Linear Regression with a Single Categorical Predictor with Related Levels

AKA, REPEATED MEASURES ANOVA / PAIRED SAMPLES T-TEST

Between-Subjects vs Within-Subjects

- Between-Subjects Design
  - When investigating the effect of an independent variable (IV) on a dependent variable (DV), and
  - The participants in each sample are *independent of* or *unrelated to* each other

- Within-Subjects Design
  - When investigating the effect of an independent variable (IV) on a dependent variable (DV), and
  - The participants in each sample are *related to* each other
    - Violates the “independence of errors” assumption

F-Statistic for Between-Subjects Design

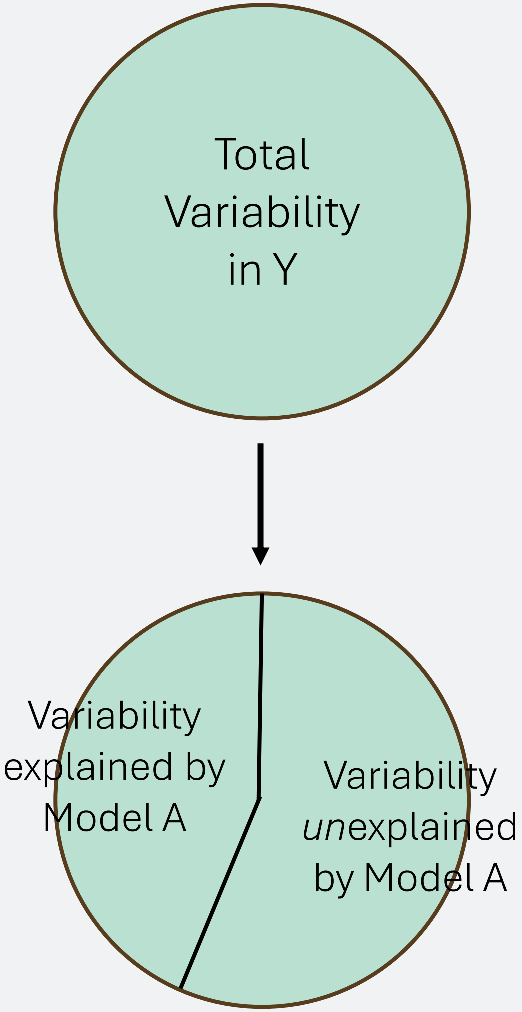
- Recall that we typically calculate an *F*-statistic to test the significance of Model A compared to Model C, where the *F*-statistic has the general form of:

$$F = \frac{\boldsymbol{variance\ explained\ by\ model\ A}}{\boldsymbol{variance\ unexplained\ by\ model\ A}}$$

- For a between-subjects design, we calculated this *F*-statistic using:

$$F = \frac{MS_{Reduced}}{MS_{ModelA}} = \frac{SSR/df_{Reduced}}{SSE(A)/df_{ModelA}}$$

- $MS_{Reduced}$ (MSR) = variance explained by Model A compared to Model C
- MS_{ModelA} (MSE) = variance left unexplained by Model A



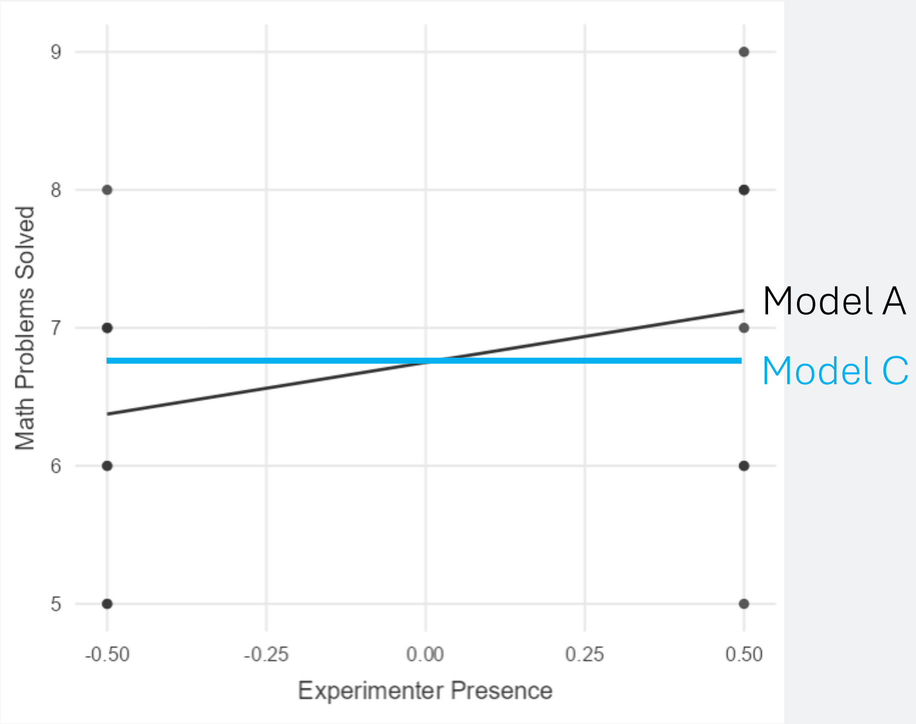
Variability Explained vs Unexplained by Model A

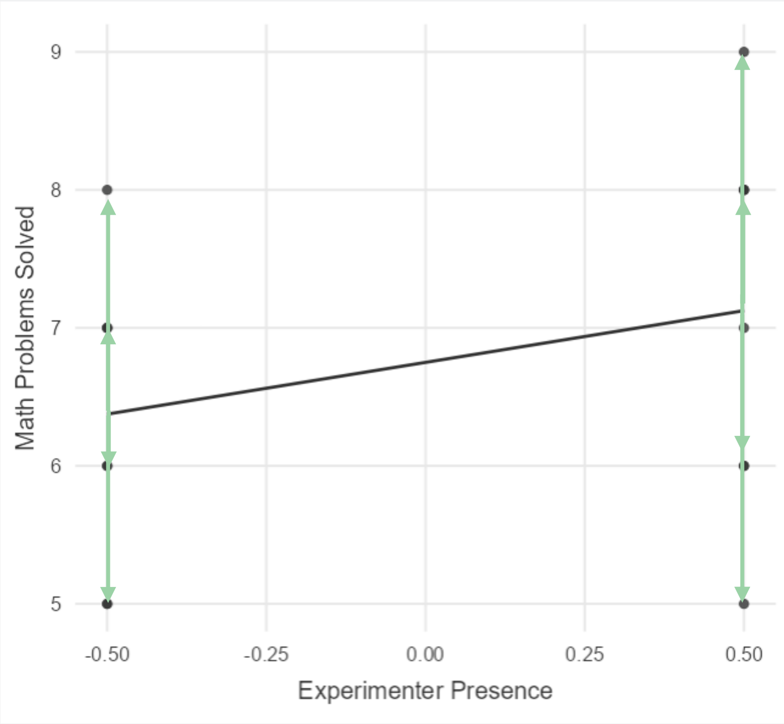
Variability explained by Model A:

- The additional variability explained by Model A compared to Model C
  - SSR = SSE(C) – SSE(A)

Variability unexplained by Model A:

- The remaining error left unaccounted for by Model A
  - SSE(A) = $\Sigma(Y_i - Y'_i)^2$



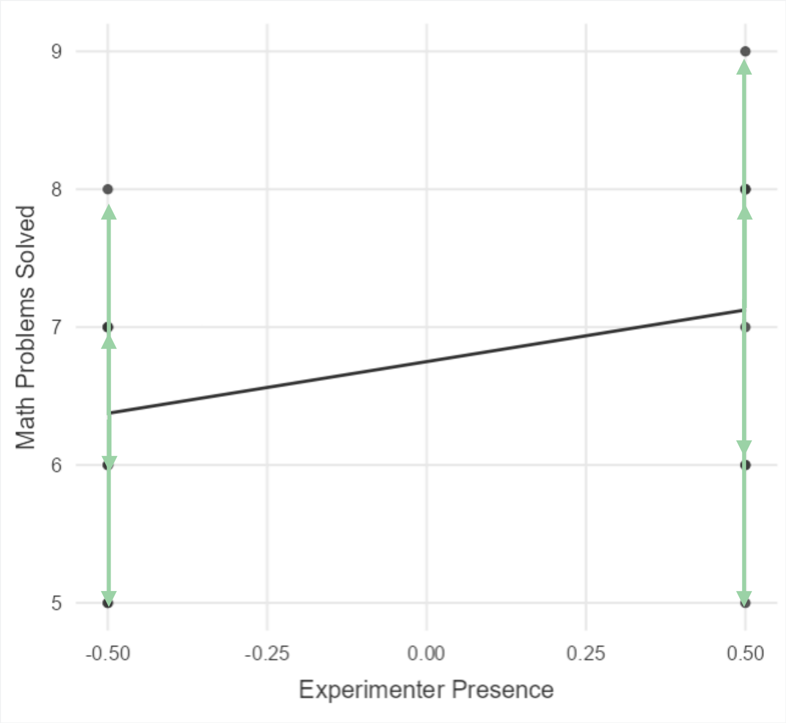


Variability Unexplained by Model A

- Q: Where is there remaining error left unaccounted for by Model A ?

- The remaining unexplained error is due to:
  - Individual differences
  - Random error

- A within-subjects design allows us to estimate the amount of the error left unexplained by Model A that is due to individual differences
  - Called “subject variability”

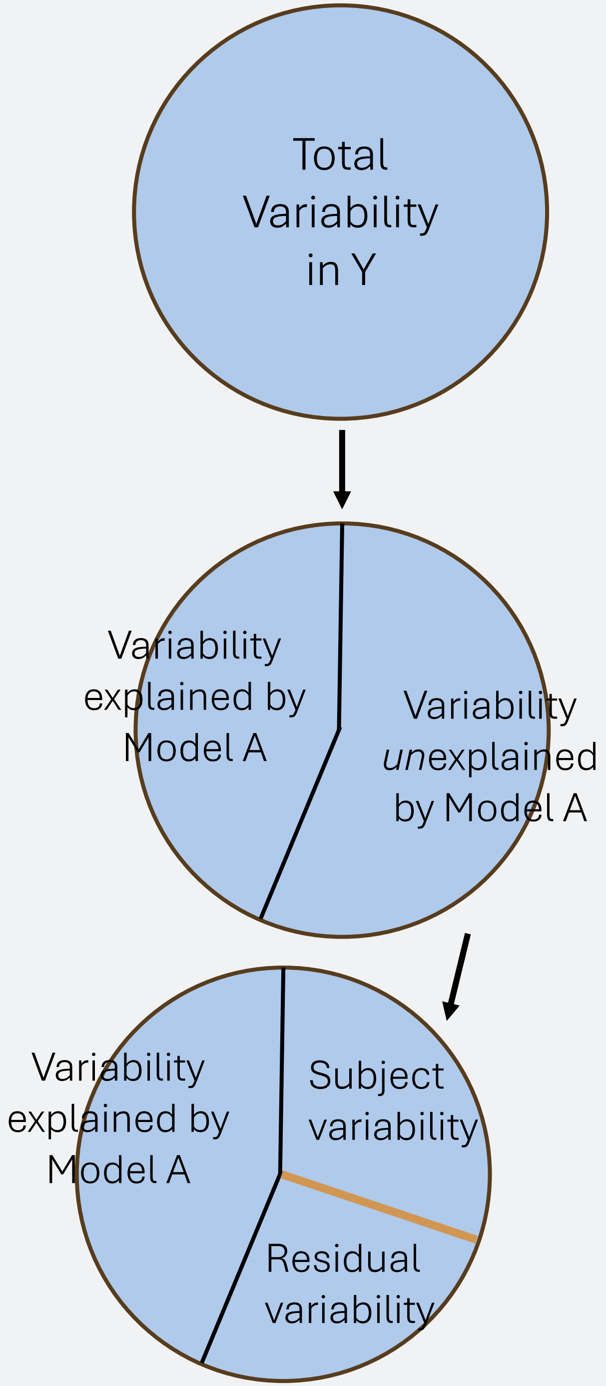


F-Statistic for Within-Subjects Design

- For a within-subjects design, the total variability in Y can be broken down into:
  - Variability explained by Model A
  - Variability *un*explained by Model A

- And the variability <u>un</u>explained by Model A can further be broken down into:
  - Subject variability
  - Residual variability

- Residual Variability
  - The part of the unexplained variability that is left over after subtracting out subject variability
  - SSE(Residual) = SSE(A) – SSE(Subjects)



F-Statistic for Within-Subjects Design

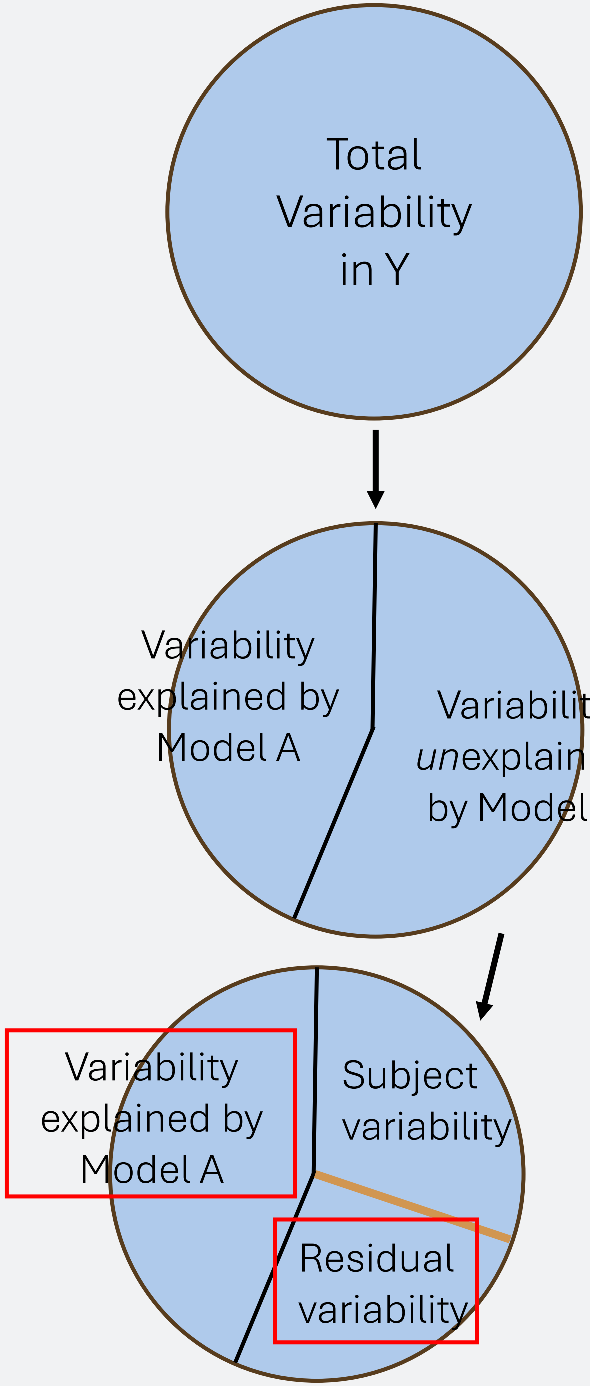
- For a within-subjects design, the F-statistic is:

$$F = \frac{variance\ explained\ by\ model\ A}{residual\ variance}$$

- Which is calculated using:

$$F = \frac{MS_{Reduced}}{MS_{Residual}} = \frac{SSR/df_{Reduced}}{SSE(Residual)/df_{Residual}}$$

- Which is more powerful: a between-subjects design or a within-subjects design?
  - Within-subjects because the denominator of the F-statistic will always be smaller



Example of Related Samples

- A researcher is interested in whether having someone present enhances people's performance on a task. To test this theory, the researcher has the same 8 participants complete a math exam with an experimenter present and again with the experimenter absent. The researcher measured the number of math problems each participant got correct in each condition.

Participant ID	Experimenter Absent	Experimenter Present
1	7	8
2	5	5
3	6	6
4	7	9
5	8	8
6	7	7
7	5	6
8	6	8
	$M_{Absent} = 6.375$	$M_{Present} = 7.125$

IV: Presence of Experimenter
- Experimenter Absent
- Experimenter Present

DV: Number of math problems correct

Contrast Coding the Categorical Predictor

IV: Presence of Experimenter

- Experimenter Absent
- Experimenter Present

	Absent	Present
ContrastCode (CC)	-1/2	+1/2

Model Comparison

Model Comparison:

Model A: $Y_i = \beta_0 + \beta_1 \text{ContrastCode}_i + \varepsilon_i$ PA = 2

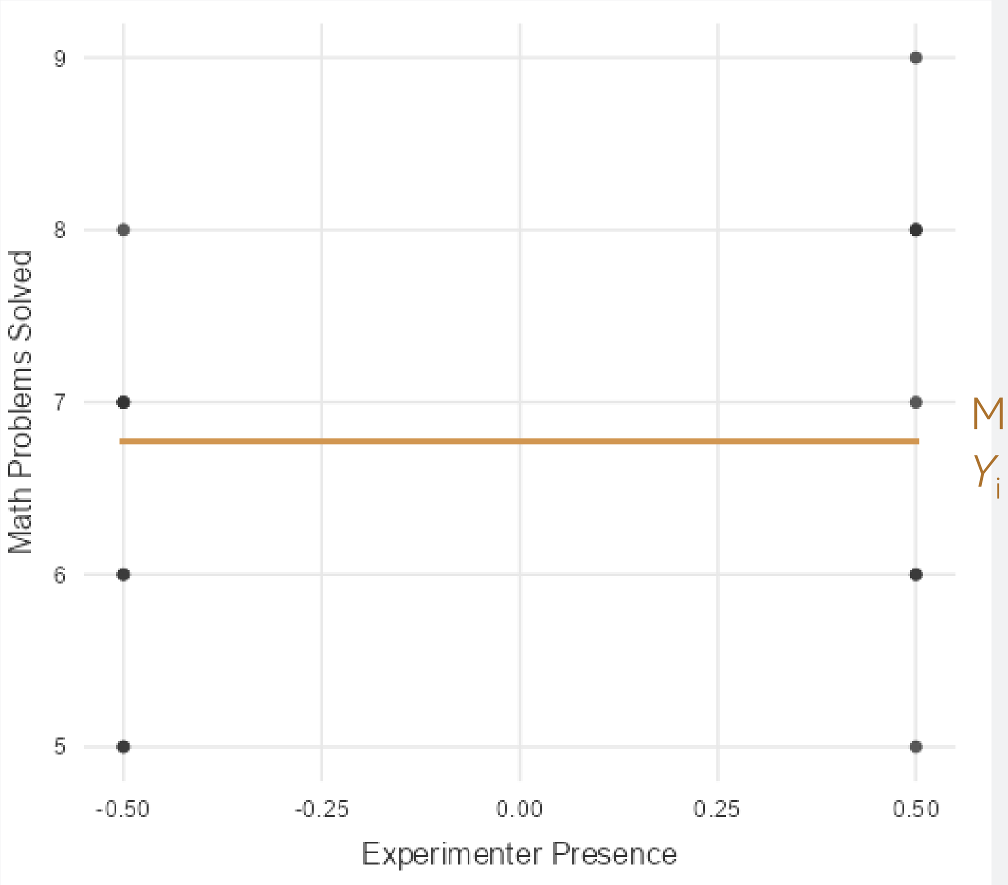
Model C: $Y_i = \beta_0 + \varepsilon_i$ PC = 1

Null & Alternative Hypotheses:

H_0: $\beta_1 = 0$

H_1: $\beta_1 \neq 0$

Estimating Model C



Model C:
$Y_i = 6.75$

<u>Model Comparison:</u>
Model A: $Y_i = \beta_0 + \beta_1 \text{ContrastCode}_i + \varepsilon_i$
Model C: $Y_i = \beta_0 + \varepsilon_i$

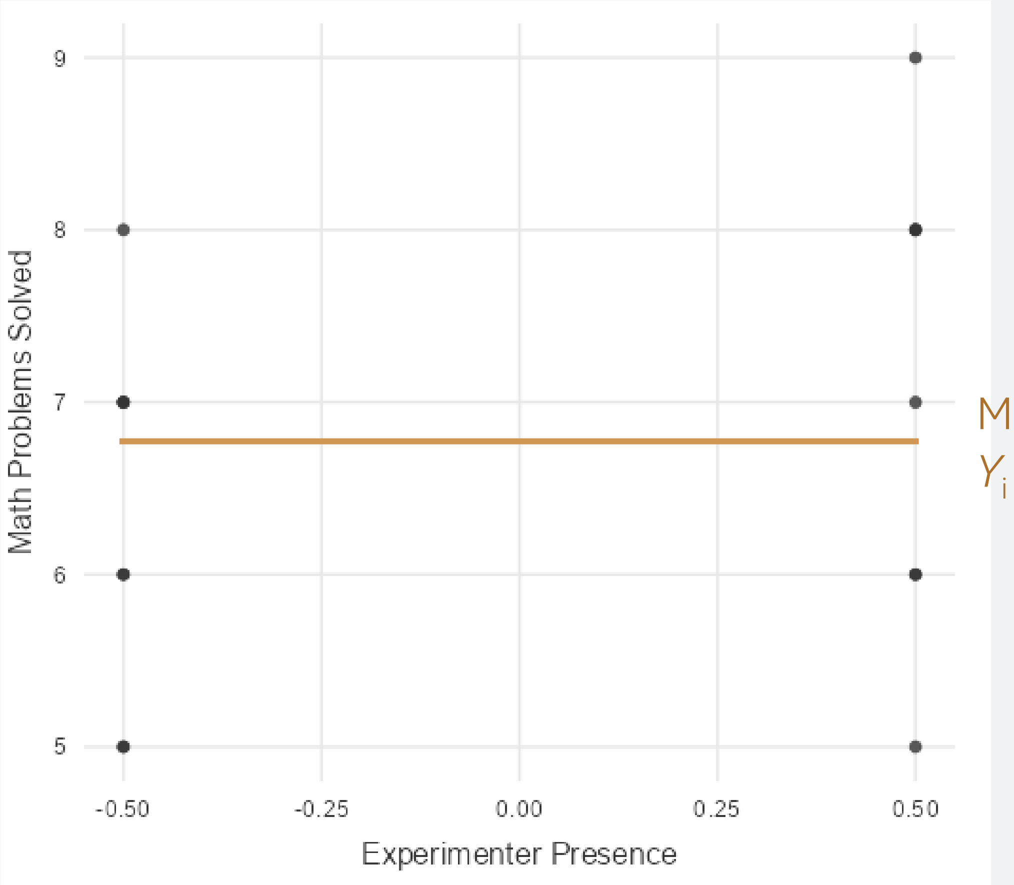
Model C: $Y_i = \beta_0 + \varepsilon_i$

- The best estimate of β_0 is the mean of Y
  - $b_0 = 6.75$

Estimate of Model C:

- $Y_i = 6.75 + e_i$

Calculating SSE(C)



Model C:
$Y_i = 6.75$

$$SSE(C) = \Sigma(Y_i - Y'_i)^2$$

ID	Y_i	$Y' = 6.75$	$(Y_i - 6.75)$	$(Y_i - 6.75)^2$
1	7	6.75	0.25	.0625
2	5	6.75	-1.75	3.0625
3	6	6.75	-0.75	.5625
4	7	6.75	0.25	.0625
5	8	6.75	1.25	1.5625
6	7	6.75	0.25	.0625
7	5	6.75	-1.75	3.0625
8	6	6.75	-0.75	.5625
1	8	6.75	1.25	1.5625
2	5	6.75	-1.75	3.0625
3	6	6.75	-0.75	.5625
4	9	6.75	2.25	5.0625
5	8	6.75	1.25	1.5625
6	7	6.75	0.25	.0625
7	6	6.75	-0.75	.5625
8	8	6.75	1.25	1.5625

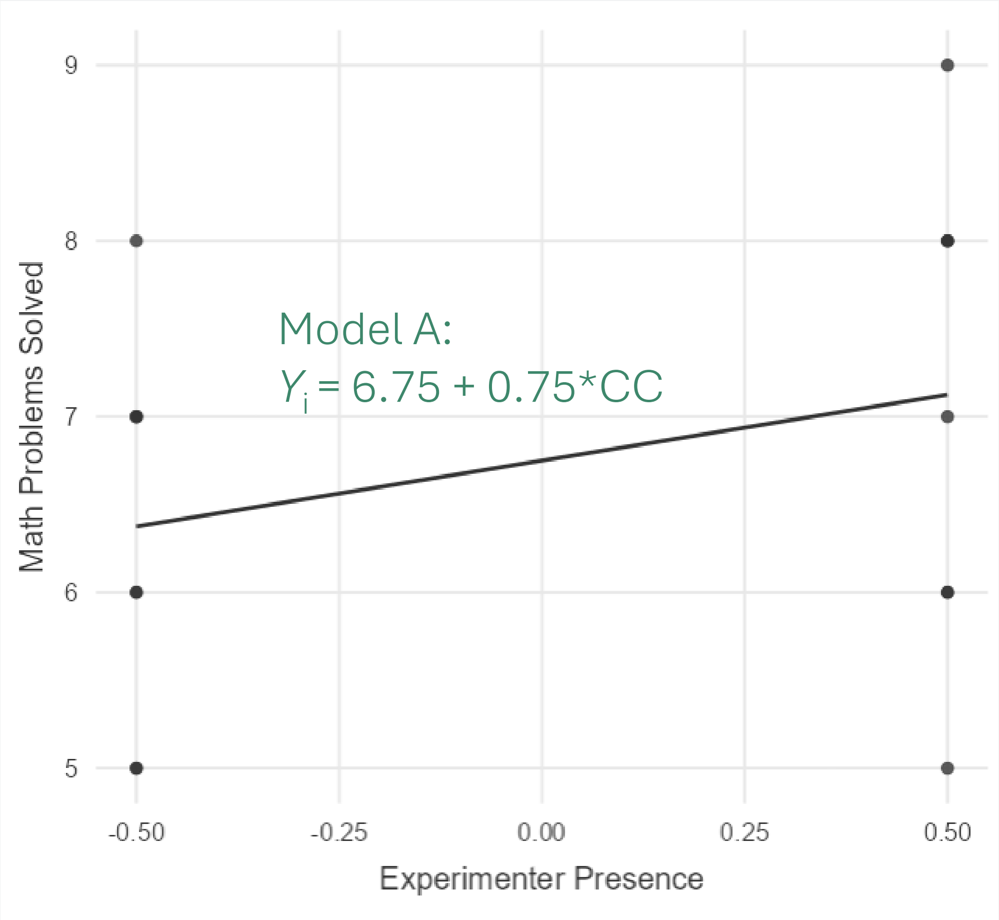
$\boxed{SSE(C) = 23}$

Model Comparison:

Model A: $Y_i = \beta_0 + \beta_1 ContrastCode_i + \varepsilon_i$

Model C: $Y_i = \beta_0 + \varepsilon_i$

Estimating Model A



	Absent	Present
CC	-1/2	+1/2

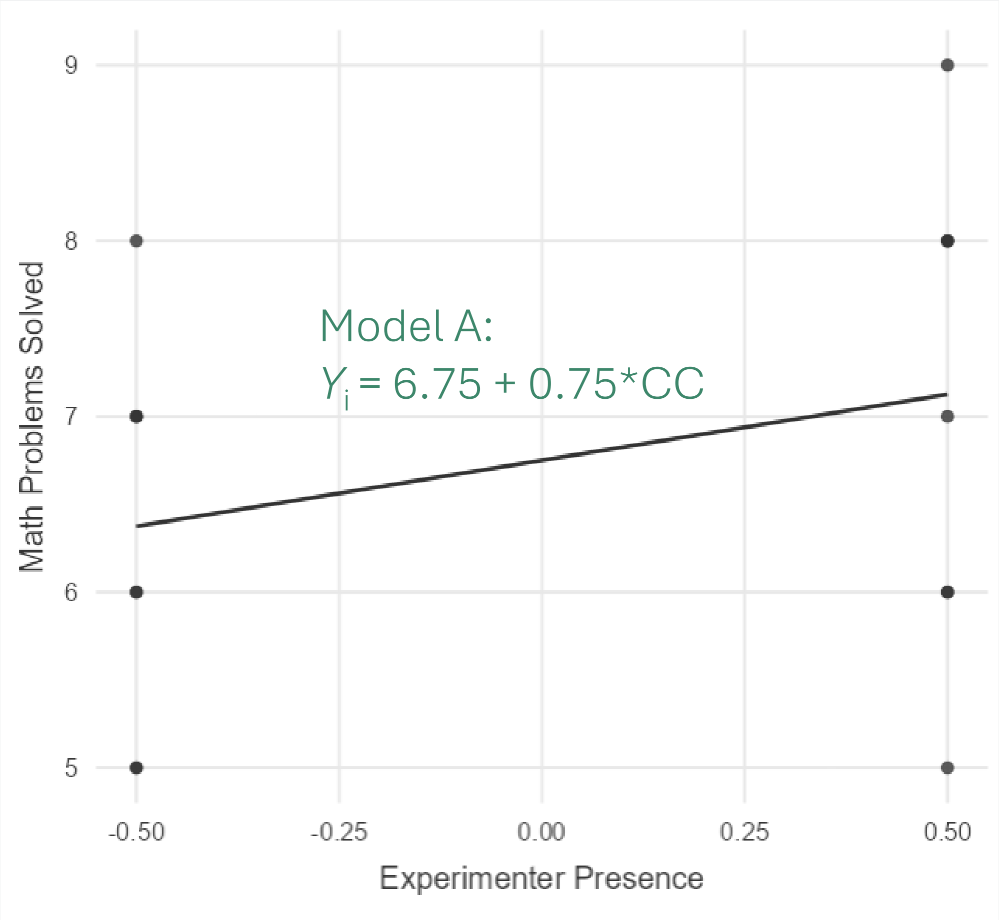
Model A: $Y_i = \beta_0 + \beta_1 CC + \varepsilon_i$

- The best-fitting line is the one that passes through the means of each group
  - b_0 is the value predicted by our model when X = 0
    - Because we used contrast codes, the y-intercept equals the mean of the group means, $b_0 = 6.75$
  - b_1 is the slope of our model (change in Y per 1-unit change in X)
    - Because we coded our predictor to space the groups 1-unit apart, the change in Y per 1-unit change in X is the difference between the group means
    - $b_1 = M_{Present} - M_{Absent} = 7.125 - 6.375 = 0.75$

Estimate of Model A:

- $Y_i = 6.75 + 0.75*CC + e_i$

Calculating SSE(A)



$SSE(A) = \Sigma(Y_i - Y'_i)^2$

	Absent	Present
CC	-1/2	+1/2

ID	Y_i	CC_i	Y' = 6.75 + 0.75*CC	$(Y_i - Y'_i)$	$(Y_i - Y'_i)^2$
1	7	-1/2	6.375	0.625	0.39
2	5	-1/2	6.375	-1.375	1.89
3	6	-1/2	6.375	-0.375	0.14
4	7	-1/2	6.375	0.625	0.39
5	8	-1/2	6.375	1.625	2.64
6	7	-1/2	6.375	0.625	0.39
7	5	-1/2	6.375	-1.375	1.89
8	6	-1/2	6.375	-0.375	0.14
1	8	1/2	7.125	0.875	0.77
2	5	1/2	7.125	-2.125	4.52
3	6	1/2	7.125	-1.125	1.27
4	9	1/2	7.125	1.875	3.52
5	8	1/2	7.125	0.875	0.77
6	7	1/2	7.125	-0.125	0.02
7	6	1/2	7.125	-1.125	1.27
8	8	1/2	7.125	0.875	0.77

$\boxed{SSE(A) = 20.75}$

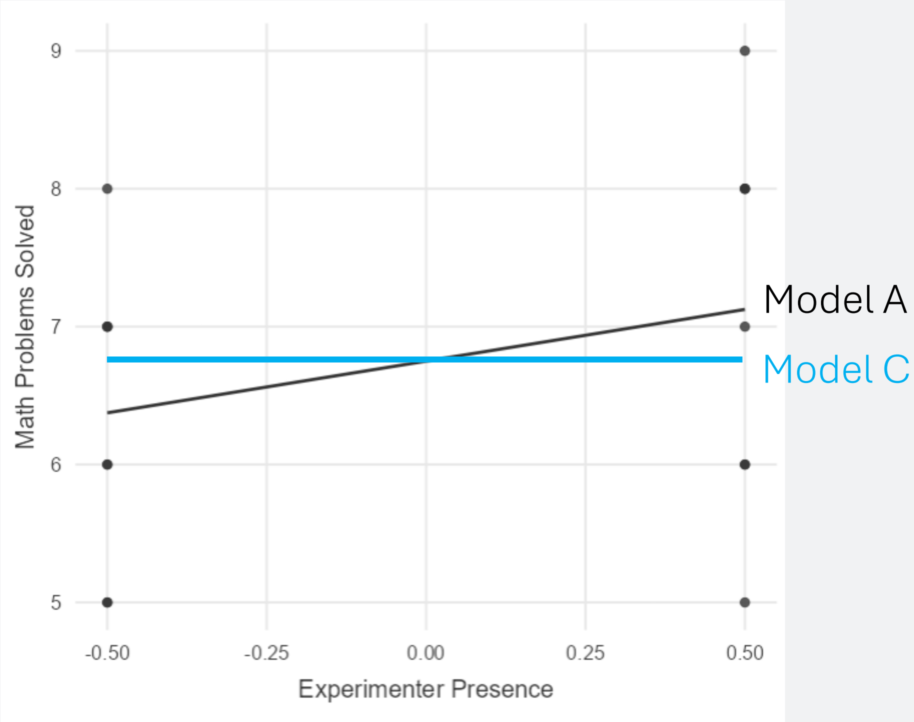
Variability Explained by Model A

Sum of Squares Reduced (SSR)

- The additional variability explained by Model A compared to Model C
  - SSR = SSE(C) – SSE(A)
  - SSR = 23 – 20.75
  - SSR = 2.25

Variance explained by Model A:

- $MS_{Reduced} = \frac{SSR}{df_{Reduced}}$, $df_{Reduced}$ = PA-PC
- $MS_{Reduced} = \frac{2.25}{1} = 2.25$



Subject Variability

- Subject variability
  - Variability due to individual differences (i.e., differences in the background characteristics of the participants)

- We can estimate subject variability because we can calculate a mean for each subject ($M_{Subject}$)
  - Across scenarios, how does each participant tend to differ from the other participants?

- $SSE(Subjects) = \Sigma m(M_{Subject} - M_{Grand})^2$
  - *m* is the number of levels of the IV

Participant ID	Experimenter Absent	Experimenter Present	$M_{Subject}$
1	7	8	7.5
2	5	5	5
3	6	6	6
4	7	9	8
5	8	8	8
6	7	7	7
7	5	6	5.5
8	6	8	7
	$M_{Absent} = 6.375$	$M_{Present} = 7.125$	$M_{Grand} = 6.75$

Subject Variability

$SSE(Subjects) = \Sigma m(M_{Subject} - M_{Grand})^2$

ID	$M_{Subject}$	M_{Grand}	$M_{Subject} - M_{Grand}$	$(M_{Subject} - M_{Grand})^2$	$m*(M_{Subject} - M_{Grand})^2$
1	7.5	6.75	0.75	0.5625	2*0.5625 = 1.125
2	5	6.75	-1.75	3.0625	6.125
3	6	6.75	-0.75	0.5625	1.125
4	8	6.75	1.25	1.5625	3.125
5	8	6.75	1.25	1.5625	3.125
6	7	6.75	0.25	0.0625	0.125
7	5.5	6.75	-1.25	1.5625	3.125
8	7	6.75	0.25	0.0625	0.125

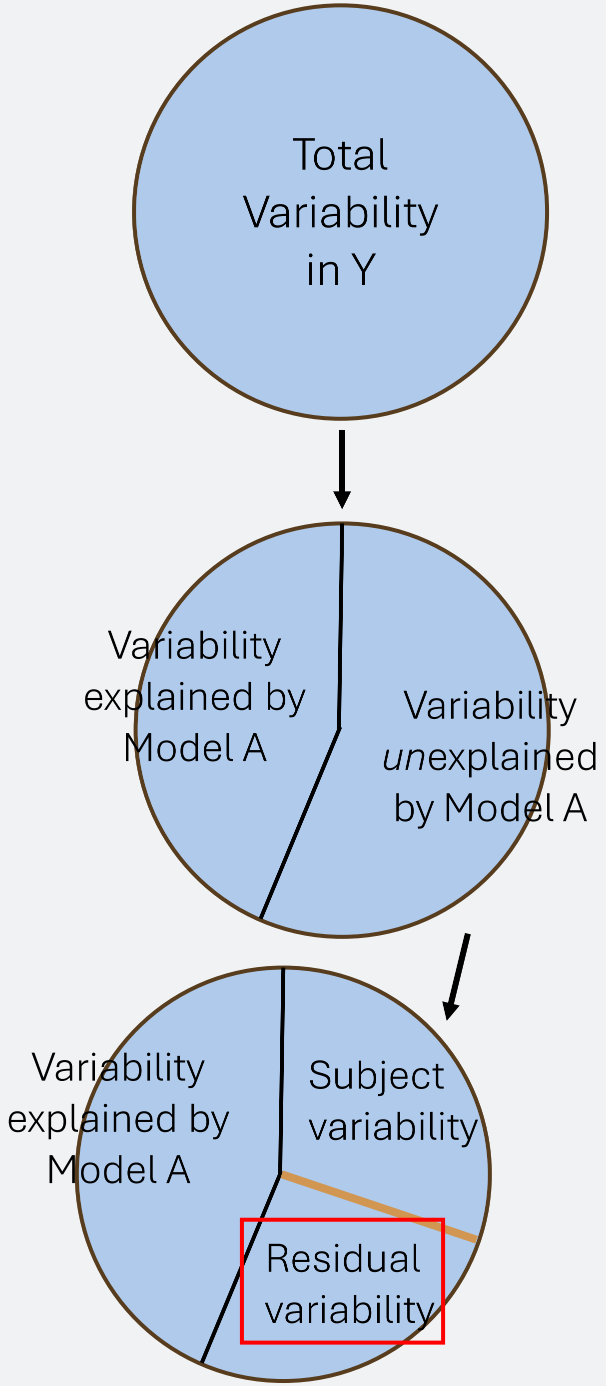
SSE(Subjects) = 18

Residual Variability

- Residual variability
  - The part of the unexplained variability that is left over after subtracting out subject variability

- SSE(Residual) = SSE(A) – SSE(Subjects)
  - SSE(A) = 20.75
  - SSE(Subjects) = 18

- SSE(Residual) = 20.78 – 18 = 2.75



Example: ANOVA Summary Table

	SS	*df*	*MS*	*F*	*p*
Reduced	2.25	PA - PC = 1	2.25	$F(1,7) = \frac{2.25}{0.3929} = 5.73$	Use R to obtain
Model A	20.75	N – PA = 16-2 = 14			
- Subjects	18	n – 1 = 8-1 = 7			
- Residual	2.75	$df_{Reduced} \times df_{Subjects} = 1*7 = 7$	0.3929		
Model C	23	N – PC = 16-1 = 15			

- N = total number of scores
- n = number of participants

ANOVA Summary Table

	SS	*df*	*MS*	*F*	*p*
Reduced	SSR = SSE(C) – SSE(A)	df_{Reduced} = PA - PC	MSR = $\frac{SSR}{df_{Reduced}}$	$F = \frac{MSR}{MS_{Residual}}$	Use R to obtain
Model A	SSE(A) = $\Sigma(Y_i - Y'_i)^2$	$df_{\text{ModelA}} = N$ - PA	$MS_{\text{ModelA}} = \frac{SSE(A)}{df_{ModelA}}$		
- Subjects	SSE(Subjects) = $\Sigma m(M_{\text{Subject}} - M_{\text{Grand}})^2$	$df_{\text{Subjects}} = n - 1$	$MS_{\text{Subjects}} = \frac{SSE(Subjects)}{df_{Subjects}}$		
- Residual	SSE(Residual) = SSE(A) – SSE(Subjects)	$df_{\text{Residual}} = df_{\text{Reduced}} \times df_{\text{Subjects}}$	$MS_{\text{Residual}} = \frac{SSE(Residual)}{df_{Residual}}$		
Model C	SSE(C) = $\Sigma(Y_i - Y'_i)^2$	$df_{\text{ModelC}} = N$ - PC			

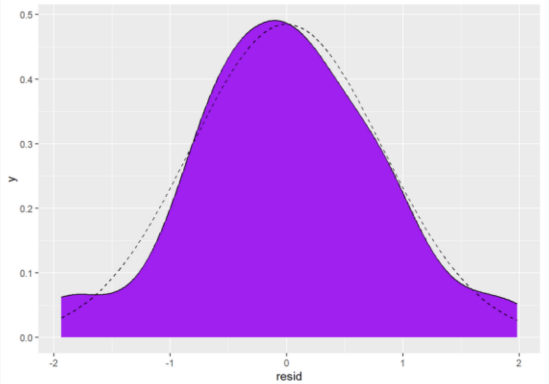
- N = total number of scores
- n = number of participants

Effect Size: PRE

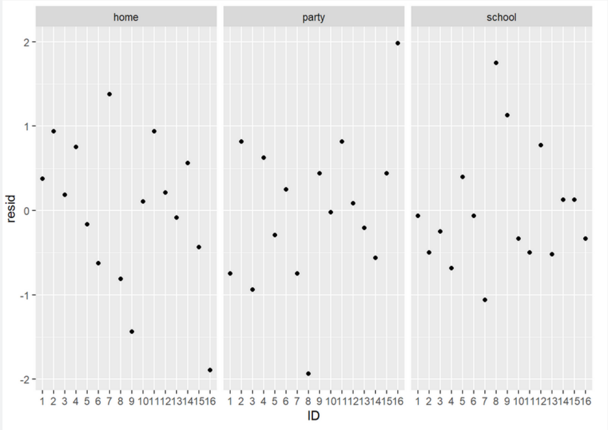
- PRE = SSR / SSE(C)
  - PRE = 2.25 / 23
  - PRE = 0.0978

- Model A accounts for approximately 9.78% for variability in math scores than Model C.

Checking the Test Assumptions



- After fitting a linear model, it's important to check whether one's test assumptions were met or violated



- The assumptions for a related samples analysis are:
  1. The errors (i.e., residuals) are normally distributed
  2. Independence of errors *within* conditions
  3. Sphericity assumption: the variance of the difference scores for all combinations of related groups are equal
     - Applicable when there's more than 2 groups

Absent	Present	Difference (Absent – Present)
7	8	-1
5	5	0
6	6	0
7	9	-2
8	8	0
7	7	0
5	6	-1
6	8	-2

```
sphericity_test$`Mauchly's Test for Sphericity`
```

```
##          Effect     W     p p<.05
## 1 environment 0.997 0.979
```

Types of Effect Size Measures

1. Proportion of variance accounted in dependent variable accounted for by a predictor variable
   - PRE, R^2, η^2

2. Size of the difference between two group means in standard deviation units
   - Cohen's *d*

- Additional considerations when calculating Cohen's *d* for a within-subjects design

Cohen's *d*

- If we pass our model to the `eff_size()` function like we used to calculate effect size for independent groups, it produces the following:
  - *d* = 1.20

```
> eff_size(means_for_effect, sigma = sigma(model), edf = df.residual(model))
 contrast          effect.size    SE   df lower.CL upper.CL
 Absent - Present         -1.2 0.556 9.09    -2.45   0.0603

sigma used for effect sizes: 0.6268
```

- Which was calculated using the formula:
  - $d = \frac{(M1 - M2)}{\sqrt{MS_{Residual}}} = \frac{(6.375 - 7.125)}{\sqrt{0.3929}} = 1.20$

	SS	*df*	*MS*	*F*	*p*
Reduced	2.25	PA - PC = 1	2.25	$F(1,7) = \frac{2.25}{0.3929} = 5.73$	Use R to obtain
Model A	20.75	*N* – PA = 16-2 = 14			
- Subjects	18	*n* – 1 = 8-1 = 7			
- Residual	2.75	$df_{Reduced} \times df_{Subjects} = 1*7 = 7$	0.3929		
Model C	23	*N* – PC = 16-1 = 15			

- What's a potential issue with calculating Cohen's *d* in this way for a within-subjects design?
  - **Overestimates** the size of the effect of the IV on the DV

Cohen's *d*

- An alternative way of calculating Cohen's *d* for a within-subjects design that better avoids the issue of overestimating the true effect size is:
  - $d = \frac{(M1 - M2)}{s_{Average}}$
    - Where $s_{Average} = \sqrt{\frac{s^2{}_1 + s^2{}_2}{2}}$
      - $s^2{}_1$ = Variance of group 1
      - $s^2{}_2$ = Variance of group 2

- $d = \frac{(M1 - M2)}{s_{Average}} = \frac{(6.375 - 7.125)}{\sqrt{\frac{1.125 + 1.839}{2}}} = 0.6161$

Participant ID	Experimenter Absent	Experimenter Present
1	7	8
2	5	5
3	6	6
4	7	9
5	8	8
6	7	7
7	5	6
8	6	8
	$M_{Absent} = 6.375$	$M_{Present} = 7.125$
	$s^2{}_{Absent} = 1.125$	$s^2{}_{Present} = 1.839$

Cohen's *d*

- Strengths of the alternative Cohen's *d* formula for calculating effect size with a within-subjects factor:

$$d = \frac{(M1 - M2)}{S_{Average}} = \frac{(M1 - M2)}{\sqrt{\frac{s^2{}_1 + s^2{}_2}{2}}}$$

- Better avoids the issue of overestimating the true effect size
- Much more similar to how effect size is calculated using independent groups
  - In fact, to get this Cohen's *d* for our model in R, run the model as a between-subjects design and use the `eff_size()` function

```
between_model <- lm(math_score ~ experimenter, data = data)

means <- emmeans(between_model, ~experimenter)

eff_size(means, sigma = sigma(between_model), edf =
df.residual(between_model))
```

```
 contrast          effect.size    SE df lower.CL upper.CL
 Absent - Present       -0.616 0.513 14    -1.72    0.485

sigma used for effect sizes: 1.217
Confidence level used: 0.95
```

Planned Comparisons vs Post-hoc Comparisons

- **Example:** Say a researcher wants to know whether people learn better when they are presented with visual information, auditory information, or tactile information. The researcher plans on having the same participants learn about the structure of molecules by 1) watching a presentation (visual), 2) listening to a presentation (auditory), and 3) putting together molecules using a hands-on building kit (tactile). After each learning method, the researcher will test how well the participants understand molecular structures.

- IV: Form in which information is presented
  - Visual
  - Auditory
  - Tactile
- DV: Understanding of the information

Planned Comparisons vs Post-hoc Comparisons

- When there are more than two groups being compared, this results in many possible comparisons between groups.
  - Visual vs Auditory
  - Visual vs Tactile
  - Auditory vs Tactile

- However, each individual significant test we perform to has a chance of making a Type I error
  - The more significance tests we conduct to make multiple comparisons, the higher the chances of making a Type I error
    - $1 – (1 - .05)^c$, where c is the number of comparisons being made
    - If $c = 3$: $1 – (1 - .05)^3 = 0.14$

- IV: Form in which information is presented
  - Visual
  - Auditory
  - Tactile

- DV: Understanding of the information

Planned Comparisons vs Post-hoc Comparisons

- **Planned comparisons** are theoretically motivated comparisons between groups that the researcher states they will analyze prior to conducting their analysis.
  - Preferred because we can specify them in our original set of contrast codes that we use to represent the categorical variable in our model

- **Ex:** Is there a significant difference in how well people learn when shown visual versus auditory information?
  - CC1: Visual = -1/2, Auditory = +1/2, Tactile = 0

- A single F-statistic is used to test the significance of the Model A – Model C comparison

- IV: Form in which information is presented
  - Visual
  - Auditory
  - Tactile
- DV: Understanding of the information

Planned Comparisons vs Post-hoc Comparisons

- **Post-hoc comparisons** are typically used to compare every combination of group means in order to ***explore*** whether there are unexpected significant differences between conditions.
  - Not necessarily theoretically motivated; performed after the data has been collected

- Because this results in *many significance tests* being conducted, post-hoc tests use different adjustment methods to prevent inflating our chances of making a Type I error.

- Post-hoc tests vary in how conservative (difficult to find significant differences) to liberal (easy to find significant differences) they are

- IV: Form in which information is presented
  - Visual
  - Auditory
  - Tactile

- DV: Understanding of the information

Post-Hoc Comparisons

- Bonferroni post-hoc test:
  - Sets new alpha level using α/c, where c is the number of pairwise comparisons being made
  - Conservative especially as number of comparisons increases

- Tukey's HSD post-hoc test:
  - Calculate HSD = $q * \sqrt{\frac{MS_{Residual}}{n}}$, where *n* is number of participants
    - HSD is the amount by which any two sample means must differ in order to be considered significantly different
  - Used for making *all* pairwise comparisons among group means; good balance on the conservative-liberal scale

- Sidak post-hoc test:
  - Adjusts alpha using $1 - (1 - \alpha)^{1/n}$, where *n* is the number of comparisons being made
  - Used when not *all* pairwise comparisons are being made; Conservative, but not as conservative as Bonferroni

A Priori Power Analysis

- **Example:** Let's go back to our example in which a researcher wanted to examine whether the format in which information is presented (visual, auditory, tactile) influences how well people understand the material. The researcher plans on using a within-subjects design by having the *same* participants receive each level of the IV's manipulation. What sample size does the researcher need to collect to have an 80% chance of detecting a small effect?

- To perform an a priori power analysis for related samples, we can either use:
  - https://pwrss.shinyapps.io/index/
  - pwrss.f.rmanova( ) function in R

- IV: Form in which information is presented
  - Visual
  - Auditory
  - Tactile
- DV: Understanding of the information

A Priori Power Analysis

- A priori power analysis using https://pwrss.shinyapps.io/index/
  - Choose Repeated Measures ANOVA
  - Choose Within
  - Request: Sample size
  - Eta-square: Fill in your best estimate of the effect size (hovering shows you the conventions for the size of eta-squared)
    - For our example, let's calculate the sample size needed to detect a small effect, eta-square = 0.26
  - Number of levels (groups): Number of groups of unrelated participants
  - Number of repeated measures: Number of measurements taken from the group of participants
  - Type I error rate: default is 0.05
  - Correlation between repeated measures: default is 0.05
  - Non-sphericity correction factor: When sphericity assumption has not been violated, use 1

- IV: Form in which information is presented
  - Visual
  - Auditory
  - Tactile
- DV: Understanding of the information

A Priori Power Analysis

- A priori power analysis using pwrss.f.rmanova( ) function in R
  - eta2 = estimated effect size
  - n.levels = number of unrelated groups
  - n.rm = number of repeated measurements
  - power = desired power level (default is 0.80)
  - alpha = willingness to make Type I error (default is 0.05)
  - corr.rm = correlation between repeated measures (default is 0.50)
  - type = “within”
    - We want to examine the effect related to the repeated measures of our within-subjects factor

```
pwrss.f.rmanova(eta2 = 0.26,  n.levels = 1, n.rm = 3,
                power = 0.80, alpha = 0.05,
                corr.rm = 0.50, type = "within")
```

```
One-way Repeated Measures
Analysis of Variance (F test)
H0: eta2 = 0 (or f2 = 0)
HA: eta2 > 0 (or f2 > 0)
------------------------------
Number of levels (groups) = 1
Number of repeated measurements = 3
------------------------------
 Statistical power = 0.8
 Total n = 7
------------------------------
Type of the effect = “within”
Numerator degrees of freedom = 2
Denominator degrees of freedom = 10.379
Non-centrality parameter = 13.048
Type I error rate = 0.05
Type II error rate = 0.2
```

ANOVA Summary Table
for Related Samples with 2+ Samples

	SS	*df*	*MS*	*F*	*p*
Reduced	SSR = SSE(C) – SSE(A)	$df_{Reduced}$ = PA - PC	MSR = $\frac{SSR}{df_{Reduced}}$	$F = \frac{MSR}{MS_{Residual}}$	Use R to obtain
Model A	SSE(A) = $\Sigma(Y_i - Y'_i)^2$	$df_{ModelA} = N$ - PA			
- Subjects	SSE(Subjects) = $\Sigma m(M_{Subject} - M_{Grand})^2$	$df_{Subjects} = n - 1$			
- Residual	SSE(Residual) = SSE(A) – SSE(Subjects)	$df_{Residual} = df_{ModelA} - df_{Subjects}$	$MS_{Residual} = \frac{SSE(Residual)}{df_{Residual}}$		
Model C	SSE(C) = $\Sigma(Y_i - Y'_i)^2$	$df_{ModelC} = N$ - PC			