Setting up models

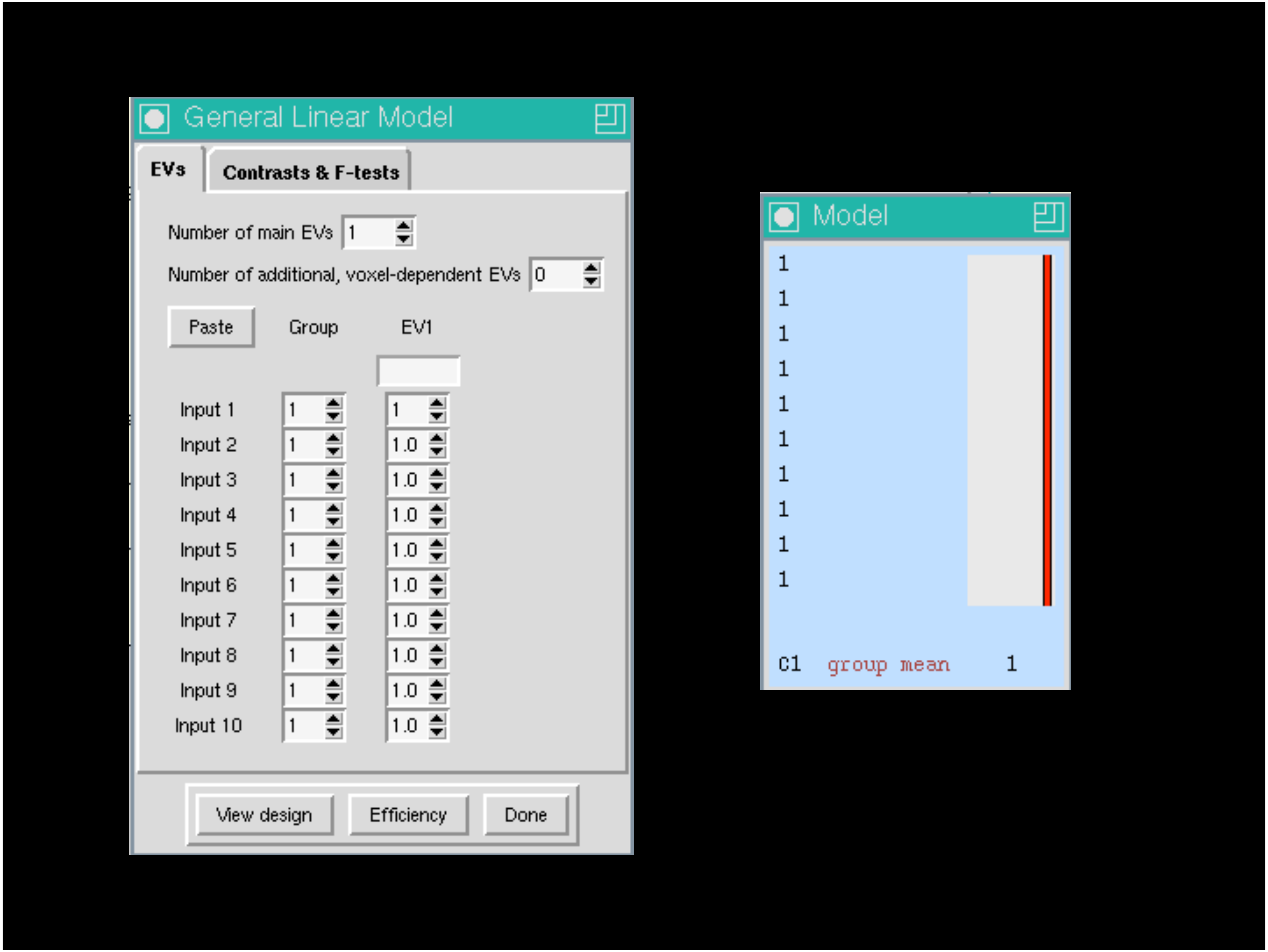
July 27, 2009

Making models!

- Warm up with some things we've already covered here
- Look at more complicated models
- Take a close look at orthogonalization

Model 1

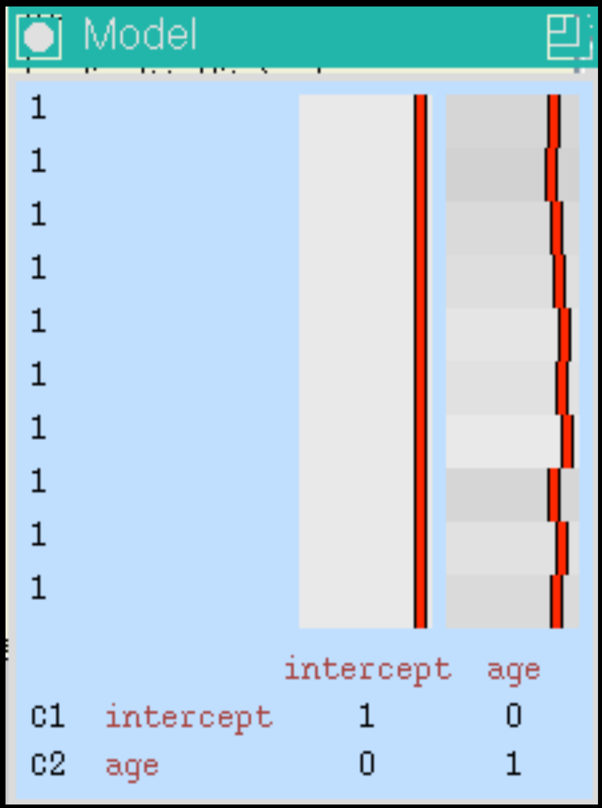
- You measure BOLD activation for some task and you would like the group mean for this activation. You have 10 subjects, what would the design look like?


General Linear Model
EVs
Contrasts & F-tests
Number of main EVs 1
Number of additional, voxel-dependent EVs 0
Paste
Group
EV1
Input 1 1 1
Input 2 1 1.0
Input 3 1 1.0
Input 4 1 1.0
Input 5 1 1.0
Input 6 1 1.0
Input 7 1 1.0
Input 8 1 1.0
Input 9 1 1.0
Input 10 1 1.0
View design
Efficiency
Done
Model
1
1
1
1
1
1
1
1
1
1
C1 group mean 1

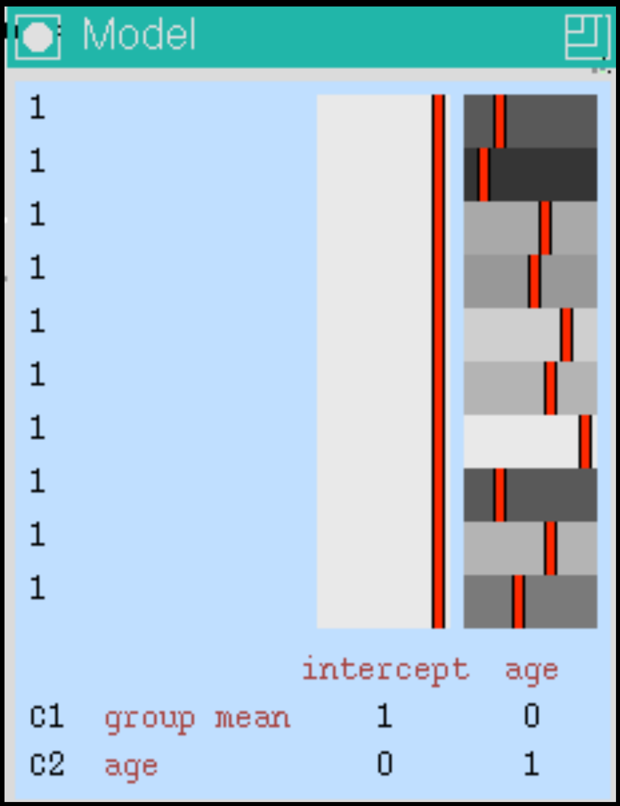
Model 2

- Building off Model 1, you also have measured age and 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?

Use age

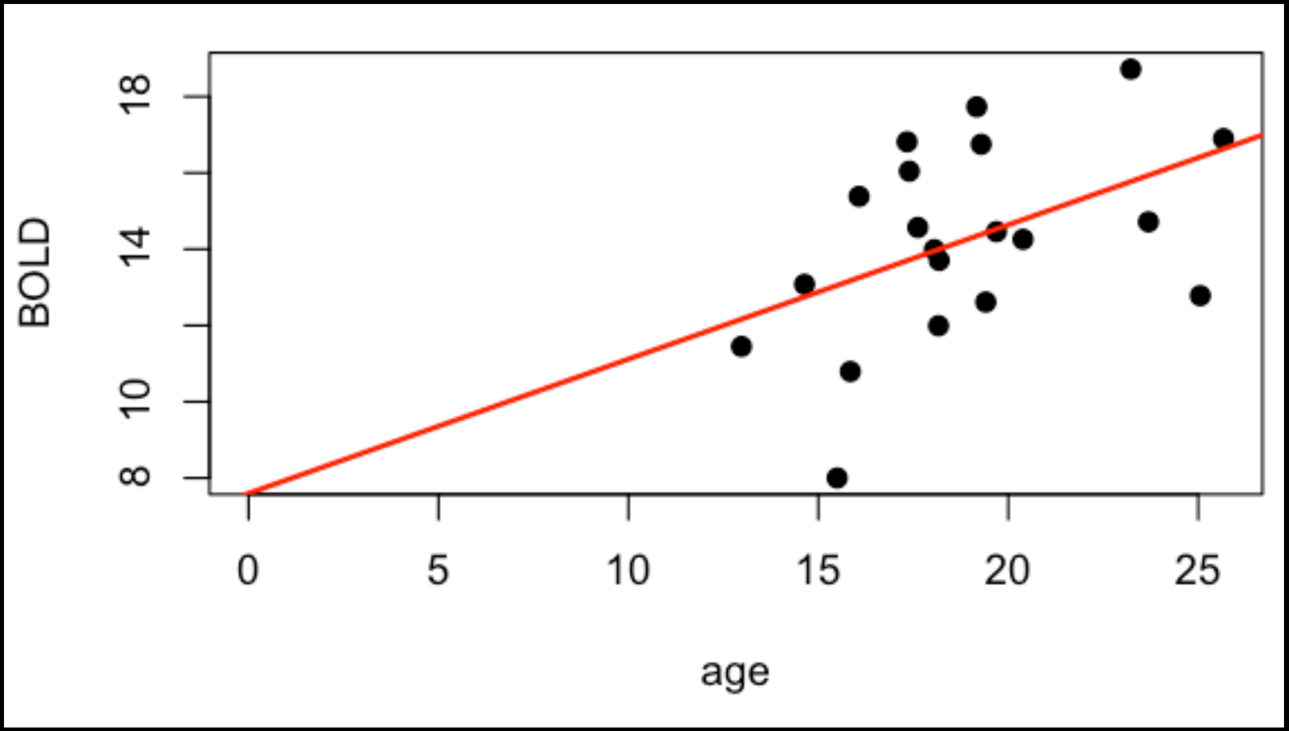


Use demeaned age



- Both models will give exactly the same result for C2, but C1 will be different.

Simulated data



Intercept ~ 7 or 8

Mean of BOLD ~ 14

Simulated data

```
Call:
lm(formula = BOLD ~ age)

Residuals:
     Min       1Q   Median       3Q      Max 
-5.04950 -1.36690 -0.01510  2.17838  3.38803 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)   7.5955     3.0545   2.487   0.0229 *
age           0.3520     0.1594   2.207   0.0405 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.359 on 18 degrees of freedom
Multiple R-squared: 0.213,  Adjusted R-squared: 0.1693 
F-statistic: 4.873 on 1 and 18 DF,  p-value: 0.0405
```

```
Call:
lm(formula = BOLD ~ age_demeaned)

Residuals:
     Min       1Q   Median       3Q      Max 
-5.04950 -1.36690 -0.01510  2.17838  3.38803 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   14.2367     0.5276  26.986 5.18e-16 ***
age_demeaned   0.3520     0.1594   2.207   0.0405 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.359 on 18 degrees of freedom
Multiple R-squared: 0.213,  Adjusted R-squared: 0.1693 
F-statistic: 4.873 on 1 and 18 DF,  p-value: 0.0405
```

Simulated data

```
Call:
lm(formula = BOLD ~ age)

Residuals:
     Min       1Q   Median       3Q      Max
-5.04950 -1.36690 -0.01510  2.17838  3.38803

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.5955     3.0545   2.487   0.0229 *
age           0.3520     0.1594   2.207   0.0405 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.359 on 18 degrees of freedom
Multiple R-squared: 0.213,  Adjusted R-squared: 0.1693
F-statistic: 4.873 on 1 and 18 DF,  p-value: 0.0405
```

```
Call:
lm(formula = BOLD ~ age_demeaned)

Residuals:
     Min       1Q   Median       3Q      Max
-5.04950 -1.36690 -0.01510  2.17838  3.38803

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   14.2367     0.5276  26.986 5.18e-16 ***
age_demeaned   0.3520     0.1594   2.207   0.0405 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.359 on 18 degrees of freedom
Multiple R-squared: 0.213,  Adjusted R-squared: 0.1693
F-statistic: 4.873 on 1 and 18 DF,  p-value: 0.0405
```

Models fit equally well

Simulated data

```
Call:
lm(formula = BOLD ~ age)

Residuals:
     Min       1Q   Median       3Q      Max 
-5.04950 -1.36690 -0.01510  2.17838  3.38803 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)   7.5955     3.0545   2.487   0.0229 *
age           0.3520     0.1594   2.207   0.0405 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.359 on 18 degrees of freedom
Multiple R-squared: 0.213,  Adjusted R-squared: 0.1693 
F-statistic: 4.873 on 1 and 18 DF,  p-value: 0.0405
```

```
Call:
lm(formula = BOLD ~ age_demeaned)

Residuals:
     Min       1Q   Median       3Q      Max 
-5.04950 -1.36690 -0.01510  2.17838  3.38803 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  14.2367     0.5276  26.986 5.18e-16 ***
age_demeaned  0.3520     0.1594   2.207   0.0405 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.359 on 18 degrees of freedom
Multiple R-squared: 0.213,  Adjusted R-squared: 0.1693 
F-statistic: 4.873 on 1 and 18 DF,  p-value: 0.0405
```

Coefficient for age and p-value don’t change

Simulated data

```
Call:
lm(formula = BOLD ~ age)

Residuals:
     Min       1Q   Median       3Q      Max
-5.04950 -1.36690 -0.01510  2.17838  3.38803

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.5955     3.0545   2.487   0.0229 *
age           0.3520     0.1594   2.207   0.0405 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.359 on 18 degrees of freedom
Multiple R-squared: 0.213,  Adjusted R-squared: 0.1693
F-statistic: 4.873 on 1 and 18 DF,  p-value: 0.0405
```

```
Call:
lm(formula = BOLD ~ age_demeaned)

Residuals:
     Min       1Q   Median       3Q      Max
-5.04950 -1.36690 -0.01510  2.17838  3.38803

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   14.2367     0.5276  26.986 5.18e-16 ***
age_demeaned   0.3520     0.1594   2.207   0.0405 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 2.359 on 18 degrees of freedom
Multiple R-squared: 0.213,  Adjusted R-squared: 0.1693
F-statistic: 4.873 on 1 and 18 DF,  p-value: 0.0405
```

- Parameter estimate and p-value changes for "intercept".
- In demeaned case, it is equal to the mean of BOLD

```
> mean(BOLD)
[1] 14.23671
```

Summary of demeaning

- 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.

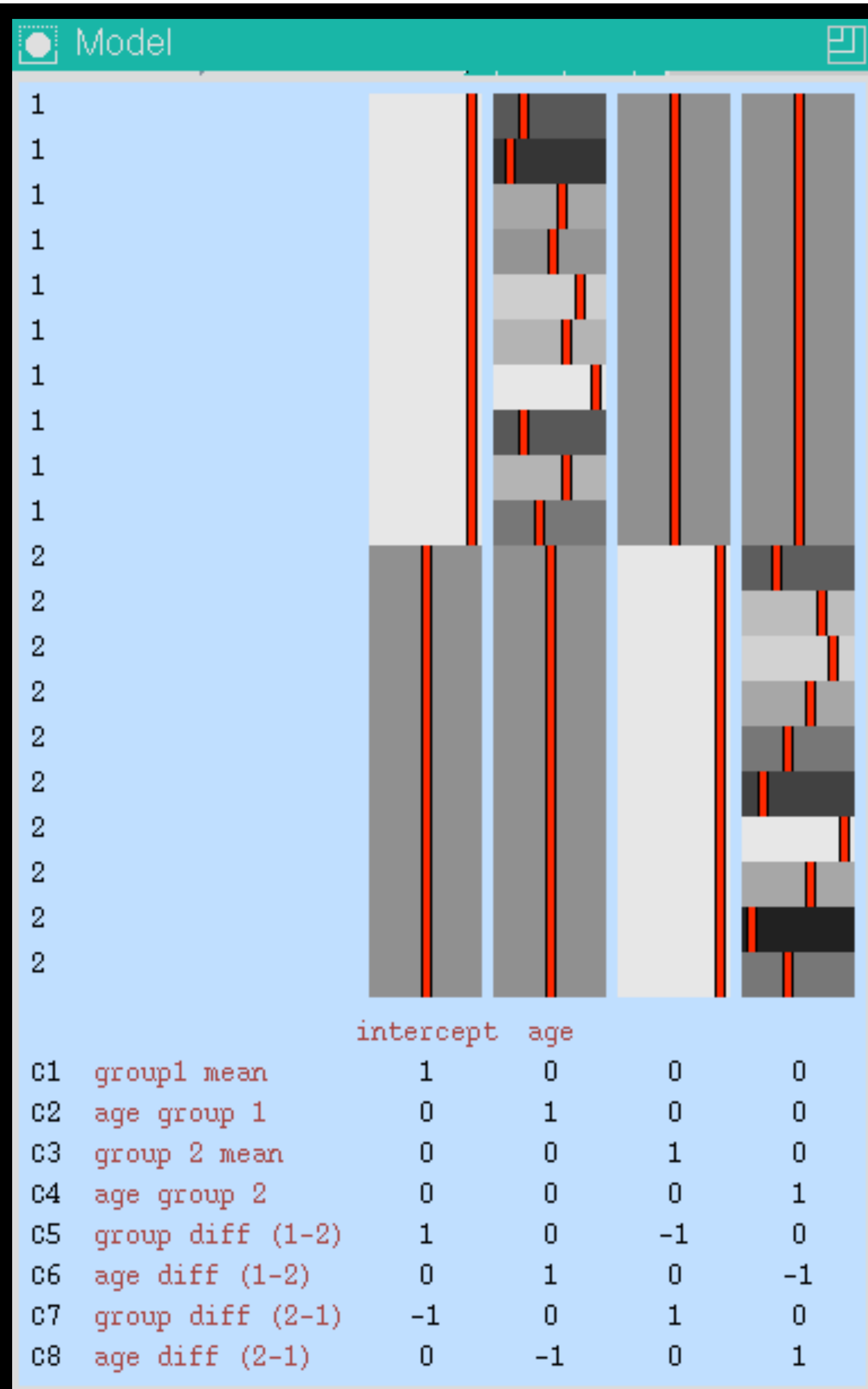
Continuous covariates between groups

- You have 2 groups and a continuous measure for all subjects (age)
  - 2 things you can look at
    - Within group trends in age: What is the relationship of age within group?
    - Effect of age between groups: How do differences in age between groups impact group differences?

Model 3

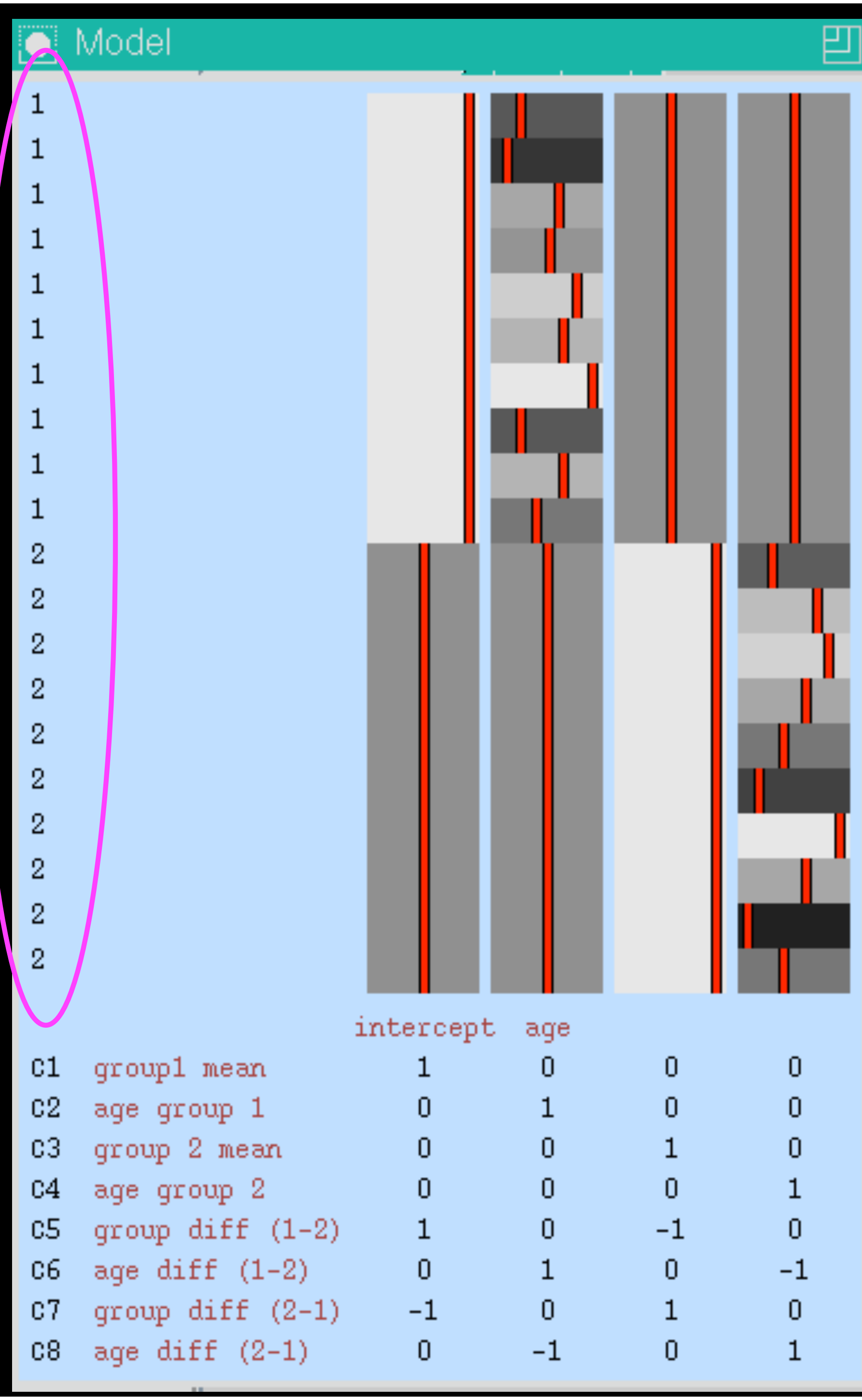
- Now you have 2 groups (10 sub each) and you'd like to see if there's a difference in overall mean between the groups and the <u>within</u> group age effect and how it compares between groups
  - What would the model look like?
  - What if your groups are expected to have different levels of variability (patients and controls)?
  - Should you demean age? How would you demean it?
  - What if the ages for your two groups are significantly different?

I demeaned age within group here since I’m interested in group means and not intercepts

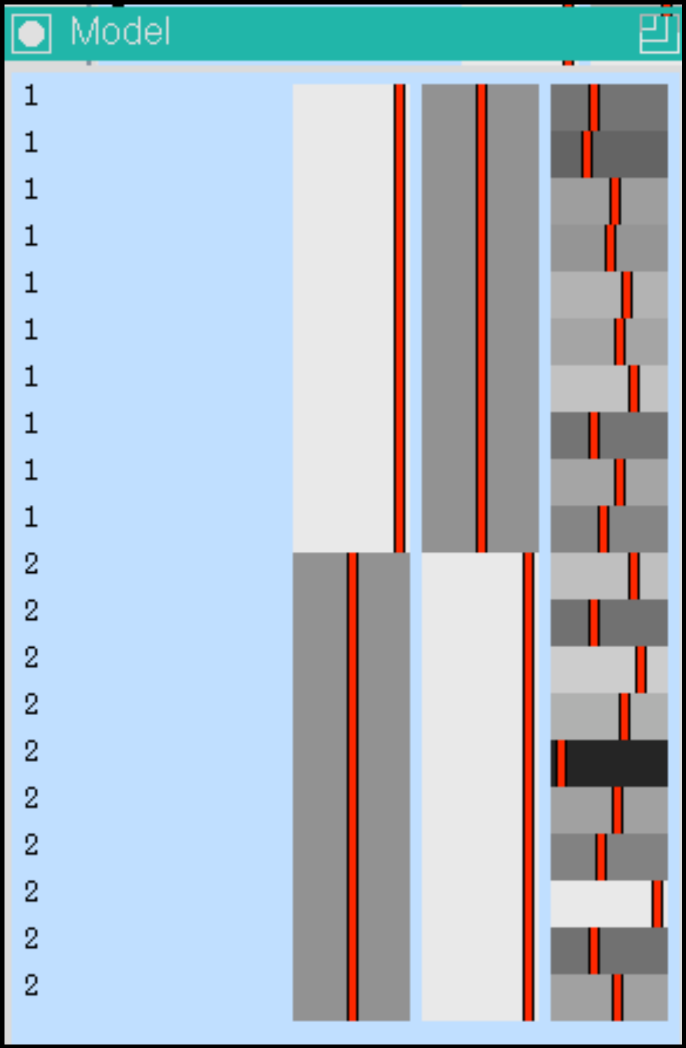


“group” column allows different variances in groups

Model must be separable!



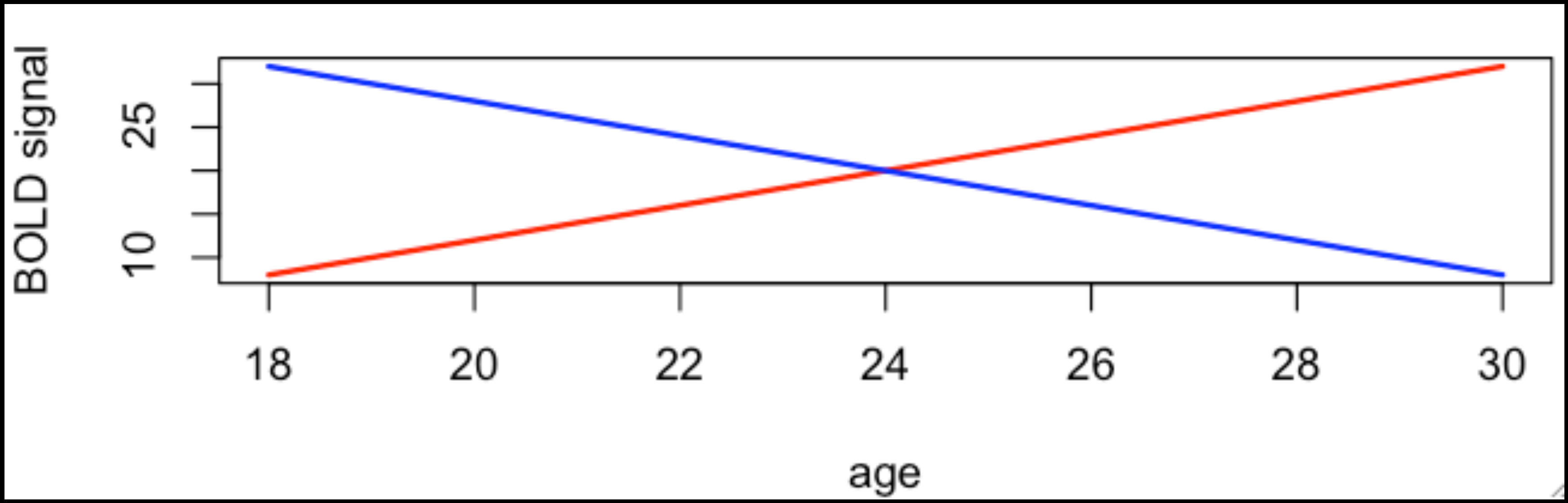
Not separable



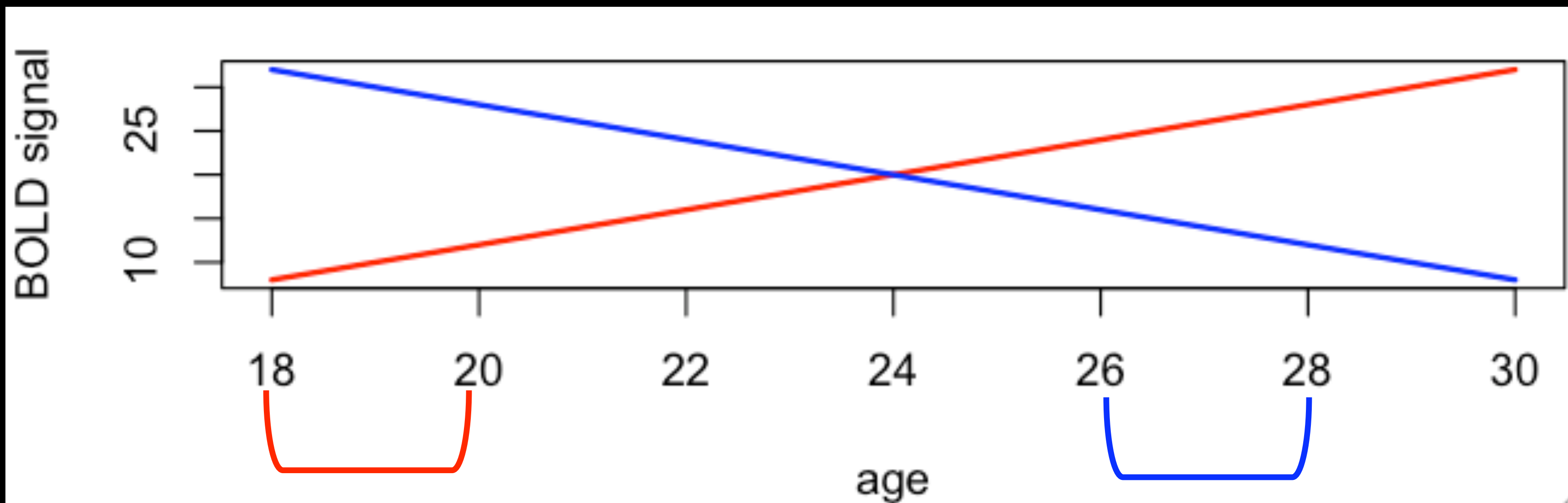
What if age is different between groups?

- Differences in slopes may only be because you're sampling different parts of the distribution
- Must assume trend is linear

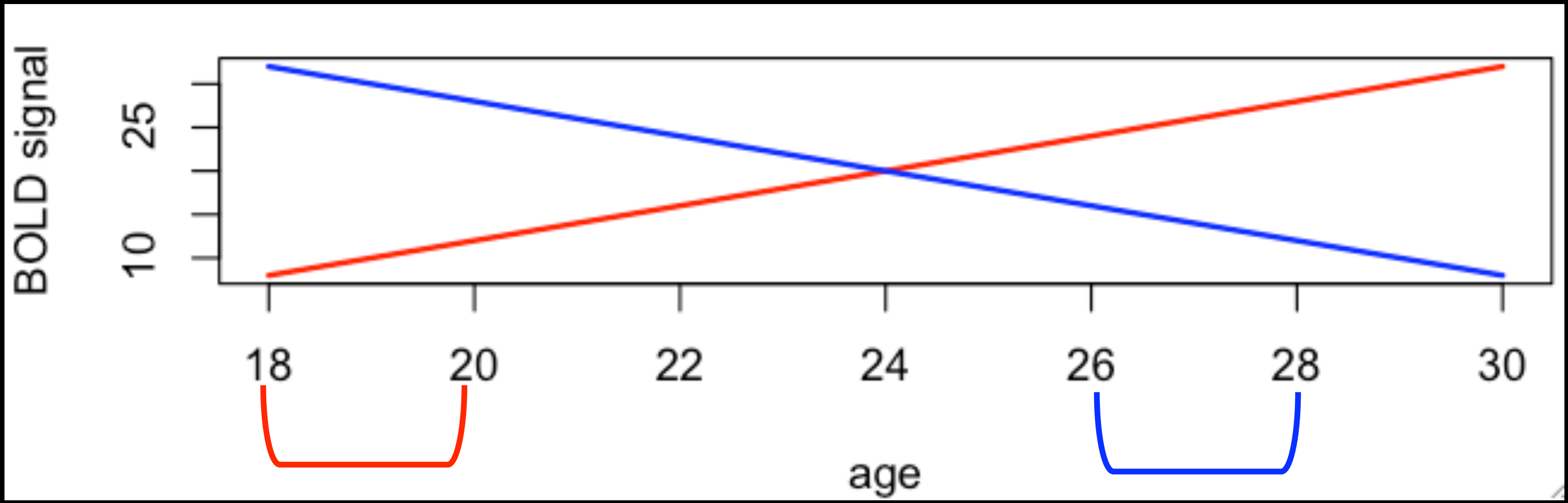
When trends are linear, it doesn't matter where I sample age, I get the same slope

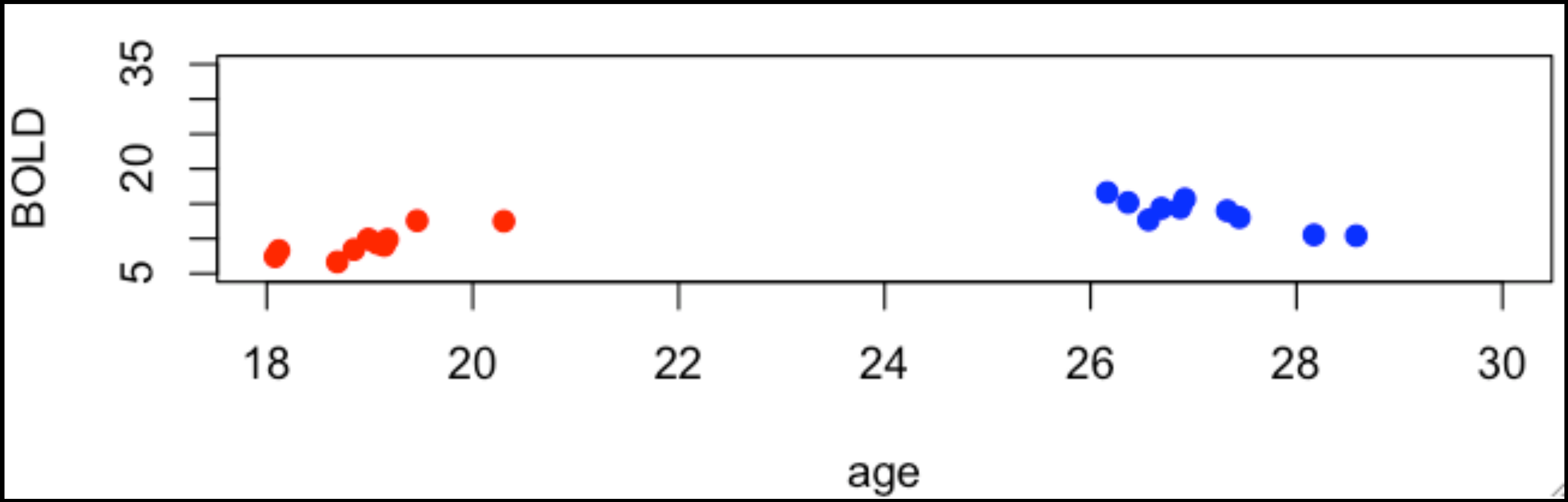


When trends are linear, it doesn't matter where I sample age, I get the same slope

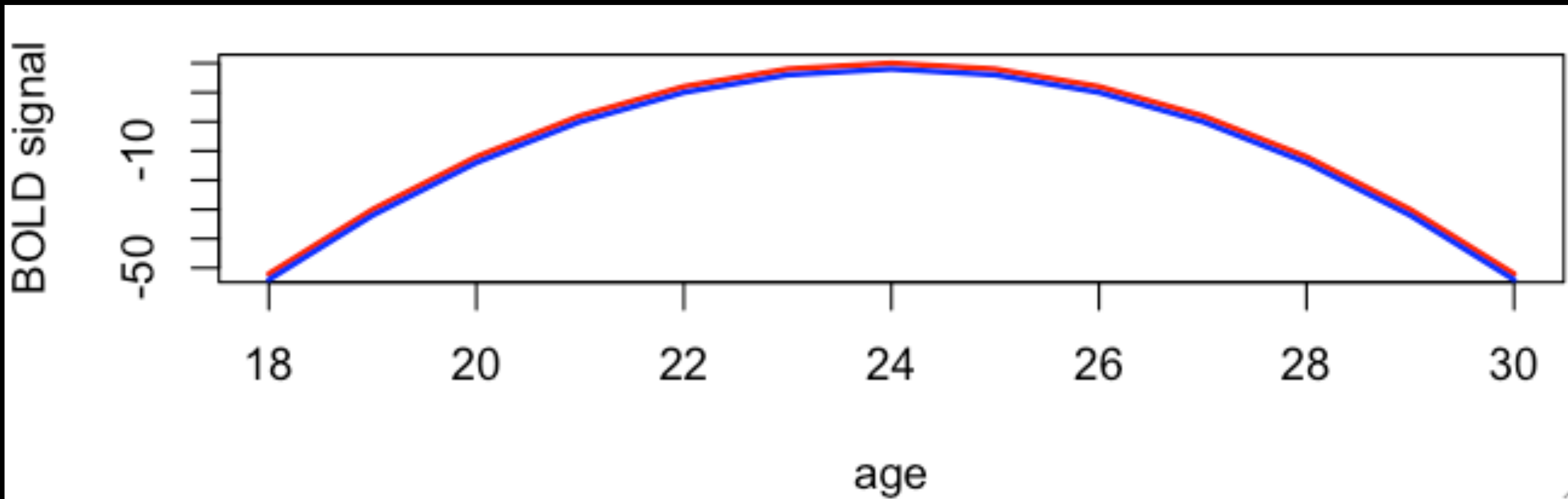


When trends are linear, it doesn't matter where I sample age, I get the same slope

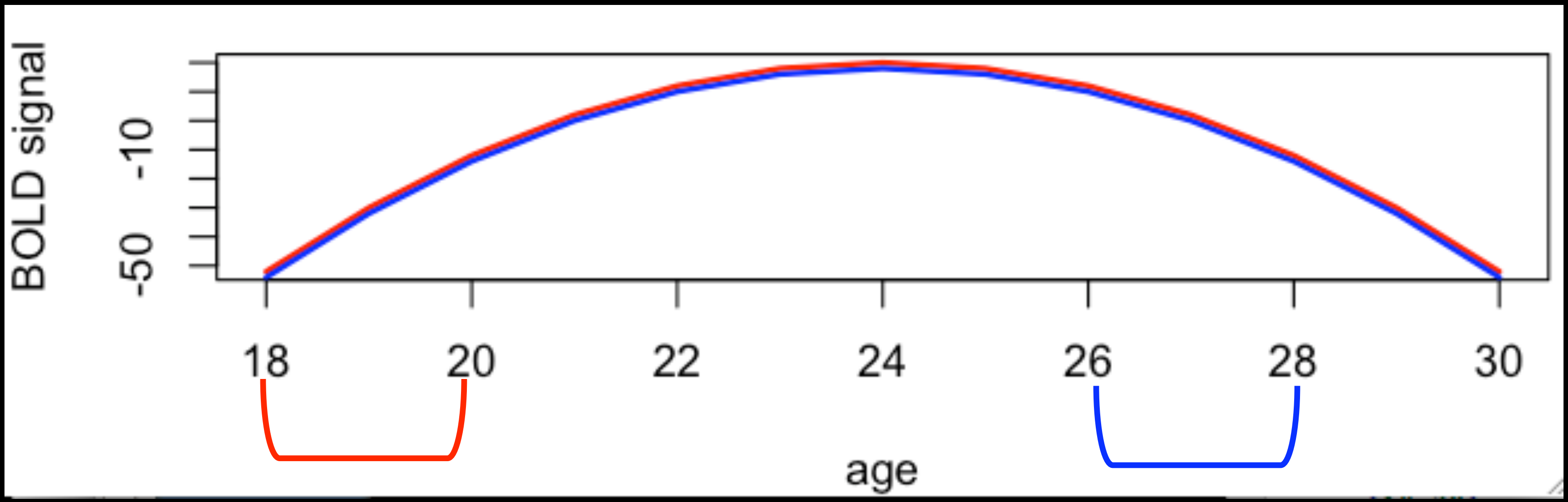




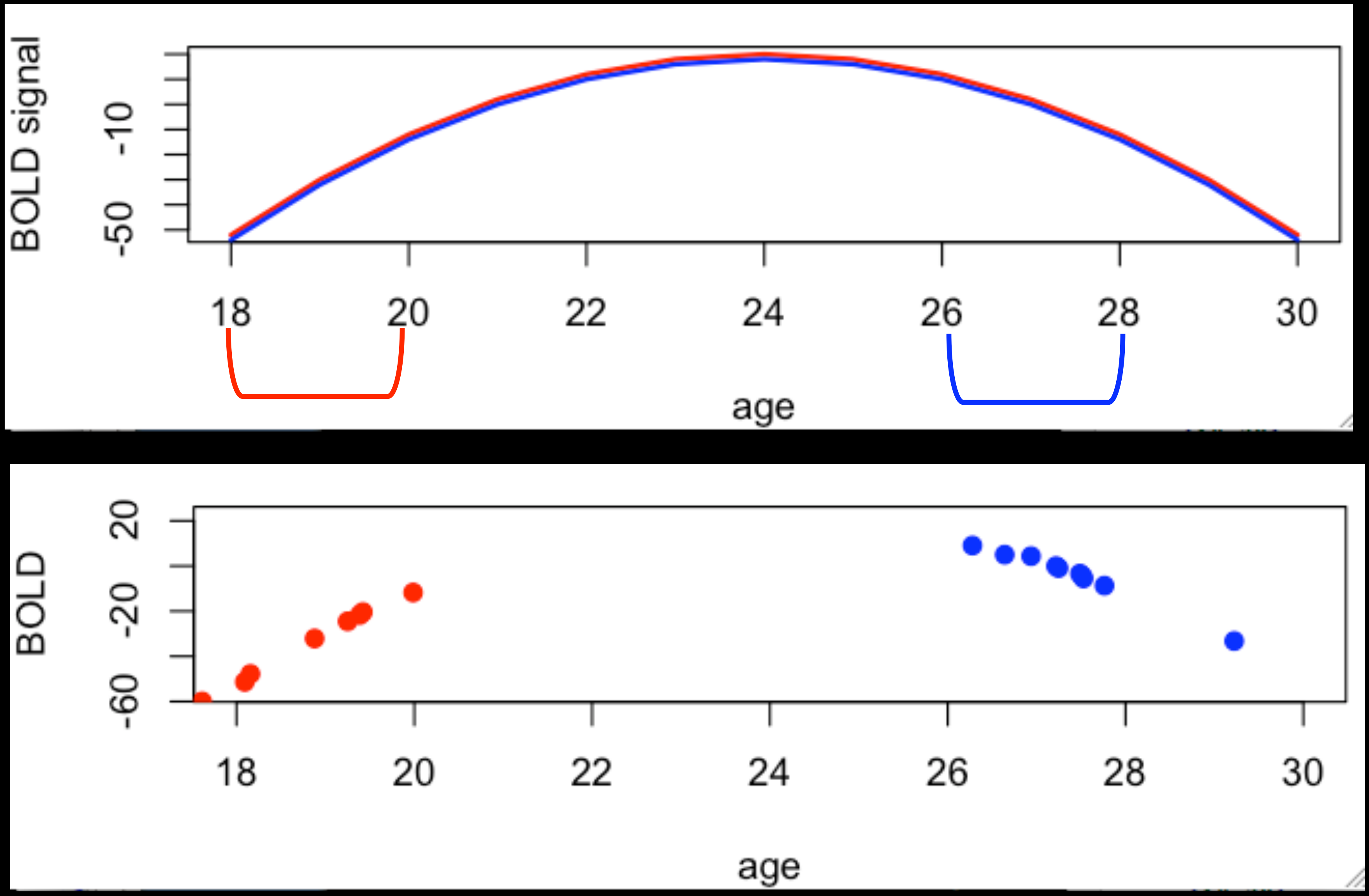
With a nonlinear trend, the trend may be the same in both groups, but sampling different ranges show different trends



With a nonlinear trend, the trend may be the same in both groups, but sampling different ranges show different trends



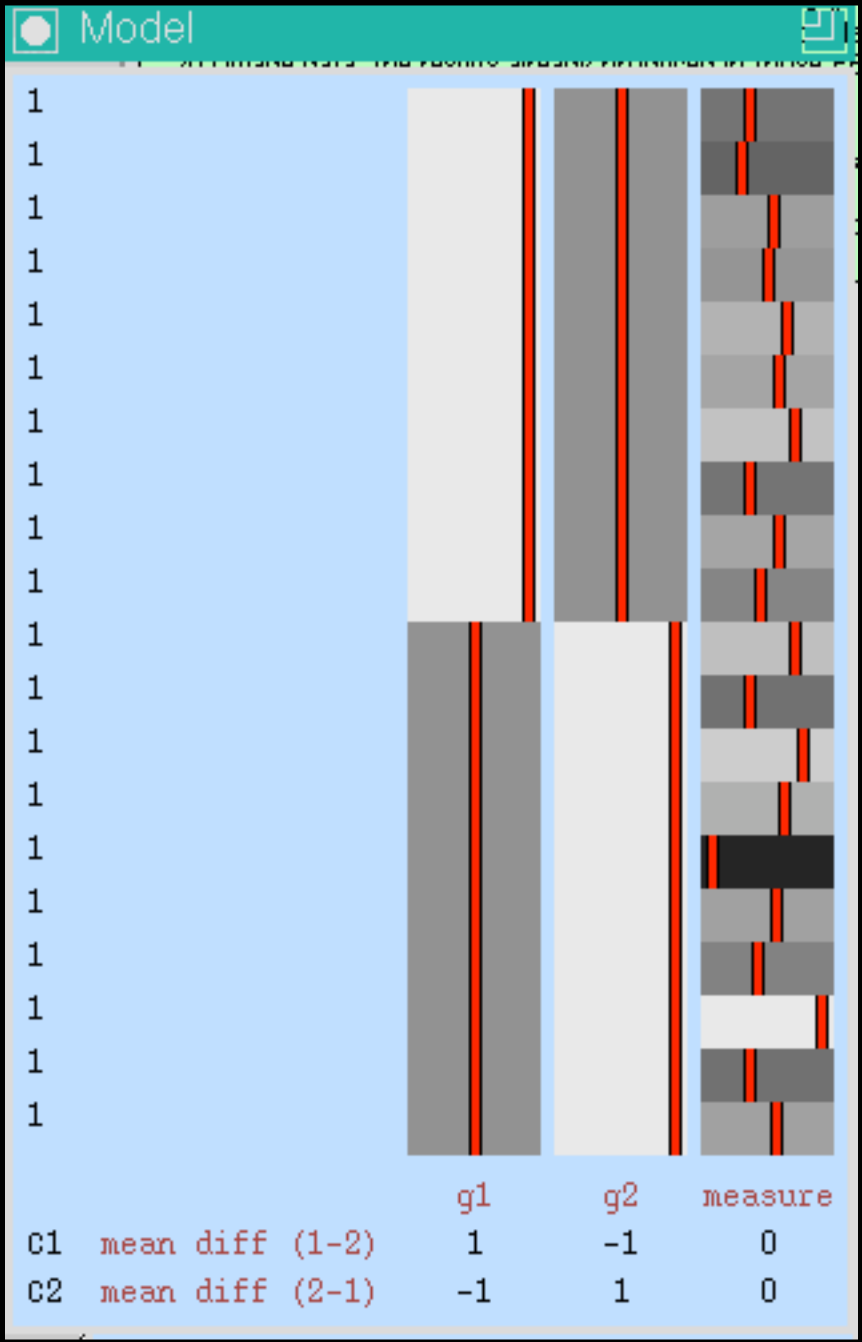
With a nonlinear trend, the trend may be the same in both groups, but sampling different ranges show different trends



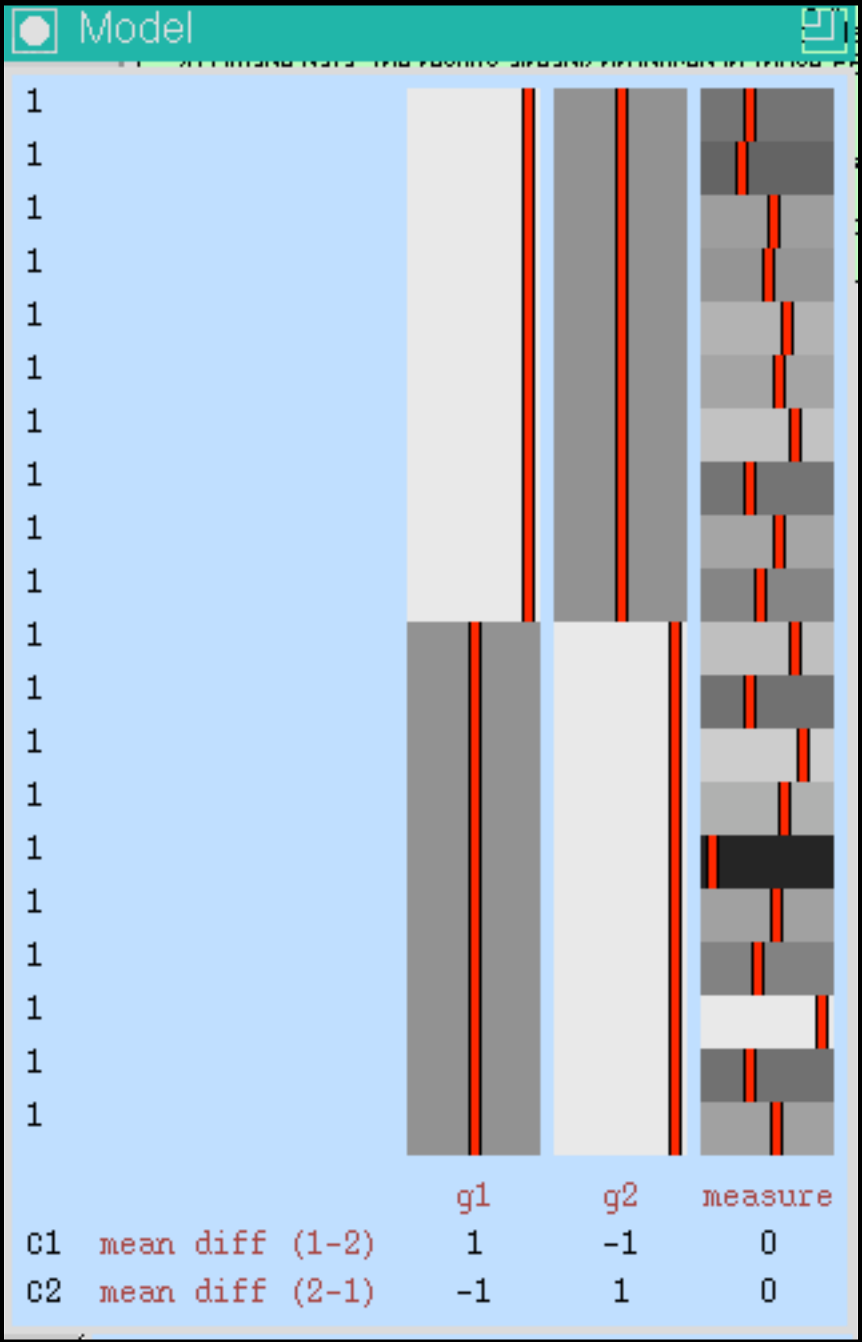
Model 4

- Similar to Model 3 we have two groups and a confounding covariate (maybe BIS or age). 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 <u>between group</u> differences in BIS, what would that model look like?
  - What are some restrictions of this model?

- Do not demean the confounding measure *especially* within group
  - This removes any confounding effect the measure might have

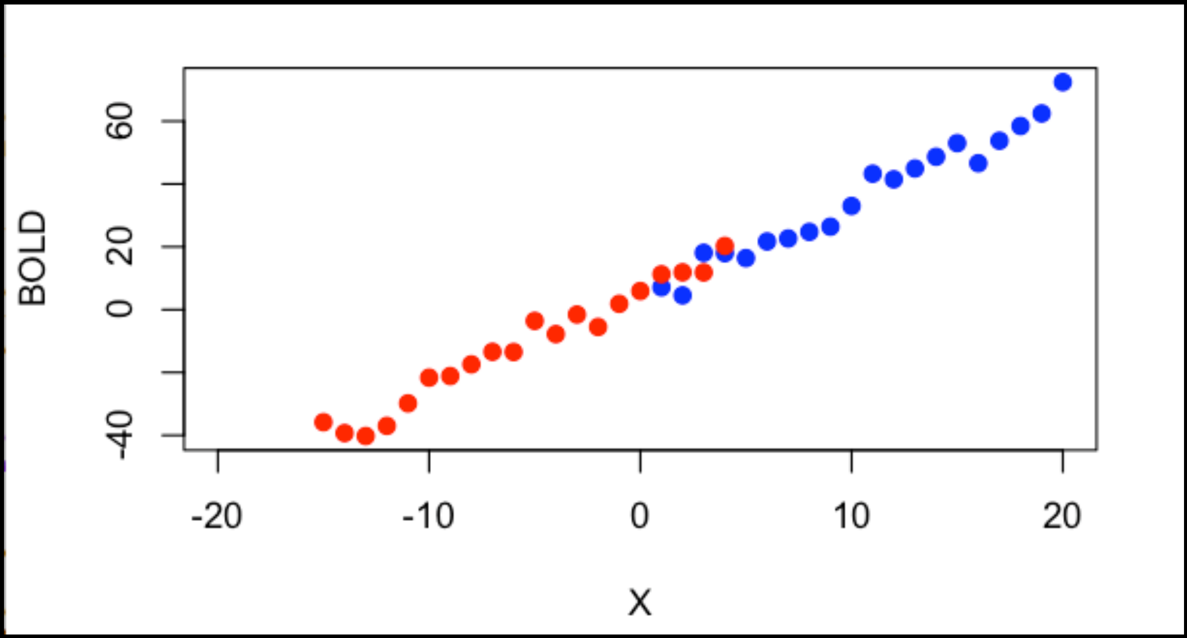


- Do not demean the confounding measure *especially* within group
  - This removes any confounding effect the measure might have
- Restriction: Cannot create a separable model, so cannot have different group variances.
  - Separable model only adjusts for the measure within group and we're interested in between group differences here.



Why you shouldn't demean in this case

- What if this is what your data look like?
  - Difference in means is clearly due to range of X sampled, not the group membership



Model comparisons

- Group difference
  - Model had column of 1s and group1 indicator

```
Call:
lm(formula = y ~ group)

Residuals:
     Min       1Q   Median       3Q      Max
-31.3533 -15.0814   0.6285  14.1005  36.5253

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -11.232      4.255  -2.640   0.0120 *
group         47.124      6.018   7.831 1.90e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 19.03 on 38 degrees of freedom
Multiple R-squared: 0.6174, Adjusted R-squared: 0.6073
F-statistic: 61.32 on 1 and 38 DF,  p-value: 1.903e-09
```

Model comparisons

- Group difference
  - Model had column of 1s and group1 indicator

```
Call:
lm(formula = y ~ group)

Residuals:
     Min       1Q   Median       3Q      Max
-31.3533 -15.0814   0.6285  14.1005  36.5253

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -11.232      4.255  -2.640   0.0120 *
group         47.124      6.018   7.831 1.90e-09 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 19.03 on 38 degrees of freedom
Multiple R-squared: 0.6174, Adjusted R-squared: 0.6073
F-statistic: 61.32 on 1 and 38 DF,  p-value: 1.903e-09
```

Groups are significantly different ($p<0.0001$)

Model comparisons

- Add in confounder (X) correctly

```
lm(formula = y ~ group + x)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.6737 -2.1005 -0.2804  1.8046  6.4766 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.16477    0.94590   6.517 1.26e-07 ***
group       -3.48417    1.89667  -1.837   0.0743 .  
x            3.16302    0.09617  32.892  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Residual standard error: 3.507 on 37 degrees of freedom
Multiple R-squared: 0.9873, Adjusted R-squared: 0.9867 
F-statistic:  1444 on 2 and 37 DF,  p-value: < 2.2e-16
```

Model comparisons

- Add in confounder (X) correctly

```
lm(formula = y ~ group + x)

Residuals:
    Min      1Q  Median      3Q     Max
-6.6737 -2.1005 -0.2804  1.8046  6.4766

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.16477    0.94590   6.517 1.26e-07 ***
group       -3.48417    1.89667  -1.837   0.0743 .
x            3.16302    0.09617  32.892  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Residual standard error: 3.507 on 37 degrees of freedom
Multiple R-squared: 0.9873, Adjusted R-squared: 0.9867
F-statistic:  1444 on 2 and 37 DF,  p-value: < 2.2e-16
```

Overall fit has improved (R^2=.61 in last model)

Model comparisons

- Add in confounder (X) correctly

```
lm(formula = y ~ group + x)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.6737 -2.1005 -0.2804  1.8046  6.4766 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.16477    0.94590   6.517 1.26e-07 ***
group       -3.48417    1.89667  -1.837   0.0743 .  
x            3.16302    0.09617  32.892  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Residual standard error: 3.507 on 37 degrees of freedom
Multiple R-squared: 0.9873, Adjusted R-squared: 0.9867 
F-statistic:  1444 on 2 and 37 DF,  p-value: < 2.2e-16
```

Group diff = -3.48417

Not significant

Model comparisons

- Separate age and demean within group

No demeaning

```
lm(formula = y ~ group + x_g1 + x_g2)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.8453 -1.9445 -0.2159  1.9138  6.1802 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   5.9932     1.0971   5.463 3.63e-06 ***
group        -3.6402     1.9808  -1.838   0.0744 .  
x_g1          3.1942     0.1377  23.201  < 2e-16 ***
x_g2          3.1318     0.1377  22.748  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.

Residual standard error: 3.55 on 36 degrees of freedom
Multiple R-squared: 0.9874, Adjusted R-squared: 0.9863 
F-statistic: 939.2 on 3 and 36 DF,  p-value: < 2.2e-16
```

With demeaning

```
lm(formula = y ~ group + x_g1_dm + x_g2_dm)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.8453 -1.9445 -0.2159  1.9138  6.1802 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -11.2319     0.7939  -14.15 2.82e-16 ***
group        47.1242     1.1227   41.97  < 2e-16 ***
x_g1_dm       3.1942     0.1377   23.20  < 2e-16 ***
x_g2_dm       3.1318     0.1377   22.75  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.

Residual standard error: 3.55 on 36 degrees of freedom
Multiple R-squared: 0.9874, Adjusted R-squared: 0.9863 
F-statistic: 939.2 on 3 and 36 DF,  p-value: < 2.2e-16
```

Model comparisons

- Separate age and demean within group

No demeaning

```
lm(formula = y ~ group + x_g1 + x_g2)

Residuals:
    Min      1Q  Median      3Q     Max
-6.8453 -1.9445 -0.2159  1.9138  6.1802

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.9932     1.0971   5.463 3.63e-06 ***
group        -3.6402     1.9808  -1.838   0.0744 .
x_g1          3.1942     0.1377  23.201  < 2e-16 ***
x_g2          3.1318     0.1377  22.748  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.

Residual standard error: 3.55 on 36 degrees of freedom
Multiple R-squared: 0.9874, Adjusted R-squared: 0.9863
F-statistic: 939.2 on 3 and 36 DF,  p-value: < 2.2e-16
```

With demeaning

```
lm(formula = y ~ group + x_g1_dm + x_g2_dm)

Residuals:
    Min      1Q  Median      3Q     Max
-6.8453 -1.9445 -0.2159  1.9138  6.1802

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -11.2310     0.7939  -14.15 2.82e-16 ***
group        47.1242     1.1227   41.97  < 2e-16 ***
x_g1_dm       3.1942     0.1377   23.20  < 2e-16 ***
x_g2_dm       3.1318     0.1377   22.75  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.:

Residual standard error: 3.55 on 36 degrees of freedom
Multiple R-squared: 0.9874, Adjusted R-squared: 0.9863
F-statistic: 939.2 on 3 and 36 DF,  p-value: < 2.2e-16
```

Group difference is quite large after demeaning

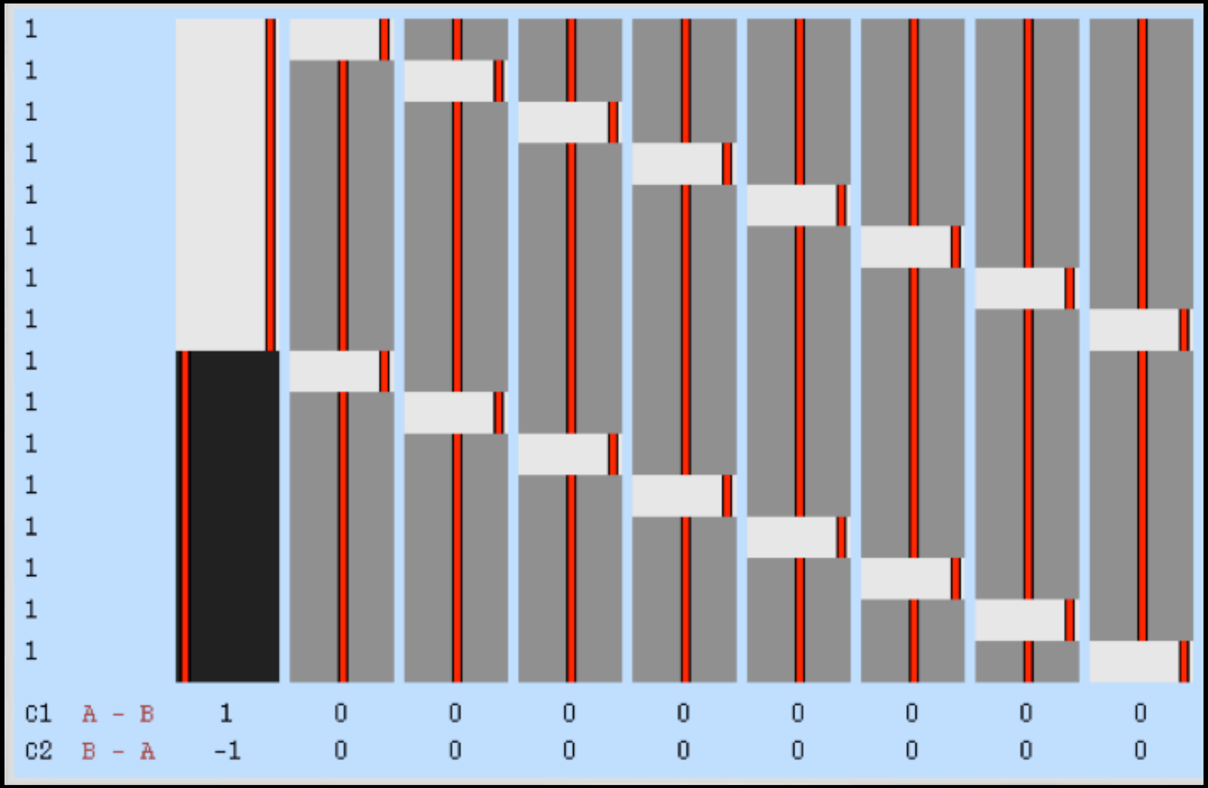
Differences between 3 & 4

- In model 3 we were interested in testing the differences in trends between two groups.
  - age modeled to show within group trends, which are then compared across groups
- In model 4 we were only interested in comparing the means between two groups and whether it was due to a confounder.
  - X modeled to show between group differences

Model 4

- You have estimated successful stop pre and post training (2 runs) for each subject. At the group level you have 8 subjects with 2 measures each, what is the appropriate model?

Paired t test



Model 5

- You have 9 subjects and all but one subject have complete pairs of data (one subject missed the second scan)
  - Can you still use a paired t test?
  - Other options?

Model 5

- Suggestions
  - If there is high within subject correlation (like in the tire example the other day) you need a paired test
    - Toss the subject with incomplete data
  - If there isn't a high correlation, a two-sample t-test may be okay
    - Be careful or else you'll be losing power
    - Use complete data to compare paired t test to 2 sample t test

Model 6

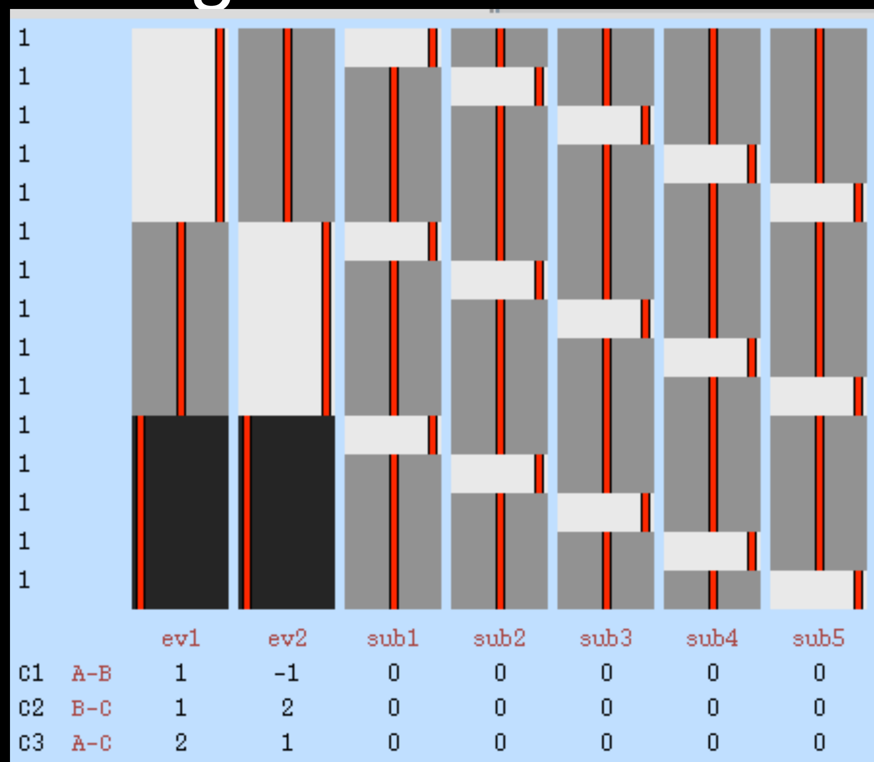
- You have 5 subjects scanned under 3 conditions (A, B & C) and you want to make all pairwise comparisons in one model
  - You must account for repeated measures
  - Construct 3 contrasts: A-B, B-C, A-C

Triple paired t test

	ev1	ev2	sub1	sub2	sub3	sub4	sub5
A1	1	0	1	0	0	0	0
A2	1	0	0	1.0	0	0	0
A3	1	0	0	0	1.0	0	0
A4	1	0	0	0	0	1.0	0
A5	1	0	0	0	0	0	1.0
B1	0.0	1.0	1	0	0	0	0
B2	0.0	1.0	0	1.0	0	0	0
B3	0.0	1.0	0	0	1.0	0	0
B4	0.0	1.0	0	0	0	1.0	0
B5	0.0	1.0	0	0	0	0	1.0
C1	-1.0	-1.0	1	0	0	0	0
C2	-1.0	-1.0	0	1.0	0	0	0
C3	-1.0	-1.0	0	0	1.0	0	0
C4	-1.0	-1.0	0	0	0	1.0	0
C5	-1.0	-1.0	0	0	0	0	1.0

Triple paired t test

- Really just a repeated measures ANOVA using factor effects



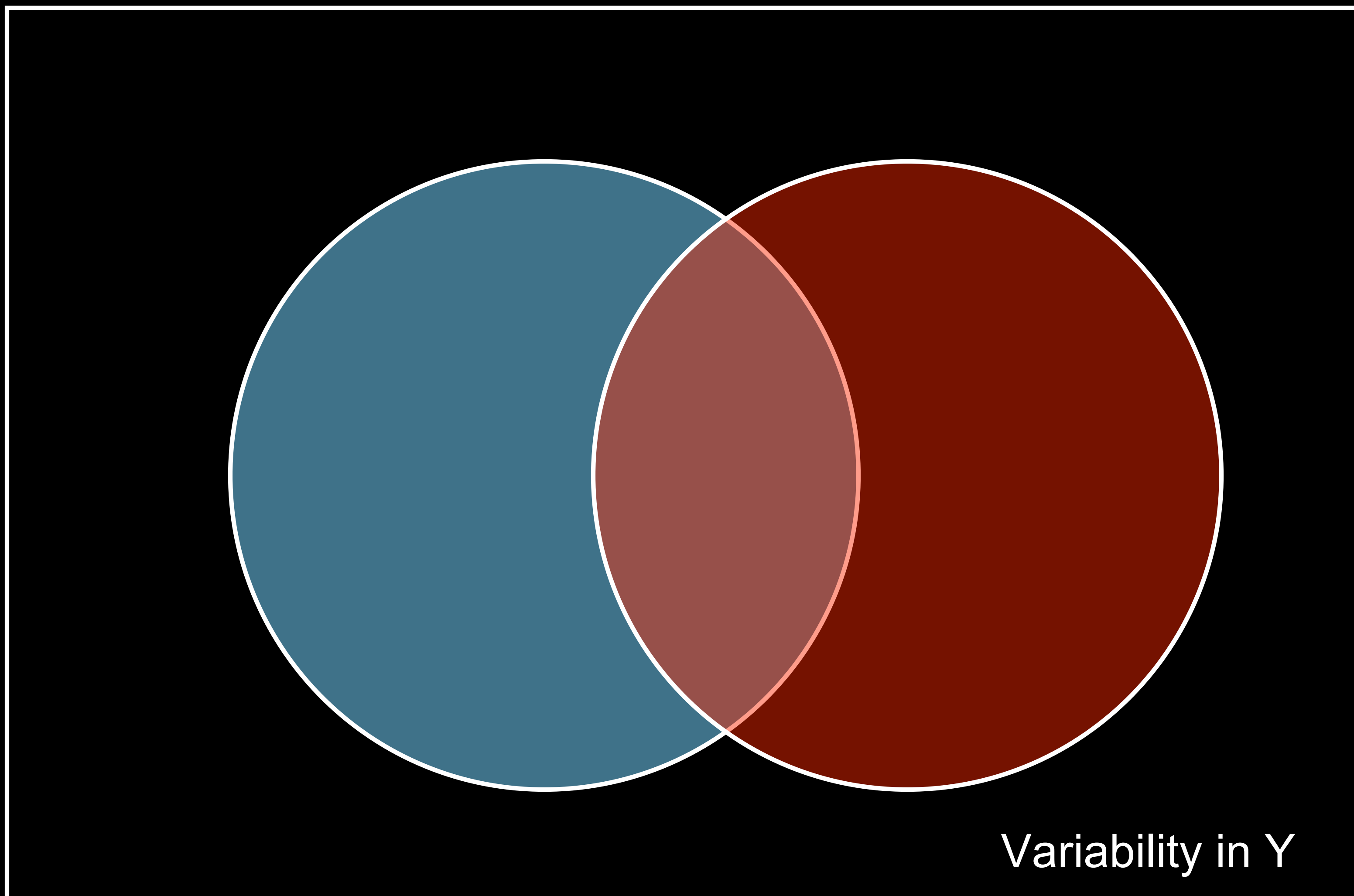
Note on repeated measures ANOVA

- As long as each subject has repeated measures for all cells of the ANOVA use factor effects and split up the intercept into subject-specific means
- Won’t work if not repeated across both factors
  - Eg. 2x2 ANOVA
    - Factor 1: Pre/post training (everybody has both)
    - Factor 2: control/unhealthy control (each subject only in one group or the other)

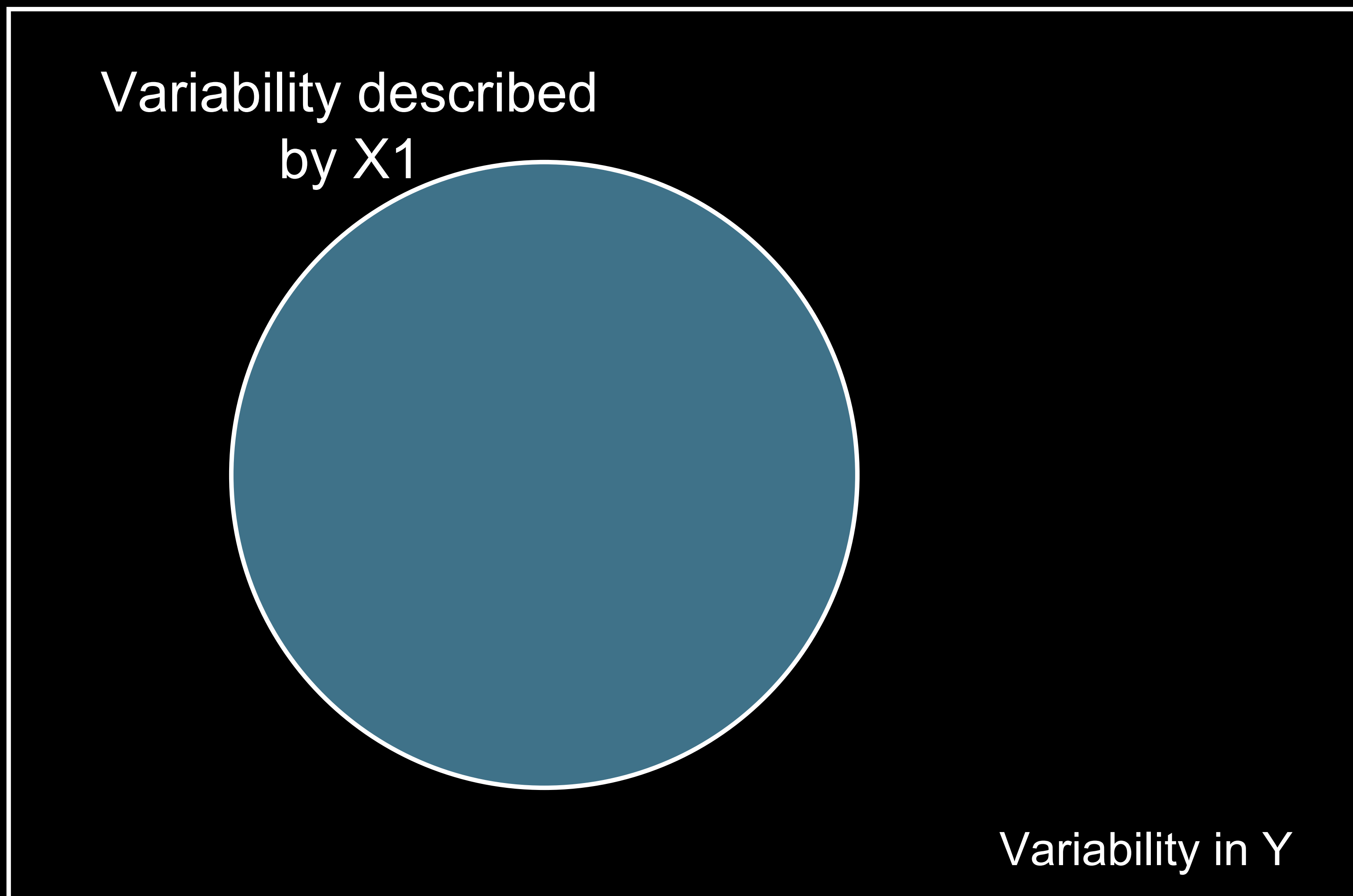
Orthogonalization…what is it?

- Demeaning is the simplest form of orthogonalization
  - Orthogonalizes your regressor to the column of 1s (earlier example)
- Generally A orth wrt B
  - What is left over in A after regressing out B

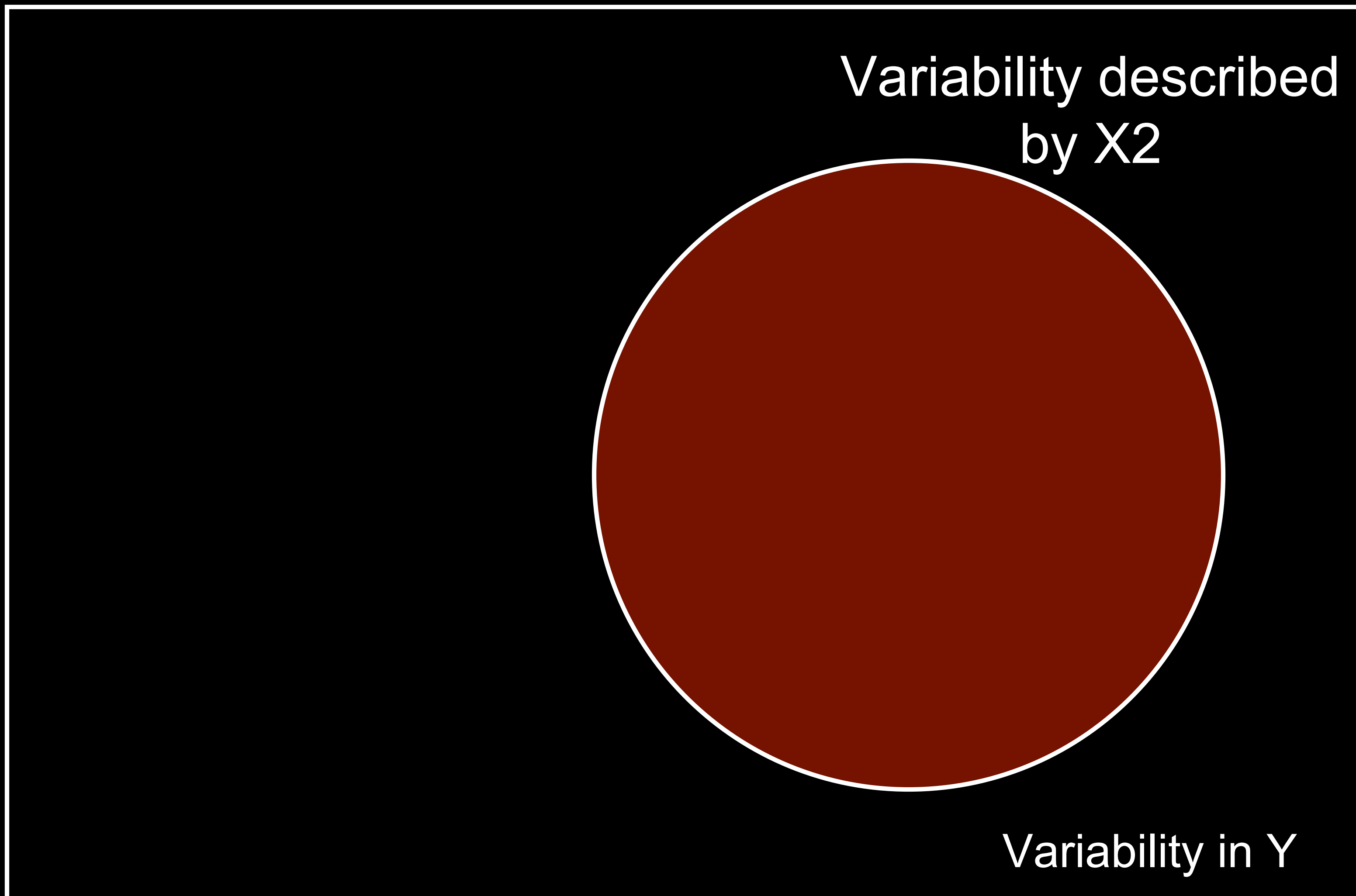
How the GLM works



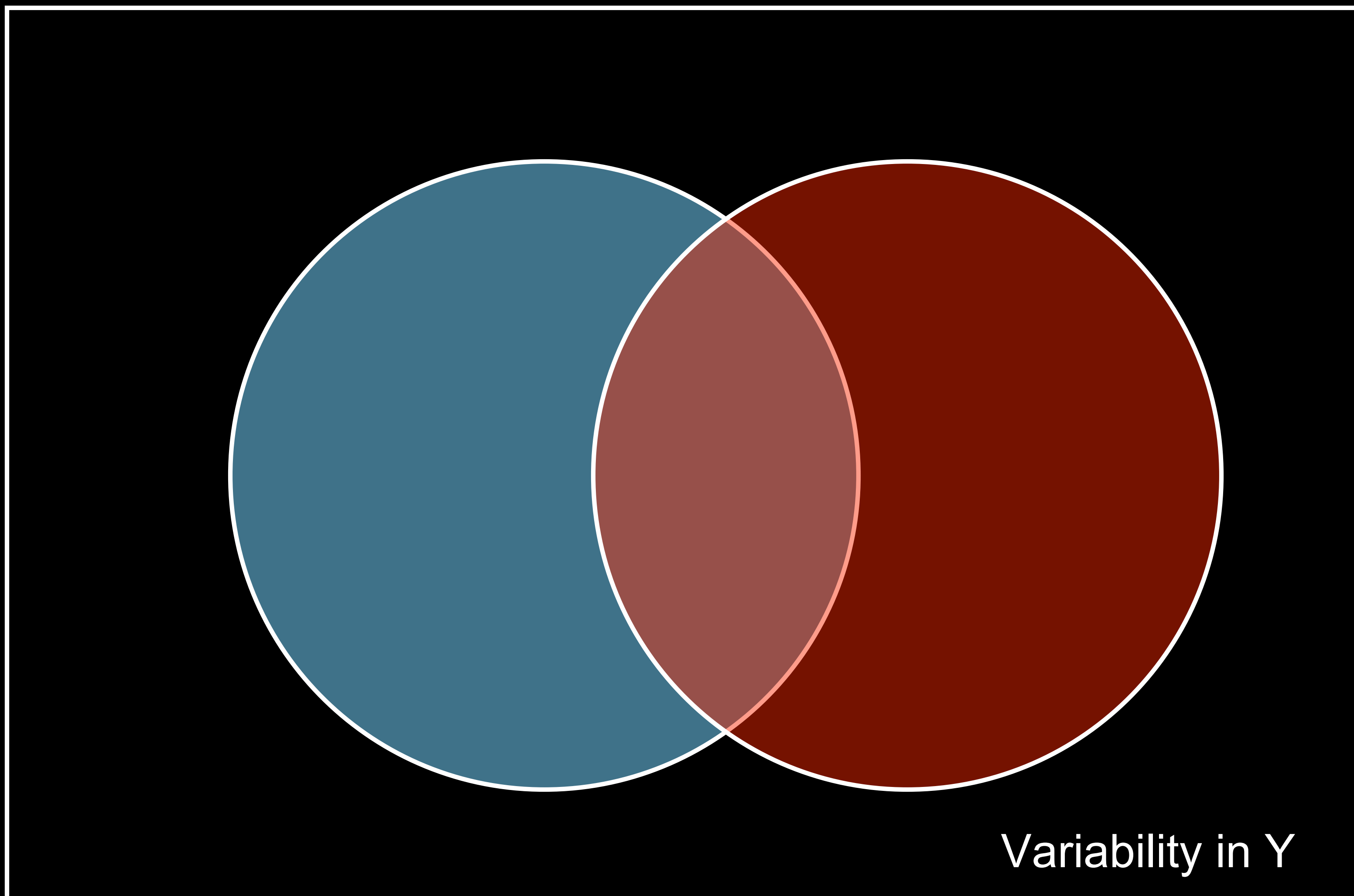
How the GLM works



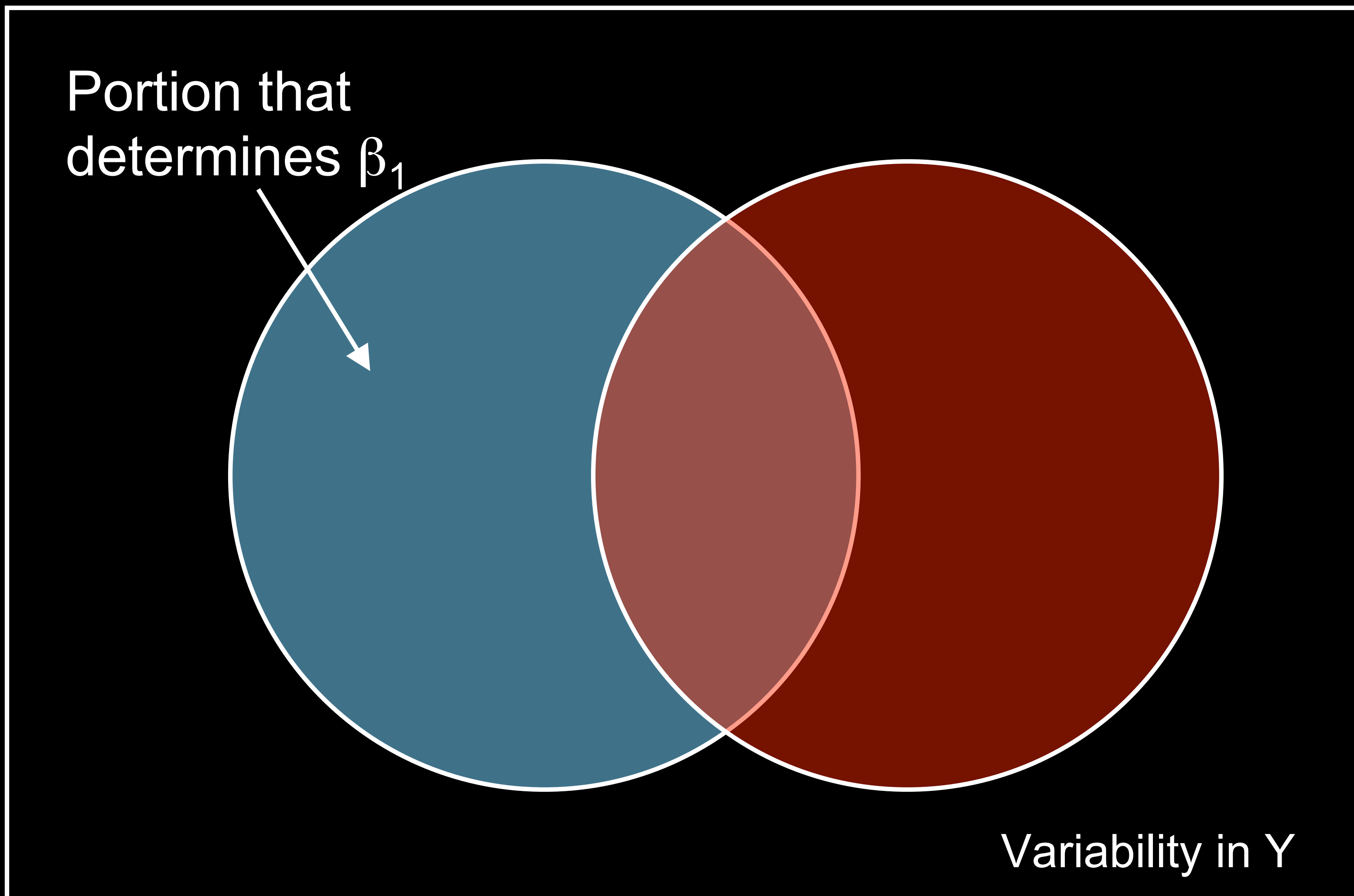
How the GLM works



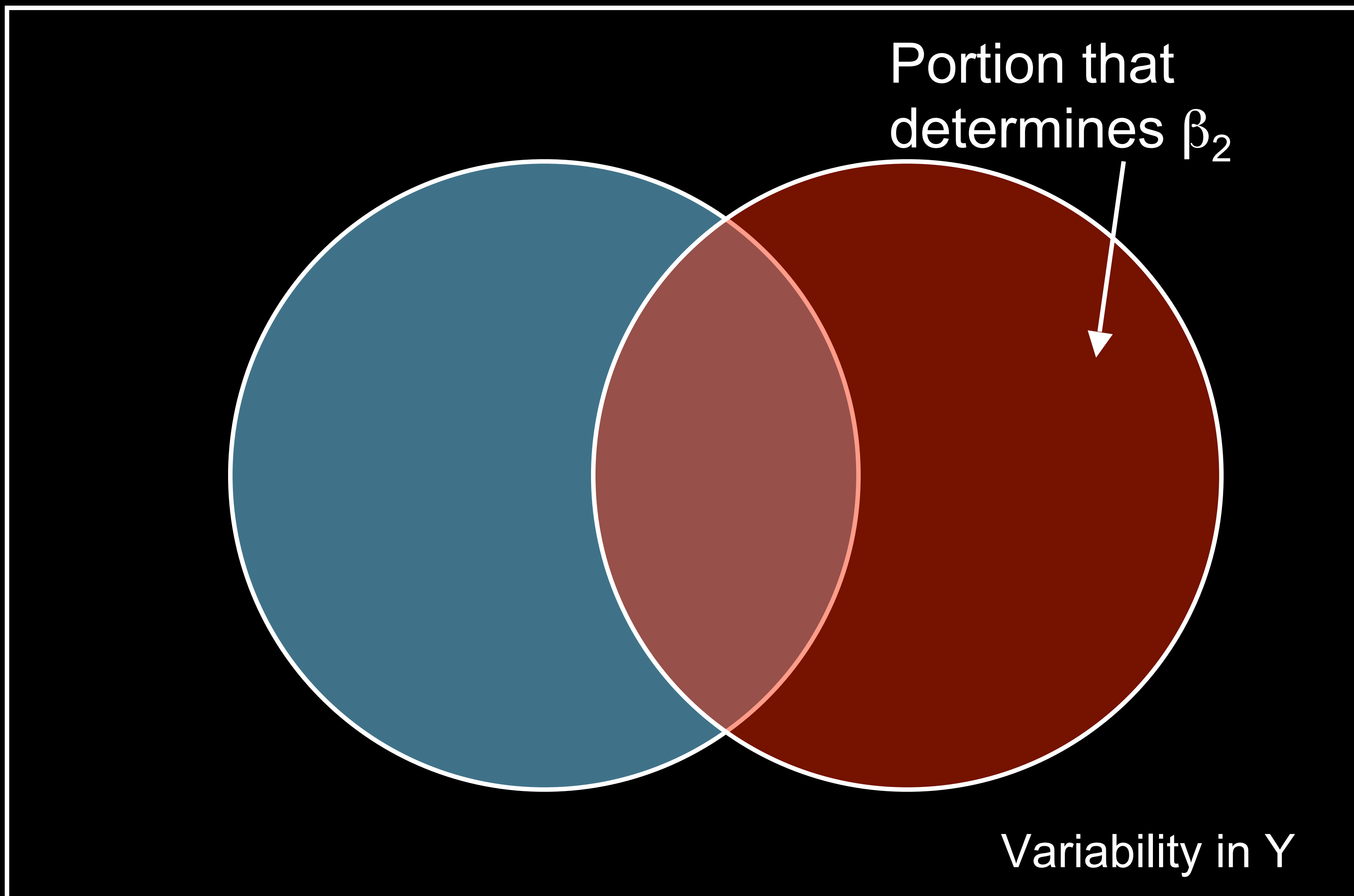
How the GLM works



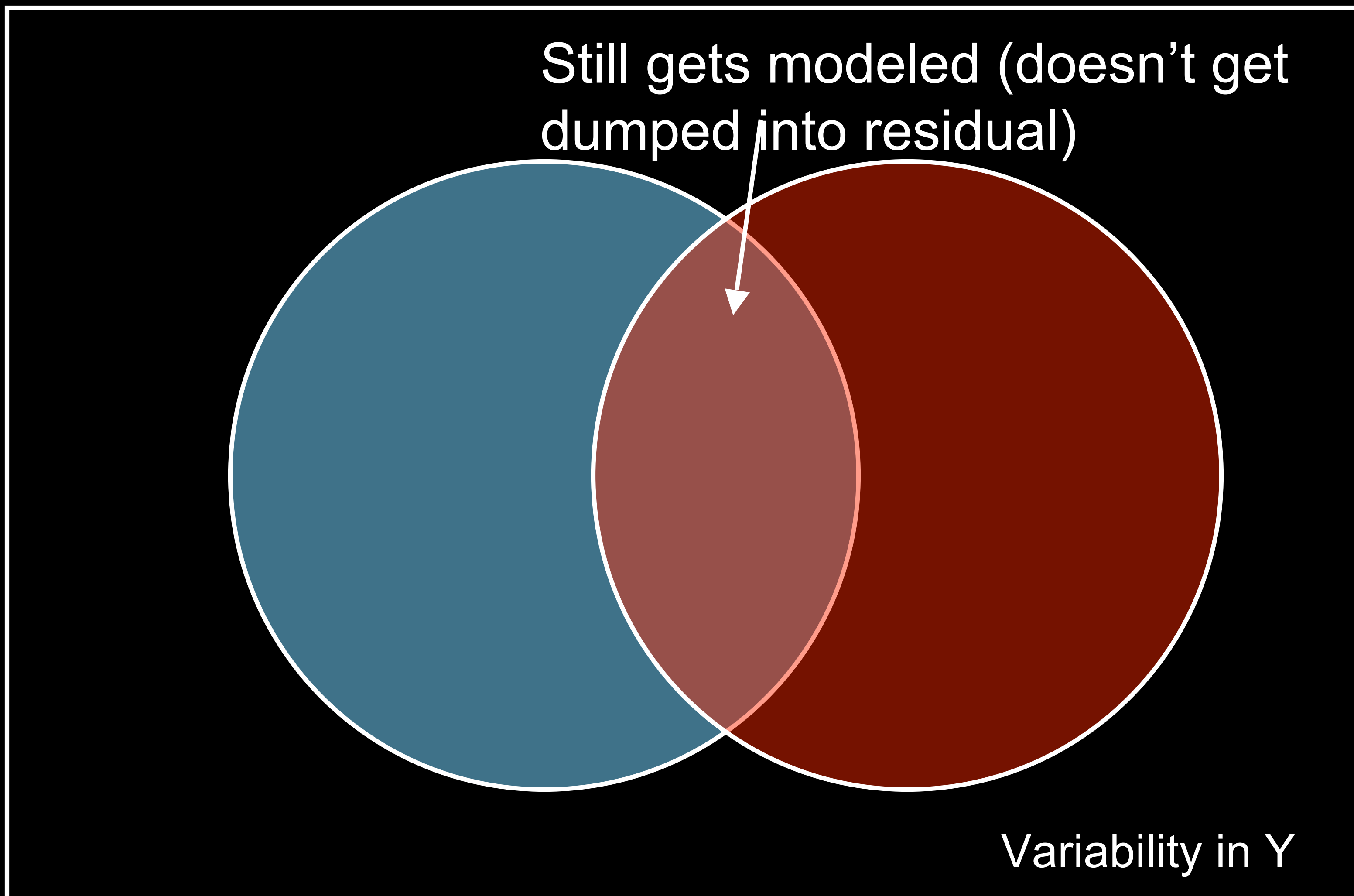
How the GLM works



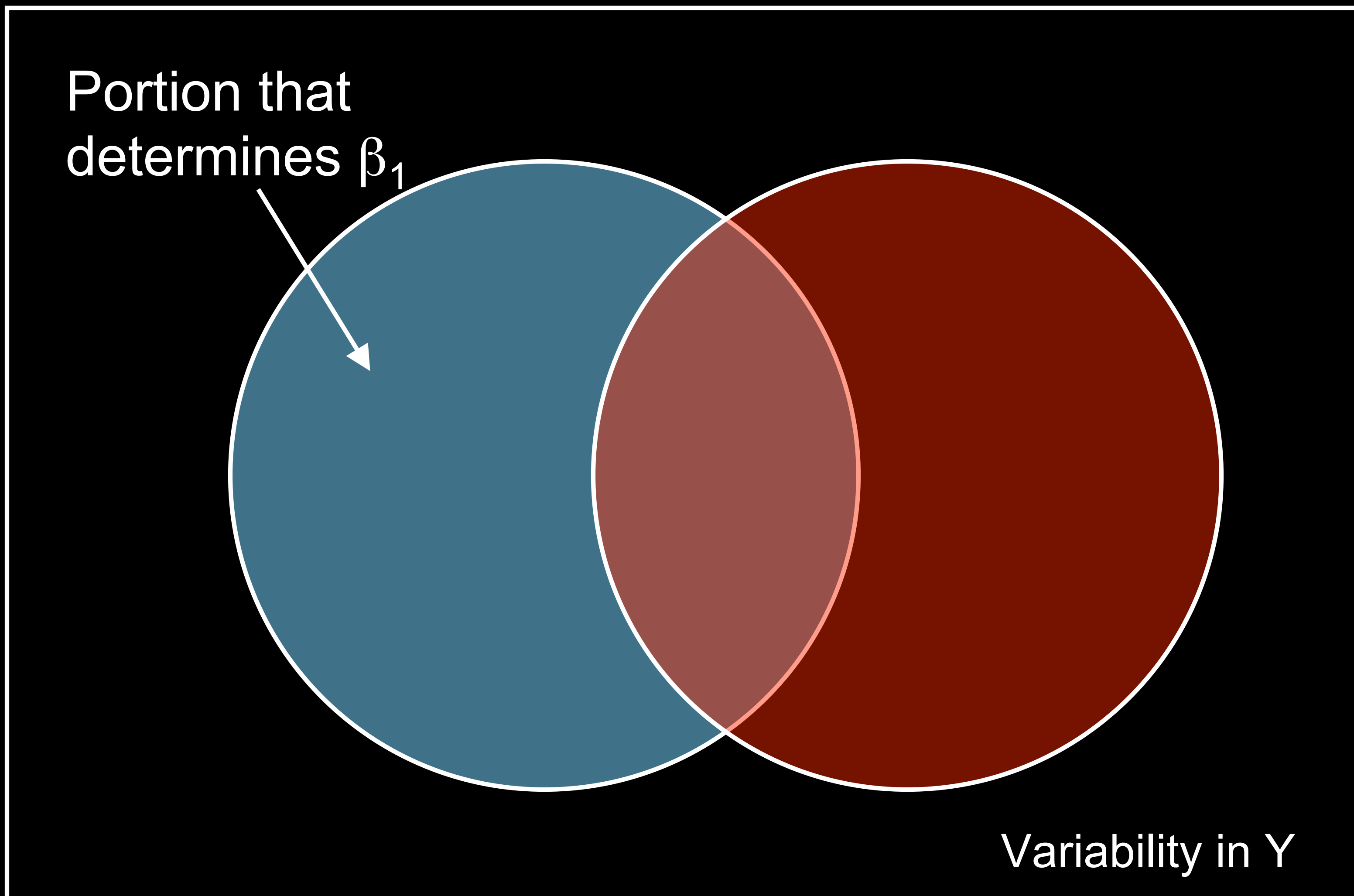
How the GLM works



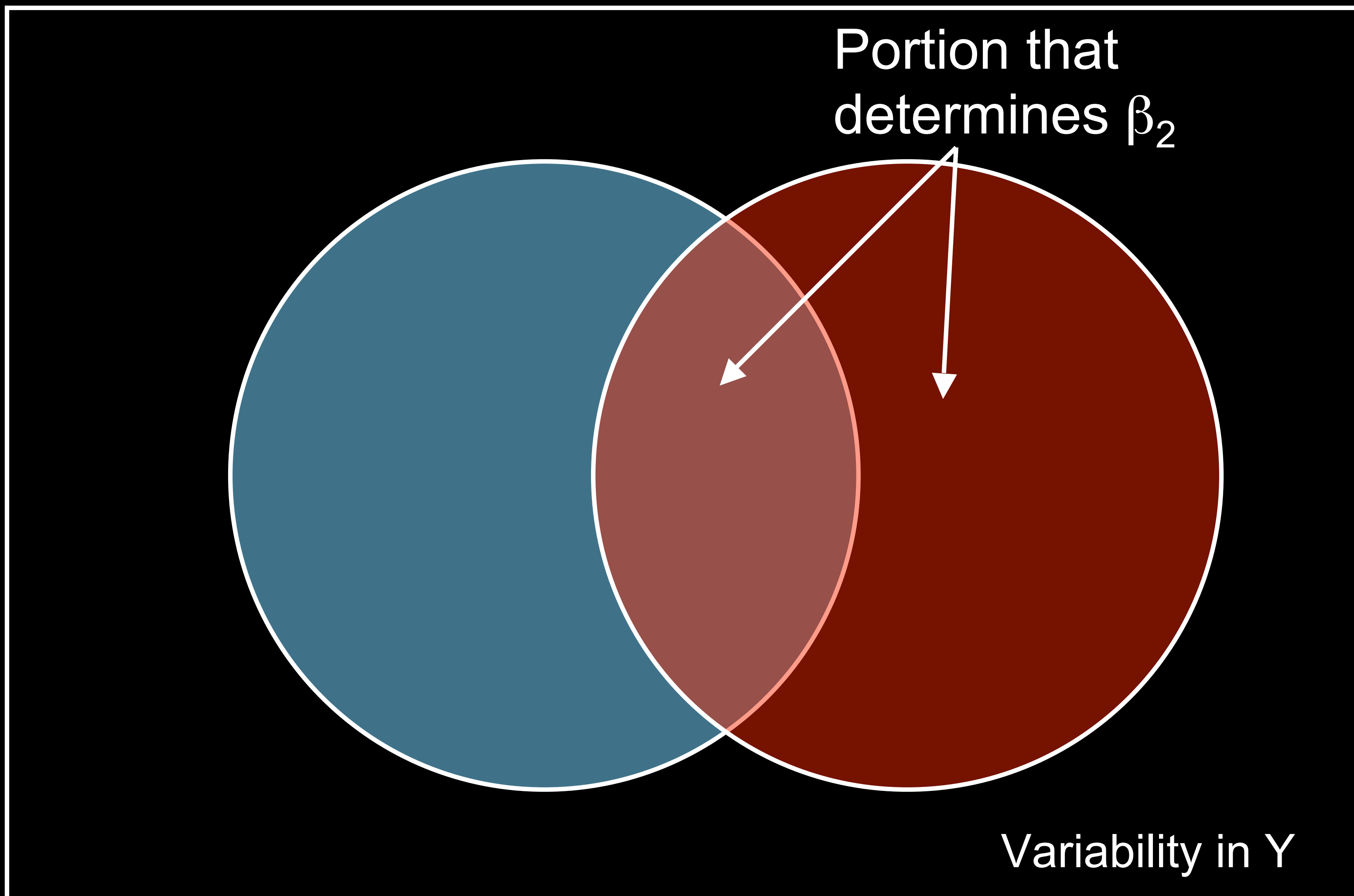
How the GLM works



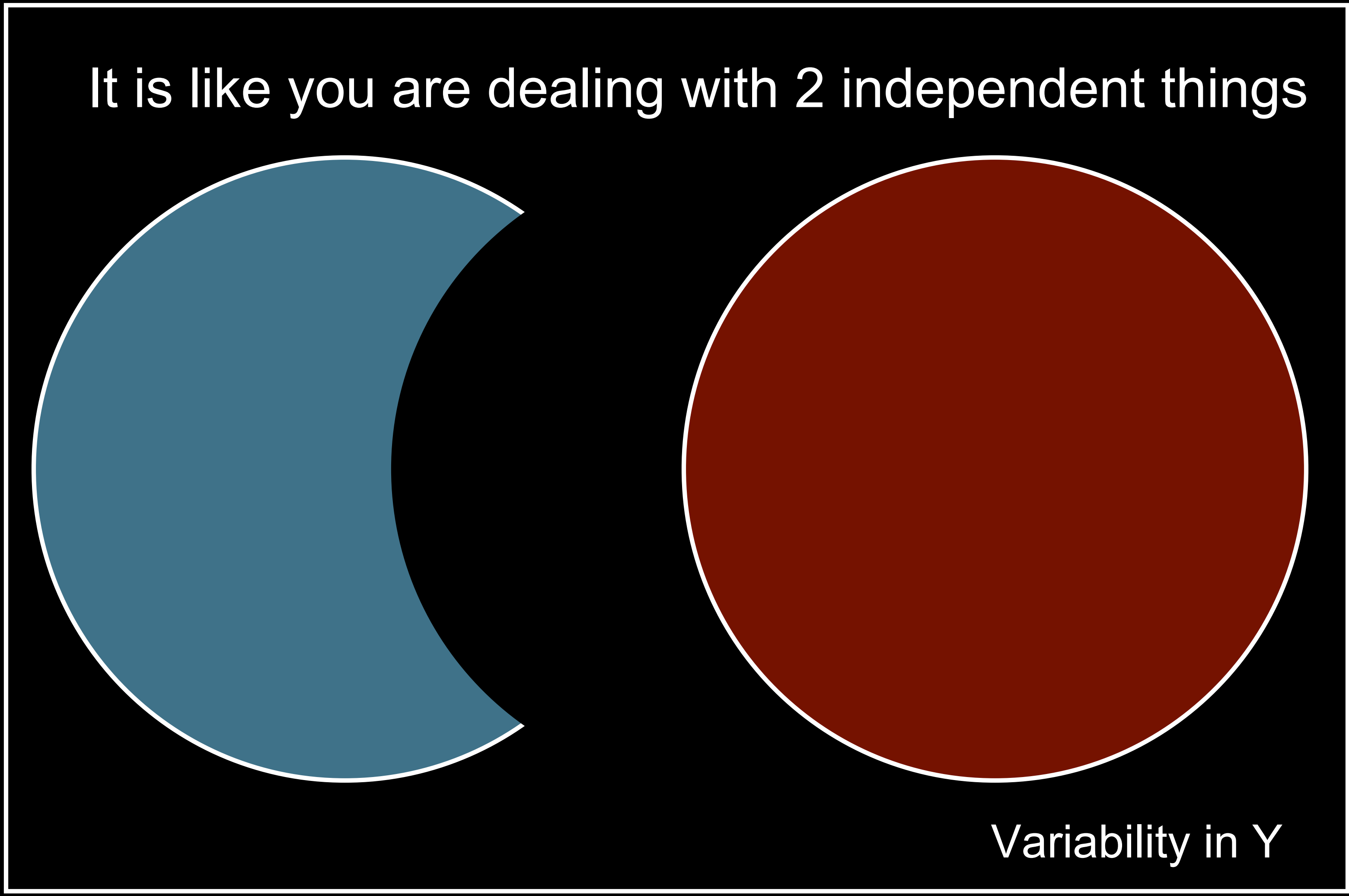
Orthogonalize X1 wrt X2



Orthogonalize X1 wrt X2



Orthogonalize X1 wrt X2



You can visualize regression using vectors

$$Y = X_1\beta_1 + X_2\beta_2$$

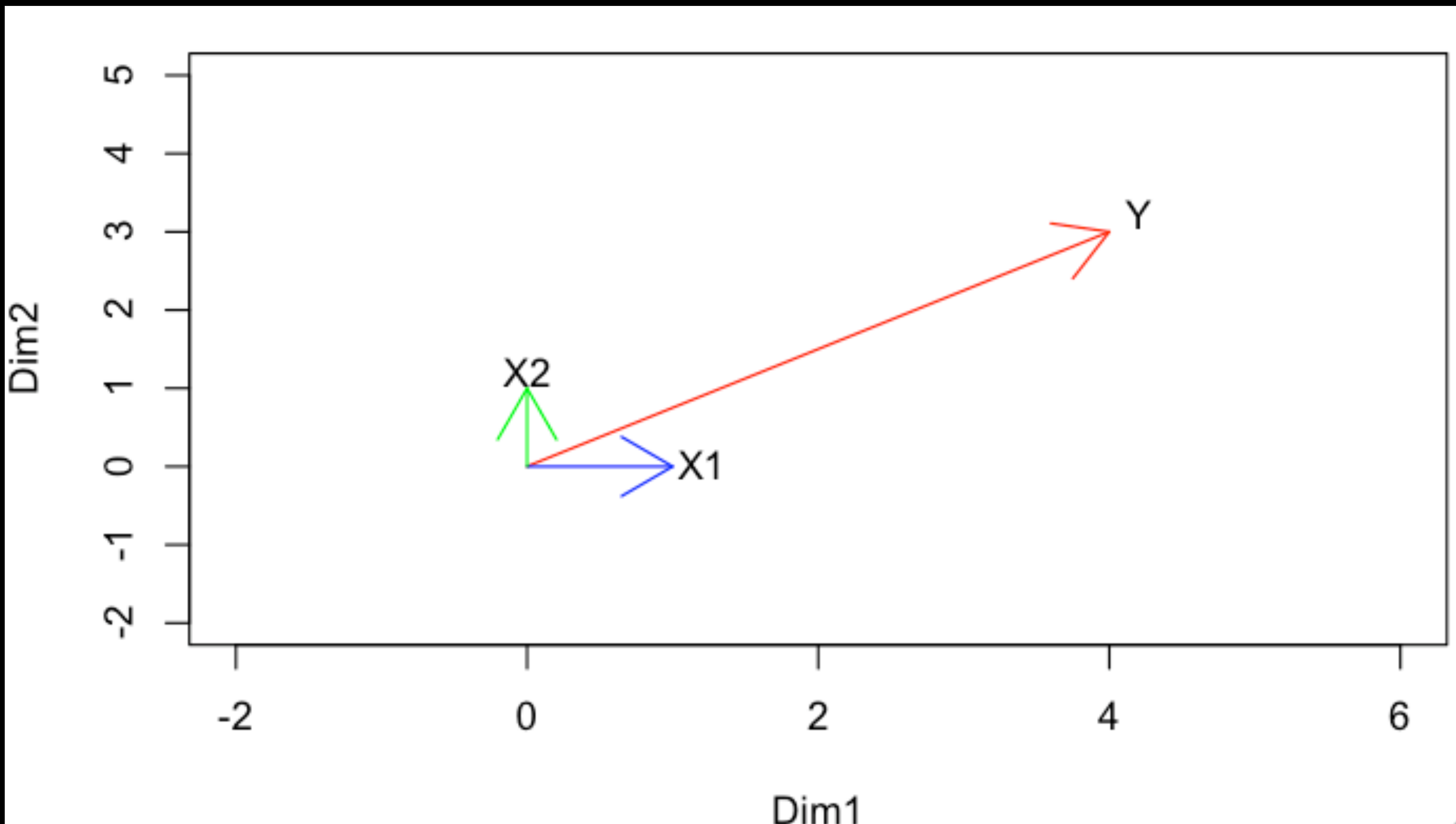
$$\begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \beta_1 + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \beta_2$$

You can visualize regression using vectors

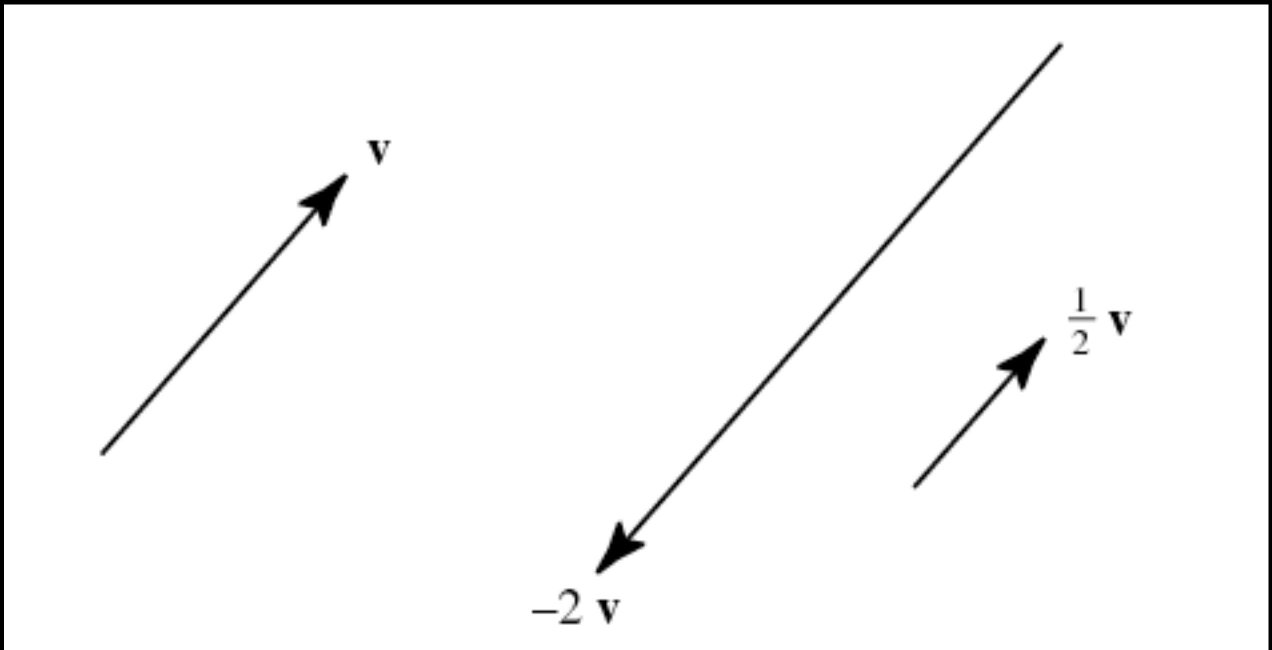
$$Y = X_1\beta_1 + X_2\beta_2$$

$$\begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \beta_1 + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \beta_2$$

β_1=3 and β_2=4

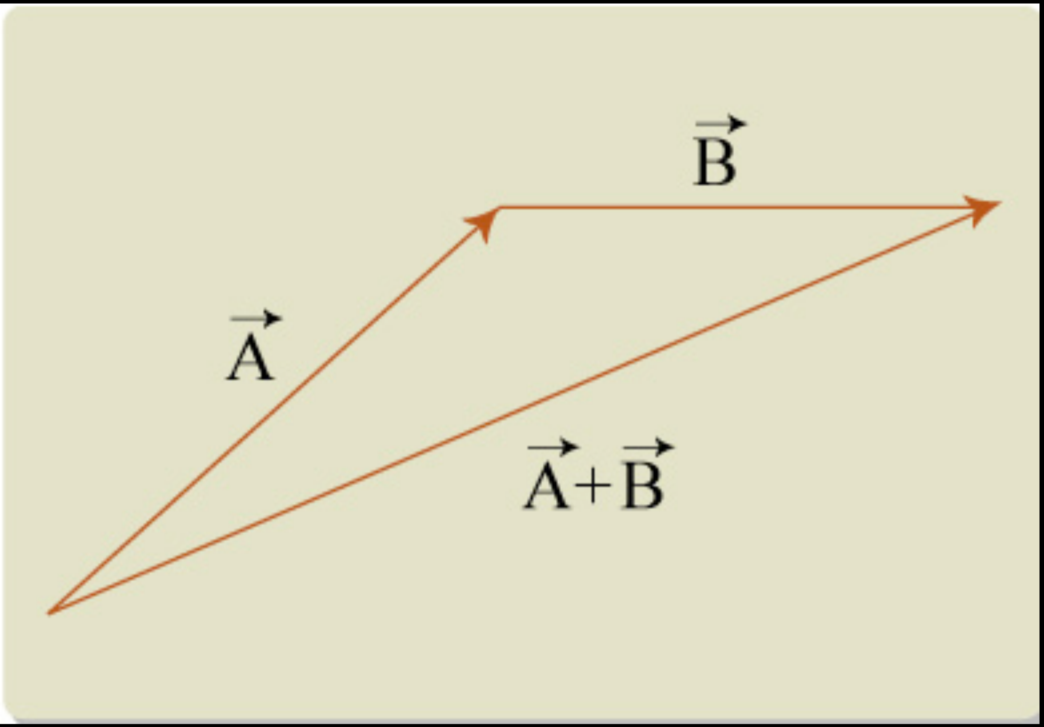

Y
X2
X1
Dim2
Dim1
-2
-1
0
1
2
3
4
5
-2
0
2
4
6

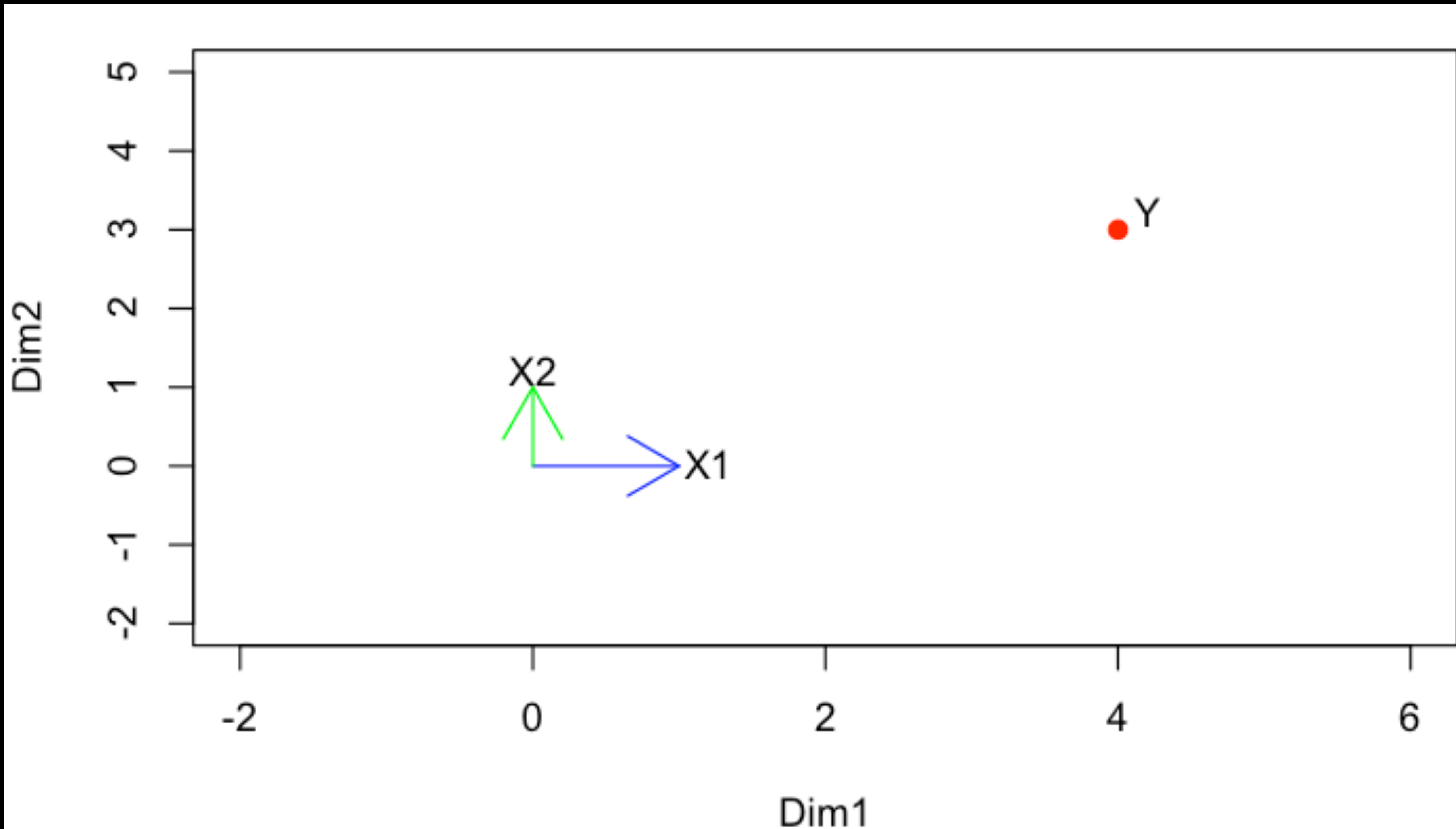
Scalar multiplication of a vector



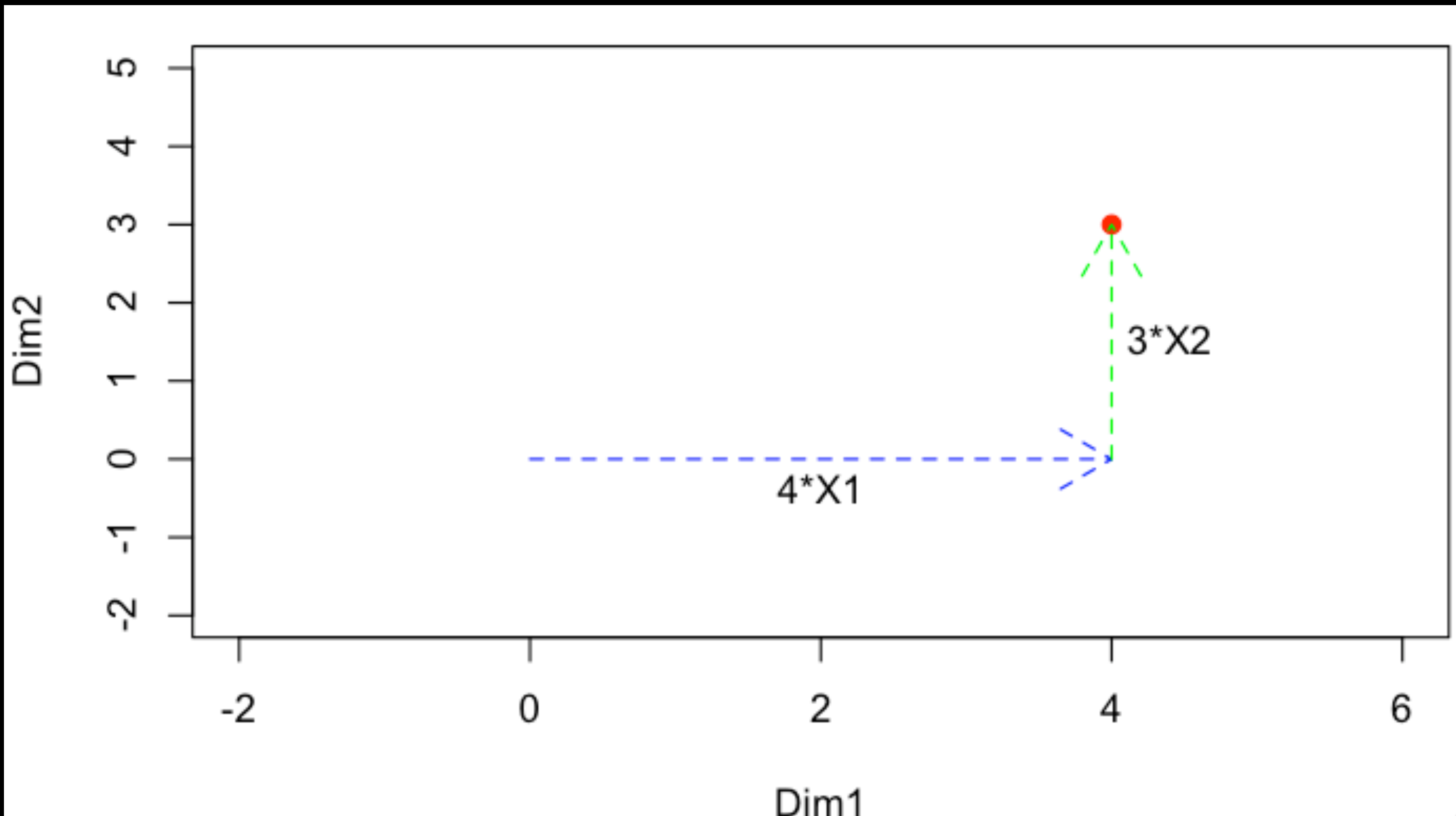
Vector addition

To add, put tail of one vector at head of another vector





Y
X2
X1
-2
-1
0
1
2
3
4
5
-2
0
2
4
6
Dim1
Dim2



3*X2
4*X1
Dim2
Dim1
-2
0
2
4
6
-2
-1
0
1
2
3
4
5

Non-orthogonal case

$$Y = X_1\beta_1 + X_2\beta_2$$

$$\begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} \beta_1 + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \beta_2$$

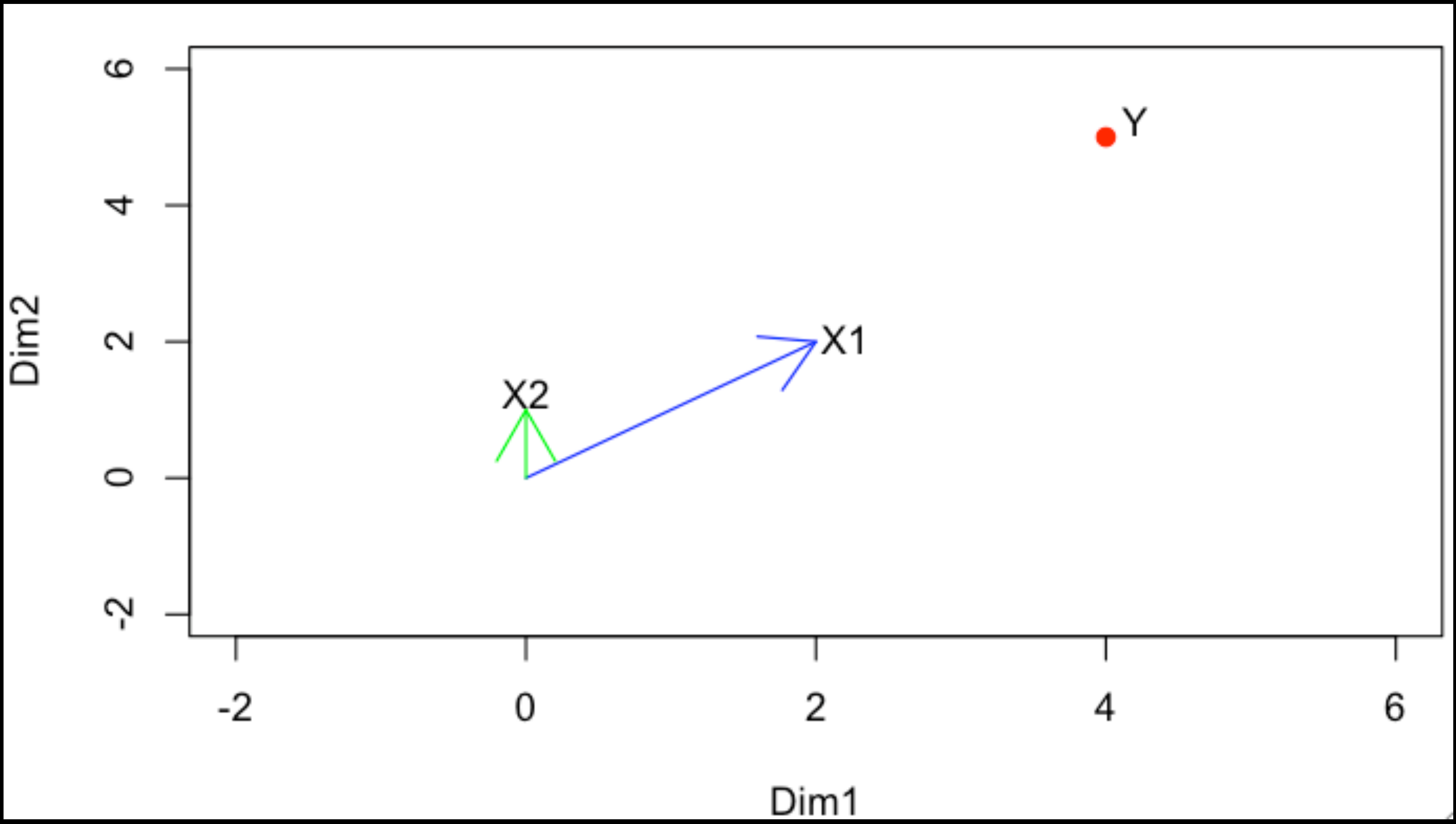
Non-orthogonal case

$$Y = X_1\beta_1 + X_2\beta_2$$

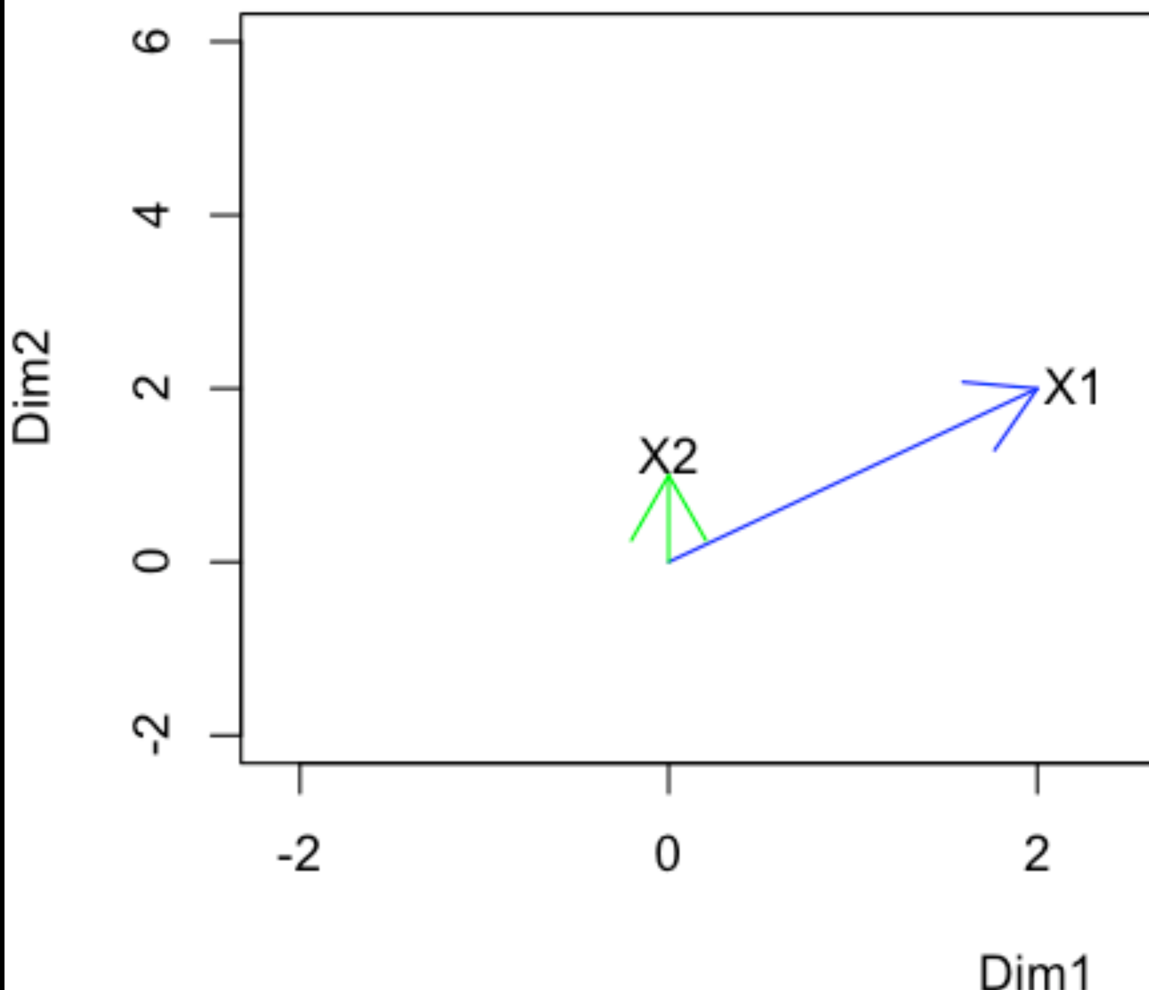
$$\begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} \beta_1 + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \beta_2$$

β_1=2 and β_2=1

Non-orthogonal case



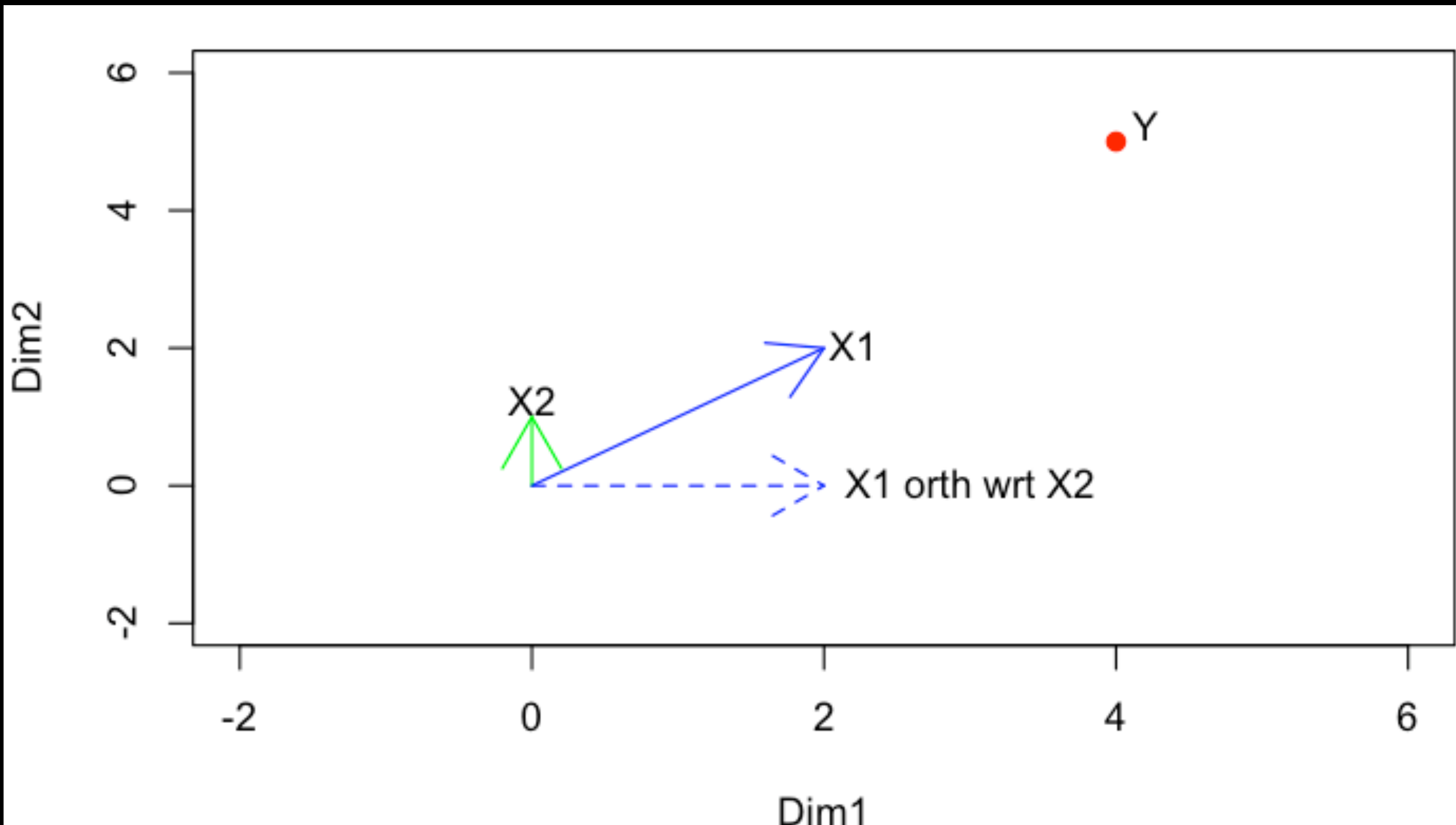
Non-orthogonal case



Only X1 describes the first dimension (Dim1)

X1 also describes the X2 direction

Non-orthogonal case



Non-orthogonal case

$Y = X_1\beta_1 + X_2\beta_2$

$$\begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} \beta_1 + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \beta_2$$

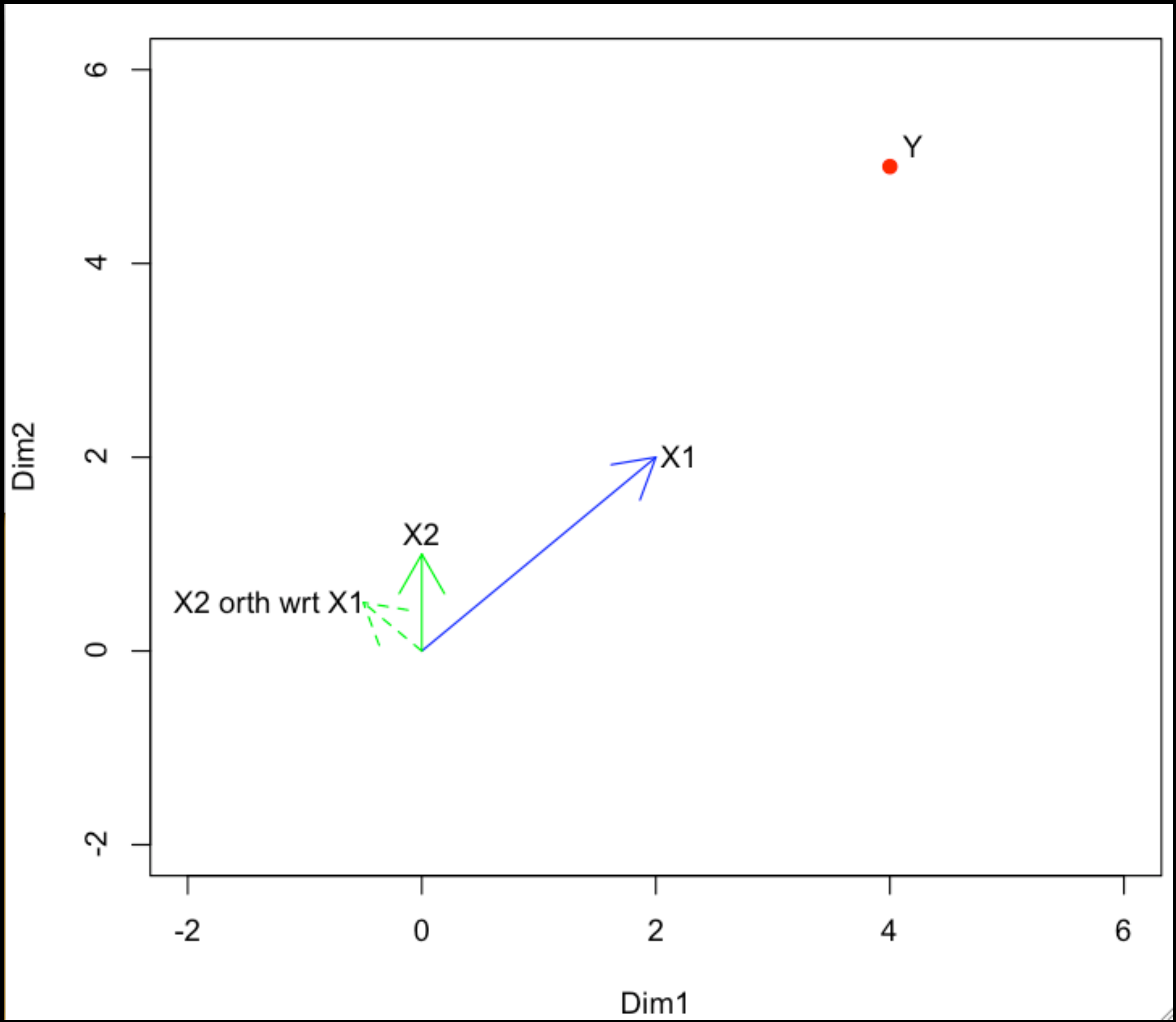
$\beta_1=2$ and $\beta_2=1$

$Y = X\_orth_1\beta_1 + X_2\beta_2$

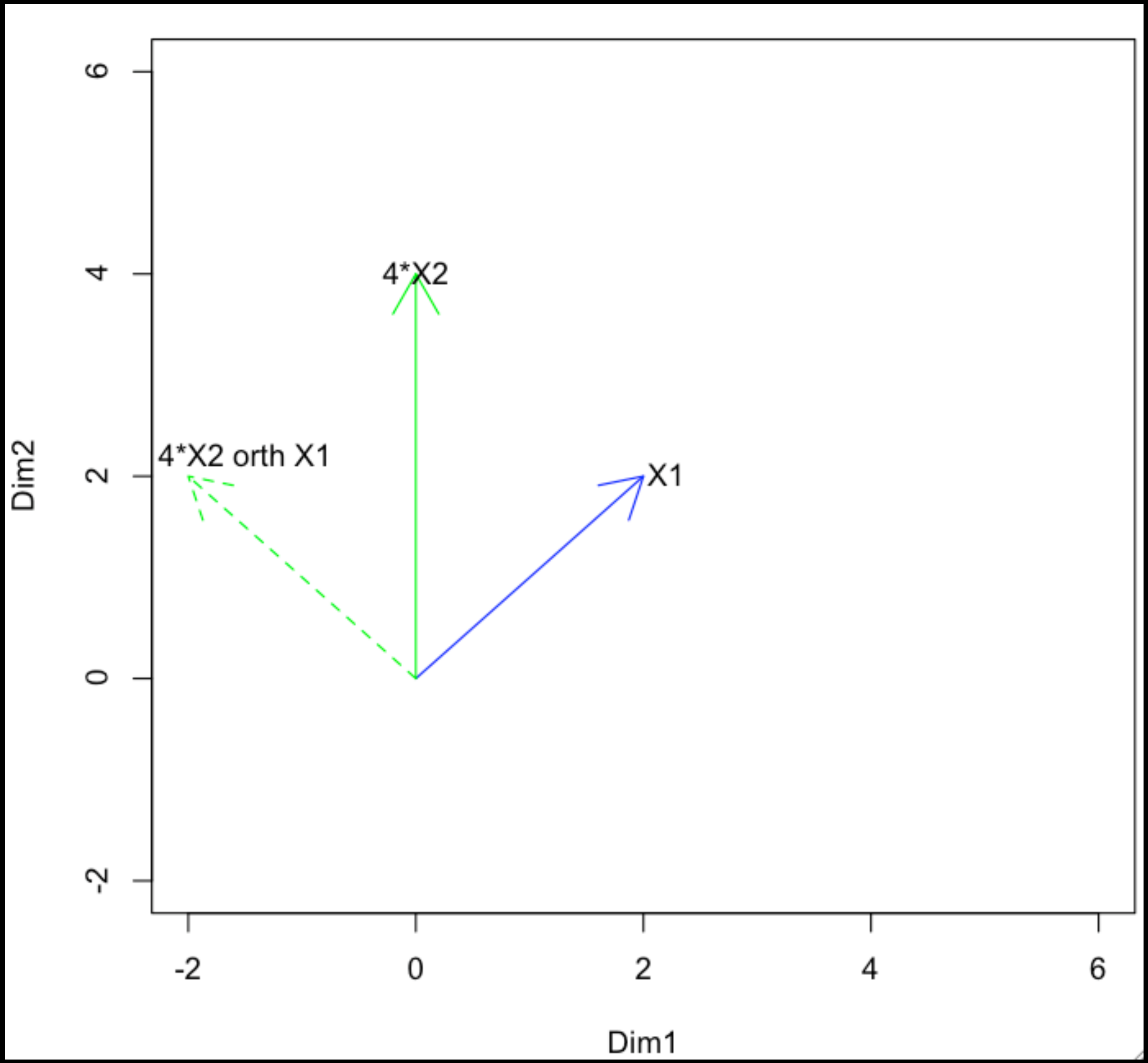
$$\begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \beta_1 + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \beta_2$$

$\beta_1=2$ and $\beta_2=5$

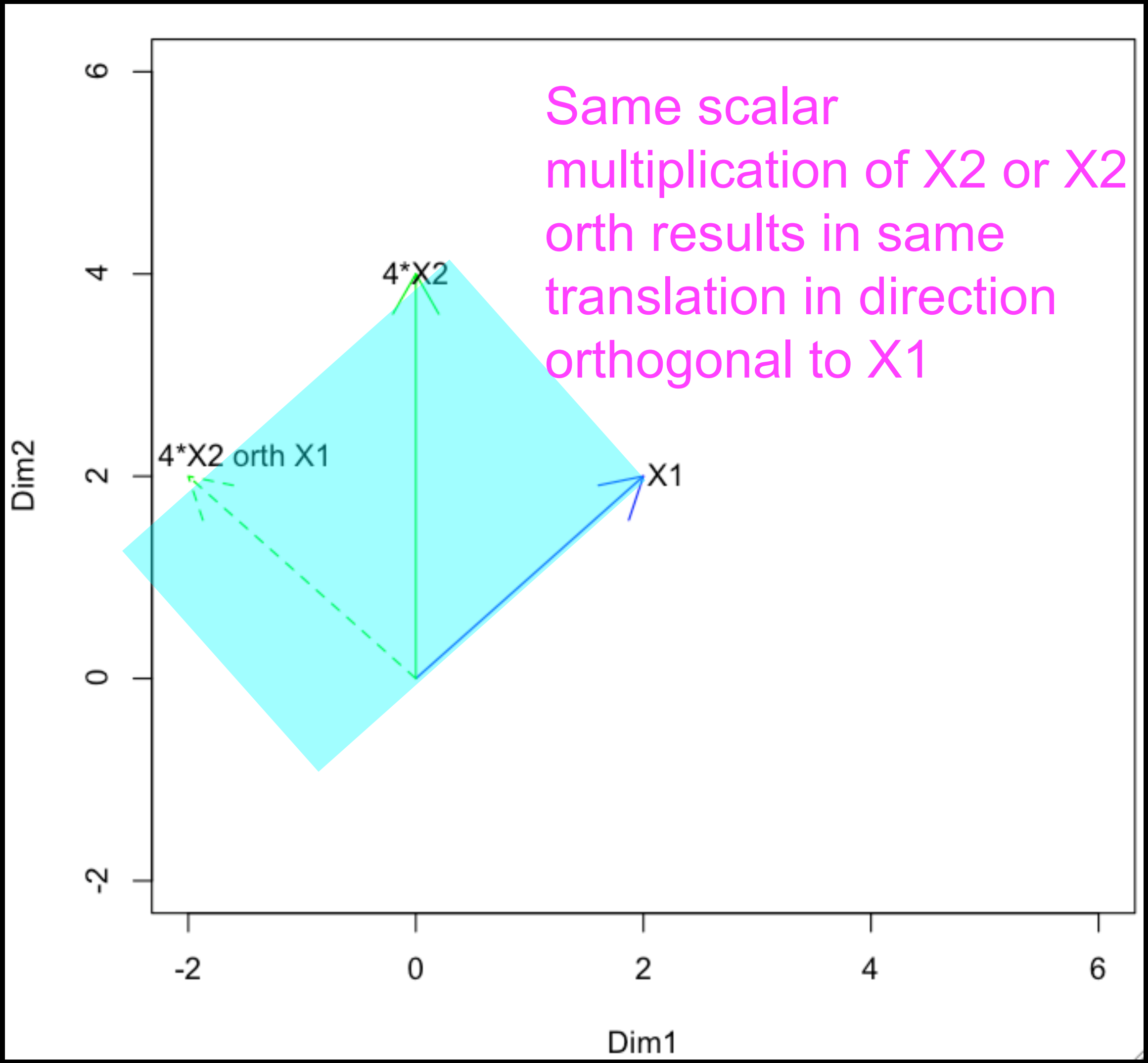
Other way around



Look at scalar multiplication



Look at scalar multiplication



Orthogonalization

- If two regressors are correlated, then they don’t explain unique “directions”
- Orthogonalizing A wrt B means you are surrendering A’s ability to explain variability in the B direction
- Eg, orthogonalizing age with respect to the mean is removing the ability of age to describe the overall mean of the data

Orthogonalization

- Be careful when you interpret parameters.
  - $Y \quad = \quad X_1\beta_1 \quad + \quad X_2\beta_2$
  - Typically you interpret β_2 as the effect of X_1 adjusting for X_2
  - Eg, the effect of age controlling for gender

- Orthogonalizing surrenders the ability of one covariate to control for another
  - If you model X1 orth X2, then X1 is only soaking up extra variability NOT also adjusting X2

Model 7

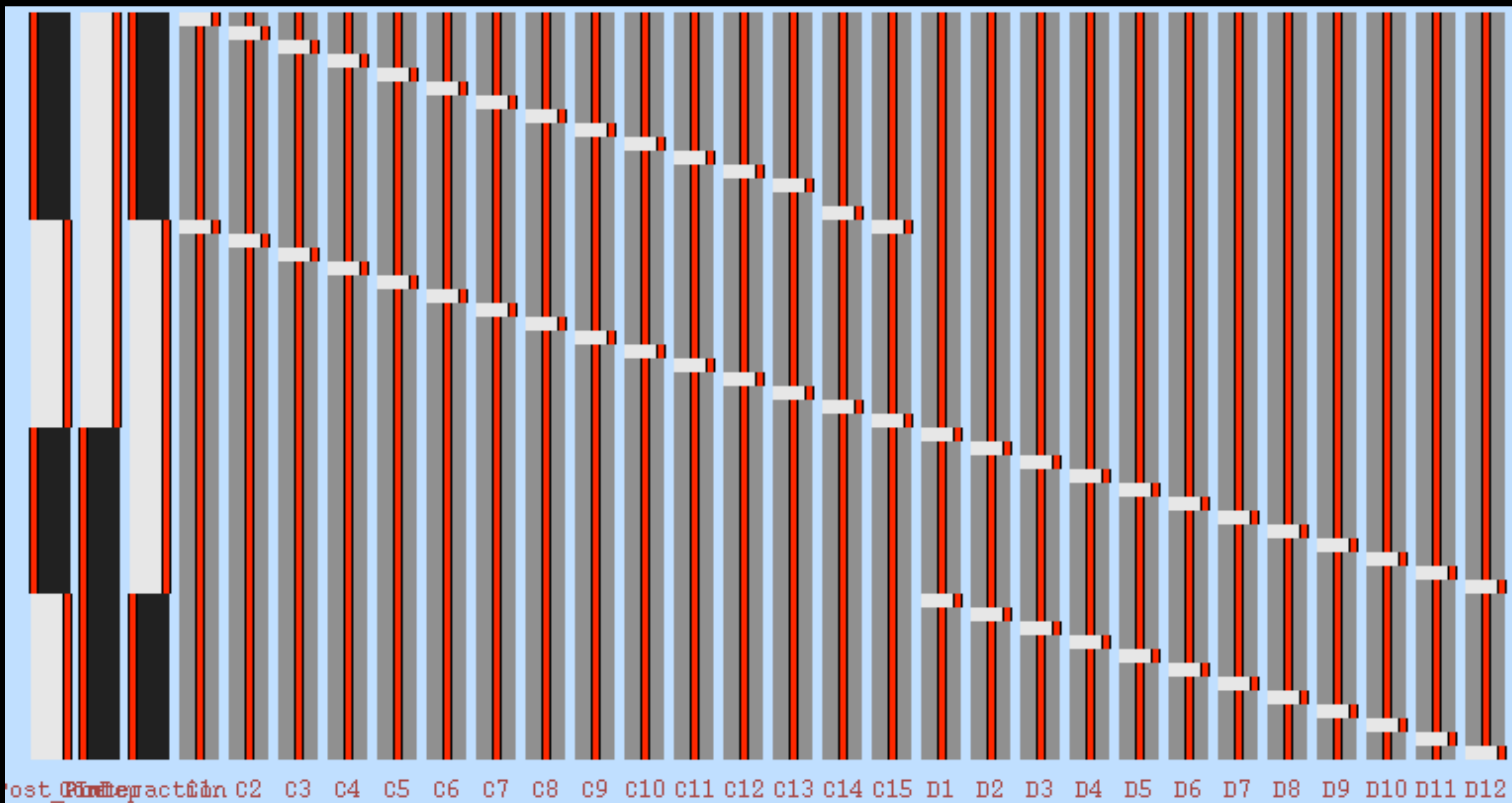
June 29, 2009, FSL help list

“We are running this model 2x2 to examine a treatment study, so all subjects (patients and controls) were scanned twice (so the model examines the GroupX Time interaction, with Time a repeated measure). We would like to examine a symptom score obtained at all scans. How would one add a symptom covariate to this model? Is it simply the addition of 4 EVs (control T1, control T2, patient T1, patient T2) of the symptom score?? If so, I assume the symptom score is demeaned seperately for each of these 4 conditions (control T1, control T2, patient T1, patient T2).”

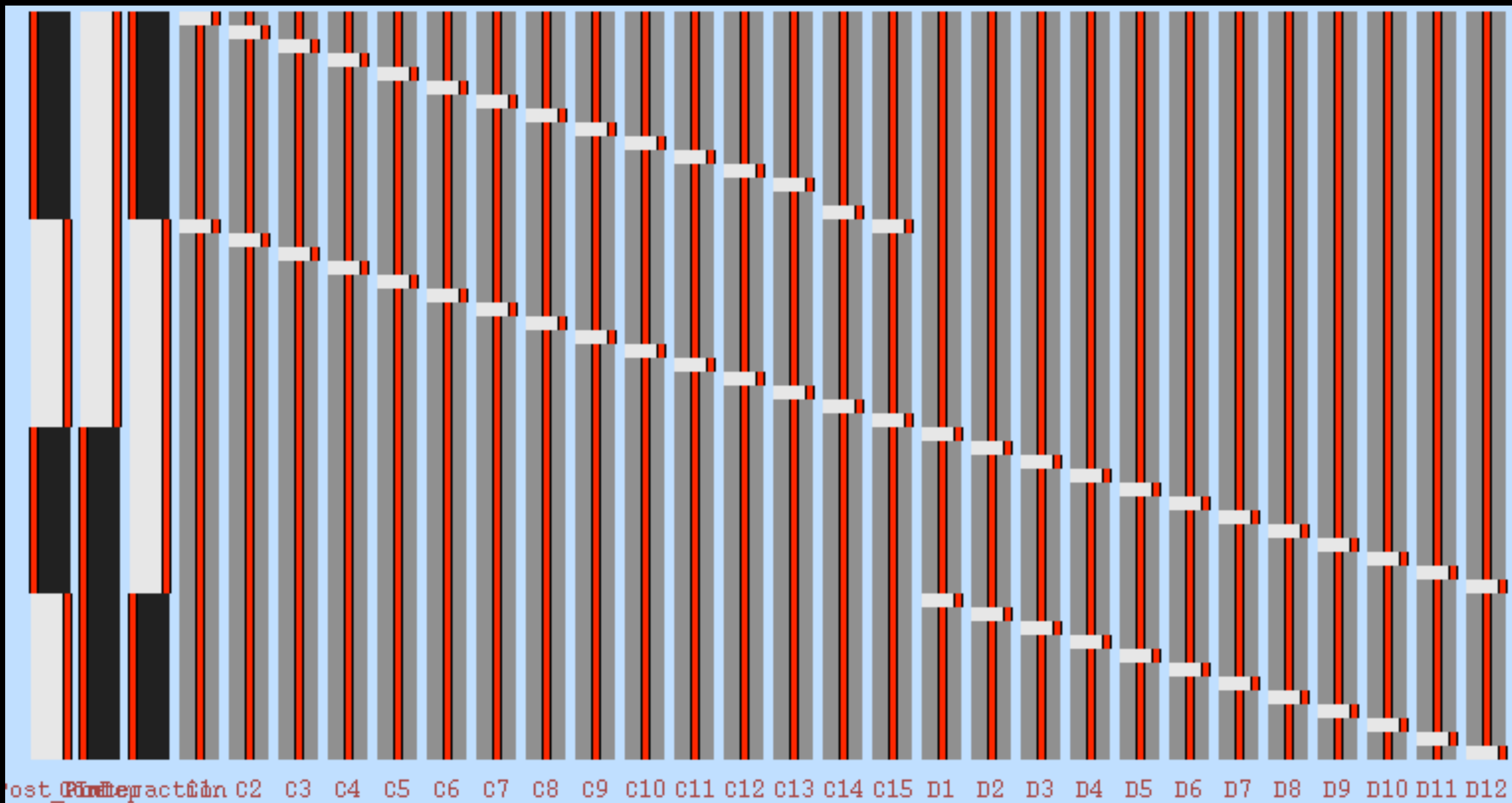
Model 7

- Summary
  - 2x2 ANOVA with additional continuous confounder
  - Repeated measure on first factor (Time)
  - Not repeated on second factor (Group)

What is wrong with this?



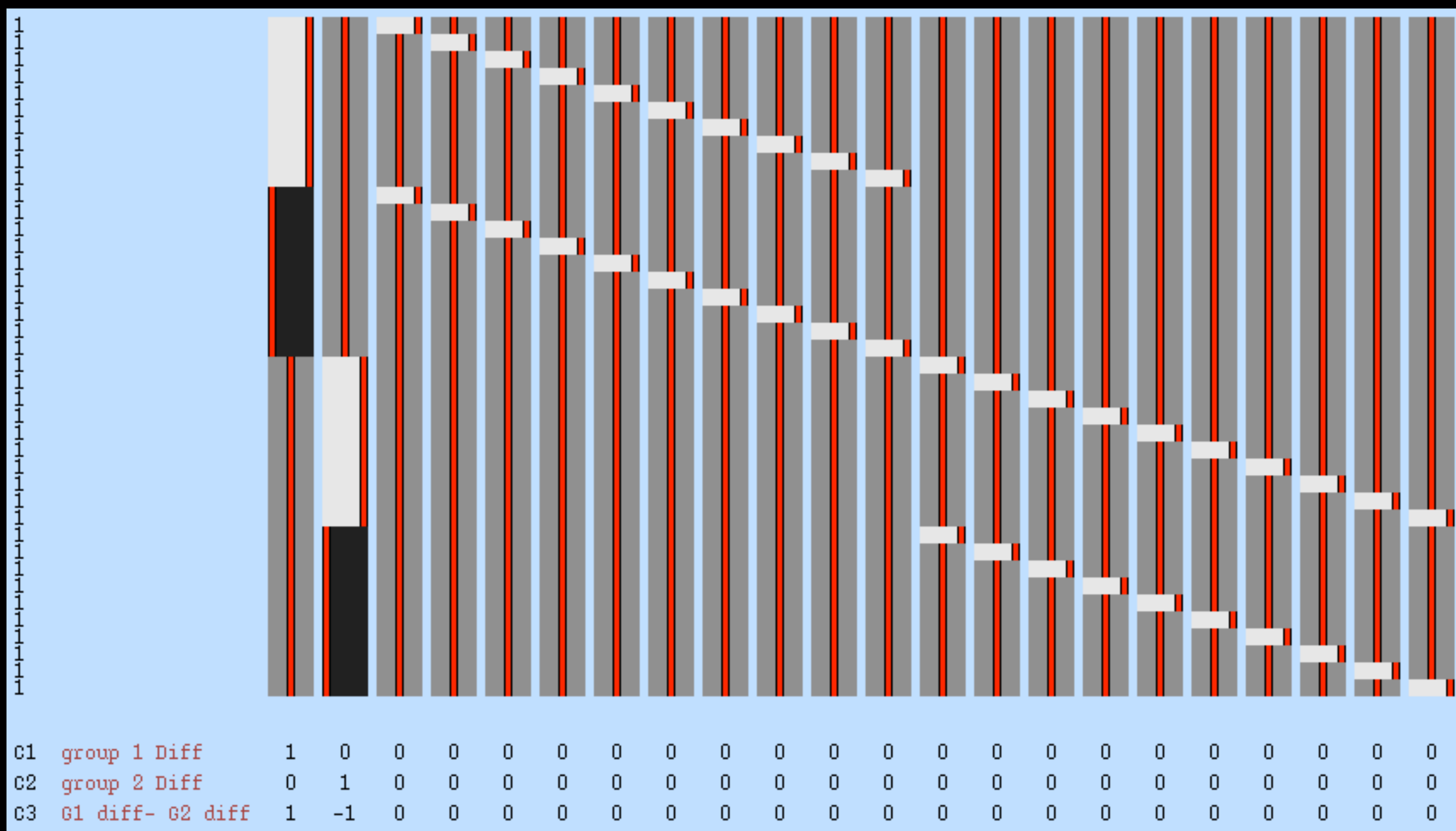
What is wrong with this?



Rank deficient…but a good effort!

Suggested model

- Extension of the paired t test
- Adjust for repeated measure within group



Adding symptom covariate

- Personally, I’d first check that it didn’t have structure that correlated with group and scan
- Orthogonalizing wrt ev1 and ev2 will adjust residuals for symptom, but not group or run differences
  - I’d only do this if there weren’t group or run differences in symptom

Questions?