Review of General Linear Model Concepts

Categorical predictors must be numerically coded

- Number of codes needed to fully represent a categorical predictor in the model:
  - $m - 1$, where m is the number of levels of the categorical predictor
- Coding options:
  - Dummy Coding
  - Contrast Coding
    - Recommended because it ensures the predictors in the model are **orthogonal/independent** of each other

Rules of Contrast Coding

Rules of Contrast Coding:

1. Each set of codes must sum to 0.
2. The sum of the products of codes in corresponding positions must equal 0.

- **Recommended for ease of interpretation**:

Put each contrast code on a scale of “1”, meaning that the span between the contrast codes is equal to 1.

The Model Comparison

- Model A:
  - The full model including all predictors (aka, all the codes created to represent the categorical predictor in the model)
- Model C:
  - A compact version of Model A that excludes ***only the predictor that it is you are wanting to test the significance of***
- The Model A – Model C comparison tests whether the predictor added to Model A makes a significant improvement compared to Model C
  - i.e., whether including the predictor accounts for significantly more variance in the DV compared to a model without the predictor

Example

1. Was there a significant difference in intentions to behave pro-environmentally between participants in the descriptive norm versus social norm condition?

	Control	Descriptive Norm	Social Norm
ContrastCode1	0	-1/2	1/2
ContrastCode2	2/3	-1/3	-1/3

Model A: $Y_i = \beta_0 + \beta_1 ContrastCode1_i + \beta_2 ContrastCode2_i + \varepsilon_i$

Model C: $Y_i = \beta_0 + \beta_2 ContrastCode2_i + \varepsilon_i$

Null Hypothesis:
$H_0: \beta_1 = 0$

Parameters per model:
PA = 3
PC = 2

Example

2. Was norm condition *overall* a significant predictor of intentions to behave pro-environmentally?
   - When testing the **overall significance of a categorical predictor**, we want to test whether *the full set of codes* used to represent the predictor are significant as a whole

	Control	Descriptive Norm	Social Norm
ContrastCode1	0	-1/2	1/2
ContrastCode2	2/3	-1/3	-1/3

Model A: $Y_i = \beta_0 + \beta_1 \text{ContrastCode1}_i + \beta_2 \text{ContrastCode2}_i + \varepsilon_i$

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

Null Hypothesis:
$H_0: \beta_1 = \beta_2 = 0$

Parameters per model:
PA = 3
PC = 1

Which model is better?

- To conclude which model, Model A or Model C, performs better, we compare how much remaining error there is after using each model to predict scores on the outcome variable

- $SSE(C) = \Sigma(Y_i - Y'_i)^2$
  - Y_i = Participants' scores on the outcome variable, Y
  - Y'_i = The score predicted for each participant on the outcome variable by Model C

- Can calculate the variance (average squared error) of Model C by dividing its SS by its df
  - df_{ModelC} = n-PC
  - $MS_{ModelC} = SSE(C)/df_{ModelC}$

Which model is better?

- To conclude which model, Model A or Model C, performs better, we compare how much remaining error there is after using each model to predict scores on the outcome variable

- SSE(A) = $\Sigma(Y_i - Y'_i)^2$
  - Y_i = Participants' scores on the outcome variable, Y
  - Y'_i = The score predicted for each participant on the outcome variable by Model A

- Can calculate the variance (average squared error) of Model A by dividing its SS by its df
  - df_{ModelA} = n-PA
  - MS_{ModelA}= SSE(A)/df_{ModelA}

Which model is better?

- Then, we calculate SSR to see how much better Model A does compared to Model C at leaving *less* unexplained error leftover

- SSR = SSE(C) – SSE(A)
  - The difference in the SSE (unexplained error remaining) between Model C and Model A
  - A smaller SSE is better – it means the model did better at predicting the data!

- Can calculate the variance accounted for by Model A (compared to Model C) by dividing the SSR by its df
  - $df_{Reduced}$ = PA-PC
  - $MS_{Reduced} = SSR/df_{Reduced}$

F-statistic

- The test statistic most often reported for testing the significance of Model A versus Model C is an *F*-statistic

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

$$F = \frac{MS_{Reduced}}{MS_{ModelA}}$$

- MS_{ModelA} also commonly called *MSE*

Anova Summary Table

Source	*SS*	*df*	*MS*	*F*	*PRE*	*p*
Reduced	SSR	$df_{Reduced}$ = PA-PC	MSR = SSR/df_{Reduce}	$F = \frac{MSR}{MSE}$	$\frac{SSR}{SSE(C)}$	Use R to obtain
Model A (Error)	SSE(A)	df_{ModelA} = n-PA	MSE = SSE(A)/df_{ModelA}			
Model C (Total)	SSE(C)	df_{ModelC} = n-PC				

What does a model mean?

The general formula for a population-level linear model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_j X_{ij} + \varepsilon_i$$

Estimate of the general linear model from our sample data:

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + \ldots + b_j X_{ij} + e_i$$

- A linear model is a geometric shape that we're fitting to the data
    - With one predictor, we're fitting a line
    - With two predictors, we're fitting a plane
    - With each additional predictor, add another dimension to the geometric shape you're fitting to the data

With one predictor

The general formula for a population-level linear model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$

Estimate of the general linear model from our sample data:

$$Y_i = b_0 + b_1 X_{i1} + e_i$$

- With one predictor:
  - b_0 = the y-intercept of the line
  - b_1 = the slope of the line corresponding to the effect of the predictor, X_1

With two predictors

The general formula for a population-level linear model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

Estimate of the general linear model from our sample data:

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + e_i$$

- With two predictors:
  - b_0 = the y-intercept of the plane
  - b_1 = the slope contribution corresponding to the effect of predictor 1, X_1
  - b_2 = the slope contribution corresponding to the effect of predictor 2, X_2

Effect Sizes

- Lakens (2013): Calculating and reporting effect sizes to facilitate cumulative science

> **DIFFERENCES AND SIMILARITIES BETWEEN EFFECT SIZES**
> As Poincaré (1952, p. 34) has said: "mathematics is the art of giving the same name to different things." Unfortunately, in the domain of effect size calculations statisticians have failed Poincare. Effect sizes have either different names although they are basically the same entity (such as referring to r^2 as η^2), or they

The amount of variance accounted for in the DV by our model (R^2):

- $R^2 = \dfrac{SSR}{SS_{Total}}$
  - R^2 Also called eta-squared (η^2)
  - SS_{Total} is the total SS in Y
    - $SS_{\text{Total}} = \Sigma(Y_i - \bar{Y})^2$
    - $\bar{Y}$ is the mean of Y

The additional proportion of variance accounted for by Model A compared to Model C (*PRE*):

- $PRE = \dfrac{SSR}{SSE(C)}$
  - *PRE* is equal to R^2 when Model C is a one parameter model: $Y_i = b_0 + e_i$
    - because $SSE(C) = \Sigma(Y_i - Y')^2$ and Y', the model predicted by our model is b_0, which is the mean of Y ($\bar{Y}$)