The General Linear Model in Functional Neuroimaging

Finn Årup Nielsen

Neurobiology Research Unit, Rigshospitalet;
Informatics and Mathematical Modelling
Technical University of Denmark

September 8, 2004

Introduction

- The general linear model has the form (Mardia et al., 1979, eq. 6.1.1)

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U}, \tag{1}$$

  where $\mathbf{Y}$(scans × voxels) is the image data, $\mathbf{X}$(scans × design variables) is the "design matrix" and $\mathbf{B}$(design variables × voxels) contains parameters to be estimated and tested. The residuals $\mathbf{U}$ are usually assumed Gaussian.

- Encapsulates many statistical models: t-test (paired, un-paired), F-test, ANOVA (one-way, two-way, main effect, factorial), MANOVA, ANCOVA, MANCOVA, simple regression, linear regression, multiple regression, multivariate regression, . . .

- Widely used in functional neuroimaging through the SPM program where it is performed in a mass-univerate setting — in parallel over the columns of $\mathbf{Y}$

Hypothesis test example with t-test

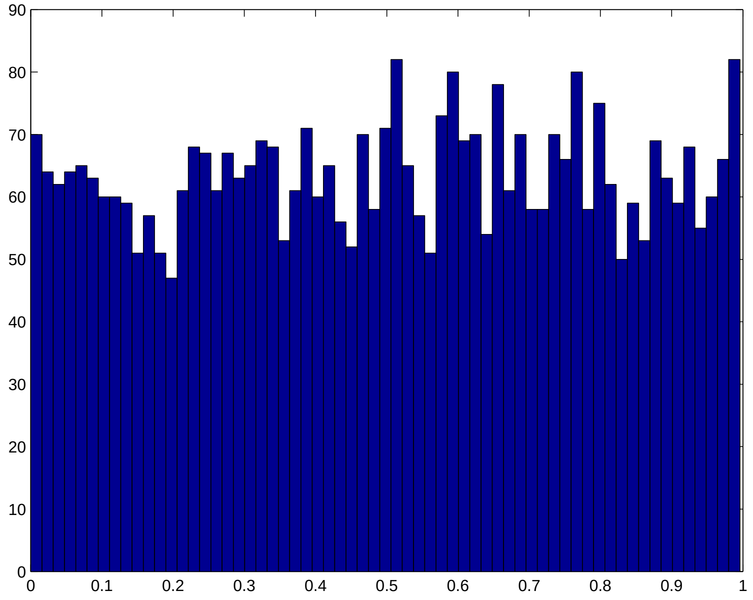

Figure 1: Histogram of the lower tail area of the t-value: $1 - p$-value.

Matlab program with a random design matrix and random image data:

```
X    = rand(12, 5);
Y    = randn(size(X,1), 4000);

B    = pinv(X) * Y;
dof  = size(X,1) - rank(X);
U    = Y - X*B;
SSE  = diag(U'*U)';
MSSE = SSE / dof;
SE   = sqrt(MSSE);

C = [ 1 -1 0 0 0 ];
T = C*B ./ (SE * sqrt(C*pinv(X'*X)*C'));
P = brede_cdf_t(T, dof);

figure
hist(P, sqrt(length(P)));
```

Hypothesis test example with F-test

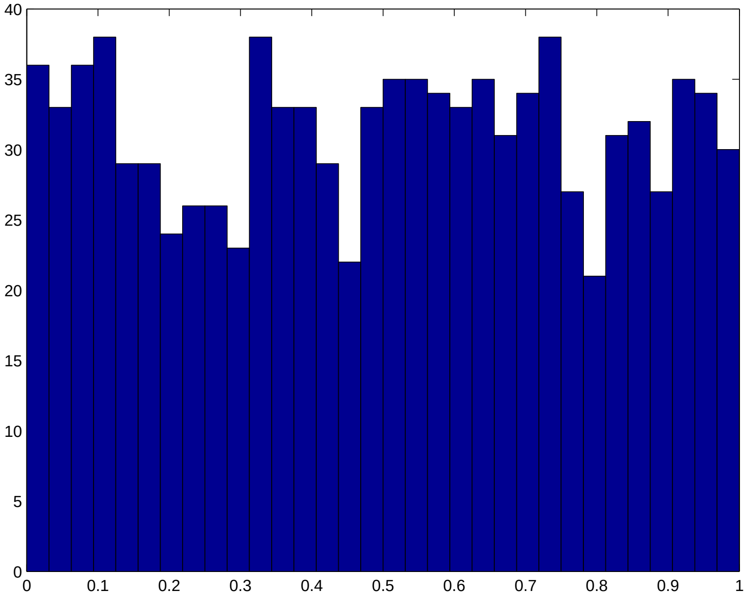

Figure 2: Histogram of the lower tail area of the F-value: $1 - p$-value.

Matlab program with a random design matrix and random image data:

```
X    = rand(12, 5);
Y    = randn(size(X,1), 1000);

B     = pinv(X) * Y;
dof   = size(X,1) - rank(X);
U     = Y - X*B;
SSE   = sum(U.^2);
MSSE = SSE / dof;

C = [ 1 0 0 0 0 ; 0 1 0 0 0 ];
F = 1/rank(C) * (diag((C*B)' * pinv(C * ...
         pinv(X'*X) * C') * (C*B))' ./ MSSE);
P = brede_cdf_f(F, rank(C), dof);

figure
hist(P, round(sqrt(length(P))));
```

Simple regression

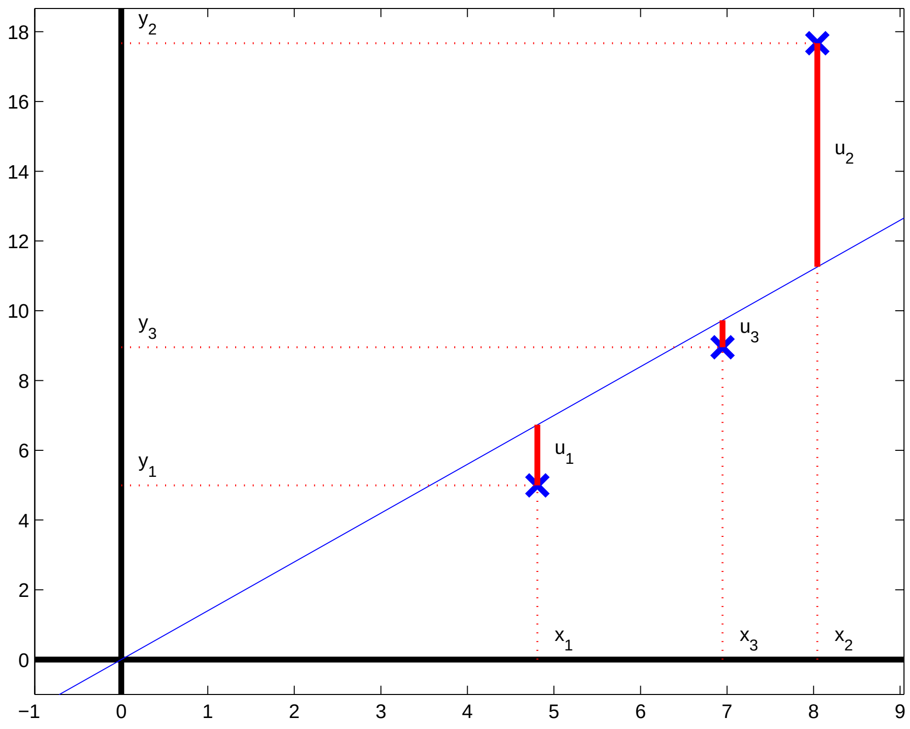

Figure 3: Simple regression.

In simple regression (e.g., one voxel) is univariate and the matrices from the general linear model become vectors or scalars: $\mathbf{Y} \rightarrow \mathbf{y}$, $\mathbf{X} \rightarrow \mathbf{x}$ and $\mathbf{B} \rightarrow b$

$$\mathbf{y} = \mathbf{x}b + \mathbf{u}, \qquad (2)$$

where $\mathbf{y}$ is the dependent variable (usually measured), $\mathbf{x}$ is the independent variable (design variable) and b is the parameter (regression coefficient).

Categorical variables

Categorical variable can be coded in two different ways:

“Sigma-restricted”, where two groups (e.g., male and female) are coded in one design variables

$$\mathbf{x}_{(1)} = \left[\begin{array}{cccccc} 1, & -1, & 1, & -1, & 1, & -1, \end{array}\right]^{\mathsf{T}}, \tag{3}$$

that leads to a design matrix with full rank.

“Overparameterized”, where two groups are coded in two design variables

$$\mathbf{X}_{(1:2)} = \left[\begin{array}{cccccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right]^{\mathsf{T}}, \tag{4}$$

that leads to a design matrix of degenerate rank.

The overparameterized version is often preferred due to better “ordnung”.

(www.statsoftinc.com)

Multiple regression

With several dependent variables

$$\mathbf{y} = \mathbf{x}_{(1)} b_1 + \mathbf{x}_{(2)} b_2 + \ldots + \mathbf{u}, \tag{5}$$

where $\mathbf{x}_{(1)}$ and $\mathbf{x}_{(2)}$ are column vectors. In matrix form with $\mathbf{X} = \left[\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \ldots\right]$ and $\mathbf{b} = [b_1, b_2, \ldots]$

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u} \tag{6}$$

Simple regression with intercept

In simple regression a parameter is usually added to model the the intercept ($\mu = b_2$) is included

$$\mathbf{y} = \mathbf{x}b_1 + b_2 + \mathbf{u}, \tag{7}$$

and changed to matrix form with $\mathbf{b} = [b_1, b_2]^\mathsf{T}$ and $\mathbf{X} = [\mathbf{x}, \mathbf{1}]$.

This is an instance of the multiple regression model.

“Spam variable”: Seems to be automatically added by SPM and SAS?

ANCOVA — ANalysis of COVAriance

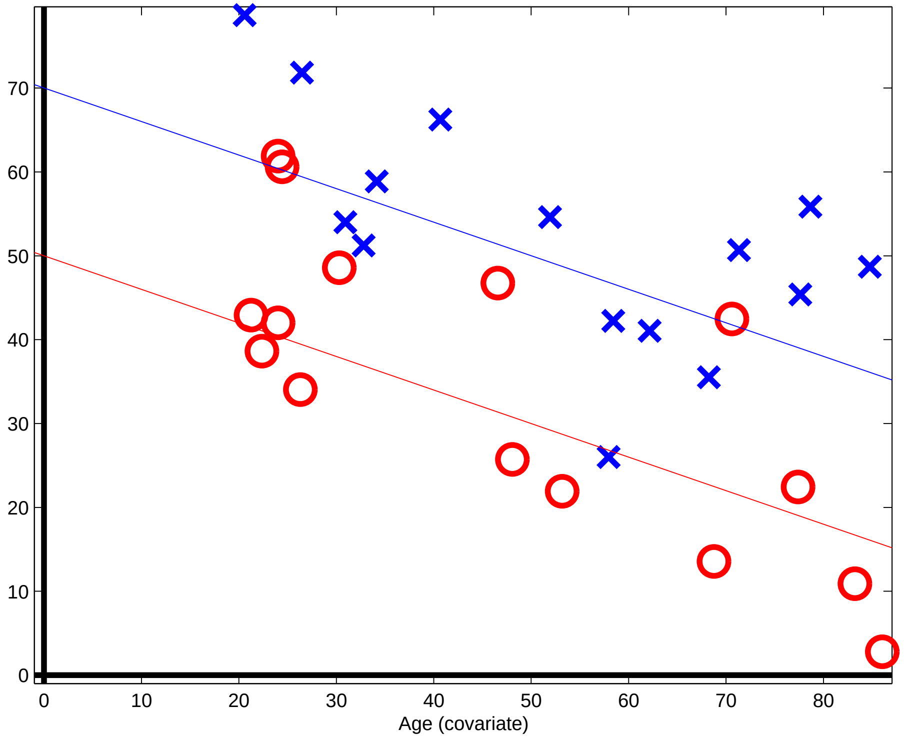

Figure 4: ANCOVA. Two groups (e.g., normals and patients) with and age-effect.

1) Model with categorical and continuous design variables.

2) Conditions + Nuisances (covariates, e.g., age)

An instance of multiple regression.

Why ANCOVA? Because the variance induced by the covariates might make the test less powerful! t-statistics for the example:

$$t_{\text{ordinary}} = -3.1 \qquad (8)$$

$$t_{\text{ANCOVA}} = -5.0 \qquad (9)$$

ANCOVA with bias

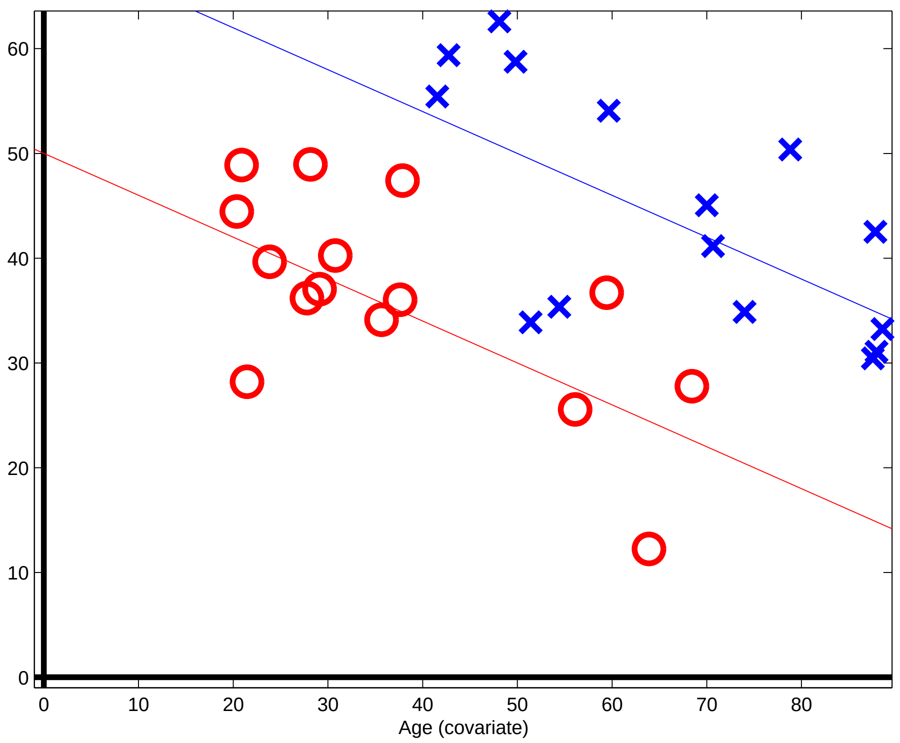

Figure 5: ANCOVA. Two groups (e.g., normals and patients) with and age-effect where the two groups have difference age.

ANCOVA is especially important with bias in the independent variable, e.g., in an uncontrolled study.

t-test statistics for the example

$$t_{\text{ordinary}} = -2.1 \qquad (10)$$

$$t_{\text{ANCOVA}} = -5.2 \qquad (11)$$

Thus it is able "to correct for bias" (Armitage and Berry, 1994, p. 301+) and remove, e.g., an age-effect.

Interactions

With “linear” interactions (aka moderator effects)

$$\mathbf{y} = \mathbf{x}_{(1)} b_1 + \mathbf{x}_{(2)} b_2 + (\mathbf{x}_{(1)} \odot \mathbf{x}_{(2)}) b_3 + \mathbf{u}, \tag{12}$$

where $\odot$ is an elementwise multiplication: $\mathbf{x}_{(3)} = \mathbf{x}_{(1)} \odot \mathbf{x}_{(2)}$.

Nonlinear effects

The are many ways (a infinite number) in which the nonlinearity can be modeled.

One of the simplest ways is by elementwise squaring of the dependent variable so the second column in the design matrix becomes

$$\mathbf{x}_{(2)} = [x_{1,1}^2, x_{2,1}^2, x_{3,1}^2, \ldots]. \tag{13}$$

There is no canonical nonlinearity: Squaring is just one way.

Paired t-test

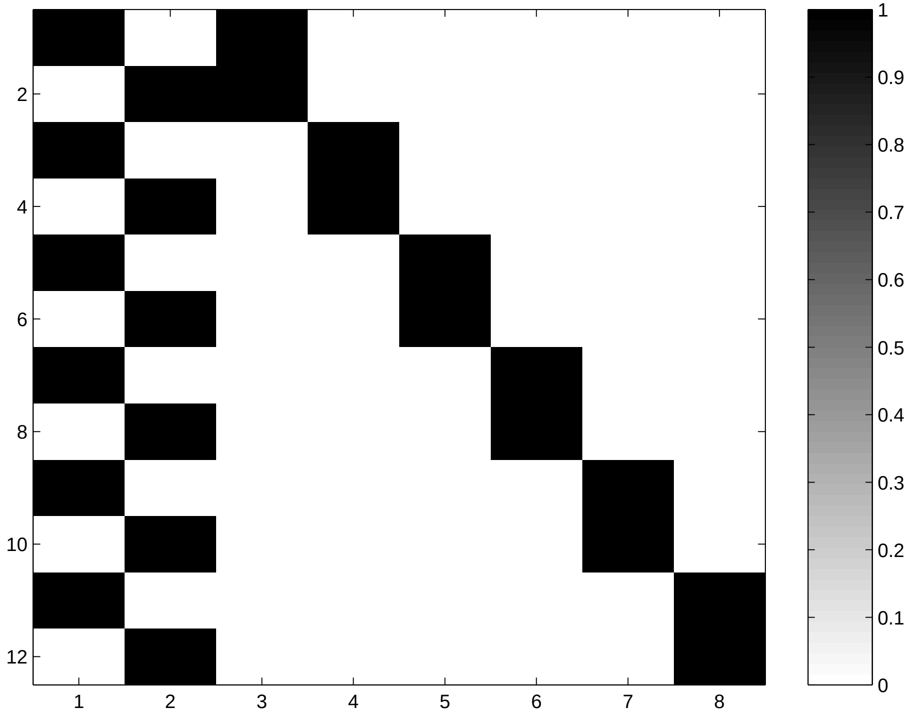

Figure 6: Design matrix $\mathbf{X}$ for paired t-test with 12 scans.

Paired t-test example

$$\mathbf{y} = \left[d_{1,2}, d_{3,4}, \ldots, d_{11,12}\right]^{\mathsf{T}}, \tag{14}$$

where, e.g., $d_{1,2} = y_1 - y_2$

Degrees of freedom is lost.

New degrees of freedom

$$\begin{aligned} r &= N - \text{rank}(\mathbf{X}) && (15) \\ &= 12 - 7 = 5 && (16) \end{aligned}$$

ANCOVA or paired t-test? An example

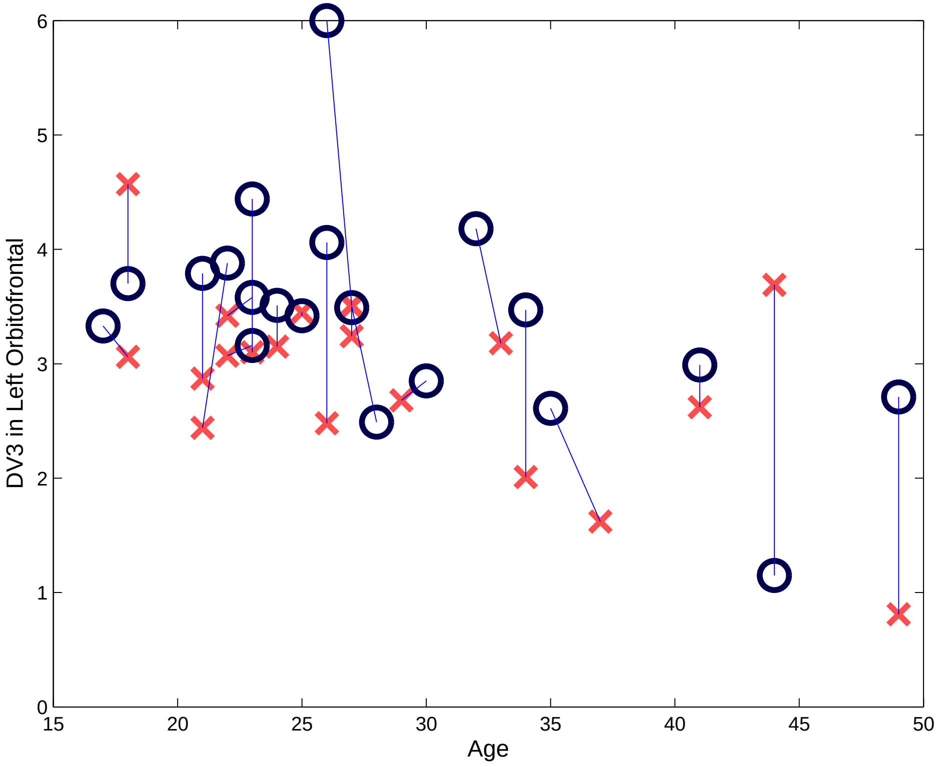

Figure 7: Distribution volume values (DV3) for left orbitofrontal cortex as a function of subject age from Steven Haugbøl's study. Tourette paired with controls.

Context: Two subject groups (Tourette and controls) and the subjects have different age. The measured variable ("altanserin DV3") is dependent on age (Adams et al., 2004).

Should ANCOVA or paired t-test be chosen as the analysis?

Paired t-test: Subjects are paired with respect to age.

ANCOVA: Subjects are not paired and the age confound is modeled by adding one row to the design matrix

ANCOVA or paired t-test?

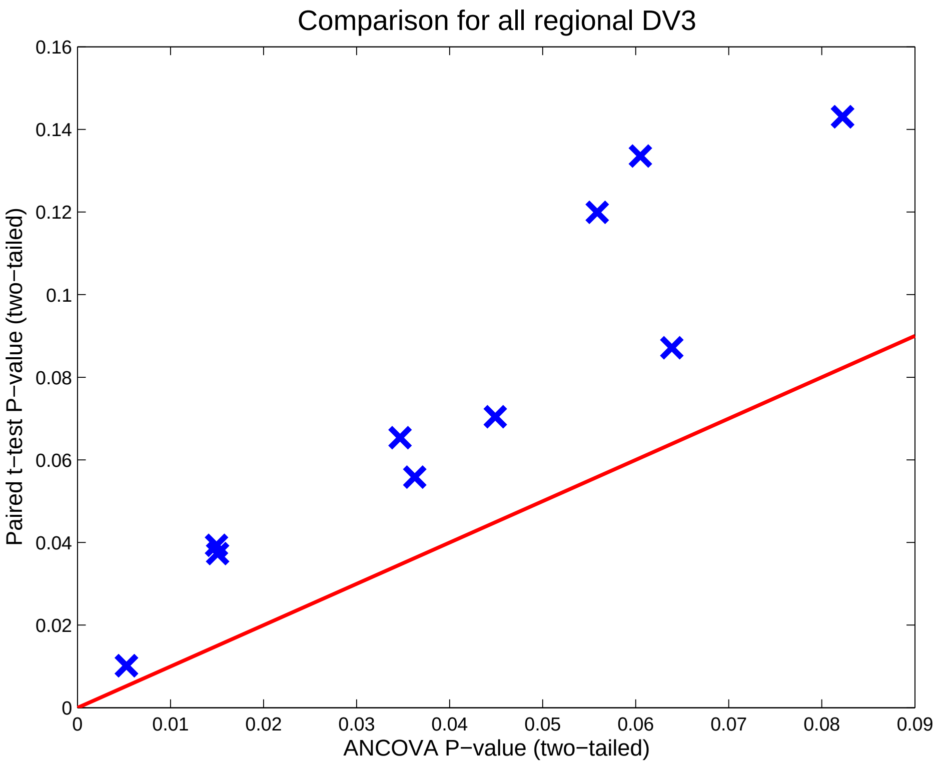

Figure 8: Comparison of two-tailed P-values between ANCOVA and paired t-test for a set of regions.

Comparison of P-values between ANCOVA and paired t-test.

Paired t-test: One design variable (intercept), 20 samples.

ANCOVA t-test: Four design variables (subject group 1, subject group 2, age, intercept), 40 samples.

(Here) ANCOVA is "better" than paired t-test. If you are allow to choose between the two analyses!

Estimation

The "normal equation" to estimate the parameters $\mathbf{B}$

$$\hat{\mathbf{B}} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{Y}, \tag{17}$$

or with the pseudo-inverse † (`pinv` in Matlab)

$$\hat{\mathbf{B}} = \mathbf{X}^{\dagger}\mathbf{Y}. \tag{18}$$

The pseudo-inverse will also work for design matrices of degenerate rank.

The "'fitted' error matrix" (Mardia)

$$\hat{\mathbf{U}} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}. \tag{19}$$

The residual sum of squares and products (SSP) matrix $\hat{\mathbf{U}}^{\mathsf{T}}\hat{\mathbf{U}}$ is a (voxels$\times$ voxels)-matrix. In a mass-univariate test only the diagonal is used

"Extra sum-of-squares" (ESS in SPM, "extra" variance not explained by the design variables)

"General Linear Hypothesis"

Most general form (Mardia et al., 1979, sec. 6.3)

$$\mathbf{CBM} = \mathbf{D} \tag{20}$$

Usually only a "null" $(\mathbf{D} = \mathbf{0})$ hypothesis is tested and with $\mathbf{M} = \mathbf{I}$

$$\mathbf{CB} = \mathbf{0} \tag{21}$$

Univariate hypothesis with an F-test

$$\mathbf{Cb} = \mathbf{0} \tag{22}$$

A univariate t-test with $\mathbf{c}$ as a row vector

$$\mathbf{cb} = 0, \tag{23}$$

Mass-univariate t-test

$$\mathbf{cB} = \mathbf{0}^{\mathsf{T}}. \tag{24}$$

Example contrasts

F-contrast for ANOVA with 3 groups encoded in an overparametrized design matrix (cf. SPM2 `spm_conman.m`)

$$\mathbf{C} = \begin{bmatrix} +1 & -1 & 0 & 0 \\ 0 & +1 & -1 & 0 \end{bmatrix} \tag{25}$$

t-contrast with 2 groups, one covariate and one grand mean

$$\mathbf{C} = \begin{bmatrix} +1 & -1 & 0 & 0 \end{bmatrix} \tag{26}$$

Testable contrasts

For design matrices of degenerate rank not all contrasts are valid: The contrast matrix $\mathbf{C}$ should be *testable* (Mardia et al., 1979, sec. 6.4).

$\mathbf{C}$ should be in the subspace of $\mathbf{X}$: $\mathcal{C} \subset \mathcal{X}$

$$\mathbf{0} = \mathbf{C} - \mathbf{C}\mathbf{X}^{\dagger}\mathbf{X} \tag{27}$$

In practice it should be numerically zero.

With rank($\mathbf{X}$)-truncated singular value decomposition of $\mathbf{X}$

$$\mathbf{X} = \mathbf{U}\mathbf{L}\mathbf{V}^{\mathsf{T}} + \mathbf{E}, \tag{28}$$

the projection can be computed from the eigenvectors $\mathbf{V}$

$$\mathbf{X}^{\dagger}\mathbf{X} = \mathbf{V}\mathbf{V}^{\mathsf{T}}. \tag{29}$$

(SPM2 `spm_sp.m` lines 973–980, 1211–1217; `spm_SpmUtil.m` line 282)

Nuisances: Simultaneous or "pre-processing"

Design matrix with interesting variables $\mathbf{X}_I$ and with uninteresting effects (nuisances/confounds) $\mathbf{X}_N$

"Simultaneous" modeling:

$$\mathbf{Y} = [\mathbf{X}_I, \mathbf{X}_N]\,\mathbf{B} + \mathbf{U} \tag{30}$$

"Pre-processed": Initial extration of confounds

$$\begin{array}{rcll} \mathbf{Y} &=& \mathbf{X}_N\mathbf{B}_N + \tilde{\mathbf{U}} & \text{"Pre-processing"} \\ \tilde{\mathbf{U}} &=& \mathbf{X}_I\mathbf{B}_I + \mathbf{U} & \text{Actual analysis} \end{array} \tag{31}$$

"Post-processed": Initial analysis of interesting effects followed by modeling of non-interesting effects.

What is the difference between the results from the three analyses?

Masked estimation (a la Goutte)

Resources

Short Course on Statistical Parametric Mapping, ftp://ftp.fil.ion.ucl.ac.uk/spm/course/notes04/slides/london2004.htm

Jonathan Taylors notes for his “stats191” course: http://www-stat.stanford.edu/~jtaylo/courses/stats191/spring.2004/

“General Linear Models” StatSoft, Inc, http://www.statsoftinc.com-/textbook/stglm.html

References

Adams, K. H., Pinborg, L. H., Svarer, C., Hasselbalch, S. G., Holm, S., Haugbol, S., Madsen, K., Frokjaer, V., Martiny, L., Paulson, O. B., and Knudsen, G. M. (2004). A database of [(18)F]-altanserin binding to 5-HT(2A) receptors in normal volunteers: normative data and relationship to physiological and demographic variables. *Neuroimage*, 21(3):1105–1113. PMID: 15006678. DOI: 10.1016/j.neuroimage.2003.10.046. ISSN 1053-8119.

Armitage, P. and Berry, G. (1994). *Statistical Methods in Medical Research*. Blackwell Science, Oxford, United Kingdom, third edition. ISBN 063203695.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Probability and Mathematical Statistics. Academic Press, London. ISBN 0124712525.