General Linear Models with Interactions: Testing Moderation!

- Topics:
  - Slopes of predictors within interactions: from unique main (marginal) effects to unique simple (conditional) effects
    - The 4 possible kinds of interactions
    - Model-implied slopes as linear combinations of model slopes
    - Regions of significance for when simple slopes "turn on or off"
    - Interactions with categorical predictors
    - Interactions with quantitative predictors with nonlinear effects
  - Special uses of interaction terms to create nested effects
    - "ANOVA with a hole in it"
    - Missing (or impossible) predictor data

GLM with an Interaction: $y_i = \beta_0 + \beta_1(x_i) + \beta_2(z_i) + \boxed{\beta_3(x_i)(z_i)} + e_i$

- **Interaction slopes (β_3 here) test "Moderation":** whether a predictor's slope **depends** on the value of an interacting predictor
  - Either predictor can be "the moderator" (is interpretive distinction only)
- **Interactions can always be evaluated** for any combination of categorical and quantitative predictors, although traditionally…
  - **In "ANOVA":** By default, all possible interactions are estimated
    - Oddly, nonsignificant interactions are usually kept in the model (even if only significant interactions are interpreted)
  - **In "ANCOVA":** Quantitative predictors ("covariates") are not included in interaction terms → this is the "homogeneity of regression assumption"
    - But you don't have to assume this—it is always a testable hypothesis!
  - **In "Regression":** No default—effects of predictors are as you specify
    - Requires most thought, but gets annoying in regression-specific programs when you have to manually create the interaction variable:
    - e.g., XZinteraction = X * Z;

Interaction variables are made on the fly in GLM! ☺

Main Effects of Predictors within Interactions

- **"Main effect" slopes** of predictors that are included in interaction terms should always remain in the model regardless of their significance
  - e.g., given $\boldsymbol{\beta_3}(\boldsymbol{x_i})(\boldsymbol{z_i})$, you must keep $\boldsymbol{\beta_1}(\boldsymbol{x_i})$ and $\boldsymbol{\beta_2}(\boldsymbol{z_i})$ in the model, too
  - Why? Because an interaction term creates an over-additive (enhancing) or under-additive (dampening) effect, so *what it is additive to* must be included for the interaction to correctly represent an "interaction"

- **Role of a two-way interaction is to adjust the "main effect" slopes of the two predictors involved… (in one of four possible ways)**
  - But the idea of a "marginal" main effect slope (that holds for everyone) no longer applies: the main effect slopes become ***simple main effect slopes*** that are ***conditional*** each interacting predictor = 0

- Note that this is a different type of conditionality than just "holding the other predictors constant" (which means constant at **any value**)
  - Simple main effect slopes are held constant (conditional on) the **0 value** of the interacting predictor(s)—these slopes would be different if 0 were defined differently by centering the interacting predictor elsewhere
  - This language can be confusing, so next is a **taxonomy** that may help…

A Taxonomy of Fixed Effect Interpretations

- In the most common statistical models, fixed effects will be either:
  - an **intercept** that provides an expected (conditional) $\boldsymbol{y_i}$ outcome,
  - or **a slope** for the difference in $\boldsymbol{y_i}$ per unit difference in $\boldsymbol{x_i}$ predictor

- **All slopes** can be described as falling within one of three categories: ***bivariate marginal***, ***unique marginal***, or ***unique conditional***
  - In models with only **one fixed slope**, that slope's main effect is ***bivariate marginal*** (is uncontrolled and applies across all persons)
  - In models with **more than one fixed slope**, each slope's main effect is ***unique*** (it controls for the overlap in contribution with each other slope)
    - If a predictor is not part of an interaction term, its ***unique effect is marginal*** (it controls for the other slopes, but its effect still applies across all persons)
    - NEW → If a predictor is part of one or more interaction terms, its ***unique effect is conditional,*** which means it is **specific to each interacting predictor = 0**
      - **Unique conditional** effects are also called **"simple main effects" (simple slopes)**

Practice Labeling Fixed Slopes—Choices:

bivariate marginal, unique marginal, or unique conditional

Model: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{w_i}) + \boldsymbol{e_i}$

- Label for $\boldsymbol{\beta_1}$ slope of $\boldsymbol{w_i}$ =

Model: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{w_i}) + \boldsymbol{\beta_2}(\boldsymbol{x_i}) + \boldsymbol{\beta_3}(\boldsymbol{z_i}) + \boldsymbol{\beta_4}(\boldsymbol{x_i})(\boldsymbol{z_i}) + \boldsymbol{e_i}$

- Label for $\boldsymbol{\beta_1}$ slope of $\boldsymbol{w_i}$ =
- Label for $\boldsymbol{\beta_2}$ slope of $\boldsymbol{x_i}$ =
- Label for $\boldsymbol{\beta_3}$ slope of $\boldsymbol{z_i}$ =
- Label for $\boldsymbol{\beta_4}$ slope of $\boldsymbol{x_i z_i}$ interaction term=

The 4 Possible Kinds of Interactions

- There are only 4 kinds of interactions: they make each of their main effect slopes more/less positive/negative
  - **More** positive or more negative → effect becomes **stronger**, known as "over-additive" interaction
  - **Less** positive or less negative → effect becomes **weaker**, known as "under-additive" interaction

- Model: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{w_i}) + \boldsymbol{\beta_2}(\boldsymbol{x_i}) + \boldsymbol{\beta_3}(\boldsymbol{z_i}) + \boldsymbol{\beta_4}(\boldsymbol{x_i})(\boldsymbol{z_i}) + \boldsymbol{e_i}$

Slope of x_i is β_2=	**Interaction Slope is β_4=**	**So β_4 makes effect of x_i ??? per unit higher z_i**
10	2	
10	-2	
-10	-2	
-10	2	

Fixed Effects: Why Centering Matters

- y_i = Student achievement (GPA as percentage out of 100)
  x_i = Parent **attitudes** about education (measured on 1–5 scale)
  z_i = Parent **education** level (measured in years of education)

$$GPA_i = \beta_0 + \beta_1(Att_i) + \beta_2(Ed_i) + \beta_3(Att_i)(Ed_i) + e_i$$
$$GPA_i = 30 + 1(Att_i) + 2(Ed_i) + 0.5(Att_i)(Ed_i) + e_i$$

- **Interpret β_0:**
- **Interpret β_1:**
- **Interpret β_2:**
- **Interpret β_3: Attitude as Moderator:**

  Education as Moderator:

- **Predicted GPA** for **attitude = 3** and **Ed = 12**?

  75 = 30 + 1*(3) + 2*(12) + 0.5*(3)*(12)

How Centering Changes Fixed Effects

- $\boldsymbol{y_i}$ = Student achievement (GPA as percentage out of 100)
  $\boldsymbol{x_i}$ = Parent **attitudes** about education (now centered at **3**)
  $\boldsymbol{z_i}$ = Parent years of **education** (now centered at **12**)

$$GPA_i = \beta_0 + \beta_1(Att_i - 3) + \beta_2(Ed_i - 12) + \beta_3(Att_i - 3)(Ed_i - 12) + e_i$$
$$GPA_i = 75 + 7(Att_i - 3) + 3.5(Ed_i - 12) + 0.5(Att_i - 3)(Ed_i - 12) + e_i$$

- **Interpret $\boldsymbol{\beta_0}$:**
- **Interpret $\boldsymbol{\beta_1}$:**
- **Interpret $\boldsymbol{\beta_2}$:**
- **Interpret $\boldsymbol{\beta_3}$: Attitude as Moderator:**

  Education as Moderator:

- But how did I know what the new fixed effects would be???

Model-Implied Predicted Outcomes

- **Predicted outcomes = expected outcomes = conditional means**
  - ALL model effects must be included (or else are assumed = 0)

$$\widehat{GPA}_i = \beta_0 + \beta_1(Att_i - 3) + \beta_2(Ed_i - 12) + \beta_3(Att_i - 3)(Ed_i - 12)$$

STATA: Each line starts with lincom, labels given in comments after //

```
lincom _cons*1 + att*__ + ed*__ + att#ed*__ // Yhat: Att=5 Ed=16
lincom _cons*1 + att*__ + ed*__ + att#ed*__ // Yhat: Att=1 Ed=9
lincom _cons*1 + att*__ + ed*__ + att#ed*__ // Yhat: Att=3 Ed=20
```

R: Values are multipliers in GLHT *in order of the fixed effects output*:

```
glhtName = glht(model=ModelName, linfct=rbind(
"Yhat: Att=5 Ed=16" c(1,__,__,__),
"Yhat: Att=1 Ed=9"  c(1,__,__,__),
"Yhat: Att=3 Ed=20" c(1,__,__,__)))
summary(glhtName, test=adjusted("none"))
```

Model-Implied Predictor Simple Slopes

- Example equation for predicted GPA using centered predictors:
  $\widehat{GPA}_i = \beta_0 + \beta_1(Att_i - 3) + \beta_2(Ed_i - 12) + \beta_3(Att_i - 3)(Ed_i - 12)$

- This model equation provides predictions for:
  - Expected outcome given any combination of predictor values
  - Any conditional (simple) main effect slopes implied by interaction term
  - **Any slope can be found as: what it is + what *modifies* it**

- Three steps to get any model-implied simple main effect slope:

1. **Identify** all terms in model involving the predictor of interest
2. **Factor out** common predictor variable to find slope linear combination
3. **Calculate** estimate and SE for slope linear combination
   - *By "calculate" of course I mean "ask the program to do this for you"*

Model-Implied Predictor Simple Slopes

- Example equation for <u>predicted GPA</u> using centered predictors:
$$\widehat{GPA}_i = \beta_0 + \beta_1(Att_i - 3) + \beta_2(Ed_i - 12) + \beta_3(Att_i - 3)(Ed_i - 12)$$

1. **Identify** all slopes in model involving the predictor of interest

   To get attitudes slope: $Est = \beta_1(Att_i - 3) + \beta_3(Att_i - 3)(Ed_i - 12)$

   To get education slope: $Est = \beta_2(Ed_i - 12) + \beta_3(Att_i - 3)(Ed_i - 12)$

2. **Factor out** predictor of interest to find <u>slope linear combination</u>

   To get attitudes slope: $Est = [\beta_1 + \beta_3\,(Ed_i - 12)]$ **that will multiply** $(Att_i - 3)$

   To get education slope: $Est = [\beta_2 + \beta_3(Att_i - 3)]$ **that will multiply** $(Ed_i - 12)$

- Btw, the SEs for the new slopes provided by the program come from:
  - SE^2 = sampling variance of slope estimate → e.g., $Var(\beta_1) = SE^2_{\beta_1}$

  attitudes slope: $SE^2 = Var(\beta_1) + Var(\beta_3)(Ed_i - 12) + 2Cov(\beta_1, \beta_3)(Ed_i - 12)$

  education slope: $SE^2 = Var(\beta_2) + Var(\beta_3)(Att_i - 3) + 2Cov(\beta_2, \beta_3)(Att_i - 3)$

Model-Implied Predictor Simple Slopes

- **To request <u>predicted simple slopes</u> (= simple main effects):**
  - **Include ONLY** the fixed effects that contain the predictor of interest

$$\widehat{GPA}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{Att_i} - \mathbf{3}) + \boldsymbol{\beta_2}(\boldsymbol{Ed_i} - \mathbf{12}) + \boldsymbol{\beta_3}(\boldsymbol{Att_i} - \mathbf{3})(\boldsymbol{Ed_i} - \mathbf{12})$$

→ attitudes slope: $\boldsymbol{Est} = [\boldsymbol{\beta_1} + \boldsymbol{\beta_3}(\boldsymbol{Ed_i} - \mathbf{12})]$ that multiplies $(\boldsymbol{Att_i} - \mathbf{3})$
→ education slope: $\boldsymbol{Est} = [\boldsymbol{\beta_2} + \boldsymbol{\beta_3}(\boldsymbol{Att_i} - \mathbf{3})]$ that multiplies $(\boldsymbol{Ed_i} - \mathbf{12})$

STATA: Each line starts with lincom, title in comments after //

```
lincom _cons*0 + att*__  + ed*__  + att#ed*__   // Att Slope if Ed=10
lincom _cons*0 + att*__  + ed*__  + att#ed*__   // Att Slope if Ed=18
lincom _cons*0 + att*__  + ed*__  + att#ed*__   // Ed Slope  if Att=2
lincom _cons*0 + att*__  + ed*__  + att#ed*__   // Ed Slope  if Att=5
```

R: Values are multipliers in GLHT in order of fixed effects:

```
glhtName = glht(model=ModelName, linfct=rbind(

"Att Slope if Ed=10" c(0,__,__,__),
"Att Slope if Ed=18" c(0,__,__,__),
"Ed Slope  if Att=2" c(0,__,__,__),
"Ed Slope  if Att=5" c(0,__,__,__)))
```

Regions of Significance for Simple Slopes

- For quantitative predictors, there may not be specific values of the moderator at which you want to know the slope's significance…

- For example, with age*woman (in which 0=man, 1=woman here):

$$\hat{y}_i = \beta_0 + \beta_1(Age_i - 85) + \beta_2(Woman_i) + \beta_3(Age_i - 85)(Woman_i)$$

→ **age slope:** $Est =$ ____________________ that multiplies $(Age_i - 85)$

→ **gender slope:** $Est =$ ____________________ that multiplies $(Woman_i)$

- Age slopes are only relevant for two specific values of *woman*:

```
lincom age85*__ woman*__ age85*woman*__   // Age Slope for Men
lincom age85*__ woman*__ age85*woman*__   // Age Slope for Women
```

- But there are many ages to request gender differences for…

```
lincom age85*__ woman*__ age85*woman*__   // Gender Diff at Age=80
lincom age85*__ woman*__ age85*woman*__   // Gender Diff at Age=90
```

Regions of Significance for Simple Slopes

- An alternative approach for continuous moderators is known as **regions of significance** (see Hoffman 2015; chapter 2; Finsaas & Goldstein, 2021)
- Rather than asking if the simple main effect of gender is still significant at an arbitrary age, we can find the **boundary ages** at which the gender slope becomes non-significant
- We know that: EST / SE = *t*-value → if |*t*| > |1.96|, then *p* < .05
- So we work backwards to find the EST and SE such that:

$$\pm \mathrm{t} = \pm 1.96 = \frac{\text{Slope Estimate}}{\sqrt{\text{Variance of Slope Estimate}}}\text{, where:}$$

$$\text{Gender Slope (Gender Difference) Estimate} = \beta_2 + \beta_3\left(\text{Age} - 85\right)$$

$$\text{Variance of Slope Estimate} = \text{Var}\left(\beta_2\right) + \boxed{2\text{Cov}\left(\beta_2\beta_3\right)\left(\text{Age} - 85\right)} + \text{Var}\left(\beta_3\right)\left(\text{Age} - 85\right)^2$$

- Need to request "asymptotic covariance matrix" (COVB)
  - Covariance matrix of fixed effect *estimates* (SE^2 is on the diagonal)

Regions of Significance for Simple Slopes

$$\pm t = \pm 1.96 = \frac{\text{Slope Estimate}}{\sqrt{\text{Variance of Slope Estimate}}}, \text{ where:}$$

$$\text{Gender Slope (Gender Difference) Estimate} = \beta_2 + \beta_3(\text{Age} - 85)$$

$$\text{Variance of Slope Estimate} = \text{Var}(\beta_2) + 2\text{Cov}(\beta_2\beta_3)(\text{Age} - 85) + \text{Var}(\beta_3)(\text{Age} - 85)^2$$

- For example, age*woman (0=man, 1=woman), age = moderator:
  $\hat{y}_i = \beta_0 + \beta_1(Age_i - 85) + \beta_2(Woman_i) + \beta_3(Age_i - 85)(Woman_i)$

- **$\beta_2 = -0.5306$* at age=85,** $Var(\beta_2) \rightarrow SE^2$ **for** β_2 **was** 0.06008
- **$\beta_3 = -0.1104$* unconditionally,** $Var(\beta_3) \rightarrow SE^2$ **for** β_3 **was** 0.00178
- **Covariance of β_2 SE and β_3 SE was 0.00111**

- **Regions of Significance for Moderator of Age = 60.16 to 79.52**
  - The gender effect β_2 is predicted to be <u>significantly negative</u> above age 79.52, <u>non-significant</u> from ages 79.52 to 60.16, and <u>significantly positive</u> below age 60.16 (because non-parallel lines will cross eventually).

When There's More than One Interaction

- Model: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{w_i}) + \boldsymbol{\beta_2}(\boldsymbol{x_i}) + \boldsymbol{\beta_3}(\boldsymbol{z_i})$
  $+ \boldsymbol{\beta_4}(\boldsymbol{x_i})(\boldsymbol{z_i}) + \boldsymbol{\beta_5}(\boldsymbol{w_i})(\boldsymbol{z_i}) + \boldsymbol{e_i}$

- Now all main effect slopes are "unique conditional" (simple):
  - $\boldsymbol{\beta_1}$ = diff in $\boldsymbol{y_i}$ per one-unit $\boldsymbol{w_i}$ specifically when $\boldsymbol{z_i = 0}$
  - $\boldsymbol{\beta_2}$ = diff in $\boldsymbol{y_i}$ per one-unit $\boldsymbol{x_i}$ specifically when $\boldsymbol{z_i = 0}$
  - $\boldsymbol{\beta_3}$ = diff in $\boldsymbol{y_i}$ per one-unit $\boldsymbol{z_i}$ specifically when $\boldsymbol{w_i = 0}$ **and** $\boldsymbol{x_i = 0}$

- Interaction slopes ($\boldsymbol{\beta_4}$ and $\boldsymbol{\beta_5}$) are "unique marginal"
  - $\boldsymbol{\beta_4}$ is now controlling for $\boldsymbol{\beta_5}$, but interpretation of $\boldsymbol{\beta_4}$ is unchanged:
    How slope of $\boldsymbol{x_i}$ is moderated by $\boldsymbol{z_i}$: $\boldsymbol{\beta_4}$ = diff in $\boldsymbol{\beta_2}$ per one-unit $\boldsymbol{z_i}$
    How slope of $\boldsymbol{z_i}$ is moderated by $\boldsymbol{x_i}$: $\boldsymbol{\beta_4}$ = diff in $\boldsymbol{\beta_3}$ per one-unit $\boldsymbol{x_i}$
  - New $\boldsymbol{\beta_5}$ has two equally correct interpretations:
    How slope of $\boldsymbol{w_i}$ is moderated by $\boldsymbol{z_i}$: $\boldsymbol{\beta_5}$ = diff in $\boldsymbol{\beta_1}$ per one-unit $\boldsymbol{z_i}$
    How slope of $\boldsymbol{z_i}$ is moderated by $\boldsymbol{w_i}$: $\boldsymbol{\beta_5}$ = diff in $\boldsymbol{\beta_3}$ per one-unit $\boldsymbol{w_i}$

When There's More than One Interaction

- Model: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{w_i}) + \boldsymbol{\beta_2}(\boldsymbol{x_i}) + \boldsymbol{\beta_3}(\boldsymbol{z_i})$
  $+ \boldsymbol{\beta_4}(\boldsymbol{x_i})(\boldsymbol{z_i}) + \boldsymbol{\beta_5}(\boldsymbol{w_i})(\boldsymbol{z_i}) + \boldsymbol{e_i}$

- Model-implied slopes of $\boldsymbol{w_i}$, $\boldsymbol{x_i}$ and $\boldsymbol{z_i}$ are **linear combinations:** (1) find common terms, (2) factor out the predictor the slope is for, and (3) then the term in brackets is model-implied predictor slope
  - Slope of $\boldsymbol{w_i}$: $\boldsymbol{\beta_1}(\boldsymbol{w_i}) + \boldsymbol{\beta_5}(\boldsymbol{w_i})(\boldsymbol{z_i}) \rightarrow [\boldsymbol{\beta_1} + \boldsymbol{\beta_5}(\boldsymbol{z_i})](\boldsymbol{w_i})$
  - Slope of $\boldsymbol{x_i}$: $\boldsymbol{\beta_2}(\boldsymbol{x_i}) + \boldsymbol{\beta_4}(\boldsymbol{x_i})(\boldsymbol{z_i}) \rightarrow [\boldsymbol{\beta_2} + \boldsymbol{\beta_4}(\boldsymbol{z_i})](\boldsymbol{x_i})$
  - Slope of $\boldsymbol{z_i}$: $\boldsymbol{\beta_3}(\boldsymbol{z_i}) + \boldsymbol{\beta_4}(\boldsymbol{x_i})(\boldsymbol{z_i}) + \boldsymbol{\beta_5}(\boldsymbol{w_i})(\boldsymbol{z_i}) \rightarrow [\boldsymbol{\beta_3} + \boldsymbol{\beta_4}(\boldsymbol{x_i}) + \boldsymbol{\beta_5}(\boldsymbol{w_i})](\boldsymbol{z_i})$

- More than one interaction will be necessary for **categorical** predictors (i.e., ordinal or nominal predictors with 3+ groups)
  - I will continue to show you the longer but more transparent way using binary-coded contrasts to represent group differences → matches model equation
  - An alternative is to let the program create the contrast for you using by listing it on CLASS in SAS (or BY in SPSS), or using i. in STATA or factor variables in R
    - Can be more convenient but more prone to misinterpretation (so I'm not doing it here)

Reviewing Categorical Predictors

Comparing outcome means across **4 groups** requires creating **3 new binary predictors** to be included **simultaneously** along with the intercept—for example, using "**indicator dummy-coded**" **predictors** so Control= Reference

Treatment Group	d1: C vs T1?	d2: C vs T2?	d3: C vs T3?
1. Control	0	0	0
2. Treatment 1	**1**	0	0
3. Treatment 2	0	**1**	0
4. Treatment 3	0	0	**1**

- Model: $y_i = \beta_0 + \beta_1(d1_i) + \beta_2(d2_i) + \beta_3(d3_i) + e_i$
  - The model gives us **the predicted outcome mean for each category** as follows:

Control (Ref) Mean	Treatment 1 Mean	Treatment 2 Mean	Treatment 3 Mean
β_0	$\beta_0+\beta_1(d1_i)$	$\beta_0+\beta_2(d2_i)$	$\beta_0+\beta_3(d3_i)$

  - Model directly provides 3 mean differences (control vs. each treatment), and indirectly provides another 3 mean differences (differences between treatments) as **linear combinations of the fixed effects**… let's see how this works

See p. 278 of: Darlington, R. B., & Hayes, A. F. (2016). *Regression analysis and linear models: Concepts, applications, and implementation*. Guilford.

Reviewing Categorical Predictors

Control (Ref) Mean = 10	Treatment 1 Mean = 12	Treatment 2 Mean = 15	Treatment 3 Mean = 19
$\boldsymbol{\beta}_0$	$\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1(\boldsymbol{d1}_i)$	$\boldsymbol{\beta}_0 + \boldsymbol{\beta}_2(\boldsymbol{d2}_i)$	$\boldsymbol{\beta}_0 + \boldsymbol{\beta}_3(\boldsymbol{d3}_i)$

Model: $\boldsymbol{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1(\boldsymbol{d1}_i) + \boldsymbol{\beta}_2(\boldsymbol{d2}_i) + \boldsymbol{\beta}_3(\boldsymbol{d3}_i) + \boldsymbol{e}_i$

Given the means above, here are the pairwise category differences:

	Alt Group		Ref Group	Difference
• C vs. T1 =	$(\beta_0 + \boldsymbol{\beta}_1)$	$-$	(β_0)	$= \boldsymbol{\beta}_1 = 2$
• C vs. T2 =	$(\beta_0 + \boldsymbol{\beta}_2)$	$-$	(β_0)	$= \boldsymbol{\beta}_2 = 5$
• C vs. T3 =	$(\beta_0 + \boldsymbol{\beta}_3)$	$-$	(β_0)	$= \boldsymbol{\beta}_3 = 9$
• T1 vs. T2 =	$(\beta_0 + \boldsymbol{\beta}_2)$	$-$	$(\beta_0 + \boldsymbol{\beta}_1)$	$= \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = 5 - 2 = 3$
• T1 vs. T3 =	$(\beta_0 + \boldsymbol{\beta}_3)$	$-$	$(\beta_0 + \boldsymbol{\beta}_1)$	$= \boldsymbol{\beta}_3 - \boldsymbol{\beta}_1 = 9 - 2 = 7$
• T2 vs. T3 =	$(\beta_0 + \boldsymbol{\beta}_3)$	$-$	$(\beta_0 + \boldsymbol{\beta}_2)$	$= \boldsymbol{\beta}_3 - \boldsymbol{\beta}_2 = 9 - 5 = 4$

Interactions Involving Categorical Predictors

- When using manual contrasts for predictors with 3 or more categories, **interactions must be specified with ALL dummy-coded predictors**
- If the program creates the dummy-coded predictors for you, all needed interaction predictors will be automatically included (but be careful!)

- e.g., **Adding an interaction of 4-category "group" with age (0=85):**
  - New predictors we must create for the model:
    - ***d1*** = 0, 1, 0, 0 → difference between Control and Treat1
    - ***d2*** = 0, 0, 1, 0 → difference between Control and Treat2
    - ***d3*** = 0, 0, 0, 1 → difference between Control and Treat3

$$
\begin{aligned}
\boldsymbol{y_i} = {} & \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{d1_i}) + \boldsymbol{\beta_2}(\boldsymbol{d2_i}) + \boldsymbol{\beta_3}(\boldsymbol{d3_i}) + \boldsymbol{\beta_4}(\boldsymbol{Age_i} - \mathbf{85}) \\
& + \boldsymbol{\beta_5}(\boldsymbol{d1_i})(\boldsymbol{Age_i} - \mathbf{85}) + \boldsymbol{\beta_6}(\boldsymbol{d2_i})(\boldsymbol{Age_i} - \mathbf{85}) + \boldsymbol{\beta_7}(\boldsymbol{d3_i})(\boldsymbol{Age_i} - \mathbf{85}) + \boldsymbol{e_i}
\end{aligned}
$$

- Multivariate Wald test would be needed to lump together the interaction contrasts ($\boldsymbol{\beta_5}$, $\boldsymbol{\beta_6}$, and $\boldsymbol{\beta_7}$) to test the "omnibus" group*age interaction
- Group difference slopes ($\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, and $\boldsymbol{\beta_3}$) are each conditional on age = 85
- Age slope ($\boldsymbol{\beta_4}$) is specific to the control group (when interactions = 0)
- But the model provides age slopes for each group, as well as group differences at any age as linear combinations of the fixed effects…

Interactions Involving Categorical Predictors

- **Adding an interaction of 4-category "group" with age (0=85):**
  - New predictors we must create for the model:
    - ***d1*** = 0, 1, 0, 0 → difference between Control and Treat1
    - ***d2*** = 0, 0, 1, 0 → difference between Control and Treat2
    - ***d3*** = 0, 0, 0, 1 → difference between Control and Treat3

$$y_i = \beta_0 + \beta_1(d1_i) + \beta_2(d2_i) + \beta_3(d3_i) + \beta_4(Age_i - 85) + \beta_5(d1_i)(Age_i - 85) + \beta_6(d2_i)(Age_i - 85) + \beta_7(d3_i)(Age_i - 85) + e_i$$

- **Equations for model-implied effects: [slope] (predictor)**
  - Effect of Age in Control group: $[\beta_4 + \beta_5(0) + \beta_6(0) + \beta_7(0)](Age_i - 85)$
  - Effect of Age in Treat1 group: $[\beta_4 + \beta_5(1) + \beta_6(0) + \beta_7(0)](Age_i - 85)$
  - Effect of Age in Treat2 group: $[\beta_4 + \beta_5(0) + \beta_6(1) + \beta_7(0)](Age_i - 85)$
  - Effect of Age in Treat3 group: $[\beta_4 + \beta_5(0) + \beta_6(0) + \beta_7(1)](Age_i - 85)$
  - Control vs. Treat1 for any age: $[\beta_1 + \beta_5(Age_i - 85)](d1_i)$
  - Control vs. Treat2 for any age: $[\beta_2 + \beta_6(Age_i - 85)](d2_i)$
  - Control vs. Treat3 for any age: $[\beta_3 + \beta_7(Age_i - 85)](d3_i)$
  - T1 vs T2 for any age: $[\beta_2 + \beta_6(Age_i - 85)](d2_i) - [\beta_1 + \beta_5(Age_i - 85)](d1_i)$
  - T1 vs T3 for any age: $[\beta_3 + \beta_7(Age_i - 85)](d3_i) - [\beta_1 + \beta_5(Age_i - 85)](d1_i)$
  - T2 vs T3 for any age: $[\beta_3 + \beta_7(Age_i - 85)](d3_i) - [\beta_2 + \beta_6(Age_i - 85)](d2_i)$

Multiple-DF Interactions More Generally

- Interactions can be tested between any predictors, including quantitative predictors that require more than one slope…
- Do piecewise education slopes differ between men and women? *(inspired by Example 4 models predicting annual income)*

$$\boldsymbol{Income_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{lessHS_i}) + \boldsymbol{\beta_2}(\boldsymbol{gradHS_i}) + \boldsymbol{\beta_3}(\boldsymbol{overHS_i}) + \boldsymbol{\beta_4}(\boldsymbol{MvW_i}) \\ + \boldsymbol{\beta_5}(\boldsymbol{MvW_i})(\boldsymbol{lessHS_i}) + \boldsymbol{\beta_6}(\boldsymbol{MvW_i})(\boldsymbol{gradHS_i}) + \boldsymbol{\beta_7}(\boldsymbol{MvW_i})(\boldsymbol{overHS_i}) + \boldsymbol{e_i}$$

  - Use SAS CONTRAST, STATA TEST/NESTREG, or R GLHT/hierarchical_lm to lump together $\boldsymbol{\beta_5}$, $\boldsymbol{\beta_6}$, and $\boldsymbol{\beta_7}$ for DF=3 F-test of interaction term
  - Simple slopes $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, and $\boldsymbol{\beta_3}$ give education effect for $MvW_i = 0$
  - Interactions $\boldsymbol{\beta_5}$, $\boldsymbol{\beta_6}$, and $\boldsymbol{\beta_7}$ give DIFF in education effect for $MvW_i = 1$
  - So simple slopes for each subsample of education for $MvW_i = 1$ are given by: $\boldsymbol{\beta_1}+\boldsymbol{\beta_5}$ for $lessHS_i$, $\boldsymbol{\beta_2}+\boldsymbol{\beta_6}$ for $gradHS_i$, and $\boldsymbol{\beta_3}+\boldsymbol{\beta_7}$ for $overHS_i$

- Btw, how many new fixed effects would be needed if we add a third sex category (e.g., nonbinary)?

What about 3-way interactions???

- Model: $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i)$
  $+ \beta_4(x_i)(z_i) + \beta_5(w_i)(z_i) + \beta_6(x_i)(w_i)$
  $+ \beta_7(w_i)(x_i)(z_i) + e_i$

- **Simple main effects make the predicted outcome higher or lower**
  - 1 possible interpretation for each simple main effect slope
  - Each simple main effect is conditional on other interacting predictors = 0
- **Each 2-way interaction (3 of them in a 3-way model) makes its simple main effect slopes (more/less) (positive/negative)**
  - So there are 2 possible interpretations for each 2-way interaction
  - Each "simple" 2-way interaction is conditional on third predictor = 0
- **The 3-way interaction makes each of its *2-way simple interaction slopes* (more/less) (positive/negative)**
  - So there are 3 possible interpretations of a 3-way interaction!
  - Is highest-order term in model, so is unconditional (marginal)

3-Way Interactions Follow the Same Rules

- Model: $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i)$
  $+ \beta_4(x_i)(z_i) + \beta_5(w_i)(z_i) + \beta_6(x_i)(w_i)$
  $+ \beta_7(w_i)(x_i)(z_i) + e_i$

- **Model-implied simple (conditional) main effect slopes:**
  - Effect of w_i: $[\beta_1 + \beta_5(z_i) + \beta_6(x_i) + \beta_7(x_i)(z_i)](w_i)$
  - Effect of x_i: $[\beta_2 + \beta_4(z_i) + \beta_6(w_i) + \beta_7(w_i)(z_i)](x_i)$
  - Effect of z_i: $[\beta_3 + \beta_4(x_i) + \beta_5(w_i) + \beta_7(w_i)(x_i)](z_i)$

- **Model-implied simple (conditional) 2-way interactions:**
  - Effect of x_i by z_i: $[\beta_4 + \beta_7(w_i)](x_i)(z_i)$
  - Effect of w_i by z_i: $[\beta_5 + \beta_7(x_i)](w_i)(z_i)$
  - Effect of x_i by w_i: $[\beta_6 + \beta_7(z_i)](x_i)(w_i)$

Intermediate Summary

- Interactions create "**moderation**": the idea that the effect (slope) of one predictor **depends** upon the value of another predictor

- Predictors' **main effect slopes** will change once they are included in an interaction term, because **they now mean different things**:
  - Former "marginal main effect slopes" become "conditional (or simple) effect slopes" specifically when the interacting predictor = 0
  - Need to have **0 as a meaningful value** for each predictor for that reason

- **Rules for interpreting conditional (or simple) fixed slopes:**
  - Predicted outcomes are conditional on (get adjusted by) main effect slopes
    - Positive slopes create higher outcomes; negative slopes create lower outcomes
  - Main effect slopes are conditional (get adjusted by) on two-way interactions
    - Interactions make main effect slopes more/less positive or more/less negative
    - Btw, three-way interactions do the same thing to two-way interactions
  - Highest-order interaction slope is unconditional—it will stay the same regardless of centering (i.e., extent of moderation is unconditional)

Categorical Predictors with Issues

- Experimental designs with fully crossed conditions lend themselves to analysis of variance-type models
- What happens when things go wrong? Two examples:
  - ANOVA with a hole in it
  - Predictors that don't apply or weren't measured for everyone
- These designs can be analyzed using **nested effects**
  - Different programs specify these differently, so I'll show them using a common language of pseudo-interaction terms
  - In specifying nested effects, what look like "interactions" actually act as switches instead to turn effects on/off…

A Traditional View of ANOVA

ANOVAs usually provide F-tests for **marginal mean differences...**

F(df=1) → a $v.$ b

F(df=2) → 1 $v.$ 2. $v.$ 3

F(df=2) → a $v.$ b * 1 $v.$ 2. $v.$ 3

Is this **really** what you want to know?

Means	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\bar{a}$	$\overline{a1}$	$\overline{a2}$	$\overline{a3}$
$\bar{b}$	$\overline{b1}$	$\overline{b2}$	$\overline{b3}$

ANOVA as a General Linear Model

$$y_i = \boxed{\beta_0} + \boxed{\beta_1(a1\ v.\ b1_i)}$$
$$+ \boxed{\beta_2(a1\ v.\ a2_i) + \beta_3(a1\ v.\ a3_i)}$$
$$+ \beta_4(a1\ v.\ b1_i)(a1\ v.\ a2_i)$$
$$+ \beta_5(a1\ v.\ b1_i)(a1\ v.\ a3_i) + e_i$$

The focus is now on **differences between *specific* conditions** as created by the β fixed effects.

Means	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{a}$	$\overline{a1}$	$\overline{a2}$	$\overline{a3}$
$\bar{b}$	$\overline{b1}$	$\overline{b2}$	$\overline{b3}$

ANOVA as a General Linear Model

- Software will find any **simple slopes (differences) you ask for**
  - TEST in SPSS MIXED (not GLM); ESTIMATE in SAS (GLM or MIXED)
  - LINCOM or MARGINS in STATA; NEW in Mplus

- Seeing research questions through linear models saves **nontraditional research designs**
  - Not fully crossed on purpose or by accident… "ANOVA with a hole in it"

Means	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{a}$	β_0	$\beta_0 + \beta_2$	$\beta_0 + \beta_3$
$\bar{b}$	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \boxed{\beta_4}$	$\beta_0 + \beta_1 + \beta_3 + \boxed{\beta_5}$

A Nontraditional ANOVA Design

$$y_i = \boxed{\beta_0} + \boxed{\beta_1(t3\ v.t1_i) + \beta_2(t2\ v.t1_i)} + \boxed{\beta_3(t1_i)(t\ v.c_i)} + \boxed{\beta_4(t2_i)(t\ v.c_i)} + e_i$$

β_3 and β_4 are not interaction terms. Instead, they are *nested* effects.

You are allowed to use any C effects you want to represent the C means, even in fully crossed designs!

Means	Cohort 1	Cohort 2	Cohort 3
Control	$\beta_0 + \beta_1 + \boxed{\beta_3}$	$\beta_0 + \beta_2 + \boxed{\beta_4}$	
Treatment	$\beta_0 + \boxed{\beta_1}$	$\beta_0 + \boxed{\beta_2}$	$\boxed{\beta_0}$

A Nested-Effects General Linear Model

- Example: predicting outcomes by dementia type and dementia timing in persons with OR without dementia
  - Type and timing do not apply to persons without dementia
  - So this requires the following new variables, created as follows:

DemType	**demYes**	**demAorV**	**demtime5**
None	0	0	0
AD (Alzheimer's)	1	−0.5	demtime−5
VA (Vascular)	1	0.5	demtime−5

  - **demYES** keeps track of diagnosis at all
  - **demAorV** distinguishes different diagnoses
  - **demtime** is years since diagnosis (centered at 5 when applicable)

A Nested-Effects General Linear Model

$$y_i = \beta_0 + \beta_1(demYes_i) + \beta_2(demAorV_i)$$
$$+\beta_3(demYes_i)(demtime_i - 5) + \beta_4(demAorV_i)(demtime_i - 5) + e_i$$

Fixed Effect	**Interpretation**
β_0: Intercept	Expected outcome for persons without dementia
β_1: demYes	Simple slope for difference between persons without dementia or with dementia at 5 years (averaged across AD and VA dementia types)
β_2: demAorV	Simple slope for outcome difference between persons with VA instead of AD type dementia (at 5 years)
β_3: demYes* demtime5	Because the main effect of demtime5 is not the model, β_3 is NOT an interaction term: Slope for outcome difference per year of dementia *only in persons with dementia* (averaged across AD and VA dementia types)
β_4: demAorV* demtime5	Because the main effect of demAorV IS in the model, β_4 IS an interaction term: Difference in slope for effect of years between persons with AD or VA type

Other Uses for GLM Nested Effects

- **Nested effects** are main effects specified to apply selectively to subsamples of the possible cases contributing to the model
- They have lots of potential—but relatively unknown—uses
  - “If and how much” effects of **semi-continuous** predictors
    - Difference between groups of “younger” and “older” adults; + slope for years of age within “older” adults (see Hoffman 2015 ch. 12)
    - Presence and severity of abuse: difference between groups of “not abused” and “abused” persons; + slope for severity of abuse within “abused” group (for which severity > 0)
  - Missing, refused to answer, or other **incomplete predictor data**:
    - Difference between groups of “incomplete” versus “complete” predictor values; + slope for predictor values in “complete” group
  - Predictor effects that only apply to one outcome in a multivariate GLM predicting multiple outcomes simultaneously…
    - Come back to my *Generalized Linear Models* class to see this usage!