MRes Advanced Brain Imaging

Revision session

Remi Gau

School of psychology
University of Birmingham

11th February 2013

General questions

- **What is an objective function when comparing 2 images?**

  It is a function that allows quantification of the difference between these images and the optimization of the parameters set to get their best realignment, usually by minimizing/maximizing its value.

- **Why are there different objective functions?**

  T1 and T2 weighted have very different ranges of values for the same tissue class.

General questions

Which objective functions are used for (1) realignment of functional images and (2) coregistration of a structural and functional image?

1. Realignment : within modality
   - Least squares : minimize
   - Normalized cross correlation : maximize
2. Registration : between modality (usually entropy based approaches)
   - Mutual information : maximize
   - Normalized Mutual information : maximize

General questions

If you resample your images after each preprocessing step, how will the images look different (as compared to when you resampled only after smoothing)?

Your images will be smoother : interpolation error increases the smoothness of your data.

Slice timing

In their simulation data, Sladky et al. (2011) found that to slice-delay effects were more pronounced for experimental designs long TRs (4 secs) compare to short TRs (1 secs), or for event-related and short block designs (10 secs) compared to long blocks (30 secs) design.

Can you explain those results?

- Longer TRs are more affected because more time goes by between the first and and last acquired slices.
- Event related and short blocks are more affected because there will be quicker temporal changes in the the BOLD signal.

Realignment

What are some the possible sources of residual errors that are still present after realignment?

1. Re-sampling can introduce interpolation errors
2. Gaps between slices can cause aliasing artifacts
3. Slices are not acquired simultaneously: rapid movements not accounted for by rigid body model
4. Image artifacts may not move according to a rigid body model
   - image distortion and realignment $\times$ inhomogeneity interaction
   - image dropout
   - Nyquist ghost
5. Spin history effect

Realignment

What can you do to minimize the effects of those residual errors?

1. Unwarp your data (if you have not already done so).
2. Include realignment parameters in the design matrix of your GLM.

Normalization

What are the reasons for normalization?

1. Try to account for inter-subject anatomical variability
2. Express results in a standard common space across studies

Normalization

After acquiring one anatomical and a series of functional images, you want to normalize all the images. **Describe two distinct processing streams that would bring the images of both modalities into a standard (e.g. MNI) space. What are the (dis)advantages of the two streams?**

1. T1 images normalized to T1 template and EPI images to the EPI template: less accurate as the lower resolution of the EPI images will mean that the registration will be driven by high-contrast feature of the brain (e.g edges)
2. Coregister the mean EPI and structural images, segment the structural image (GM/WM/CSF) which implies a normalization step and will produce normalization parameters that can then be used to normalize the EPI images.

Normalization

Most spatial normalization procedures involves two steps, affine and non-linear transformation. Why?

1. Affine : rigid body transformation
2. Non-linear : warping for better match between structures

Smoothing

- **What are the main reasons for spatial smoothing ?**

  1. Compensate for inter-subject variability
  2. Matched filter theorem
  3. Sufficient smoothness of the data is necessary if random field theory has to be used for multiple comparison

- **What is the matched filter theorem?**

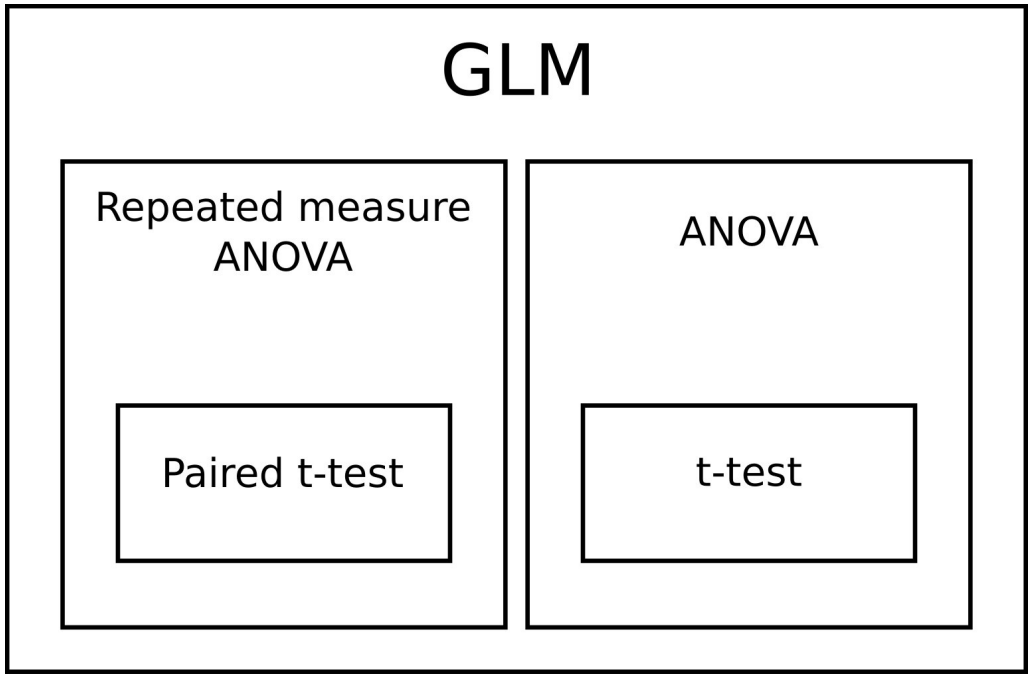
  Using a filter that has the same shape as the signal will give the best signal to noise ratio.

- **If we expect your signal to be of Gaussian shape with a kernel of 8 mm FWHM, which smoothing kernel would you use?**

  A gaussian shape with a kernel of 8 mm FWHM.

General questions

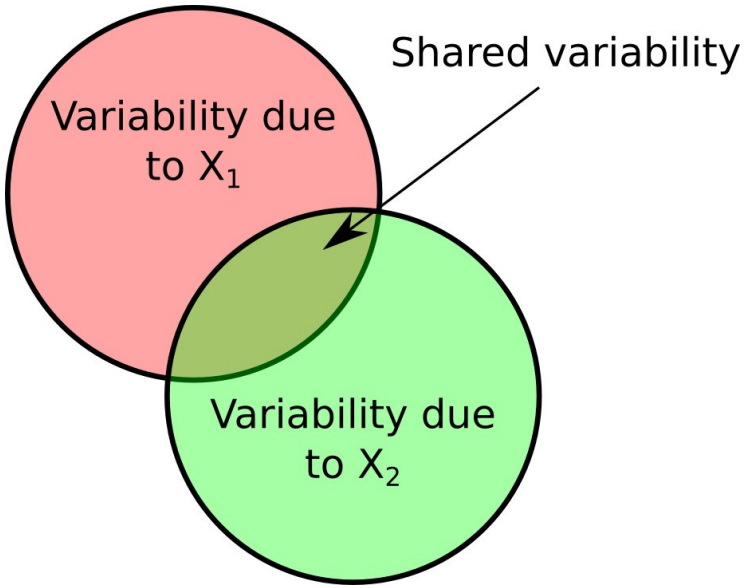
Draw a Venn diagram of: t-test, paired t-test, ANOVA, repeated measures ANOVA, General Linear model.



General questions

If two regressors X_1 and X_2 are correlated, what do the β estimates represent?

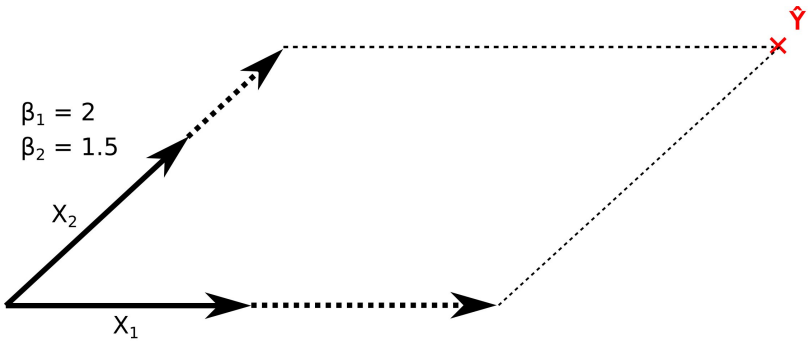
Only the variability unique to a regressor drives the parameter estimate of that regressor.

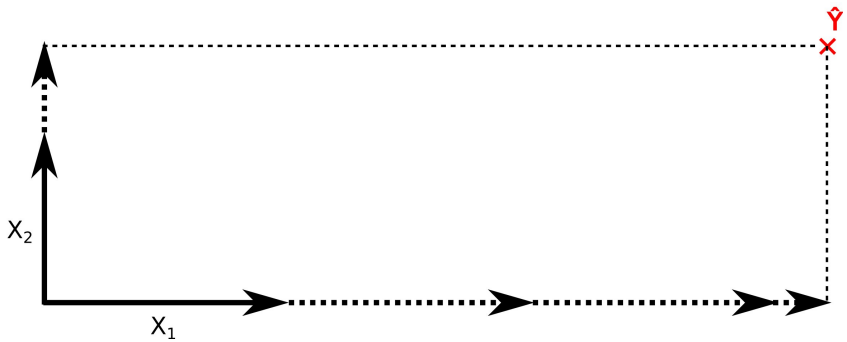


General questions

If regressor 1 and 2 are correlated and you orthogonalize 2 with respect to 1 which, if any, β will change and why? Explain via geometric perspective.

The β associated to the regressor 1 will change.





General questions

The General Linear Model described by $Y = X\beta + \varepsilon$. **Define the design matrices (X) of the following** :

- Two-sample t-test with 3 subjects in group A and B

$$
\begin{aligned}
Y_{A1} &= 1 \times \beta_A + 0 \times \beta_B + \varepsilon_{A1} \\
Y_{A2} &= 1 \times \beta_A + 0 \times \beta_B + \varepsilon_{A2} \\
Y_{A3} &= 1 \times \beta_A + 0 \times \beta_B + \varepsilon_{A3} \\
Y_{B1} &= 0 \times \beta_A + 1 \times \beta_B + \varepsilon_{B1} \\
Y_{B2} &= 0 \times \beta_A + 1 \times \beta_B + \varepsilon_{B2} \\
Y_{B3} &= 0 \times \beta_A + 1 \times \beta_B + \varepsilon_{B3}
\end{aligned}
$$

General questions

The General Linear Model described by $Y = X\beta + \varepsilon$. **Define the design matrices (X) of the following** :

- Two-sample t-test with 3 subjects in group A and B

$$\begin{bmatrix} Y_{A1} \\ Y_{A2} \\ Y_{A3} \\ Y_{B1} \\ Y_{B2} \\ Y_{B3} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} \beta_A \\ \beta_B \end{bmatrix} + \begin{bmatrix} \varepsilon_{A1} \\ \varepsilon_{A2} \\ \varepsilon_{A3} \\ \varepsilon_{B1} \\ \varepsilon_{B2} \\ \varepsilon_{B3} \end{bmatrix}$$

General questions

The General Linear Model described by $Y = X\beta + \varepsilon$. **Define the design matrices (X) of the following** :

- Paired t-test with 2 conditons (A and B) with 3 subjects

$$
\begin{aligned}
Y_{A1} &= 1 \times \beta_A + 0 \times \beta_B + 1 \times \beta_1 + 0 \times \beta_2 + 0 \times \beta_3 + \varepsilon_{A1} \\
Y_{A2} &= 1 \times \beta_A + 0 \times \beta_B + 0 \times \beta_1 + 1 \times \beta_2 + 0 \times \beta_3 + \varepsilon_{A2} \\
Y_{A3} &= 1 \times \beta_A + 0 \times \beta_B + 0 \times \beta_1 + 0 \times \beta_2 + 1 \times \beta_3 + \varepsilon_{A3} \\
Y_{B1} &= 0 \times \beta_A + 1 \times \beta_B + 1 \times \beta_1 + 0 \times \beta_2 + 0 \times \beta_3 + \varepsilon_{B1} \\
Y_{B2} &= 0 \times \beta_A + 1 \times \beta_B + 0 \times \beta_1 + 1 \times \beta_2 + 0 \times \beta_3 + \varepsilon_{B2} \\
Y_{B3} &= 0 \times \beta_A + 1 \times \beta_B + 0 \times \beta_1 + 0 \times \beta_2 + 1 \times \beta_3 + \varepsilon_{B3}
\end{aligned}
$$

General questions

The General Linear Model described by $Y = X\beta + \varepsilon$. **Define the design matrices (X) of the following** :

- Paired t-test with 2 conditons (A and B) with 3 subjects

$$\begin{bmatrix} Y_{A1} \\ Y_{A2} \\ Y_{A3} \\ Y_{B1} \\ Y_{B2} \\ Y_{B3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \beta_A \\ \beta_B \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{A1} \\ \varepsilon_{A2} \\ \varepsilon_{A3} \\ \varepsilon_{B1} \\ \varepsilon_{B2} \\ \varepsilon_{B3} \end{bmatrix}$$

General questions

The General Linear Model described by $Y = X\beta + \varepsilon$. **Define the design matrices (X) of the following** :

- ANOVA with three groups of subjects and 3 subjects in each group

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

General questions

The General Linear Model described by $Y = X\beta + \varepsilon$. **Define the design matrices (X) of the following** :

- Repeated measures ANOVA for 3 subjects and with three within subject levels

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

General questions

The General Linear Model described by $Y = X\beta + \varepsilon$. **Define the design matrices (X) of the following** :

- Regression

$$
\begin{aligned}
Y_1 &= x_1 \times \beta_1 + 1 \times \beta_0 + \varepsilon_1 \\
Y_2 &= x_2 \times \beta_1 + 1 \times \beta_0 + \varepsilon_2 \\
Y_3 &= x_3 \times \beta_1 + 1 \times \beta_0 + \varepsilon_3 \\
Y_4 &= x_4 \times \beta_1 + 1 \times \beta_0 + \varepsilon_4 \\
Y_5 &= x_5 \times \beta_1 + 1 \times \beta_0 + \varepsilon_5 \\
Y_6 &= x_6 \times \beta_1 + 1 \times \beta_0 + \varepsilon_6
\end{aligned}
$$

General questions

The General Linear Model described by $Y = X\beta + \varepsilon$. **Define the design matrices (X) of the following** :

- Regression

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \\ x_5 & 1 \\ x_6 & 1 \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix}$$

Experimental design

- **List at least 6 parametric factors.**

  Speed, coherence, angle, intensity...

- **What is the advantage of having a parametric factor with 3 rather than 2 levels?**

  With 2 levels anything you fit will be linear.

- **What is the disadvantage of having 3 (or 5 or 7) levels?**

  In the GLM, the parametric regressor is demeaned to avoid any correlation with the modulated regressor. Therefore the median value of the parametric regressor will not explain any of the variance.

General questions

In the experiment, you present 4 flashes with the following onset vector.

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

You assume a simple HRF.

$$\begin{bmatrix} 1 & 3 & 5 & 1 & -2 & 0 \end{bmatrix}$$

Generate a predictor for your BOLD-response in the experiment.

General questions

Generate a predictor for your BOLD-response in the experiment.

$$\begin{bmatrix} 1 & 3 & 5 & 1 & -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & 5 & 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 & 5 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 3 & 5 & -2 & 0 \end{bmatrix}$$

Then sum.

$$\begin{bmatrix} 1 & 3 & 6 & 5 & 6 & 6 & 2 & 3 & -2 & 0 \end{bmatrix}$$

General questions

- **A**n experiment was performed with a TR $= 3$ secs and an SOA $= 3$ secs. How would you change the experimental parameters and why?

  If the SOA is a multiple of your TR then you will always sample the same time point of the HRF and might miss the peak. You could alternatively change the SOA or jitter the onsets.

- **What are the assumptions inherent in using convolution for modeling the BOLD response?**

  The system you study must be LTI : Linear and Time Independent.

General questions

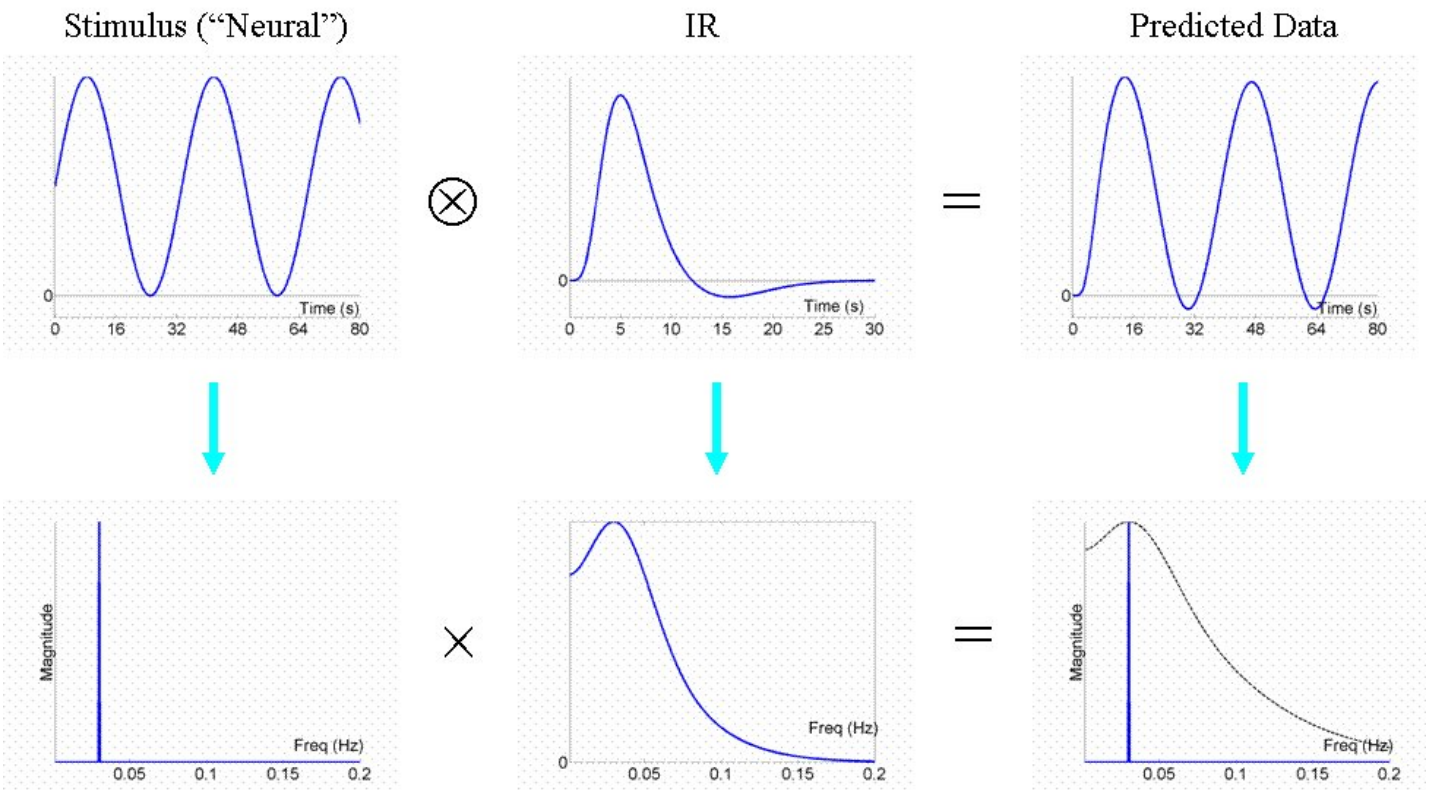
Explain the concept of design efficiency from the perspective of maximal bandpassed energy.

Any fMRI experiment can be viewed as trying to record a signal subjected to a bandpass filter.
The *high pass filter* is the one we explicitly apply as part of the GLM to get rid of low frequency scanner noise.
There is an *implicit low pass filter* in the experiment: that of the HRF. It will greatly attenuate high frequencies.
Only the energy of the signal going through the bandpass defined by these two filters will be of interest.

General questions

What is the most efficient fMRI design and why?

An experiment with a sinusoidal signal with a frequency equal to the frequency less filtered by the HRF.



Basis functions

Which experiments make *a priori* optimization of design efficiency difficult?

- If some conditions depend on subject's responses.
- If you must lock stimuli presentation to TR trigger.

Basis functions

- **Explain the concept of basis functions for modeling the HRF.**

  They are functions or set of functions that you can linearly combine to model the response function of the brain you are interested in.

- You have performed a complex working memory experiment. You are not sure about the shape of the activation functions. **Which basis sets might be useful for modeling the hemodynamic response? What are their advantages and disadvantages?**

  You could use the finite impulse response (FIR) or a Fourrier set. These sets are much less constrained in terms of the kind of HRF shape they can model.
  This means they can also model physiologically implausible HRF.
  Besides they also involves more regressors, which means that the same number of data will have to explain more parameters increasing the variability of each parameter's estimate.

Basis functions

- **Why would you model your events using the informed basis functions set (i.e. what is the function of the derivatives)?**

  The derivatives can account for BOLD responses that depart from the canonical HRF.

  Temporal derivative can help model temporal shifts in the HRF.
  Dispersion derivative can help model wider or sharper HRF.

- **How can you test whether the informed basis functions set is really sufficient for modeling the HRF?**

  Model your protocol with the informed basis functions set and also with a finite impulse response. With a F-test, you can then check if the full model (informed set + FIR) explains more variance the reduced model (informed set alone).

Basis functions

What conditions need to be fulfilled in order to perform selective averaging rather then using a convolution model for estimating the averaged response to an event class?

Either the events have to be very far apart in time so that there is no overlap between the responses to the different events. But this is very inefficient.
Or you could have overlap between the different responses but you must then make sure that you have equal balancing of ordering of the different conditions.

General questions

- **What is the iid assumption?**

  Independent and Identically Distributed errors

- **What is heteroscedasticity?**

  The variances are unequal but not correlated.

- **What are the two assumptions underlying sphericity? Define them precisely by referring to the properties of the error covariance matrix.**

  The different error terms have equal variances and no correlation between them. The error covariance matrix is a diagonal matrix and the values along the diagonal are all the same.

General questions

- **Which two sources of variability need to be considered in group studies?**

  Inter and intra subject variability

- **When is the summary statistic approach (in)valid (i.e. which requirements need to be fulfilled?)**

  The summary statistic only takes the mean activation for each subject at the second level for the group analysis. Mainly, it requires intra-subject variability to be equivalent.
  It also requires that the first level design is the same for each subject and to only have one value (contrast) per subject.

General questions

- The one sample t-test is generally regarded as the gold standard for 2nd level analysis. **Describe two classes of analyses (= statistical test, inferences) that cannot be performed in a 2nd level one-sample t-test.**

  If you want to test for interaction or if you have more than 2 levels.

- **Which variability is relevant for generalization to the population level?**

  Inter-subject variability.

General questions

How does non-sphericity emerge at 1st and 2nd level?

First level	There is an auto-correlation from one scan to the next due to the fact that we have a time series.
Second level	Different groups (patient VS control) might have unequal variances.
Second level	In repeated measures designs, some values come from the same subject and are likely to be correlated.

General questions

- **What is the meaning of a p-value in classical statistics? E.g. $p < \alpha$ means that:**

  Think of it as a surprise factor: it is the probability of getting these data under H_0

- **In the case where $p > \alpha$, does it provide evidence to say that H_0 is true?**

  No because the computation of the p value start by assuming that H_0 is true.

- **A t-test is a ratio of?**

  It is a signal to noise ratio.

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} - \frac{\hat{\sigma}_2^2}{n_2}}}$$

$$t = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}}$$

General questions

Does the t-statistic depend on the scaling of the regressors?

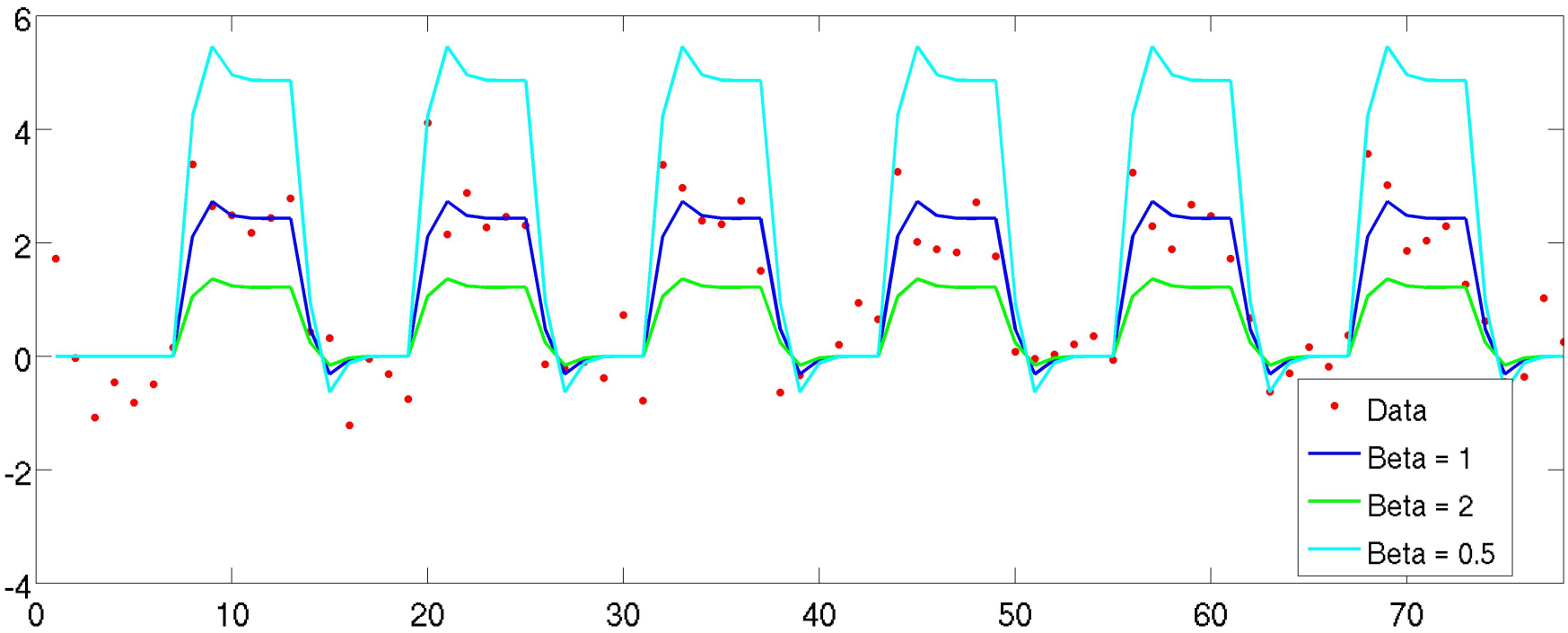
No because the scaling is going to affect the signal and the noise in the same direction.
You have a design matrix X and some associated $\hat{\beta}$ values. If you have A as a scaling factor, your new scaled design matrix will be $X_s = AX$. And your new scaled $\hat{\beta}_s$ will be equal to $\frac{\hat{\beta}}{A}$.

$$\hat{\beta}_s = \frac{X_s^T y}{X_s^T X_s} = \frac{AX^T y}{AX^T AX} = \frac{X^T y}{AX^T X} = \frac{1}{A}\hat{\beta}$$

General questions

Does the t-statistic depend on the scaling of the regressors?

If you have A as a scaling factor, your new scaled $\hat{\beta}_s$ will be equal to $\frac{\hat{\beta}}{A}$.



General questions

Does the t-statistic depend on the scaling of the regressors?

No because the scaling is going to affect the signal and the noise in the same direction.
You have a design matrix X and some associated $\hat{\beta}$ values. If you have A as a scaling factor, your new scaled design matrix will be $X_s = AX$. And your new scaled $\hat{\beta}_s$ will be equal to $\frac{\beta}{A}$.
Your new t value for a given contrast then becomes :

$$t_s = \frac{c^T \hat{\beta}_s}{\sqrt{\hat{\sigma}^2 c^T (X_s^T X_s)^{-1} c}} = \frac{c^T A^{-1} \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (A X_s^T A X_s)^{-1} c}} = \frac{c^T A^{-1} \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T A^{-2} (X_s^T X_s)^{-1} c}} = \frac{c^T A^{-1} \hat{\beta}}{A^{-1} \sqrt{\hat{\sigma}^2 c^T (X_s^T X_s)^{-1} c}} = \frac{c^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 c^T (X^T X)^{-1} c}} = t$$

Inferences

Define precisely the three levels of inference and their constraints: voxel, cluster and set. What are their advantages and disadvantages?

There a trade off between an increase of sensitivity VS loss of spatial specificity when you from voxel to set level. TANSTAAFL

Voxel Just requires to specify the voxel level threshold. You can have *p* values and confidence intervals.

Cluster Requires to specify threshold at the voxel level and the minimum cluster size. You can only have *p* values.

Set Requires to specify threshold at the voxel level, the minimum cluster size and the minimum set size. You can only have *p* values.

Inferences

How can you obtain p- and t-values from a uni-dimensional F-test?

In uni-dimensional condition :

A t contrast looks for regions whose activity is positively correlated to the regressor you are testing.

A F contrast looks for regions whose activity is positively *OR* negatively correlated to the regressor you are testing. The F-statistic is simply the square of the corresponding t-statistic.

The p-value of the t-test will be half that of the F-test.

Inferences

Explain the extra-sum of square principle for the F-testing

Given two nested models for the same response, with a reduced model that includes m regressors and a full model containing $p = m + n$ regressors, the extra sum of squares tests the following null hypothesis:

$$H_0: \quad \beta_{m+1} = \beta_{m+2} = \cdots = \beta_p = 0$$

Inferences

Explain the extra-sum of square principle for the F-testing

It tests whether a set of regressors is necessary given that another set has already been included in the model. It looks at the change in the residual sum of squares (SSR) between the full and reduced models and comparing it with the mean SSR from the full model. The F-statistic is:

$$F \propto \frac{SSR_{full} - SSR_{reduced}}{SSR_{full}}$$

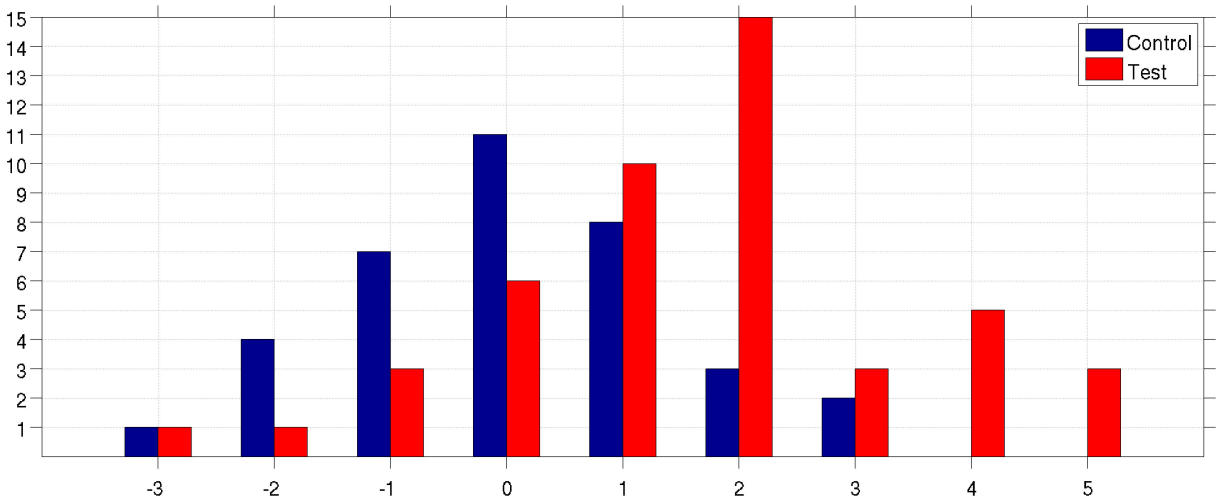
If the extra regressors do not change the SSR, then

$$(SSR_{full} - SSR_{reduced}) \approx 0$$

and the ratio will be small. Alternatively, if the additional regressors change the SSR, then the ratio will be large and hence these regression still explain some of the variance and should be considered for the model.

Inferences

Given the following outcomes and a threshold $\mu_1 = 0.5$:

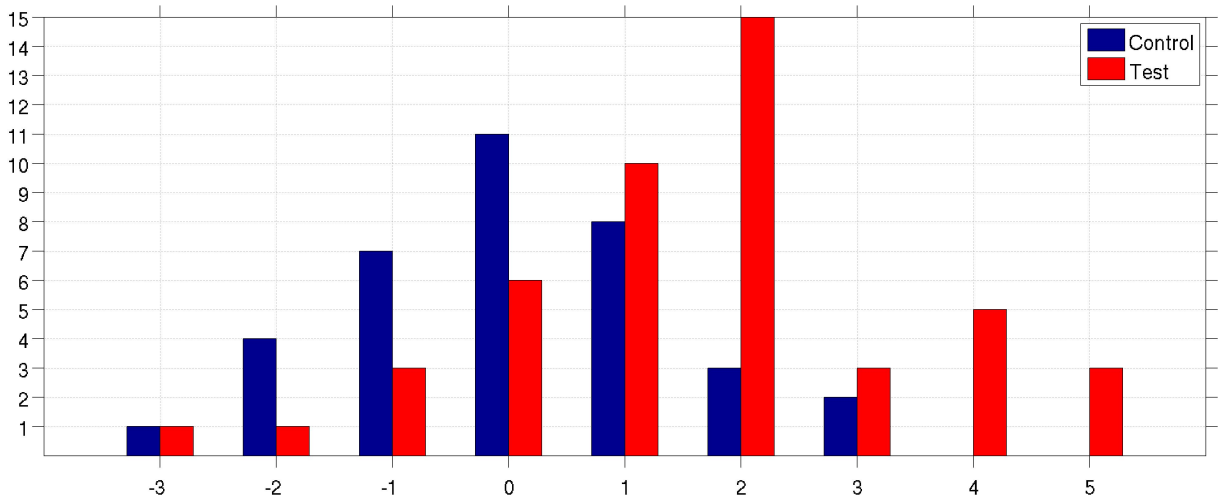


Compute:

- True positives or Hits (TP) :
- True negatives or Correct Rejections (TN)
- False positives or False alarms (FP):
- False negatives or Missed (FN):

Inferences

Given the following outcomes and a threshold $\mu_1 = 0.5$:



Compute:

- Sensitivity or Power: $\frac{TP}{TP+FN}$
- Specificity: $\frac{TN}{TN+FP}$
- False discovery rate - Type II error: $\frac{FN}{FN+TP}$
- False positive rate - Type I error: $\frac{FP}{FP+TN}$

Family wise error

Suppose you have a statistical parametric map with 100 000 voxels.
But you have smoothed your data with a ridiculously large FWHM (e.g 1 meter!). You are looking for activation with $\alpha = 0.05$.

What threshold would a Bonferonni correction give you? Why is that not threshold not appropriate? What would a better threshold be?

$$\alpha_c = \frac{\alpha}{N} = \frac{0.05}{100000} \text{ with N} = \text{number of tests}$$

This threshold does not take into account that the tests you are making are not independent.
Indeed all your voxels are so massively dependent that they have pretty much the same value, so you just have one big voxel with one value.
You in fact do not have a multiple comparison problem and you can keep your uncorrected α.

Family wise error

The SPM results page gives you the following values for the estimated smoothness of your data : FWHM = 12.6 11.9 10.9 (in mm). Yet you have smoothed your data with a [8 8 8] mm FWHM Gaussian kernel.

Is that an error? What could be the differences due to?

This is perfectly normal. The differences can be accounted for by :

1. The intrinsic smoothness of your data. Your data have a structure and are not random values.
2. Remember that the interpolation error of your preprocessing will increase the smoothness of your data.

General questions

Define the concepts of functional segregation and integration.

functional segregation Functional localization implies that a function can be localized in a cortical area, whereas specialization suggests that a cortical area is specialized for some aspects of perceptual or motor processing, and that this specialization is anatomically segregated within the cortex.

functional integration The cortical infrastructure supporting a single function may involve many specialized areas whose union is mediated by the functional integration among them.

General questions

What are the differences between structural, functional and effective connectivity?

structural connectivity It studies the connectivity at the anatomical level that can be studied with tracer studies or DTI.

functional connectivity It studies the correlation between the activation level of different brain areas.

effective connectivity It studies the effect that one neuronal system can have on another one.

General questions

Which methodological approaches can be used to analyze your functional imaging data from the perspective of functional integration (at least three)?

1. Seed voxel correlation
2. Psychophysiological interaction
3. Structural equation modeling
4. Granger Causality
5. Dynamic causal modeling

General questions

Describe the main limitations of a psychophysiological interaction analysis?

1. Correlational
2. Static
3. Only stays at the BOLD level