

# Review of Fixed Effects within General Linear Models (and especially interaction terms)

- Topics:
  - Fixed slopes: Interpretation and significance
  - Scaling predictor variables: Centering and coding
    - Categorical predictors: Manual vs. program-automated coding
    - Semi-continuous predictor coding: If and how much (piecewise/spline)
    - Testing multiple slopes (for a single predictor or multiple predictors)
  - Linear models with interaction terms
    - Taxonomy terminology: Bivariate marginal, unique marginal, or unique conditional fixed slopes
    - Interpreting interaction slopes as modifiers of main effect slopes

# Naming Conventions in the GLM

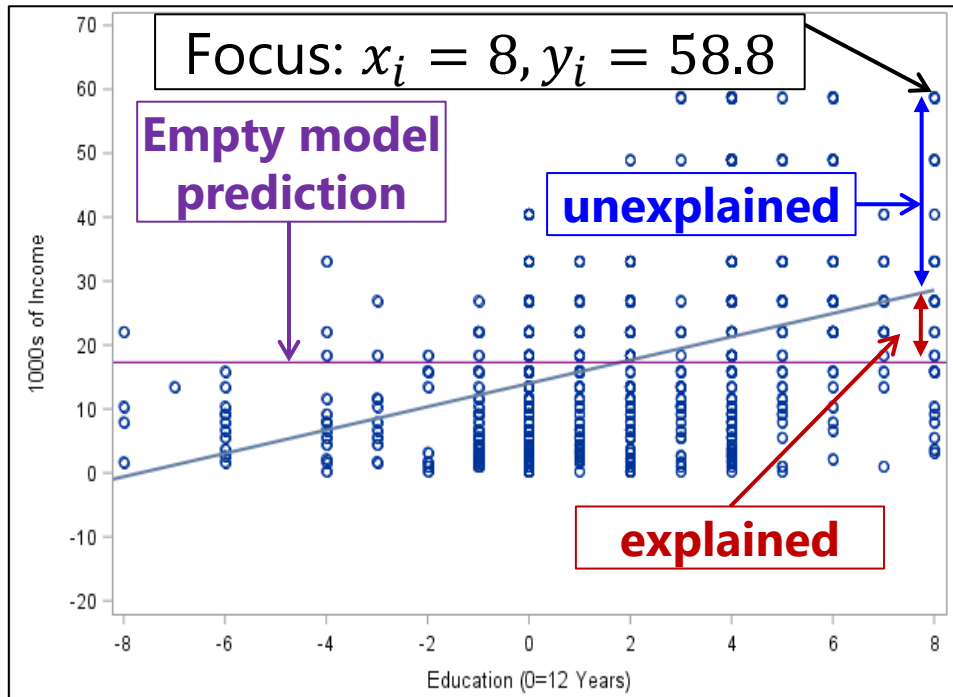
- The **general linear model** incorporates many different labels of single-level analyses (for **independent** obs) under 1 unifying term:

	Categorical Predictors	Quantitative Predictors	Both Types of Predictors
Univariate (one outcome)	"ANOVA"	"Regression"	"ANCOVA"
Multivariate (2+ outcomes)	"MANOVA"	"Multivariate Regression"	"MANCOVA"

- What these models all have in common is the use of a normal conditional distribution (i.e., for the *residuals* that remain after creating conditional outcomes from the model predictors)
- Btw, predictors do NOT have distributional assumptions!
- The use of these model labels **almost always implies estimation using "least squares"** (LS), aka "**ordinary least squares**" (OLS)

# A One-Slope GLM Example

The  $\beta$  formulas result from the goal of minimizing the squared residuals across the sample—this is called “**ordinary least squares estimation**”—let’s see what happens for one example person



Empty Model for  $y_i = \text{income}$ :

$$y_i = \beta_0 + e_i$$

$$\hat{y}_{Focus} = 17.3$$

$$y_{Focus} = 17.3 + 41.5$$

$$\text{Variance: } \sigma_e^2 = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N-1} = 190.2$$

→ 190.2 is **all** the  $y_i$  variance

Add Education as Predictor:

$$y_i = \beta_0 + \beta_1 (\text{Educ}_i - 12) + e_i$$

$$\hat{y}_{Focus} = 14.0 + 1.8(8) = 28.4$$

$$y_{Focus} = 28.4 + 30.4$$

$$\text{Variance: } \sigma_e^2 = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N-2} = 162.3$$

→ 162.3 is **leftover**  $y_i$  variance

$$\rightarrow R_{adj}^2 = \frac{190.2 - 162.3}{190.2} = .17$$

# General Linear Models, More Generally

- A **General Linear Model (GLM\*)** for outcome  $y_i$  looks like this:
  - actual  $y_i = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + \dots \beta_p(xp_i) + e_i$
  - predicted  $\hat{y}_i = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + \dots \beta_p(xp_i)$
  - The “ $i$ ” subscript denotes **variables** (that are individual-specific)
  - The  $\beta$  (“beta”) terms are the model **fixed effects** → **constants** whose subscripts range from 0 up to  $p$  as the last fixed effect):
    - $\beta_0 = \text{intercept}$  = expected  $y_i$  when all  $x_i$  predictors are 0
    - $\beta_1 = \text{slope of } x1_i$  = difference in  $y_i$  per one-unit difference in  $x1_i$
    - $\beta_2 = \text{slope of } x2_i$  = difference in  $y_i$  per one-unit difference in  $x2_i$
    - ...
    - $\beta_p = \text{slope of } xp_i$  = difference in  $y_i$  per one-unit difference in  $xp_i$

\* GLM may also stand for *Generalized* Linear Models, which includes General as one type (ugh)

# Significance Tests of Fixed Slopes

- Each  **$\beta$  fixed slope** has 6 relevant characteristics (\*essential to report):
  - **\*Estimate** = best guess for the fixed slope from our data (ML → tallest answer)
  - **\*Standard Error = SE** = average distance of sample slope from population slope → expected *inconsistency* of slope across samples
  - **t-value** =  $(\text{Estimate} - H_0) / SE$  = test-statistic for fixed slope against  $H_0 (= 0)$
  - **Denominator DF** =  $N - k$  (where  $k$  = total number of fixed effects)
  - **p-value** = (two-tailed) probability of fixed slope estimate *as or more extreme* IF  $H_0$  is true → how unexpected our result is on a  $t$ -distribution with  $0 = H_0$ ,  $SD = SE$
  - **(95%) Confidence Interval = CI** =  $\text{Estimate} \pm t_{\text{critical}} * SE$  = range in which true (population) value of estimate is expected to fall across 95% of samples
- Compare  $t$  test-statistic to  $t$  critical-value at pre-chosen level of significance (where % unexpected = alpha level): this is a “**univariate Wald test**”
  - Btw, if denominator DF are not used, then  $t$  is treated as a  $z$  instead (same value)
  - Btw, whether the  $p$ -value is found using a  **$t$ - or  $z$ -distribution** will differ by program and variant in generalized linear models...

# Significance of Each Fixed Slope

- Standard Error (SE) for fixed effect estimate  $\beta_x$  in a one-predictor model (remember, SE is like the SD of the estimated parameter):

$$SE_{\beta_x} = \sqrt{\frac{\text{residual variance of } y_i}{\text{Var}(x_i) * (N - k)}}$$

$N$  = sample size  
 $k$  = number of fixed effects

- When more than one predictor is included, SE turns into:

$$SE_{\beta_x} = \sqrt{\frac{\text{residual variance of } y_i}{\text{Var}(x_i) * (1 - R_x^2) * (N - k)}}$$

$R_x^2$  =  $x_i$  variance accounted for by other predictors, so  
 **$1 - R_x^2$  = unique  $x_i$  variance**

- So all things being equal, SE is smaller when:
  - More of the outcome variance has been reduced (better predictive model)
    - This means fixed effects can become significant later if  $R^2$  is higher at that point
  - The predictor has less covariance with other predictors
    - Best case scenario:  $X$  is uncorrelated with all other predictors
- If SE is smaller  $\rightarrow$   $t$ -value or  $z$ -value is bigger  $\rightarrow$   $p$ -value is smaller

# Effect Size of Each Fixed Slope

- Beyond just reporting testing of significance (against  $H_0 = 0$ ), every fixed slope should also have a reported **effect size**
  - Conveys **absolute size** of a slope relationship in more intuitive scale
  - Helps **compare** across predictors (and studies in meta-analysis)
  - Helps inform statistical **power** for similar effects in future research
- Most **common effect size metrics** in general linear models:
  - "**d**" family: for two-group difference in standard deviation units
  - "**r**" (aka, "eta"  $\eta$ ) family: for quantitative slopes in correlation metric
  - "Standardized slopes" are problematically ambiguous (as [explained here](#))
- A "partial" **d** or **r** [can be found](#) via the  $t$ -value for a fixed slope:
  - $d = \frac{2t}{\sqrt{DF_{den}}}$  ,  $r = \frac{t}{\sqrt{t^2 + DF_{den}}}$  ,  $d = \sqrt{\frac{4r^2}{1-r^2}}$  ,  $r = \sqrt{\frac{d^2}{4+d^2}}$
  - Squared version of "partial" conveys unique effect relative to unexplained variance (whereas "semi-partial" is relative to total variance instead)

# Scaling of Predictor Variables

- Get in the habit of rescaling all **predictors** so **0 is meaningful value**
  - **Why?** To maintain a **meaningful fixed intercept** in ALL models
  - For meaningful **conditional fixed slopes** within interactions (stay tuned)
  - (To avoid estimation problems in multilevel models with random slopes)
- For **quantitative** predictors, this is called (constant) "**centering**"
  - Center by subtracting a constant: sample mean is a common choice, but any meaningful value is good (e.g., known reference, minimum)
- For **categorical** predictors, this is called "**coding**"
  - Create  $C - 1$  **slopes** to describe  $C$  **categories** using values of 0 or 1 ("dummy coding") or values of 0, 1,  $-1$  ("effect coding") in a pattern that creates the desired interpretation of group differences
    - Will perfectly re-create all category means and mean differences using either **fixed effects directly** or **linear combinations** of fixed effects
    - I prefer **dummy coding**, in which 1 chosen category is the "**reference**" for which **all predictors = 0** (instead of reference = overall mean)



# Coding Strategies for Categorical Predictors

**Indicator coding:** Each non-ref category has a 1 value in **1 predictor only** to represent its mean difference from reference (good for **nominal**)

Group	(Intercept): A mean	AvsB: Diff for A vs B	AvsC: Diff for A vs C
A	1	0	0
B	1	1	0
C	1	0	1

**Either way**, all possible category means and mean differences not directly provided by the model fixed effects can be found from linear combinations of them...

**Sequential coding:** Each non-ref category can have multiple 1 values → predictors then give mean differences between sequential categories (good for **ordinal**)

Happy	(Intercept): 1 Mean	h1v2: 1→2 Diff	h2v3: 2→3 Diff	h3v4: 3→4 Diff	h4v5: 4→5 Diff
1	1	0	0	0	0
2	1	1	0	0	0
3	1	1	1	0	0
4	1	1	1	1	0
5	1	1	1	1	1

**Sequential coding** can be used to test whether an ordinal predictor can be treated as interval—whether it has a linear slope in predicting an outcome—by testing differences between the sequential slopes

# Bonus: Dummy vs. Effect Coding

**TABLE 10.3.** Four Ways of Coding Age Cohort and the Group Means Defined in Terms of the Regression Coefficients and Regression Constant

Age cohort by increasing age	$D_1$	$D_2$	$D_3$	Mean of $Y$
<b>Indicator coding</b>				
Generation Y	1	0	0	$\bar{Y}_1 = b_0 + b_1$
Generation X	0	1	0	$\bar{Y}_2 = b_0 + b_2$
Baby boomer	0	0	1	$\bar{Y}_3 = b_0 + b_3$
Pre-baby boomer	0	0	0	$\bar{Y}_4 = b_0$
<b>Sequential coding</b>				
Generation Y	0	0	0	$\bar{Y}_1 = b_0$
Generation X	1	0	0	$\bar{Y}_2 = b_0 + b_1$
Baby boomer	1	1	0	$\bar{Y}_3 = b_0 + b_1 + b_2$
Pre-baby boomer	1	1	1	$\bar{Y}_4 = b_0 + b_1 + b_2 + b_3$
<b>Helmert coding</b>				
Generation Y	-3/4	0	0	$\bar{Y}_1 = b_0 - \frac{3}{4}b_1$
Generation X	1/4	-2/3	0	$\bar{Y}_2 = b_0 + \frac{1}{4}b_1 - \frac{2}{3}b_2$
Baby boomer	1/4	1/3	-1/2	$\bar{Y}_3 = b_0 + \frac{1}{4}b_1 + \frac{1}{3}b_2 - \frac{1}{2}b_3$
Pre-baby boomer	1/4	1/3	1/2	$\bar{Y}_4 = b_0 + \frac{1}{4}b_1 + \frac{1}{3}b_2 + \frac{1}{2}b_3$
<b>Effect coding</b>				
Generation Y	1	0	0	$\bar{Y}_1 = b_0 + b_1$
Generation X	0	1	0	$\bar{Y}_2 = b_0 + b_2$
Baby boomer	0	0	1	$\bar{Y}_3 = b_0 + b_3$
Pre-baby boomer	-1	-1	-1	$\bar{Y}_4 = b_0 - b_1 - b_2 - b_3$

- **Indicator** and **sequential** coding each use one designated category as the reference
- **Helmert coding** “quantifies the difference in means between one group and the mean of the means in all higher-coded groups”
- **Effect coding** uses the grand mean across (equally weighted categories) as the reference; slopes give mean differences relative to grand mean
- **Others are possible, too!**

Table 10.3 on p. 278 of: Darlington, R. B., & Hayes, A. F. (2016). *Regression analysis and linear models: Concepts, applications, and implementation*. Guilford.

# Categorical Predictors: Manual Indicator Coding

- Model:  $y_i = \beta_0 + \beta_1(d1_i) + \beta_2(d2_i) + \beta_3(d3_i) + e_i$ 
  - Variable called “**group**”: Control=0, Treat1=1, Treat2=2, Treat3=3
  - 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
  - These interpretations only hold if all three new predictors are included!
- How does the model give us **all possible group differences**?  
By determining each group’s mean, and then each difference...

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

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

# Categorical Predictors: Manual Indicator Coding

- Model:  $y_i = \beta_0 + \beta_1(d1_i) + \beta_2(d2_i) + \beta_3(d3_i) + e_i$

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

Alt Group      Ref Group      Difference

- Control vs. T1 =  $(\beta_0 + \beta_1) - (\beta_0) = \beta_1$
- Control vs. T2 =  $(\beta_0 + \beta_2) - (\beta_0) = \beta_2$
- Control vs. T3 =  $(\beta_0 + \beta_3) - (\beta_0) = \beta_3$
- T1 vs. T2 =  $(\beta_0 + \beta_2) - (\beta_0 + \beta_1) = \beta_2 - \beta_1$
- T1 vs. T3 =  $(\beta_0 + \beta_3) - (\beta_0 + \beta_1) = \beta_3 - \beta_1$
- T2 vs. T3 =  $(\beta_0 + \beta_3) - (\beta_0 + \beta_2) = \beta_3 - \beta_2$

# 2 Ways to Include Categorical Predictors

1. **Manually create** and include dummy-coded predictors
  - Need  $C - 1$  predictors for  $C$  categories, added all at once, **treated as quantitative** (WITH in SPSS, by default in SAS and R, c. in STATA)
  - **We are going to do it this way**, in part because it corresponds directly to a linear model representation → transparency!
  - You have complete control of what your predictors represent!
2. **Let the program** create and include predictors for you
  - **Treated as categorical**: BY in SPSS, CLASS in SAS, i. in STATA, factor in R
    - SPSS and SAS: reference = highest/last; STATA/R: reference = lowest/first
  - Can be more convenient in GLMs to get predicted means if you have many categories, want many differences, or have interactions among categorical predictors—but not in all generalized linear models
  - And it marginalizes over other program-categorical predictors for their main effect  $F$ -tests, creating two sets of results (and confusion) ☹

# Btw, Program-Created Indicator Predictors

- Designate a predictor as “**categorical**” in program syntax
  - Use CLASS in SAS; BY in SPSS; i. prefix in STATA; factor variable in R
- For a predictor with  $C$  categories, the program automatically then creates  $C$  new dummy codes, for example “group” with  $C = 4$ :

<b>New Predictors Created Internally Mean This:</b>	<b>Control</b>	<b>Treat1</b>	<b>Treat2</b>	<b>Treat3</b>
IsControl	<b>1</b>	0	0	0
IsTreat1	0	<b>1</b>	0	0
IsTreat2	0	0	<b>1</b>	0
IsTreat3	0	0	0	<b>1</b>

- It then figures out how many of these internal predictor variables are needed—if using an intercept (the default), then it’s  $C - 1$ , not all  $C$
- It enters them until it hits that criterion—if it leaves the last one out (as when you have an intercept), then last category becomes your reference
- Everywhere in syntax you refer to the categorical predictor, you must tell the program what to do with EACH of these internal predictor variables

# What about Semi-Continuous Predictors?

- Some predictors contain both “kinds” and “amount” information
  - “Kinds” → mixtures of populations
  - “Amount” → severity within some (*nested* effect within sub-kind)
- Solution: an “if and how much” coding scheme, as shown →
  - “piecewise slopes” or “linear splines”

cig: # Daily Cigarettes	smoker: 0=no, 1=yes	smkamt: If smoker, how much >1?
0	0	0
1	1	0
2	1	1
3	1	2
4	1	3
5	1	4
6	1	5

**R:**

```
data$smoker = NA # Make 2 empty vars
data$smkamt = NA
data$smoker[which(data$cig==0)]=0
data$smkamt[which(data$cig==0)]=0
data$smoker[which(data$cig>0)]=1
data$smkamt[which(data$cig>0)]=
  data$cig[which(data$cig>0)]-1
```

**SAS:** \* Do not really need 2 empty vars first;

```
smoker=.; smkamt=.;
  IF cig=0 THEN DO; smoker=0; smkamt=0; END;
ELSE IF cig>0 THEN DO; smoker=1; smkamt=cig-1; END;
```

**STATA:**

```
gen smoker=. // Make 2 empty vars
gen smkamt=.
replace smoker=0 if cig==0
replace smkamt=0 if cig==0
replace smoker=1 if cig>0
replace smkamt=cig-1 if cig>0
```

# How Many Fixed Slopes Per Predictor?

- **“Linear”** in GLM refers to “slope\*variable + slope\*variable” format
  - This means the  $x_i$  predictors can also be nonlinear terms (e.g.,  $x_i^2$  to make a curved slope for  $x_i$ ), which is called **“nonlinear in the variables”**
  - The alternative, **“nonlinear in the parameters”** would have a nonlinear form, e.g., this exponential model:  $\hat{y}_i = \beta_0 + \beta_1[\exp(\beta_2(x_i))]$
- The **role of each predictor  $x_i$**  in creating a custom expected outcome  $y_i$  can be described through one or more fixed slopes
  - **One slope** is sufficient to capture the mean difference between two categories for a **binary  $x_i$**  or to capture a **linear effect** of a quantitative  $x_i$  (or exponential for  $\log x_i$  or logistic for  $\text{logit } x_i$ )
  - **More than one slope** may be needed to capture other nonlinear effects of a quantitative  $x_i$  (e.g., **quadratic** or **piecewise** trends)
  - **$C - 1$  slopes** are needed to capture the mean differences in the outcome across a **categorical predictor** with  **$C$  categories**
  - When **multiple slopes** are needed to describe the effect of a predictor, you will likely want a **joint hypothesis test** for all of them together...



# Multivariate Wald Tests of Fixed Effects

- General test for significance of **multiple fixed effects** at once (can be requested via SAS CONTRAST, STATA TEST; GLHT in R for GLMs)—you have likely already seen these special cases...
- GLM: Whether a set of fixed slopes significantly explains  $y_i$  variance (i.e., if  $R^2 > 0$ ) is tested via "**Multivariate Wald Test**" or "**F-test**"
  - $F(DF_{num}, DF_{den}) = \frac{SS_{model}/(k-1)}{SS_{residual}/(N-k)} = \frac{(N-k)R^2}{(k-1)(1-R^2)} = \frac{\text{weighted known info}}{\text{weighted unknown info}}$
  - **F-test** evaluates model  $R^2$  per DF spent to get to it and DF leftover
  - $R^2 = \frac{SS_{total} - SS_{residual}}{SS_{total}}$  = square of  $r$  between predicted  $\hat{y}_i$  and  $y_i$
- e.g., "Omnibus" F-test for the slopes of the main effect of a variable with  $C > 2$  categories (or for its interaction with other predictors)
- e.g., Model  $R^2$  change F-test in hierarchical regression (for grouping sets of predictors together and testing their joint contribution)
- Btw, without denominator DF,  $F$  is replaced by  $\chi^2 (= F * DF_{num})$
- Btw, when testing only 1 slope instead,  $t^2 = F$  and  $z^2 = \chi^2$

# A Taxonomy of Fixed Effect Interpretations

- In the most common statistical models, **fixed effects will be either**:
  - An **intercept** that provides an expected (conditional)  $y_i$  outcome,
  - Or a **slope** for the difference in  $y_i$  per one-unit difference in  $x_i$  predictor
    - Slopes for quantitative and categorical predictors are treated the same
- **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 effect is ***bivariate marginal*** (is uncontrolled and applies across all persons)
  - In models with **more than one fixed slope**, each slope's 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)
    - 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**)

# Fixed Slope Interpretations: Example

- Model:  $y_i = \beta_0 + \beta_1(w_i) + e_i$ 
  - $\beta_1$  is “bivariate marginal”: difference in  $y_i$  per unit  $w_i$  (uncontrolled)
- Model:  $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i) + e_i$ 
  - $\beta_1$  is “unique marginal”: diff in  $y_i$  per unit  $w_i$ , controlling for  $x_i$  and  $z_i$
  - $\beta_2$  is “unique marginal”: diff in  $y_i$  per unit  $x_i$ , controlling for  $w_i$  and  $z_i$
  - $\beta_3$  is “unique marginal”: diff in  $y_i$  per unit  $z_i$ , controlling for  $w_i$  and  $x_i$
- Model:  $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i) + \beta_4(x_i)(z_i) + e_i$ 
  - $\beta_1$  is “unique marginal”: diff in  $y_i$  per unit  $w_i$ , controlling for  $x_i$  and  $z_i$
  - $\beta_2$  is “unique conditional”: diff in  $y_i$  per unit  $x_i$ , controlling for  $w_i$  and  $z_i$ , specifically when  $z_i = 0$  (i.e.,  $\beta_2$  is a “simple” main effect slope)
  - $\beta_3$  is “unique conditional”: diff in  $y_i$  per unit  $z_i$ , controlling for  $w_i$  and  $x_i$ , specifically when  $x_i = 0$  (i.e.,  $\beta_3$  is a “simple” main effect slope)
  - $\beta_4$  is “unique marginal” (unconditional), but how do we interpret it???

# Interpreting Interaction Terms

- Model:  $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i) + \beta_4(x_i)(z_i) + e_i$ 
  - $\beta_4$  is “unique marginal” → interaction slope controlling for other slopes
  - Rather than talk about what happens to the predicted outcome  $y_i$ , interaction slopes are described by **what they do to their main effects**
- **A two-way** interaction has **two equally correct** interpretations:
  - How slope of  $x_i$  is moderated by  $z_i$ :  $\beta_4 =$  difference in  $\beta_2$  per unit  $z_i$
  - How slope of  $z_i$  is moderated by  $x_i$ :  $\beta_4 =$  difference in  $\beta_3$  per unit  $x_i$
- So the model-implied slopes of  $x_i$  and  $z_i$  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
  - Model-implied slope of  $x_i$ :  $\beta_2(x_i) + \beta_4(x_i)(z_i) \rightarrow [\beta_2 + \beta_4(z_i)](x_i)$
  - Model-implied slope of  $z_i$ :  $\beta_3(z_i) + \beta_4(x_i)(z_i) \rightarrow [\beta_3 + \beta_4(x_i)](z_i)$
  - Result can be found using SAS ESTIMATE, STATA LINCOM, or R GLHT
  - Many of our examples this semester will have interaction terms!...

# ... But There Are Only 4 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:  $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i) + \beta_4(x_i)(z_i) + e_i$

Slope of $x_i$ is $\beta_2 =$	Interaction Slope is $\beta_4 =$	So the effect of $x_i$ is ??? per unit higher $z_i$
10	2	more positive (by $\beta_4$ )
10	-2	less positive (by $\beta_4$ )
-10	-2	more negative (by $\beta_4$ )
-10	2	less negative (by $\beta_4$ )

# When There's More than One Interaction

- Model:  $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i) + \beta_4(x_i)(z_i) + \beta_5(w_i)(z_i) + e_i$
- Now all main effect slopes are “unique conditional” (simple):
  - $\beta_1$  = diff in  $y_i$  per one-unit  $w_i$  specifically when  $z_i = 0$
  - $\beta_2$  = diff in  $y_i$  per one-unit  $x_i$  specifically when  $z_i = 0$
  - $\beta_3$  = diff in  $y_i$  per one-unit  $z_i$  specifically when  $w_i = 0$  and  $x_i = 0$
- Interaction slopes ( $\beta_4$  and  $\beta_5$ ) are “unique marginal”
  - $\beta_4$  is now controlling for  $\beta_5$ , but interpretation of  $\beta_4$  is unchanged:
    - How slope of  $x_i$  is moderated by  $z_i$ :  $\beta_4$  = diff in  $\beta_2$  per one-unit  $z_i$
    - How slope of  $z_i$  is moderated by  $x_i$ :  $\beta_4$  = diff in  $\beta_3$  per one-unit  $x_i$
  - New  $\beta_5$  has two equally correct interpretations:
    - How slope of  $w_i$  is moderated by  $z_i$ :  $\beta_5$  = diff in  $\beta_1$  per one-unit  $z_i$
    - How slope of  $z_i$  is moderated by  $w_i$ :  $\beta_5$  = diff in  $\beta_3$  per one-unit  $w_i$

# When There's More than One Interaction

- Model:  $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i) + \beta_4(x_i)(z_i) + \beta_5(w_i)(z_i) + e_i$
- Model-implied slopes of  $w_i$ ,  $x_i$  and  $z_i$  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
  - Effect of  $w_i$ :  $\beta_1(w_i) + \beta_5(w_i)(z_i) \rightarrow [\beta_1 + \beta_5(z_i)](w_i)$
  - Effect of  $x_i$ :  $\beta_2(x_i) + \beta_4(x_i)(z_i) \rightarrow [\beta_2 + \beta_4(z_i)](x_i)$
  - Effect of  $z_i$ :  $\beta_3(z_i) + \beta_4(x_i)(z_i) + \beta_5(w_i)(z_i) \rightarrow [\beta_3 + \beta_4(x_i) + \beta_5(w_i)](z_i)$
- For quantitative moderators, **regions of significance** (see Hoffman 2015 ch. 2; [Finsaas & Goldstein, 2021](#)) can identify **moderator boundary values** for direction and significance of main effect slope
  - e.g., at what values of moderator  $z_i$  does the effect of  $w_i$  go from:
    - (a) significantly negative to nonsignificant?
    - (b) nonsignificant to significantly positive?

# 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

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

- Multivariate Wald test would be needed to lump together the interaction contrasts ( $\beta_5$ ,  $\beta_6$ , and  $\beta_7$ ) to test the “omnibus” group\*age interaction
- Group difference slopes ( $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ ) are each conditional on age = 85
- Age slope ( $\beta_4$ ) 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:  $[\beta_3 + \beta_7(Age_i - 85)](d3_i) - [\beta_2 + \beta_6(Age_i - 85)](d2_i)$

# 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 the 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  $w_i$ :  $[\beta_6 + \beta_7(z_i)](x_i)(w_i)$

# Interpreting Interactions: Summary

- Interactions represent “moderation” – the idea that the effect of one predictor depends upon the level of the other(s)
- The main effect slopes **WILL CHANGE** once their predictors are part of an interaction, because they now mean different things:
  - Main effect → Simple effect specifically when interacting predictor(s) = 0
  - Need to have 0 as a meaningful value for each predictor for that reason
- Rules for interpreting conditional (simple) fixed slopes:
  - Intercepts are conditional on (i.e., get adjusted by) main effect slopes
  - Main effects are conditional on two-way interactions
  - Two-way interactions are conditional on three-way interactions
  - Highest-order term is unconditional → same regardless of centering