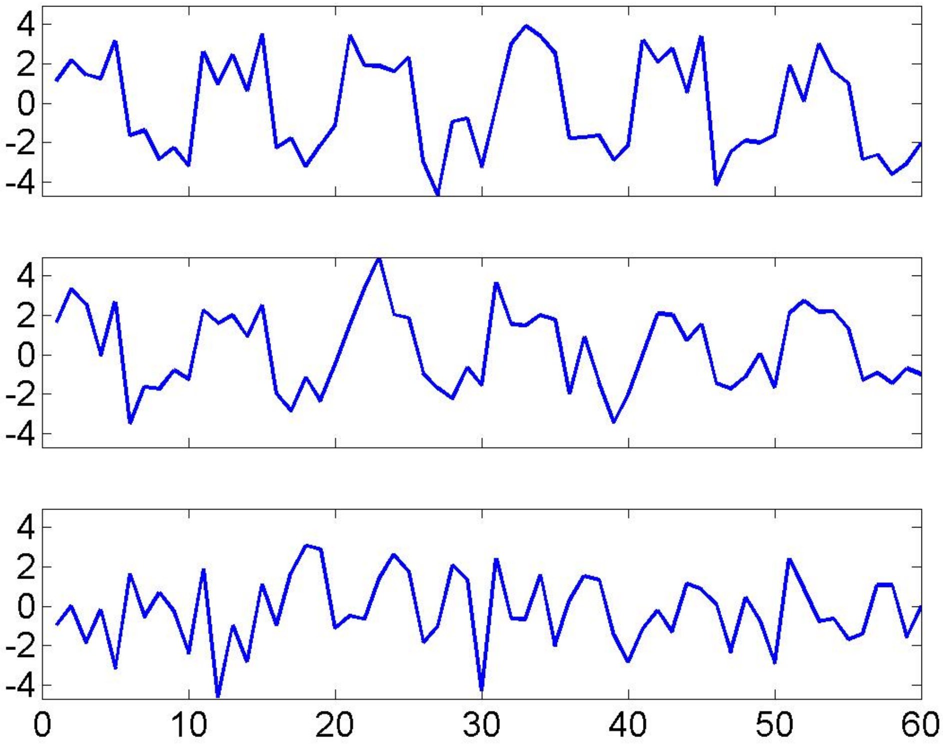

Figure 2: *Simulated data for random effects analysis. Three representative time series produced from the two-level hierarchical model. The first two time-series contain stimulus-related activity but the third does not.*

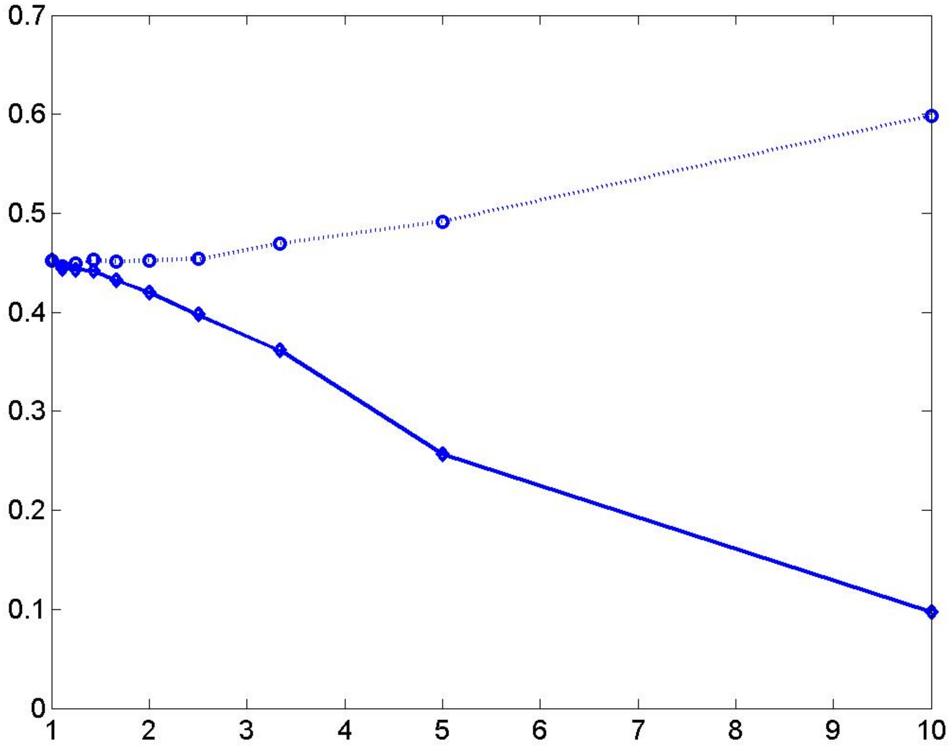

Figure 3: *A plot of the error in estimating the population mean* $E = <(\hat{\mu} - \mu)^2>$ *versus the observation noise level for the third subject,* $\sigma_w^2(3)$*, for the Parametric Empirical Bayes approach (solid line) and the Summary-Statistic approach (dotted line).*

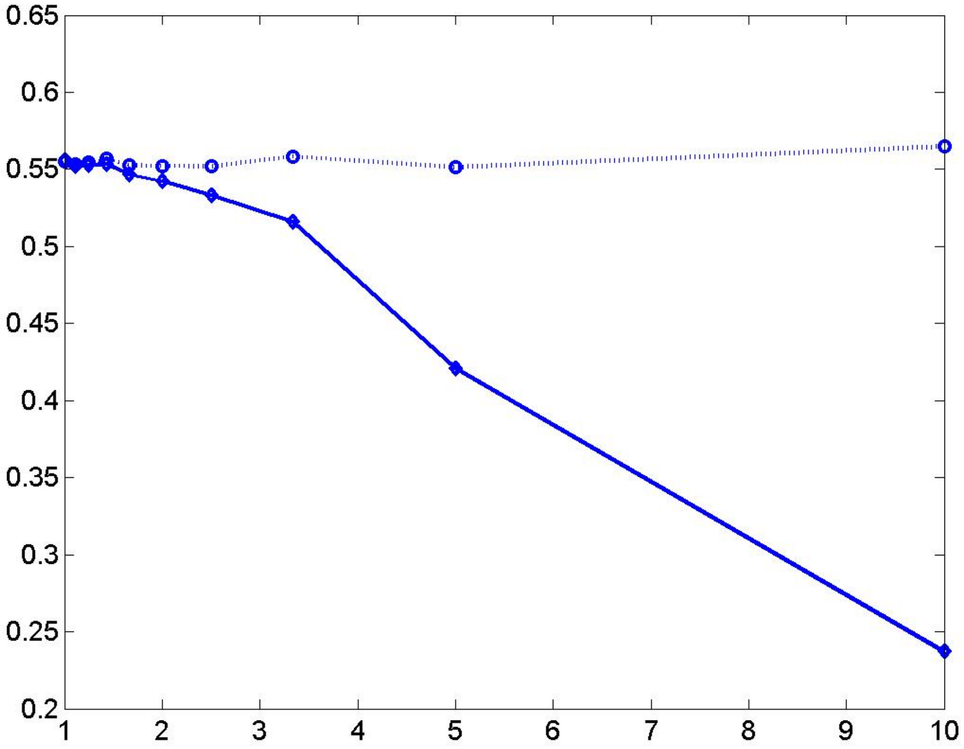

Figure 4: *A plot of the estimated variance of the population mean,* σ_μ^2*, versus the observation noise level for the third subject,* $\sigma_w^2(3)$*, for the Parametric Empirical Bayes approach (solid line) and the Summary-Statistic approach (dotted line).*

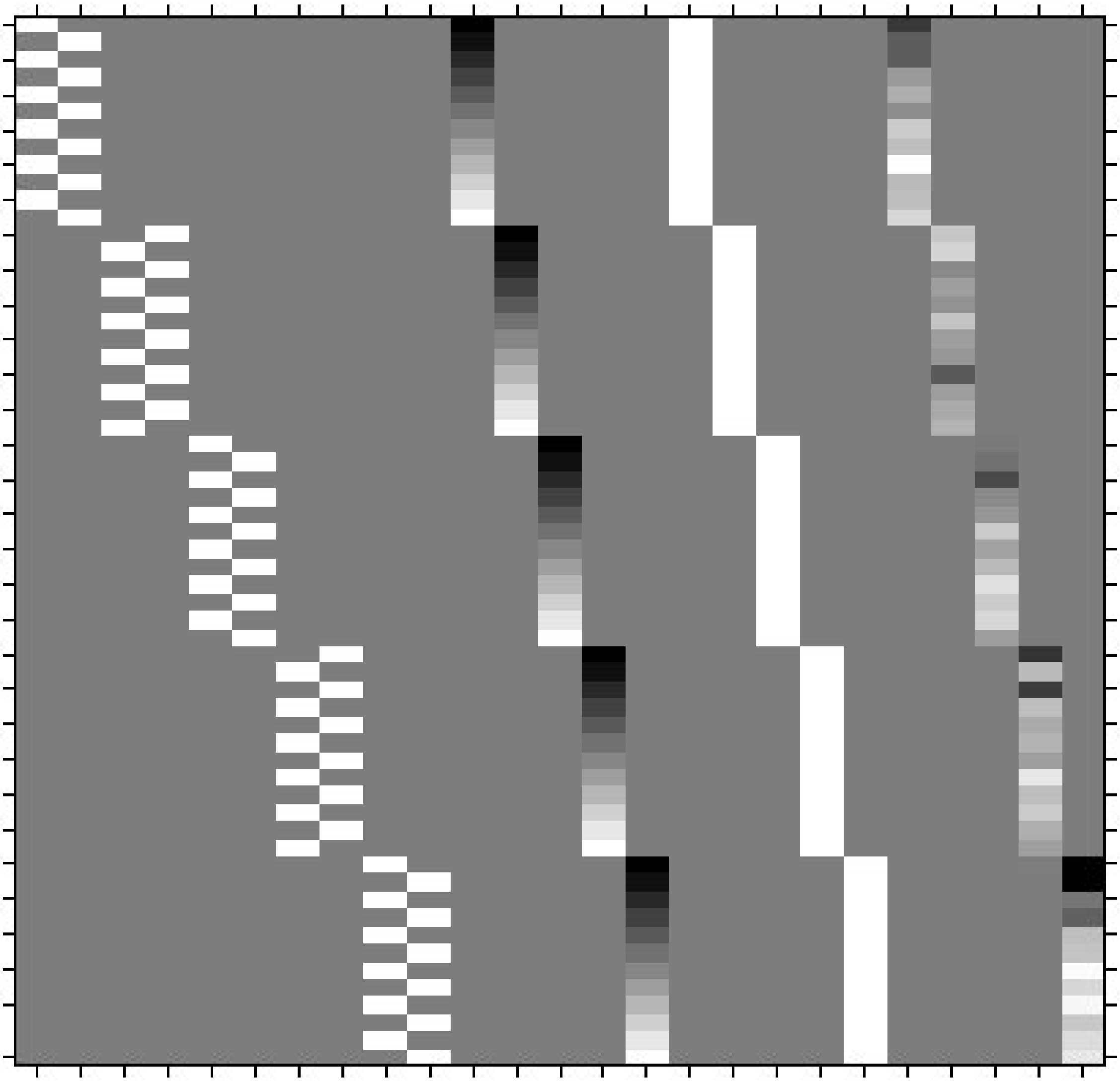

Figure 5: *Design matrix for the five-subject FFX analysis of PET data. There are 60 rows, 12 for each subject. The first ten columns contain indicator variables showing which condition (word shadowing or word generation) relates to which scan. Columns 11 to 15 contain time variables, columns 16 to 20 subject-specific offsets and the last 5 columns the global effect at each scan.*

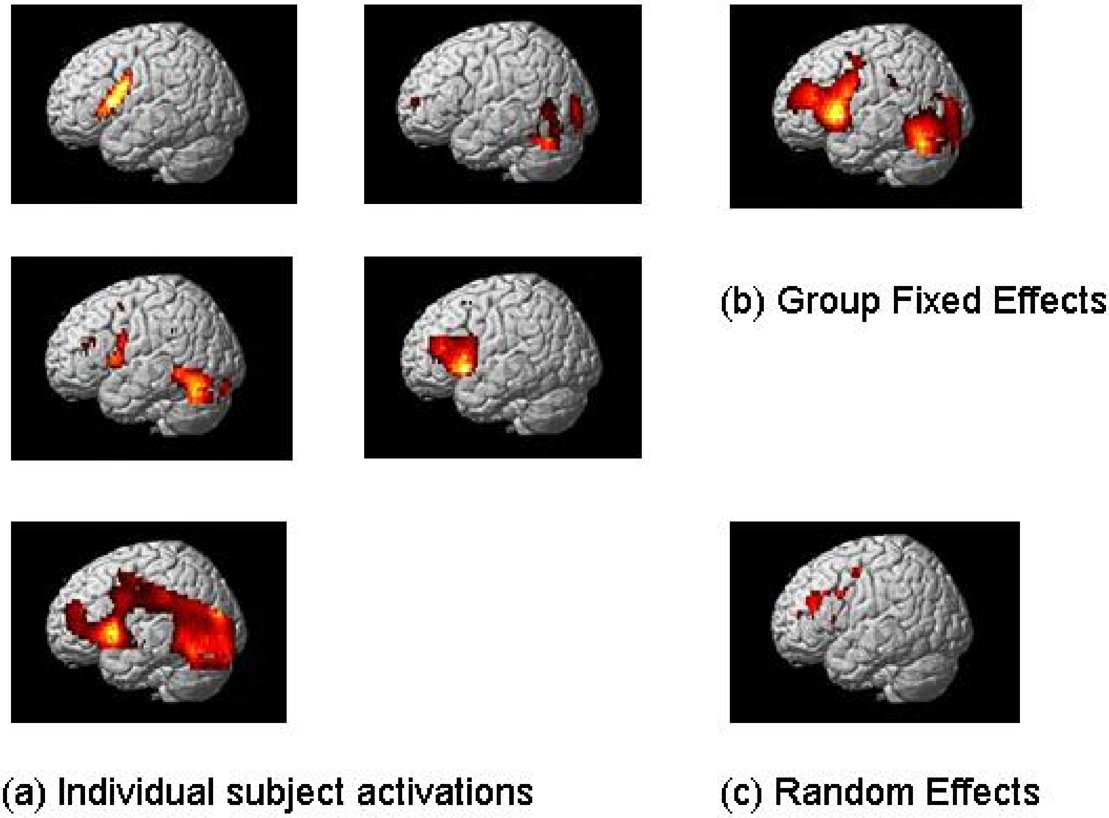

Figure 6: *Analysis of PET data showing active voxels ($p < 0.001$ uncorrected).The maps in (a) show the significance of subject-specific effects whereas map (b) shows the significance of the average effect over the group. Map (c) shows the significance of the population effect from an RFX analysis.*

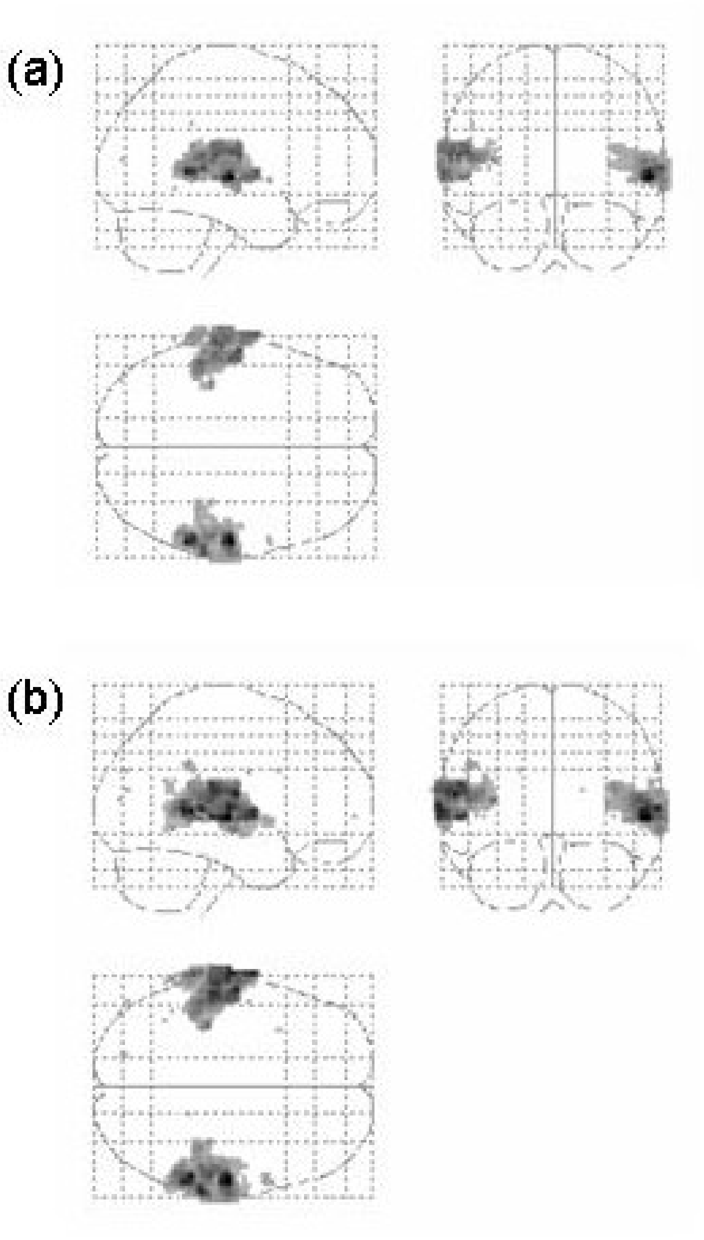

Figure 7: *SPMs showing the effect of words in the population using (a) SS and (b) PEB approaches.*

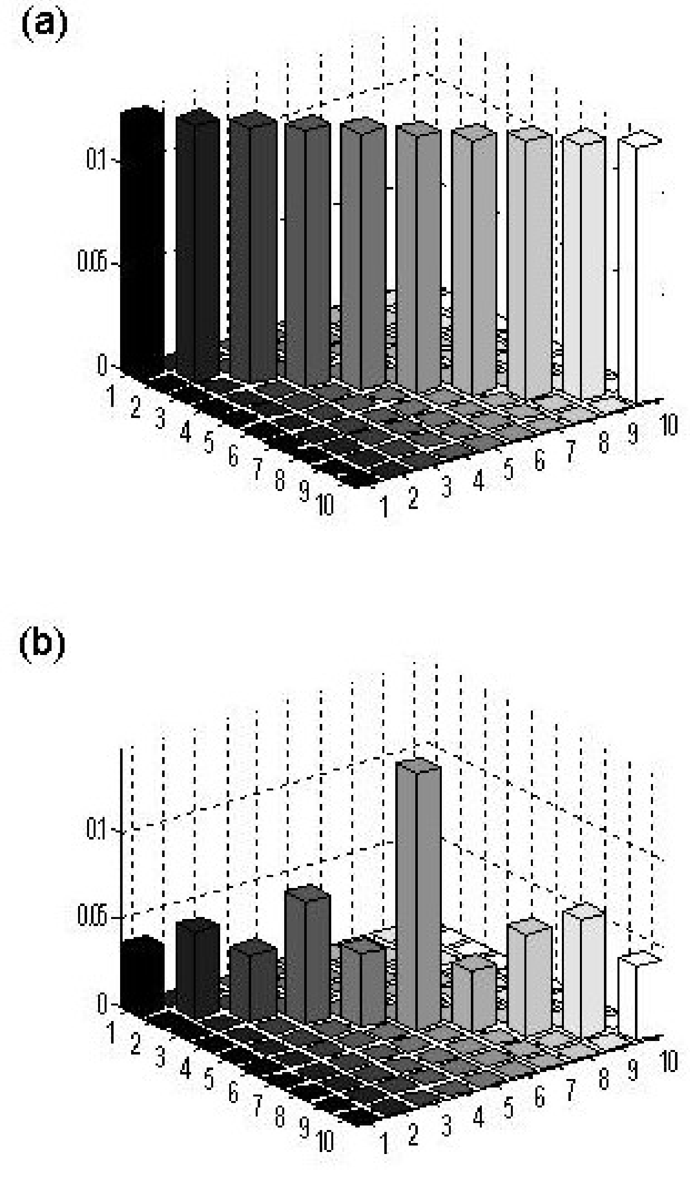

Figure 8: *Within-session variance as (a) assumed by SS and (b) estimated using PEB. This shows that within-session variance can vary by up to a factor of four, although this makes little difference to the final inference (see Figure 7).*