Setting up group models
Part 1
NITP, 2011
What is coming up

• Crash course in setting up models
  – 1-sample and 2-sample t-tests
  – Paired t-tests
  – ANOVA!
• Mean centering covariates
• Identifying rank deficient matrices
Recall

• GLM is flexible
  – One Sample T Test
  – ANOVA
  – Two sample T Test
  – Paired T test

• What do the models look like?
1-Sample T Test

\[ X\beta = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \beta \]

\[ H_0 : c\beta = 0 \quad \text{where} \quad c = [1] \]
2-Sample T Test

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_2 \\
A_2 \\
A_2 \\
A_2
\end{pmatrix}
= \begin{pmatrix}
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2
\end{pmatrix}
\]

\[H_0 : c\beta = 0 \text{ where } c = [1 \quad -1]\]
2-Sample T Test

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_1 \\
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_2 \\
A_2 \\
A_2 \\
\end{pmatrix}
\begin{pmatrix}
1 & 1 \\
1 & 1 \\
1 & 1 \\
1 & 1 \\
1 & 1 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
\end{pmatrix}
= 
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\end{pmatrix}
\]
Understanding a model

• If you’re unsure about a model or the contrasts
  – Plug in numbers
  – Look at graphs (fMRI data)

• Always ask yourself if your model is doing what you want it to
For example…

- For the 2 sample T test
  - Set \( \beta_1 = 3 \) \( \beta_2 = 5 \)
  - Then G1=8 and G2=3
  - So \( \beta_1 \) is the mean of group 2 and \( \beta_2 \) is the difference between the groups
  - What are the contrasts to test
    - Mean of G2
    - Mean of G1
    - G1-G2

\[
X\beta = \begin{pmatrix}
1 & 1 \\
1 & 1 \\
1 & 1 \\
1 & 1 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2
\end{pmatrix}
\]
For example...

- For the 2 sample T test
  - Set $\beta_1 = 3$ $\beta_2 = 5$
  - Then G1=8 and G2=3
  - So $\beta_1$ is the mean of group 2 and $\beta_2$ is the difference between the groups
  - What are the contrasts to test
    - Mean of G2 $c = [1 \ 0]$
    - Mean of G1 $c = [1 \ 1]$
    - G1-G2

$$X\beta = \begin{pmatrix}
1 & 1 \\
1 & 1 \\
1 & 1 \\
1 & 1 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0
\end{pmatrix} \begin{pmatrix}
\beta_1 \\
\beta_2
\end{pmatrix}$$
For example…

• For the 2 sample T test
  – Set $\beta_1 = 3$  $\beta_2 = 5$
  – Then G1=8 and G2=3
  – So $\beta_1$ is the mean of group 2 and $\beta_2$ is the difference between the groups
  – What are the contrasts to test
    • Mean of G2 $c = [1 \ 0]$
    • Mean of G1 $c = [1 \ 1]$
    • G1-G2 $c = [0 \ 1]$
Paired T Test

- A common mistake is to use a 2-sample t test instead of a paired test
- Tire example

<table>
<thead>
<tr>
<th>Automobile</th>
<th>Tire A</th>
<th>Tire B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.6</td>
<td>10.2</td>
</tr>
<tr>
<td>2</td>
<td>9.8</td>
<td>9.4</td>
</tr>
<tr>
<td>3</td>
<td>12.3</td>
<td>11.8</td>
</tr>
<tr>
<td>4</td>
<td>9.7</td>
<td>9.1</td>
</tr>
<tr>
<td>5</td>
<td>8.8</td>
<td>8.3</td>
</tr>
</tbody>
</table>

- 2-sample T test p=0.58
- Paired T test p<0.001
Why so different?
Why so different?

![Graph showing tire wear comparison between Mean A and Mean B across different automobiles.](image-url)
Why so different?

Difference is OK
Why so different?

Residuals are HUGE!
Paired T Test

Adjust for the mean of each pair
Paired T Test

![Graph showing paired T Test results with mean A and mean B indicated.](image-url)
Paired T Test

- Difference is same
- Residual variance much smaller
Paired T Test GLM

\[
\begin{pmatrix}
A_1 \\
B_1 \\
A_2 \\
B_2 \\
A_3 \\
B_3 \\
A_4 \\
B_4 \\
A_5 \\
B_5
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 \\
-1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 \\
-1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 \\
-1 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6
\end{pmatrix}
\]

\[H_0 : \text{Paired difference} = 0\]
\[H_0 : c\beta = 0, \quad c = [1 \ 0 \ 0 \ 0 \ 0 \ 0]\]
ANOVA

1-way ANOVA

\[ \mu_1 \]
\[ \mu_2 \]
\[ \vdots \]
\[ \mu_N \]

2-way ANOVA

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>\mu_{11}</td>
<td>\mu_{12}</td>
</tr>
<tr>
<td>\mu_{21}</td>
<td>\mu_{22}</td>
</tr>
<tr>
<td>\mu_{31}</td>
<td>\mu_{32}</td>
</tr>
</tbody>
</table>

![Graphs showing measure and levels of A and B](image)
Modeling ANOVA with GLM

• Cell means model
  – Model a mean for each “cell”
Modeling ANOVA with GLM

• Factor effects model
  – Model each factor as a set of regressors
  – One regressor for overall mean, other regressors describe how factor effects relate to the overall mean
Modeling ANOVA with GLM

- Factor effects model
  - Model the overall mean and have regressors for each factor
  - Hypothesis tests from this model correspond to standard ANOVA hypotheses
    - Eg, if group (2 levels) and stimulus type (3 levels) are modeled you can test for a main group effect, main stimulus type effect and interaction effect
1 Way ANOVA - Cell Means

$$
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\end{pmatrix}
$$

$$H_0 : G2 - G3 = 0$$
1 Way ANOVA - Cell Means

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4
\end{pmatrix}
=
\begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4
\end{pmatrix}
\]

\[H_0 : G2 - G3 = 0\]
\[H_0 : c\beta = 0 \quad \text{where} \quad c = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix}\]
1 Way ANOVA - Cell Means

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\end{pmatrix}
\]

\[H_0 : G1 = G2 = G3 = G4 = 0\]
1 Way ANOVA - Cell Means

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\end{pmatrix}
\]

\[H_0 : G1 = G2 = G3 = G4 = 0\]

\[H_0 : c\beta = 0 \quad \text{where} \quad c = 
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}\]
1 Way ANOVA factor effects

- Always start with overall mean (column of 1s)
- Number of regressors for a specific factor is number of levels – 1
  - Why?
1 Way ANOVA factor effects

• Always start with overall mean (column of 1s)
• Number of regressors for a specific factor is number of levels – 1
  – Why?
  – If I know the sum of 4 numbers is 10, then I only need to know 3 of the numbers to figure out what the fourth is.
In general
- # of regressors for a factor = # levels - 1
- Factor with 4 levels

\[ X_i = \begin{cases} 
1 & \text{if case from level } i \\
-1 & \text{if case from level 4} \\
0 & \text{otherwise}
\end{cases} \]
1 Way ANOVA - Factor Effects

• In general
  – # of regressors for a factor = # levels-1
  – Factor with 4 levels

\[ X_i = \begin{cases} 
1 & \text{if case from level } i \\
-1 & \text{if case from level 4} \\
0 & \text{otherwise}
\end{cases} \]

Note: I’m assuming a balanced design!!
1 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4
\end{pmatrix}
= \begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4
\end{pmatrix}
\text{mean}
\]
### 1 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\end{pmatrix}
\]

\[
G_1 = \beta_1 + \beta_2 \\
G_2 = \beta_1 + \beta_3 \\
G_3 = \beta_1 + \beta_4 \\
G_4 = \beta_1 - \beta_2 - \beta_3 - \beta_4 \\
\]

**mean**
1 Way ANOVA - Factor Effects

Mean g1
Mean g2
Mean g3
Mean g4
1 Way ANOVA - Factor Effects

- Mean g1
- Mean g2
- Mean g3
- Mean g4
1 Way ANOVA - Factor Effects

- Mean g1
- Mean g2
- Mean g3
- Mean g4
1 Way ANOVA - Factor Effects

Mean g1

Mean g2

Mean g3

Mean g4
1 Way ANOVA - Factor Effects

Mean g1
Mean g2
Mean g3
Mean g4

\[ \beta_1 \]
\[ \beta_2 \]
\[ \beta_3 \]
\[ \beta_4 \]

\[ -\beta_2 - \beta_3 - \beta_4 \]
1 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\end{pmatrix}
\]

Test main effect of factor A?

\[
c = 
\begin{pmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}
\]
1 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4
\end{pmatrix}

H_0 : \text{mean of G1} = 0
1 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\end{pmatrix}
\]

\[H_0 : \text{mean of G1} = 0\]

\[H_0 : c\beta = 0 \quad \text{where} \quad c = [1 \ 1 \ 0 \ 0 \ 0]\]
1 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\end{pmatrix}

H_0 : G1 - G4 = 0
1 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1 \\
A_1 \\
A_2 \\
A_2 \\
A_3 \\
A_3 \\
A_4 \\
A_4 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 \\
1 & -1 & -1 & -1 \\
1 & -1 & -1 & -1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\end{pmatrix}
\]

\[H_0 : G1 - G4 = 0\]

\[c = (1 \ 1 \ 0 \ 0) - (1 \ -1 \ -1 \ -1 \ -1) = (0 \ 2 \ 1 \ 1)\]
2 Way ANOVA (3x2)

\[
\begin{pmatrix}
A_1B_1 \\
A_1B_1 \\
A_1B_2 \\
A_1B_2 \\
A_2B_1 \\
A_2B_1 \\
A_2B_2 \\
A_2B_2 \\
A_3B_1 \\
A_3B_1 \\
A_3B_2 \\
A_3B_2
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6
\end{pmatrix}
\]

\(H_0: \text{main factor A effect} = 0\)
No effect means the marginals would be the same

Null: A1=A2=A3  equivalently A1-A3=0 and A2-A3=0

2 Way ANOVA (3x2)
2 Way ANOVA (3x2)

\[
\begin{pmatrix}
A_1B_1 \\
A_1B_1 \\
A_1B_2 \\
A_1B_2 \\
A_2B_1 \\
A_2B_1 \\
A_2B_2 \\
A_2B_2 \\
A_3B_1 \\
A_3B_1 \\
A_3B_2 \\
A_3B_2
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6
\end{pmatrix}
\]

\[H_0: \text{main factor A effect} = 0\]
2 Way ANOVA (3x2)

\[
\begin{pmatrix}
A_1B_1 \\
A_1B_1 \\
A_1B_2 \\
A_1B_2 \\
A_2B_1 \\
A_2B_1 \\
A_2B_2 \\
A_2B_2 \\
A_3B_1 \\
A_3B_1 \\
A_3B_2 \\
A_3B_2 \\
A_3B_2 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6 \\
\end{pmatrix}
\]

\[H_0 : \text{main factor A effect} = 0\]

\[H_0 : c\beta = 0 \text{ where } c = \begin{pmatrix}
1 & 1 & 0 & 0 & -1 & -1 \\
0 & 0 & 1 & 1 & -1 & -1 \\
\end{pmatrix}\]
2 Way ANOVA (3x2)

\[
\begin{pmatrix}
A_1B_1 \\
A_1B_1 \\
A_1B_2 \\
A_1B_2 \\
A_2B_1 \\
A_2B_1 \\
A_2B_2 \\
A_2B_2 \\
A_3B_1 \\
A_3B_1 \\
A_3B_2 \\
A_3B_2 \\
\end{pmatrix}
= \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6 \\
\end{pmatrix}
\]

\[H_0 : \text{interaction effect} = 0\]
2 Way ANOVA (3x2)

\[ H_0 : \text{interaction effect} = 0 \]

No effect means the lines would be parallel
2 Way ANOVA (3x2)

\[ H_0 : \text{interaction effect} = 0 \]

No effect means the lines would be parallel

2 Way ANOVA (3x2)

\[ H_0 : \text{interaction effect} = 0 \]

No effect means the lines would be parallel

2 Way ANOVA (3x2)

\[
\begin{pmatrix}
A_1B_1 \\
A_1B_1 \\
A_1B_2 \\
A_1B_2 \\
A_2B_1 \\
A_2B_1 \\
A_2B_2 \\
A_2B_2 \\
A_3B_1 \\
A_3B_1 \\
A_3B_2 \\
A_3B_2
\end{pmatrix}
= \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6
\end{pmatrix}

H_0 : interaction effect = 0

H_0 : c\beta = 0 \text{ where } c = \begin{pmatrix}
1 & -1 & 0 & 0 & -1 & 1 \\
0 & 0 & 1 & -1 & -1 & 1
\end{pmatrix}
By the way…

• If you feel confused by that last example, that was my goal!
2 Way ANOVA - Factor Effects

• Recall for factor effects, a factor with $n$ levels has regressors set up like

$$X_i = \begin{cases} 
1 & \text{if case from level } i \\
-1 & \text{if case from level } n \\
0 & \text{otherwise}
\end{cases}$$

• $A$ has 3 levels, so 2 regressors
• $B$ has 2 levels, so 1 regressor
2 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1B_1 \\
A_1B_1 \\
A_1B_2 \\
A_1B_2 \\
A_2B_1 \\
A_2B_1 \\
A_2B_2 \\
A_2B_2 \\
A_3B_1 \\
A_3B_1 \\
A_3B_2 \\
A_3B_2 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6 \\
\end{pmatrix}
\]
2 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1 B_1 \\
A_1 B_1 \\
A_1 B_2 \\
A_1 B_2 \\
A_2 B_1 \\
A_2 B_1 \\
A_2 B_2 \\
A_2 B_2 \\
A_3 B_1 \\
A_3 B_1 \\
A_3 B_2 \\
A_3 B_2 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6 \\
\end{pmatrix}
\]

\(H_0 :\) main factor A effect = 0
2 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1B_1 \\
A_1B_2 \\
A_1B_1 \\
A_2B_1 \\
A_2B_2 \\
A_2B_1 \\
A_3B_1 \\
A_3B_2 \\
A_3B_1
\end{pmatrix}
= \begin{pmatrix}
1 & 1 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 & 1 \\
1 & 0 & 1 & -1 & 0 & -1 & 1 \\
1 & 0 & 1 & -1 & 0 & -1 & 1 \\
1 & -1 & -1 & 1 & -1 & -1 & 1 \\
1 & -1 & -1 & 1 & -1 & -1 & 1 \\
1 & -1 & -1 & -1 & 1 & 1 & 1 \\
1 & -1 & -1 & -1 & 1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6
\end{pmatrix}
\]

\[H_0: \text{main factor A effect} = 0\]

\[H_0: c\beta = 0 \quad \text{where} \quad c = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}\]
2 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1B_1 \\
A_1B_1 \\
A_1B_2 \\
A_1B_2 \\
A_2B_1 \\
A_2B_1 \\
A_2B_2 \\
A_2B_2 \\
A_3B_1 \\
A_3B_1 \\
A_3B_2 \\
A_3B_2 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6 \\
\end{pmatrix}
\]

\( H_0 : \text{interaction effect} = 0 \)
2 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1 B_1 \\
A_1 B_1 \\
A_1 B_2 \\
A_1 B_2 \\
A_2 B_1 \\
A_2 B_1 \\
A_2 B_2 \\
A_2 B_2 \\
A_3 B_1 \\
A_3 B_1 \\
A_3 B_2 \\
A_3 B_2 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6 \\
\end{pmatrix}
\]

\[H_0 : \text{interaction effect} = 0\]

\[H_0 : c\beta = 0 \text{ where } c = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}\]
2 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1B_1 \\
A_1B_1 \\
A_1B_2 \\
A_1B_2 \\
A_2B_1 \\
A_2B_1 \\
A_2B_2 \\
A_2B_2 \\
A_3B_1 \\
A_3B_1 \\
A_3B_2 \\
A_3B_2 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6 \\
\end{pmatrix}
\]

\[H_0 : \text{mean cell } A_1B_1 = 0\]
### 2 Way ANOVA - Factor Effects

\[
\begin{pmatrix}
A_1B_1 \\
A_1B_1 \\
A_1B_2 \\
A_1B_2 \\
A_2B_1 \\
A_2B_1 \\
A_2B_2 \\
A_2B_2 \\
A_3B_1 \\
A_3B_1 \\
A_3B_2 \\
A_3B_2 \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 1 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 & 0 \\
1 & 1 & 0 & -1 & -1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & 0 & 1 & -1 & 0 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & -1 & -1 & 1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6 \\
\end{pmatrix}
\]

\[H_0 : \text{mean cell } A_1B_1 = 0\]

\[H_0 : c\beta = 0 \text{ where } c = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}\]
For more examples

• The FSL folks have a bunch of great examples
  – [http://www.fmrib.ox.ac.uk/fsl/feat5/detail.html](http://www.fmrib.ox.ac.uk/fsl/feat5/detail.html)

• Check the FSL help list regularly
  – Subscribe at jiscmail
  – Often others have already asked your questions!
Mean centering covariates

• Why do we mean center?
• When should we mean center?
• What does it do to the parameter estimate interpretation?
Single group with covariate

- You have a single group of subjects and you also have measured age. You would like to see if there is an age effect.
  - What would the model look like?
  - What contrasts would you specify for the age effect?
  - Can you still obtain the overall mean from this model?
Both models will give exactly the same result for C2, but C1 will be different.
Simulated data
Summary of mean centering

• Only really necessary if you want your PE of column of 1s to be the overall mean
• Often people have rounding errors after demeaning. Double-check this when you do it.
We have two groups and a confounding covariate (depression). Our primary interest is in the difference of means between the two groups.
- What is the model to simply test the difference in means?
- If I wanted to make sure this difference wasn’t due to between group differences in depression, what would that model look like?
- What are some restrictions of this model?
• Do not demean the confounding measure within group
  • This removes any confounding effect the measure might have
Why you shouldn’t demean within group

- What if this is what your data look like?
  - Difference in means is clearly due to range of X sampled, not the group membership
Will mean centering across all subjects change anything?
Will mean centering across all subjects change anything?

Nope!
Testing the interaction

• Same as previous example, but you’d like to know if the relationship with depression score differs between your groups
• What would the design matrix and contrasts look like?
Design matrix

No mean centering!
What contrasts make sense?

- Do comparisons of means make any sense?
- What is interesting about a significant interaction?
Compare means with a significant interaction?

- For what value of depression is the difference in means interesting?
What is interesting when there’s an interaction effect?
Mean centering when testing an interaction

• It doesn’t make much sense
  – Difference between means varies by continuous covariate when there’s an interaction effect
• Testing the mean difference in general doesn’t make sense
• Only interesting contrast tests difference in slope for continuous covariates between groups
## Mean centering guide

- [http://mumford.fmripower.org/mean_centering/](http://mumford.fmripower.org/mean_centering/)

<table>
<thead>
<tr>
<th>Design matrix $X_0$</th>
<th>What does the fitted model look like?</th>
<th>Contrast</th>
<th>Is mean centering necessary?</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Design matrix" /></td>
<td><img src="image2" alt="Fitted model" /></td>
<td>$a^{(1)} b$</td>
<td>Yes. This case is interested in the mean of the independent variable. Note the shifts with this factor. The height of the line varies with $a$. All y-values change in the same way and are not equal to the average value. We are looking at an average subject with the average value of $X$. Without mean centering, you are looking at the mean of all the group when this is a continuous variable. Mean centering a covariate will not change the contrast unless an alternative to a continuous variable (e.g., a categorical variable) is modeled.</td>
</tr>
<tr>
<td>$X_1$</td>
<td><img src="image3" alt="Fitted model" /></td>
<td>$a^{(-1)} b$</td>
<td>No. The lines are parallel and you are testing if the difference between the two fitted lines is significantly different from 0. This serves as the same for all values of $X$.</td>
</tr>
<tr>
<td>$X_2$</td>
<td><img src="image4" alt="Fitted model" /></td>
<td>$a^{(0)} b$</td>
<td>No. Mean centering a covariate will change the difference for (not-covariate) image (the image is translated).</td>
</tr>
<tr>
<td><img src="image5" alt="Design matrix" /></td>
<td><img src="image6" alt="Fitted model" /></td>
<td>$a^{(0)} b^{(-1)}$</td>
<td>No. In this case, power is being tested in the mean of an individual group. Since this is a different slope, the contrast is different in the mean centering. The contrast is the same as the regression coefficients.</td>
</tr>
<tr>
<td>$X_3$</td>
<td><img src="image7" alt="Fitted model" /></td>
<td>$a^{(0)} b^{(0)}$</td>
<td>No. This is a difficult contrast to interpret if the interaction is not significant (0, 0, 0). This situation is when the slopes are parallel and the interactions are not significant. This means that the interaction is not significant which is not what you want to test when you have a continuous variable.</td>
</tr>
<tr>
<td><img src="image8" alt="Design matrix" /></td>
<td><img src="image9" alt="Fitted model" /></td>
<td>$a^{(0)} b^{(0)}$</td>
<td>No. This is in the case or power of an individual group. Since this changes with the value of $a$, the slope is not the same. The contrast is the same as the regression coefficients.</td>
</tr>
<tr>
<td><img src="image10" alt="Design matrix" /></td>
<td><img src="image11" alt="Fitted model" /></td>
<td>$a^{(0)} b^{(0)}$</td>
<td>No. This is in the case of searching for the group when this is a factor. The contrast is the same as the regression coefficients.</td>
</tr>
</tbody>
</table>

### Notes on Mean Centering

- **Mean centering** is a technique used in statistical analysis to adjust for potential confounding variables. It involves subtracting the mean of a variable from each of its values, thus centering the data around zero. This can help in reducing multicollinearity and improving the interpretability of regression coefficients.
- **When to use mean centering** depends on the specific research question and the nature of the variables involved. It is particularly useful when dealing with continuous variables or when interactions are being investigated.
- **Contrasts** refer to specific comparisons between groups or conditions in a statistical model. Mean centering can influence the interpretation of contrasts, especially when interactions are present.

---

**Image Notes**
- Image 1: Illustration of a design matrix with a fitted model showing the relationship between the independent variable and the dependent variable, with a focus on how mean centering affects the contrast.
- Image 2: Graph showing the fitted model with different contrasts and the implications of mean centering.
- Image 3: Diagram comparing different scenarios of contrast analysis with and without mean centering.

---

*This document provides a comprehensive guide to mean centering, including examples of design matrices and fitted models to illustrate the importance and application of mean centering in statistical analysis.*
Important assumption

• That the effect tested is linear!
• Relationship between BOLD activation and depression score is linear for both groups
• If your continuous covariate is significantly different between your groups, you must be very careful with interpretation
When trends are linear, it doesn’t matter where I sample age, I get the same slope.
When trends are linear, it doesn’t matter where I sample age, I get the same slope.
When trends are linear, it doesn’t matter where I sample age, I get the same slope
With a nonlinear trend, the trend may be the same in both groups, but sampling different ranges show different trends.
With a nonlinear trend, the trend may be the same in both groups, but sampling different ranges show different trends
With a nonlinear trend, the trend may be the same in both groups, but sampling different ranges show different trends.
Rank deficient designs

• I highly recommend becoming an expert on spotting rank deficient design matrices
  – Don’t email the FSL list to ask why you got the error message "at least one EV is (close to) a linear combination of the others"
Rank deficient designs

• I highly recommend becoming an expert on spotting rank deficient design matrices
  – Don’t email the FSL list to ask why you got the error message “at least one EV is (close to) a linear combination of the others”
    • (the answer to the problem is in the error message!!)

Rank deficient

• One column is a linear combination of some of the other columns
• Linear combination of one set of columns equals a linear combination of another set of columns
Rank deficient

\[
\begin{pmatrix}
1 & 0 & 7 \\
1 & 0 & 7 \\
0 & 1 & 0 \\
0 & 1 & 0
\end{pmatrix}
\quad \begin{pmatrix}
1 & 0 & 2 \\
1 & 0 & 2 \\
1 & 1 & 5 \\
1 & 1 & 5
\end{pmatrix}
\]
Rank deficient

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Rank deficient?

<table>
<thead>
<tr>
<th>Group</th>
<th>EV1</th>
<th>EV2</th>
<th>EV3</th>
<th>EV4</th>
<th>EV5</th>
<th>EV6</th>
<th>EV7</th>
</tr>
</thead>
<tbody>
<tr>
<td>subj1</td>
<td>subj2</td>
<td>time1</td>
<td>time2</td>
<td>time3</td>
<td>time4</td>
<td>time5</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Rank deficient?
That’s it!

• Next lecture we’ll look at more models. If you have a design that has you stumped, let me know what your data look like and we can all try to figure it out!