Two Between-Subjects Factors with Interactions

Last time

Twoway ANOVA

- 2 between subjects factors
- Can have as many levels as necessary
    - 3x5 ANOVA still has 2 independent variables
- Explored main effects

This time

- What happens when those 2 variables impact each other (not only impact the dependent variable)?

Math is hard

Here are our data...

	Normal Equal Sign Problems				Tricky Equal Sign Problems			
8-9 year olds	3	4	3	4	0	0	0	1
9-10 year olds	4	4	4	3	2	3	2	1

Math is hard: Null Hypotheses for Main Effects

$H_{0.1}$: The two age groups perform similarly

$$H_{0.1} : \mu_8 = \mu_9$$

$H_{0.2}$: Children perform similarly on normal and tricky math problems

$$H_{0.2} : \mu_{\text{normal}} = \mu_{\text{tricky}}$$

Math is hard

	Normal Equal Sign Problems				Tricky Equal Sign Problems				
8-9 year olds	3	4	3	4	0	0	0	1	μ_8
9-10 year olds	4	4	4	3	2	3	2	1	μ_9
	μ_{Normal}				μ_{Tricky}				

Adding a 3rd Null Hypothesis

$H_{0.3}$: *The differences between age groups are the same regardless of problem type (tricky or normal)*

$H_{0.3}$: *Differential performance on problem types is independent of age group*

☝️ These are the same! They are symmetric.

Math is hard: Means on means

	Normal Equal Sign Problems	Tricky Equal Sign Problems
8-9 year olds	3.50	0.25
9-10 year olds	3.75	2.00
Difference between Rows	**0.25**	**1.75**

Math is hard: Means on means

	Normal Equal Sign Problems	Tricky Equal Sign Problems	Difference between Column
8-9 year olds	3.50	0.25	**3.25**
9-10 year olds	3.75	2.00	**1.75**

Equations

Previously, we've talked about the restricted and the full model. We can write those as formal equations:

$$Restricted: Y_{ij} = \mu + e_{ij}$$

$$Full: Y_{ij} = \bar{Y}_j + e_{ij}$$

Oneway ANOVA

$$Restricted: Y_{ij} = \mu + e_{ij}$$

$$Full: Y_{ij} = \bar{Age}_j + e_{ij}$$

Three Restricted Models

When testing the Age main effect...

$$Restricted: Y_{ij} = \mu + \text{Problem}_k + AP_{jk} + e_{ijk}$$

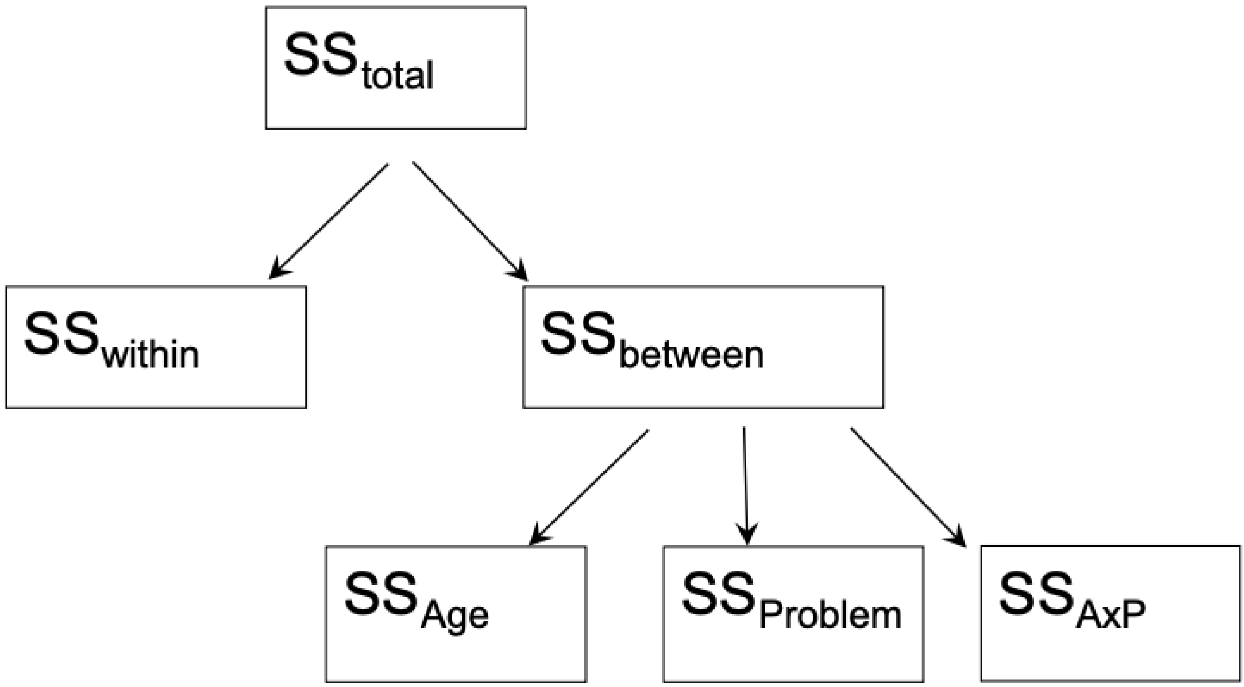
When testing the Problem main effect...

$$Restricted: Y_{ij} = \mu + \text{Age}_j + AP_{jk} + e_{ijk}$$

When testing the **interaction** between Age & Problem...

$$Restricted: Y_{ij} = \mu + \text{Age}_j + \text{Problem}_k + e_{ijk}$$

Diagrams



SS_{within} is our Ef; SS_{between} is our Er - Ef; MS_{within} is the denominator for all *F*-tests

ANOVA (Source) Table

	SS	df	MS	F
Age	SS_{Age}	df_{Age}	$\frac{SS_{Age}}{df_{Age}}$	$\frac{MS_{Age}}{MS_W}$
Problem	$SS_{Problem}$	$df_{Problem}$	$\frac{SS_{Problem}}{df_{Problem}}$	$\frac{MS_{Problem}}{MS_W}$
Age x Problem	SS_{AxP}	df_{AxS}	$\frac{SS_{AxS}}{df_{AxS}}$	$\frac{MS_{AxP}}{MS_W}$
Error / Residuals	SS_W	df_W	$\frac{SS_W}{df_W}$	

ANOVA (Source) Table

Fill in the easy stuff first

	SS	df	MS	F
Age	SS_{Age}	$(A - 1)$	$\frac{SS_{Age}}{df_{Age}}$	$\frac{MS_{Age}}{MS_W}$
Problem	$SS_{Problem}$	$(P - 1)$	$\frac{SS_{Problem}}{df_{Problem}}$	$\frac{MS_{Problem}}{MS_W}$
Age x Problem	SS_{AxP}	$(A - 1) \times (P - 1)$	$\frac{SS_{AxS}}{df_{AxS}}$	$\frac{MS_{AxP}}{MS_W}$
Error / Residuals	SS_W		$\frac{SS_W}{df_W}$	

Df for the Full Model

Always the (Number of Participants - Number of Estimated Parameters in Full Model)

How many means did we calculate?

$$AB(n-1)$$
$$2 \times 2(4-1)$$
$$4 \times 3$$
$$12$$

ANOVA (Source) Table

	SS	df	MS	F
Age	SS_{Age}	**(A – 1)**	$\frac{SS_{Age}}{df_{Age}}$	$\frac{MS_{Age}}{MS_W}$
Problem	$SS_{Problem}$	**(P – 1)**	$\frac{SS_{Problem}}{df_{Problem}}$	$\frac{MS_{Problem}}{MS_W}$
Age x Problem	SS_{AxP}	**(A - 1) x (P - 1)**	$\frac{SS_{AxS}}{df_{AxS}}$	$\frac{MS_{AxP}}{MS_W}$
Error / Residuals	SS_W	**AB(n -1)**	$\frac{SS_W}{df_W}$	

ANOVA (Source) Table

	SS	df	MS	F
Age	SS_{Age}	**1**	$\frac{SS_{Age}}{1}$	$\frac{MS_{Age}}{MS_W}$
Problem	$SS_{Problem}$	**1**	$\frac{SS_{Problem}}{1}$	$\frac{MS_{Problem}}{MS_W}$
Age x Problem	SS_{AxP}	**1**	$\frac{SS_{AxS}}{1}$	$\frac{MS_{AxP}}{MS_W}$
Error / Residuals	SS_W	**12**	$\frac{SS_W}{12}$	

How to calculate SS_{Age}

$SS_{Age} = E_R - E_F$

$Restricted : Y_{ij} = \mu + \text{Problem}_k + AP_{jk} + e_{ijk}$

$Full : Y_{ij} = \mu + \text{Age}_j + \text{Problem}_k + AP_{jk} + e_{ijk}$

Get the E_R for Age

Data	Prediction	Error	Squared Error	
3	2.375	-0.625	0.390625	
4	2.375	-1.625	2.640625	
3	2.375	-0.625	0.390625	
4	2.375	-1.625	2.640625	
0	2.375	2.375	5.640625	
0	2.375	2.375	5.640625	
0	2.375	2.375	5.640625	E_R = 35.75
1	2.375	1.375	1.890625	
4	2.375	-1.625	2.640625	
4	2.375	-1.625	2.640625	
4	2.375	-1.625	2.640625	
3	2.375	-0.625	0.390625	
2	2.375	0.375	0.140625	
3	2.375	-0.625	0.390625	
2	2.375	0.375	0.140625	
1	2.375	1.375	1.890625	

Get the E_F for Age

Data	Prediction	Error	Squared Error	
3	1.875	-1.125	1.265625	
4	1.875	-2.125	4.515625	
3	1.875	-1.125	1.265625	
4	1.875	-2.125	4.515625	
0	1.875	1.875	3.515625	
0	1.875	1.875	3.515625	
0	1.875	1.875	3.515625	**E_R = 31.75**
1	1.875	0.875	0.765625	
4	2.875	-1.125	1.265625	
4	2.875	-1.125	1.265625	
4	2.875	-1.125	1.265625	
3	2.875	-0.125	0.015625	
2	2.875	0.875	0.765625	
3	2.875	-0.125	0.015625	
2	2.875	0.875	0.765625	
1	2.875	1.875	3.515625	

How to calculate SS_{Age}

$$SS_{Age} = E_R - E_F$$

```r
35.75 - 31.75
```

```
## [1] 4
```

	SS	df	MS	F
Age	**Same E_R - E_F as if there was one factor**	1	$\frac{SS_{Age}}{1}$	$\frac{MS_{Age}}{MS_W}$
Problem	**Same E_R - E_F as if there was one factor**	1	$\frac{SS_{Problem}}{1}$	$\frac{MS_{Problem}}{MS_W}$
Age x Problem	SS_{AxP}	1	$\frac{SS_{AxS}}{1}$	$\frac{MS_{AxP}}{MS_W}$
Error / Residuals	SS_W	12	$\frac{SS_W}{12}$	

ANOVA (Source) Table

	SS	df	MS	F
Age	**4.00**	1	$\frac{SS_{Age}}{1}$	$\frac{MS_{Age}}{MS_W}$
Problem	**25.00**	1	$\frac{SS_{Problem}}{1}$	$\frac{MS_{Problem}}{MS_W}$
Age x Problem	SS_{AxP}	1	$\frac{SS_{AxS}}{1}$	$\frac{MS_{AxP}}{MS_W}$
Error / Residuals	SS_W	12	$\frac{SS_W}{12}$	

ANOVA (Source) Table

	SS	df	MS	F
Age	**4.00**	1	**4.00**	$\frac{4.00}{MS_W}$
Problem	**25.00**	1	**25.00**	$\frac{25.00}{MS_W}$
Age x Problem	SS_{AxP}	1	$\frac{SS_{AxS}}{1}$	$\frac{MS_{AxP}}{MS_W}$
Error / Residuals	SS_W	12	$\frac{SS_W}{12}$	

Calculating the Interaction (SS_{AxP})

$H_{0.3}$: *The differences between age groups are the same regardless of problem type (tricky or normal)*

$H_{0.3}$: *Differential performance on problem types is independent of age group*

What does the restricted model say?

Your score = grand mean + some age effect + some problem effect + error

Predicted Score = grand mean + (diffrence between age effect and grand mean) + (difference between problem effect and grand mean)

Predicted Score = Grand Mean + (MeanRow – Grand Mean) + (MeanColumn – Grand Mean)

Predicted Score = Grand Mean + MeanRow – Grand Mean + MeanColumn – Grand Mean

Predicted Score = MeanRow + MeanColumn – Grand Mean

Difference between Restricted and Full Models = Squared deviations for each cell from this predicted score

Interaction Effect

Difference between Restricted and Full Models = Squared deviations for each cell from this predicted score

We need to get the deviations + squared deviations of each cell mean from the predicted score

$$\Sigma(\text{Cell mean - Predicted Score})^2$$

$$\Sigma(\text{Cell mean - (Mean row} + \text{Mean column - Grand mean}))^2$$

$$\Sigma(\text{Cell mean - Mean row - Mean column} + \text{Grand mean})^2$$

OG Data

	Normal Equal Sign Problems				Tricky Equal Sign Problems				
8-9 year olds	3	4	3	4	0	0	0	1	μ_8
9-10 year olds	4	4	4	3	2	3	2	1	μ_9
	μ_{Normal}				μ_{Tricky}				

Marginal Means

	Normal Equal Sign Problems	Tricky Equal Sign Problems	
8-9 year olds	3.50	0.25	**1.875**
9-10 year olds	3.75	2.00	**2.875**
	3.625	**1.125**	**2.375**

Calculating the Interaction Effect

Cell Mean	Subtract Row Mean	Subtract Column Mean	Add Grand Mean	Total
3.50	1.875	3.625	2.375	0.375
0.25	1.875	1.125	2.375	-0.375
3.75	2.875	1.125	2.375	-0.375
2.00	2.875	3.625	2.375	0.375

Sum of deviations is 0!

Calculating the Interaction Effect

Cell Mean	Subtract Row Mean	Subtract Column Mean	Add Grand Mean	Total	Squared
3.50	1.875	3.625	2.375	0.375	.140625
0.25	1.875	1.125	2.375	-0.375	.140625
3.75	2.875	1.125	2.375	-0.375	.140625
2.00	2.875	3.625	2.375	0.375	.140625

Sum them up = 0.5625

Multiply by n per group = 0.5625 * 4 = 2.25

What we just did

- Found the means for each cell in our 2x2 ANOVA
- Took away the effect of each of the 2 factors (Age and Problem)
- Added back the grand mean
- *The extent to which each cell differs from its respective column and row is the difference between a model without an interaction (Restricted) and a model with one (Full)*

ANOVA (Source) Table

	SS	df	MS	F
Age	4.00	1	4.00	$\frac{4.00}{MS_W}$
Problem	25.00	1	25.00	$\frac{25.00}{MS_W}$
Age x Problem	**2.25**	1	**2.25**	$\frac{2.25}{MS_W}$
Error / Residuals	SS_W	12	$\frac{SS_W}{12}$	

SS_{within}

Error from the full model

The full model allows for the Age effect, the Problem effect, and the Interaction effect.

So from the full model perspective, the best guess for observation is...

SS_{within}

Score	Predicted Value	Deviation	Squared Deviation	
3	3.5	0.5	0.25	
4	3.5	-0.5	0.25	
3	3.5	0.5	0.25	
4	3.5	-0.5	0.25	
0	0.25	0.25	0.0625	
0	0.25	0.25	0.0625	**E_F = 4.5**
0	0.25	0.25	0.0625	
1	0.25	-0.75	0.5625	
4	3.75	-0.25	0.0625	
4	3.75	-0.25	0.0625	
4	3.75	-0.25	0.0625	
3	3.75	0.75	0.5625	
2	2	0	0	
3	2	-1	1	
2	2	0	0	
1	2	1	1	

ANOVA (Source) Table

	SS	df	MS	F
Age	4.00	1	4.00	$\frac{4.00}{MS_W}$
Problem	25.00	1	25.00	$\frac{25.00}{MS_W}$
Age x Problem	2.25	1	2.25	$\frac{2.25}{MS_W}$
Error / Residuals	4.50	12	$\frac{4.50}{12} = 0.375$	

ANOVA (Source) Table

	SS	df	MS	F
Age	4.00	1	4.00	$\frac{4.00}{0.375}$
Problem	25.00	1	25.00	$\frac{25.00}{0.375}$
Age x Problem	2.25	1	2.25	$\frac{2.25}{0.375}$
Error / Residuals	4.50	12	0.375	

ANOVA (Source) Table

	SS	df	MS	F
Age	4.00	1	4.00	10.667
Problem	25.00	1	25.00	66.667
Age x Problem	2.25	1	2.25	6.00
Error / Residuals	4.50	12	0.375	

```r
qf(p = .05, df1 = 1, df2 = 12, lower.tail = FALSE)
```

```
## [1] 4.747225
```

What do we conclude?

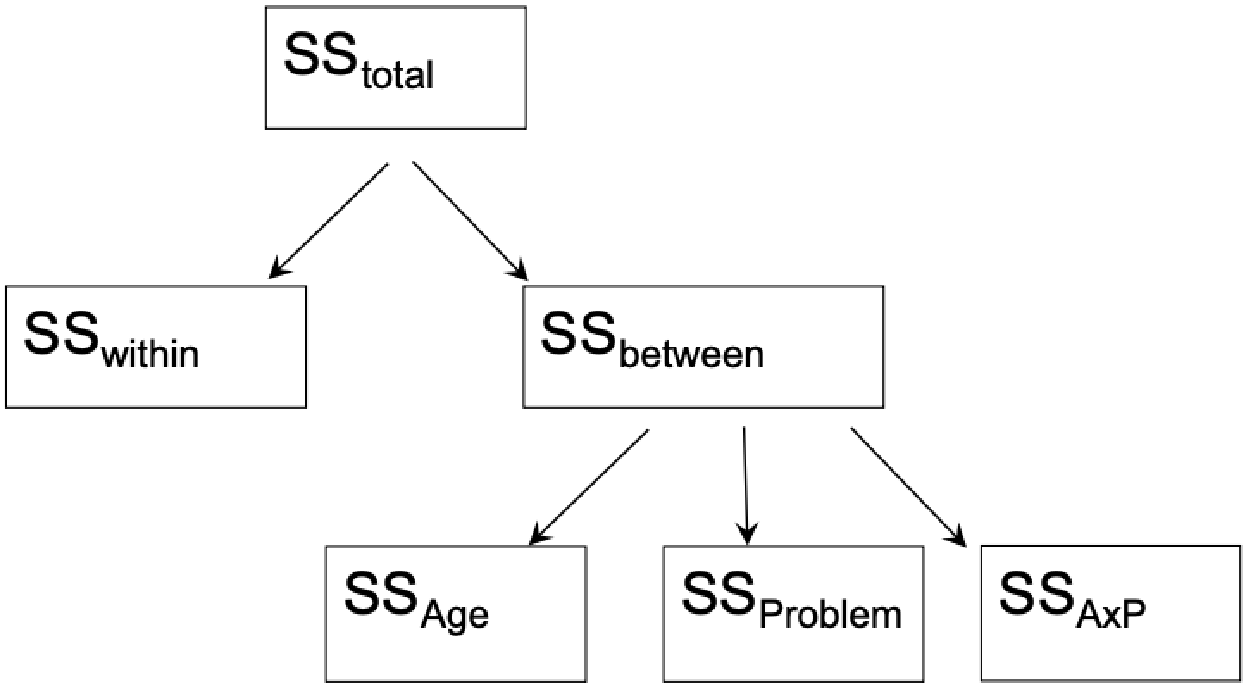
We have 3 rejected null hypotheses. We interpret:

1) There is a main effect of age; the means of age groups are not equal

2) There is a main effct of problem type; the means of normal vs. tricky math problems are not equal

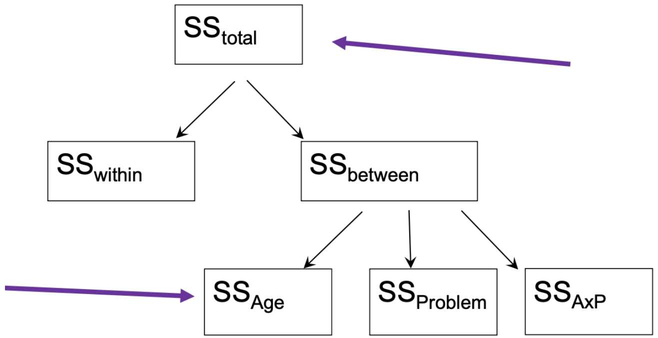
3) The difference in age groups differs by problem type (and vice versa)

Effect Sizes



Eta-squared

$$\eta^2 = \frac{SS_{Effect}}{SS_{Total}}$$

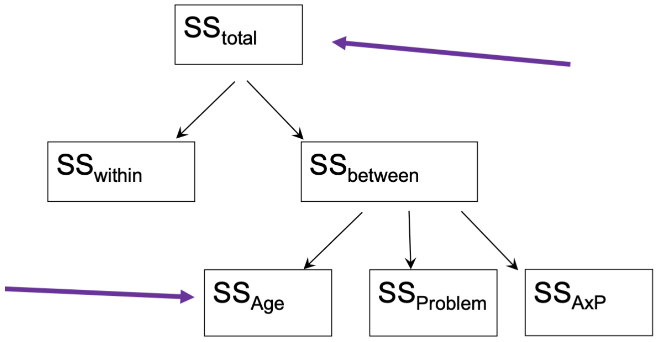


$$\eta^2 = \frac{4}{35.75} = .112$$

Proportion of the variation in Y that is associated with membership of the different groups defined by X. How much variance in our outcome is associated with Age? 11.2% of *total* variance can be accounted for by Age.

Eta-squared

$$\eta^2 = \frac{SS_{Effect}}{SS_{Total}}$$

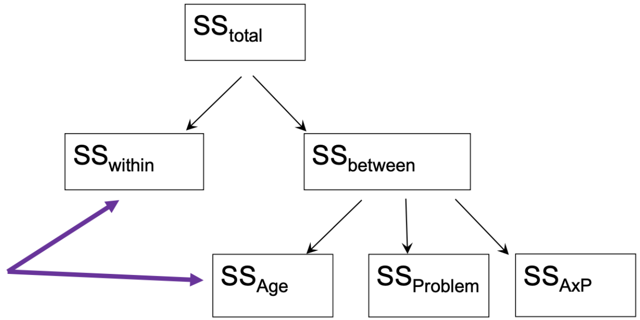


$$\eta^2 = \frac{4}{35.75} = .112$$

Hard to compare across studies bc dependent on study's total variance.

Partial eta-squared

$$\eta_p^2 = \frac{SS_{Effect}}{SS_{Effect}+SS_{Error}}$$



$$\eta^2 = \frac{4}{4+4.5} = .471$$

Proportion of the variation in Y that can be explained by X, after accounting for all other variables in th model. How much variance in our outcome is associated with Age, after accounting for Problem, and their interaction? 47.1%.

Same rules of thumb for η^2; calculations differ when sample sizes are unequal or when there are more than 2 factors

In R

```r
library(here)
mathproblem = read.csv(here("data/mathproblem.csv"))
mathproblem
```

```
##    Score   Age Problem
## 1      3 Eight  Normal
## 2      4 EIght  Normal
## 3      3 Eight  Normal
## 4      4 Eight  Normal
## 5      4  Nine  Normal
## 6      4  Nine  Normal
## 7      4  Nine  Normal
## 8      3  Nine  Normal
## 9      0 Eight  Tricky
## 10     0 Eight  Tricky
## 11     0 Eight  Tricky
## 12     1 Eight  Tricky
## 13     2  Nine  Tricky
## 14     3  Nine  Tricky
## 15     2  Nine  Tricky
## 16     1  Nine  Tricky
```

In R - No Interaction

```
summary(aov(Score ~ Age + Problem, data = mathproblem))
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## Age           2  9.161   4.580   9.464  0.00341 ** 
## Problem       1 20.782  20.782  42.939 2.72e-05 ***
## Residuals    12  5.808   0.484                     
## ---
## Signif. codes:  
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What's wrong with this?

In R - No Interaction

```
mathproblem[2,2] <- "Eight"
mathproblem$Age <- factor(mathproblem$Age)
mathproblem$Problem <- factor(mathproblem$Problem)

summary(aov(Score ~ Age + Problem, data = mathproblem))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## Age          1   4.00   4.000   7.704   0.0157 *  
## Problem      1  25.00  25.000  48.148 1.02e-05 ***
## Residuals   13   6.75   0.519                     
## ---
## Signif. codes:  
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In R - With Interaction

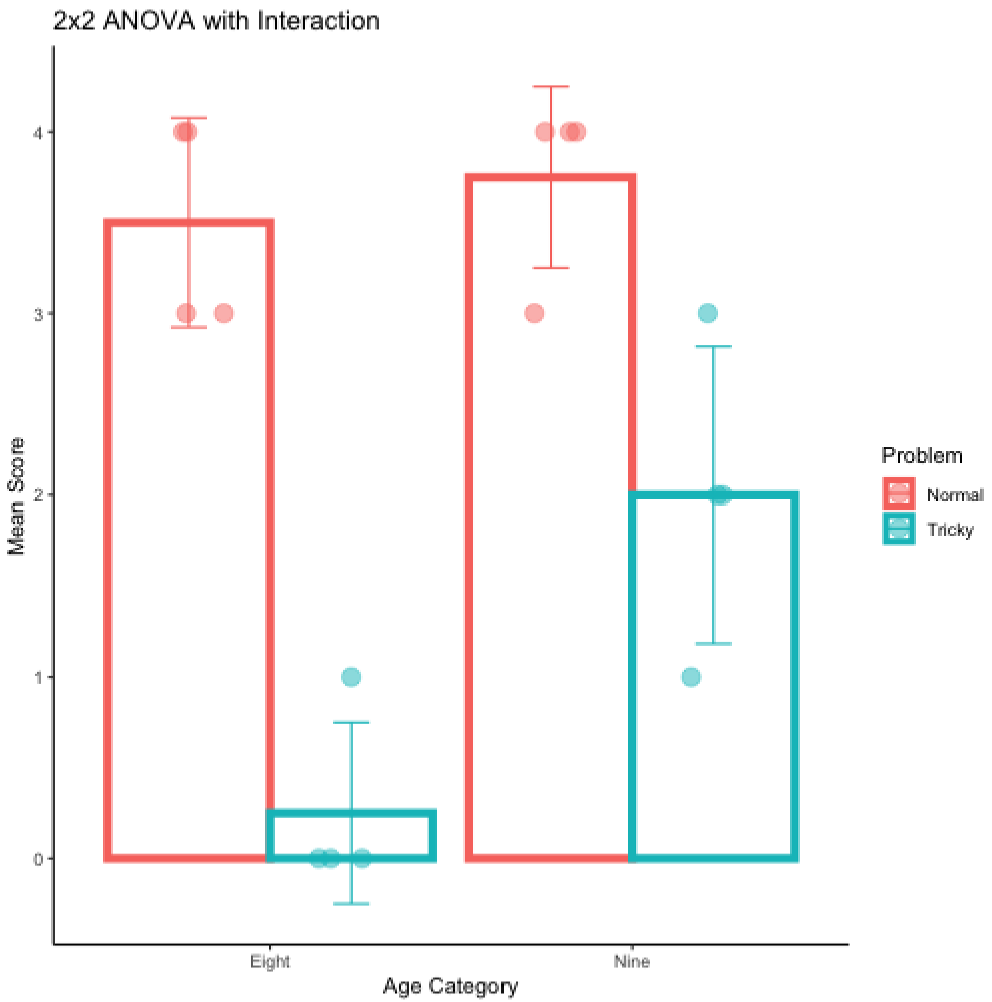
```
summary(aov(Score ~ Age + Problem + Age*Problem, data = mathprobl
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## Age           1   4.00   4.000   10.67  0.00675 ** 
## Problem       1  25.00  25.000   66.67 3.05e-06 ***
## Age:Problem   1   2.25   2.250    6.00  0.03062 *  
## Residuals    12   4.50   0.375                     
## ---
## Signif. codes:  
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(Score ~ Age * Problem, data = mathproblem))
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)    
## Age           1   4.00   4.000   10.67  0.00675 ** 
## Problem       1  25.00  25.000   66.67 3.05e-06 ***
## Age:Problem   1   2.25   2.250    6.00  0.03062 *  
## Residuals    12   4.50   0.375                     
## ---
## Signif. codes:  
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plotting


2x2 ANOVA with Interaction
Mean Score
4
3
2
1
0
Eight
Nine
Age Category
Problem
Normal
Tricky

Next time

- Within-subjects designs
- Validity
- Exam 3 review
- Teary goodbyes (until next semester)