

A Re-Introduction to General Linear Models (GLM)

- Topics:
 - You do know the GLM
 - Estimation (where the numbers in the output come from):
From least squares to residual maximum likelihood (REML)
 - Reviewing specification of fixed effects in GLMs
 - Centering and interpreting effects of quantitative predictors
 - Two ways of including categorical predictors
 - Fun with interaction terms!

You do know the General Linear Model

- The **general linear model** incorporates many different labels of related single-level analyses under one unifying umbrella term:

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

- The labels across columns are not actually helpful—they create artificial distinctions among what is really just one kind of model
- 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/expected outcomes from the model predictors)
- Note: Model predictors do NOT have distributional assumptions!

The Two Sides of Any Model

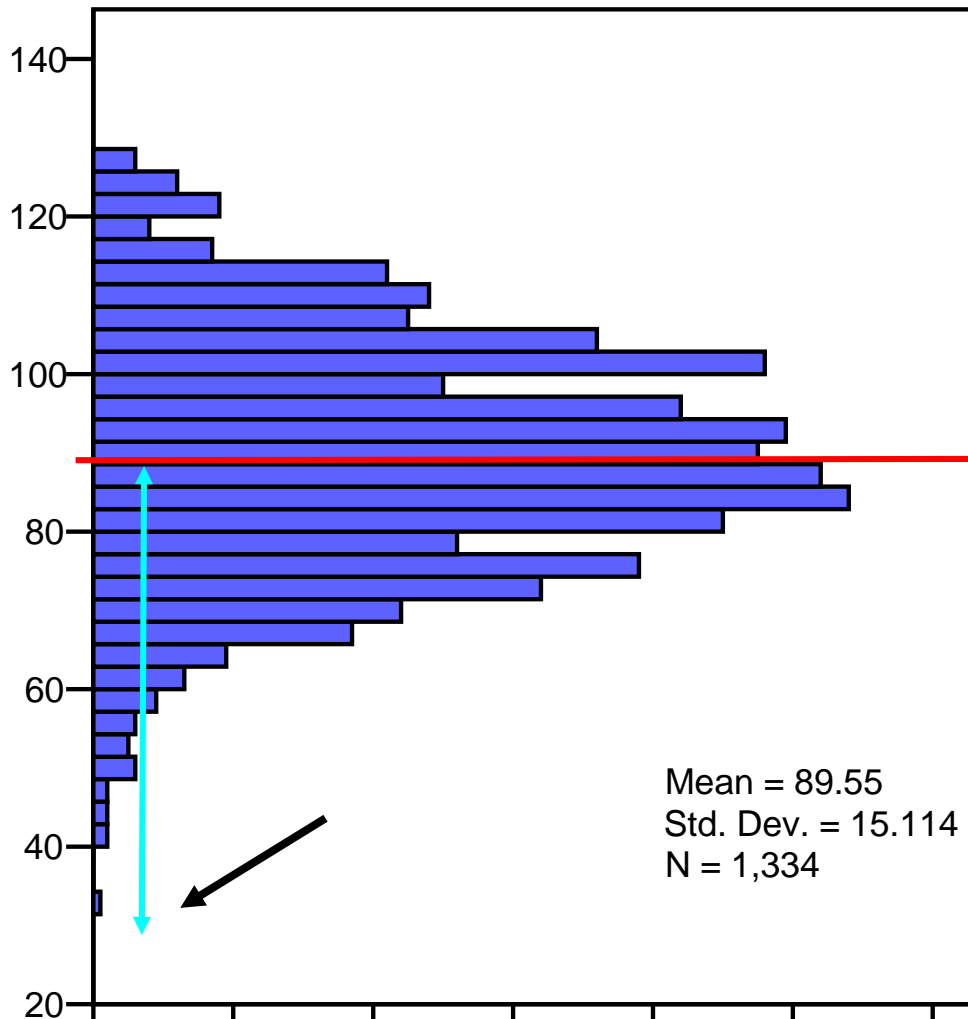
• Model for the Means:

- *Aka* **Fixed Effects**, Structural Part of Model
- What you are used to **caring about for testing hypotheses**
- How the expected outcome for a given observation varies as a weighted function of its values on the model predictor variables
 - Fixed slopes are **estimated constants** that multiply predictors

• Model for the Variance:

- *Aka* **Random Effects and Residuals**, Stochastic Part of Model
- How residuals are distributed and related across observations
- What you are used to **making assumptions about** instead...
- **For the GLM**, that residuals come from a **normal** distribution, are **independent** across persons, and have **constant variance** across persons and predictors (“identically distributed”)

The Simplest Possible Model: The “Empty Means” GLM



$$y_i = \beta_0 + e_i$$

Filling in values:

$$32 = \underbrace{90}_{\hat{y}_i} + -58$$

\hat{y}_i

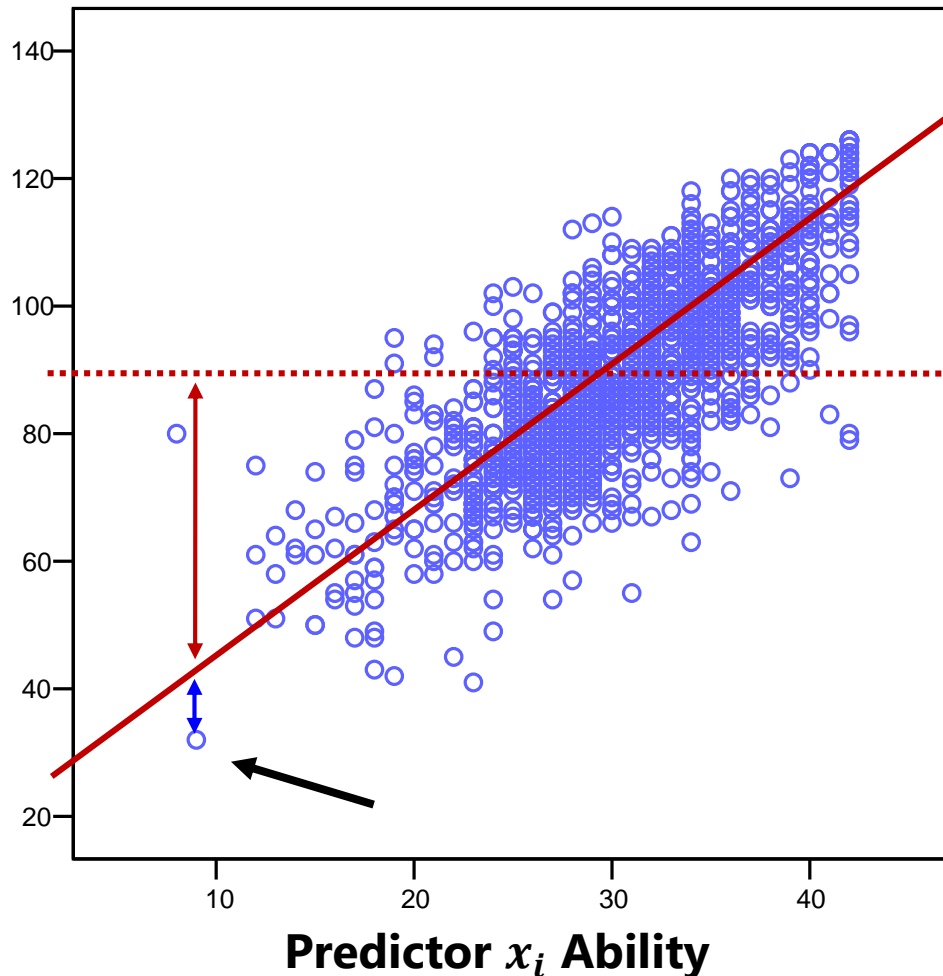
\hat{y}_i = “y-hat” model-predicted outcome

Model for the Means

y_i residual (“error”) variance:

$$\frac{\sum (y_i - \hat{y}_i)^2}{N - 1}$$

“Linear Regression” Model with a Quantitative Predictor ($x_i = \text{Ability}$)



$$y_i = \beta_0 + \beta_1(x_i) + e_i$$

Empty Model:

$$32 = 90 + \quad -58$$

Ability Model ($x_i = 9$):

$$32 = \underbrace{29 + 2(9)}_{\hat{y}_i} + -15$$

\hat{y}_i

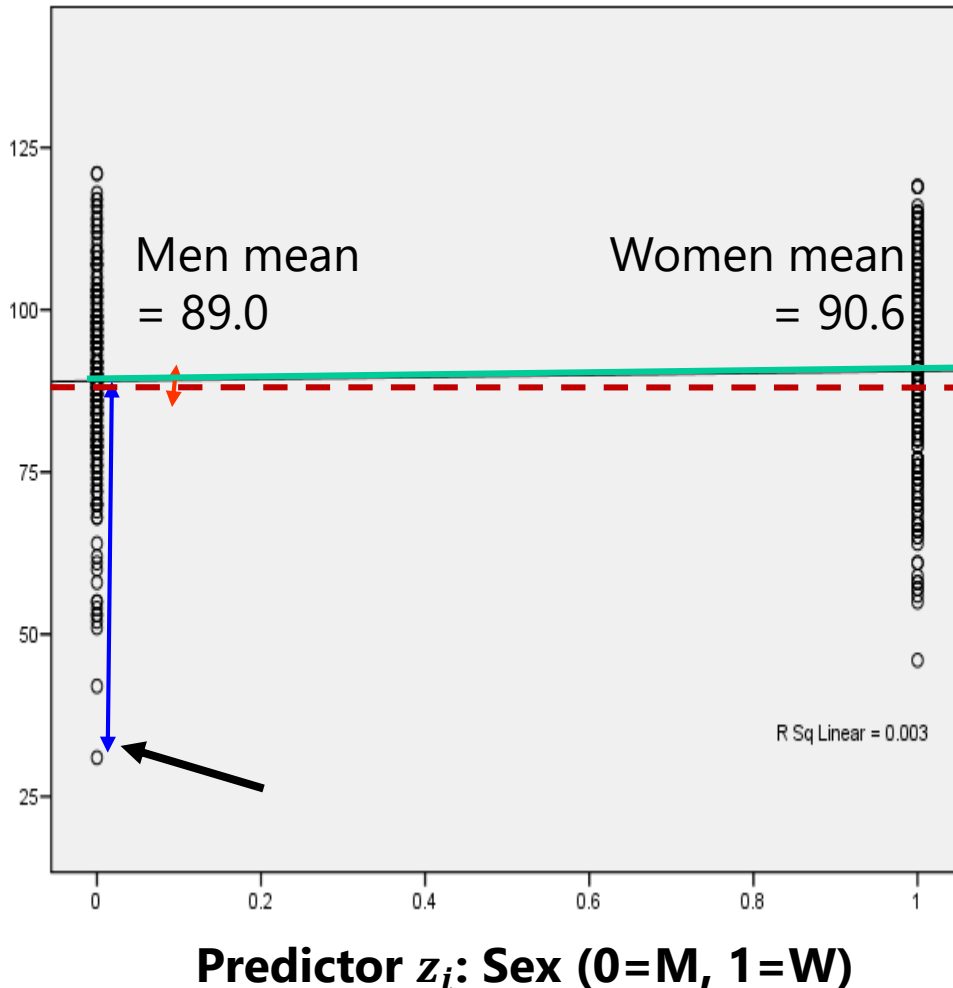
y_i residual
("error")

variance:

$$\frac{\sum (y_i - \hat{y}_i)^2}{N - 2}$$

Model
for the
Means

“ANOVA” Model with a Categorical Predictor ($z_i = \text{Sex}$)



$$y_i = \beta_0 + \beta_1(z_i) + e_i$$

Empty Model:

$$32 = 90 + -58$$

Sex Model ($z_i = 0$):

$$32 = \underbrace{89 + 1.6(0)}_{\hat{y}_i} + -57$$

y_i residual
("error")
variance:

$$\frac{\sum (y_i - \hat{y}_i)^2}{N - 2}$$

Model
for the
Means

The Two Sides of a General Linear Model

$$y_i = \beta_0 + \beta_1(x_i) + \beta_2(z_i) + \dots + e_i$$

• Model for the Means → Predicted Values:

Our focus today

- Each person's expected (predicted) outcome is a weighted linear function of his/her values on x_i and z_i (and any other predictors), each measured once per person (i.e., this is a univariate model)
- **Estimated parameters are called fixed effects** (here, β_0 , β_1 , and β_2)
- Number of fixed effects will show up in formulas as k (so $k = 3$ here)

• Model for the Variance:

- $e_i \sim N(0, \sigma_e^2) \rightarrow$ ONE source of residual (unexplained) error
- In the GLM, e_i has a mean of 0 with some estimated constant variance σ_e^2 , is normally distributed, is unrelated to x_i and z_i , and is unrelated across all observations (which is just one per person here)
- **Estimated parameter is the residual variance only** (not each e_i)
- Proportion of variance reduced relative to *empty means* model = R^2

See? You do know the GLM!

- The **general linear model** incorporates many different labels of related single-level analyses under one 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 (for the *residuals* that remain after creating conditional outcomes from the model predictors)
- The use of these words **almost always implies estimation using "least squares"** (LS), aka "**ordinary least squares**" (OLS)

How Estimation Works (In Brief)

- Most statistical estimation routines do one of three things:
- Minimize Something: Typically found with names that have “least” in the title. Forms of **least squares** include “Generalized”, “Ordinary”, “Weighted”, “Diagonally Weighted”, “WLSMV”, and “Iteratively Reweighted.” **Typically the estimator of last resort**
- Maximize Something: Typically found with names that have “maximum” in the title. Forms include “**Maximum likelihood**” (ML), “Residual Maximum Likelihood” (REML), “Robust ML”. Typically the gold standard of estimators, and what we will use this semester. **REML is the same thing as least squares for complete data**
- Use Simulation to Sample from Something: more recent advances in simulation use resampling techniques. Names include “**Bayesian Markov Chain Monte Carlo**”, “Gibbs Sampling”, “Metropolis Hastings”, “Metropolis Algorithm”, and “Monte Carlo”. Used for complex models in which ML is not available or feasible

Ordinary Least Squares (OLS) Estimation

Source of Outcome Information	Sums of Squares (each summed from $i = 1$ to N)	Degrees of Freedom	Mean Square
Model (known because of predictor slopes)	$SS_{model}: (\hat{y}_i - \bar{y})^2$	$DF_{num}: k - 1$	$MS_{model}: \frac{SS_{model}}{k-1}$
Residual (leftover after predictors; still unknown)	$SS_{residual}: (y_i - \hat{y}_i)^2$	$DF_{den}: N - k$	$MS_{residual}: \frac{SS_{residual}}{N-k}$
"Corrected" Total (all original information in y_i)	$SS_{total}: (y_i - \bar{y})^2$	$DF_{total}: N - 1$ (not shown)	$MS_{total}: \frac{SS_{total}}{N-1}$ (not shown)

- MS_{model} = how much info was captured per added fixed slope
- $MS_{residual}$ = how much info is leftover, per possible new fixed slope
→ **residual variance** = "MSE" error variance
- **Variance explained** by model fixed slopes: $R^2 = \frac{SS_{total} - SS_{residual}}{SS_{total}}$
- R^2 = square of correlation between model-predicted \hat{y}_i and actual y_i
- **F test-statistic** for significance of $R^2 > 0$? is given in two equivalent ways:
 $F(DF_{num}, DF_{den}) = \frac{MS_{model}}{MS_{residual}}$ or $F(DF_{num}, DF_{den}) = \frac{(N-k)R^2}{(k-1)(1-R^2)}$

Ordinary Least Squares (OLS) Estimation

- OLS uses fixed effect estimates that minimize:
 - $\sum_{i=1}^N (e_i^2)$ = sum of squared residuals across persons
 - Invented c. 1840, can be done via matrix algebra, so it will always work
- Has “closed form” solution (= easy formula) when used for general linear models (GLM) for single outcomes given:
 - $e_i \sim \mathbf{N}(\mathbf{0}, \sigma_e^2)$ → residuals are normally distributed, independent, and have constant variance
- For GLM for multiple outcomes, OLS becomes useless...
 - Cannot have missing outcomes (listwise-deletes entire person instead)
 - Only two options for modeling covariance between outcomes
 - Then why do it this way? Dogma + lack of awareness of alternatives...
- For non-normal outcomes, OLS can't be used at all... !

Maximum Likelihood to the Rescue

- Maximum likelihood estimation is the better way of finding the model estimates using all the data, and it comes in 2 flavors:
- **“Residual (or restricted) maximum likelihood”**
 - Only available for general linear models or general linear mixed models (key: based on normally distributed residuals at all levels of analysis)
 - **REML = OLS** given complete outcomes, but it doesn't require them
 - Estimates variances the same way as in OLS (accurate) →
$$\frac{\sum (y_i - \hat{y}_i)^2}{N - k}$$
- **“Maximum likelihood” (ML; also called FIML*)**
 - Is more general, is available for all of the above, as well as for non-normal outcomes and models with latent variables (CFA/SEM/IRT/DCM)
 - Is NOT the same as OLS: it under-estimates variances by not accounting for number of estimated fixed effects →
$$\frac{\sum (y_i - \hat{y}_i)^2}{N}$$
- **FI = Full information → it uses all original data (they both do)*

Maximum Likelihood to the Rescue

- Even though REML = OLS for complete outcomes, we will switch to software that uses REML or ML instead of OLS
 - So “sums of squares” and “mean squares” are no longer a thing
 - In SPSS, SAS, or STATA: one routine called “**MIXED**” that subsumes and expands the capabilities of GLM (SPSS, SAS) or REGRESS (STATA)
 - In R: There are lots of packages instead of just “MIXED”! Stay tuned
- Why use MIXED (even when it is the same as OLS)?
 - **Same conveniences:** Like SAS (but not SPSS) GLM or STATA REGRESS, MIXED can still produce estimates and SEs for linear combinations of fixed effects (e.g., pairwise comparisons, simple slopes of interactions), as well as joint F -tests for multiple slopes (e.g., for change in R^2)
 - **Generalizability:** We can estimate **univariate or multivariate** models for normal outcomes using the same MIXED routine
 - For **non-normal outcomes**, there are parallel routines in SAS (GLIMMIX) and STATA (several), but not in SPSS (“pseudo-ML” estimation only)

Intermediate Summary

- **What is *not* new:**

- We will be starting with the same kind of univariate general linear models for single outcomes per person you already know (regression, ANOVA, ANCOVA)
- Up next: we will see how their model for the variance must differ when predicting multiple outcomes per person (that are likely correlated)

- **What *is* new:**

- Rather than finding the fixed effects and residual variance through OLS (which yields sums of squares, mean squares, and so forth), the program will find them using residual maximum likelihood, of which OLS is a special case with limited applicability for real-life multivariate data
- MLMs will have way more fixed effects needed to address the effects of predictor variables, and will often include interactions
- So let's make sure we are comfortable with fixed effects first!

Testing Significance of Fixed Effects

- Each **β fixed effect** has 6 relevant characteristics (report at least Est, SE):
 - **Estimate** = best guess for the fixed slope from our data
 - **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 $M=H_0$, $SD=SE$
 - **(95%) Confidence Interval** = **CI** = $\text{Estimate} \pm t_{critical} * SE$ = 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 alpha level (where % unexpected = alpha): this is a “**univariate Wald test**”
 - Btw, we will only be using two-tailed tests for fixed effects
 - Btw, if denominator DF are not used, then **t** is treated as a **z** instead
 - In MLM, whether the p -value is based on **t** or **z** varies by program

Multivariate Wald Tests of Fixed Effects

- “Multivariate Wald Test” is a general way to test for significance of **multiple fixed effects** at once (i.e., by lumping fixed slopes together to test a joint null hypothesis that all slopes = 0)
- Special cases of this you have likely already seen:
 - F -test of model R^2 or change in R^2 after adding new predictor slopes (neither of may be provided for you anymore in MLM software)
 - “Omnibus” F -test for the effect of a categorical variable (e.g., for 3 dummy-coded contrasts that create differences between 4 groups)
 - Implies numerator $DF > 1$ (otherwise it would just be a t -test for $DF = 1$)
- Available for sets of fixed effects via CONTRAST (in SAS or STATA) or TEST (in STATA, SPSS, or Mplus), or GLHT (in R; CONTRAST too)
 - To indicate separate numerator DF :
 - SAS CONTRAST: Separate each fixed effect by commas
 - STATA TEST: List fixed effects in separate sets of parentheses
 - We will use this in nearly every example this semester!

Significance of Fixed Effects in MLM*

	Denominator DF is infinite (Proper Wald test)	Denominator DF is estimated instead ("Modified" Wald test)
Numerator DF = 1 (<i>test one fixed effect</i>) is Univariate Wald Test	use z distribution (Mplus, STATA default)	use t distribution (SAS, SPSS, STATA with dfmethod option)
Numerator DF > 1 (<i>test 2+ fixed effects</i>) is Multivariate Wald Test	use χ^2 distribution (Mplus, STATA default)	use F distribution (SAS, SPSS, STATA with dfmethod option)
Options for estimating Denominator DF (DDF)	not applicable	SAS, STATA 14+: Kenward-Roger SAS, STATA 14+, SPSS: Satterthwaite

* Btw, I am still figuring this out in R... default and optional DDF behavior seem to vary by package (like everything else)

Standard Errors for GLM* Fixed Effects

- Standard Error (SE) for the fixed effect estimate β_x in a one-predictor GLM

$$SE_{\beta_x} = \sqrt{\frac{\text{residual variance of Y}}{\text{variance of } x_i * (N - k)}} = \sqrt{\frac{\sigma_e^2}{\sigma_x^2 * (N - k)}}$$

- Btw, SE \approx SD of the estimated parameter across samples
- When more than one predictor is included, SE turns into:

$$SE_{\beta_x} = \sqrt{\frac{\sigma_e^2}{\sigma_x^2 * (1 - R_x^2) * (N - k)}}$$

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

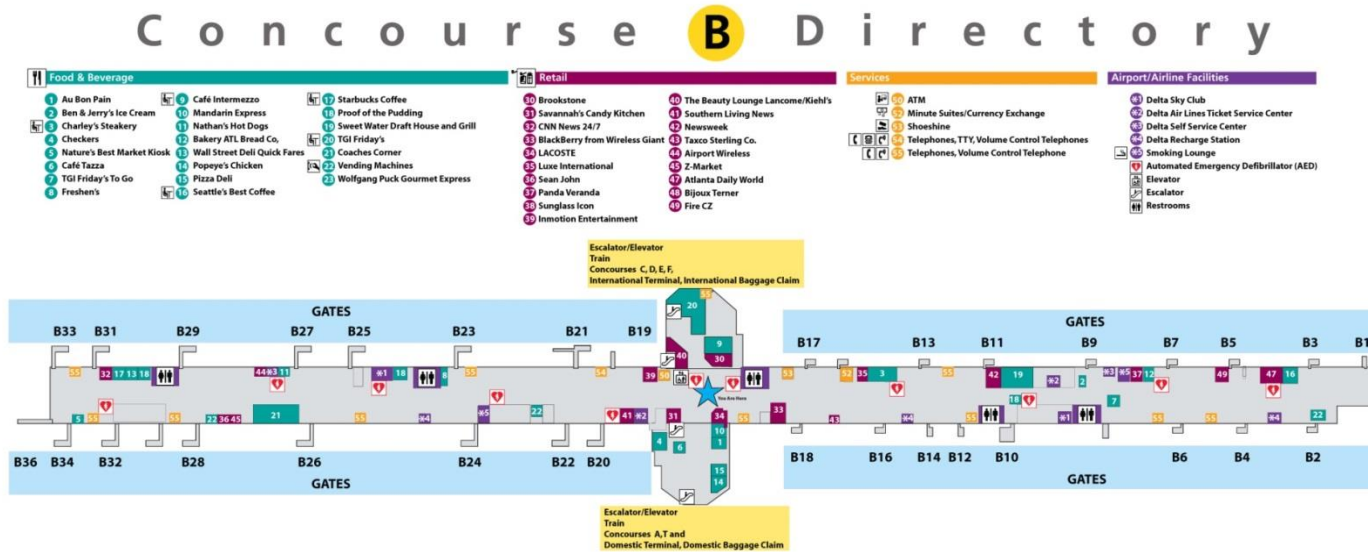
- So all things being equal, SE is smaller when:
 - More of the outcome variance has been reduced (better predictive model)
 - The predictor has less covariance with other predictors (less collinearity)
- If SE is smaller \rightarrow t -value (or z -value) is bigger \rightarrow p -value is smaller

* In MLM, the SE numerator will differ by predictor level!

Representing the Effects of Predictors

- From now on, we will think carefully about exactly how the **predictor variables** are entered into the **model for the means** (i.e., by which a predicted outcome is created for each person)
- Why don't people always care? Because the scale of predictors:
 - Does NOT affect the amount of outcome variance accounted for (R^2)
 - Does NOT affect the outcome values predicted by the model for the means (so long as the same predictor fixed effects are included)
- Why should this matter to us?
 - ***Because the Intercept = expected outcome when all predictors = 0***
 - Can end up with nonsense values for intercept if $x = 0$ isn't in the data, so we need to **center**: change the scale of the predictors to include 0
 - Centering is even more important once interactions are included or once random intercepts are included (i.e., variability around fixed intercept)

Why the Intercept β_0 *Should* Be Meaningful...

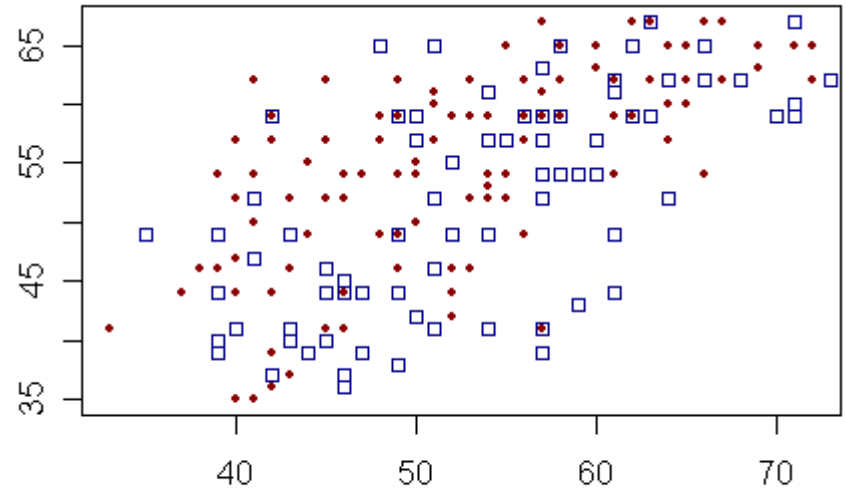


This is a very detailed map...
But what do we need to know
to be able to use the map at all?

What the Intercept β_0 *Should* Mean to You...

The model for the means will describe what happens to the predicted outcome y_i "as x_i increases" or "as z_i increases" and so forth...

But you won't know what the predicted outcome is supposed to be unless you know where the predictor variables are starting from!



Therefore, the **intercept** is the "**YOU ARE HERE**" sign in the map of your data... so it should be somewhere in the map*!

* There is no *wrong* way to center (or not), only *weird*...

Centering Quantitative Predictors

- Centering = subtract a constant to create a new "0" for the predictor
 - e.g., "mean-center" using predictor's mean: $x_{mi} = x_i - \bar{x}$
 - e.g., center using meaningful constant C : $x_{ci} = x_i - C$

Using **original** years of education:

$x_i = \text{education}$, $y_i = \text{income}$

$$y_i = \beta_0 + \beta_1(x_i) + e_i$$

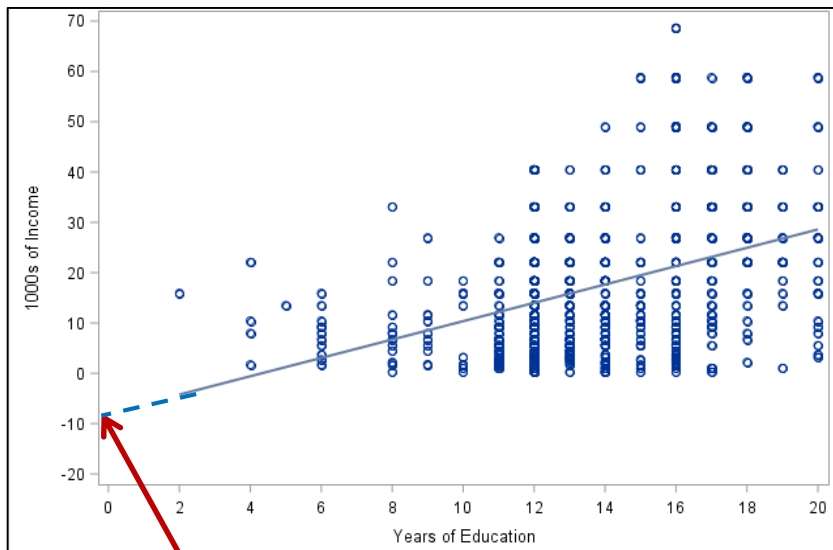
$$\hat{y}_i = -7.89 + 1.82(x_i)$$

Using **education centered at 12**:

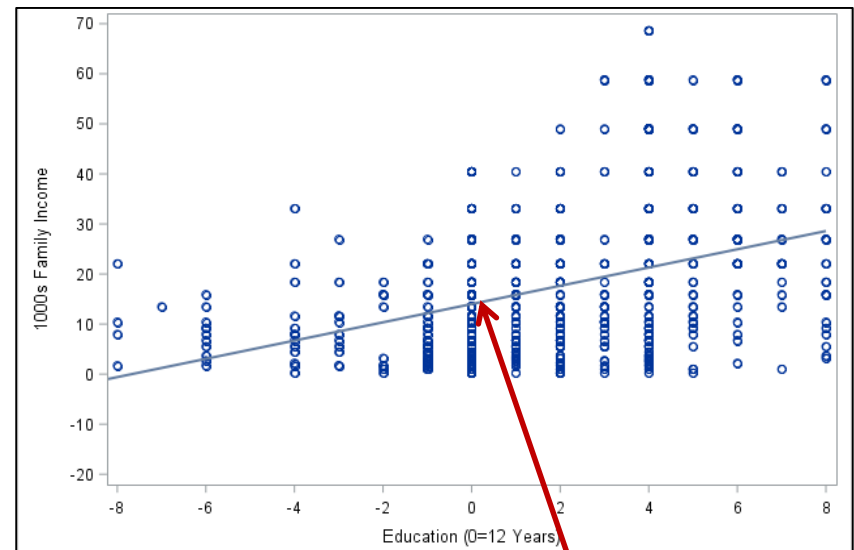
$x_i = \text{educ} - 12$, $y_i = \text{income}$

$$y_i = \beta_0 + \beta_1(x_i) + e_i$$

$$\hat{y}_i = 14.00 + 1.82(x_i)$$



Intercept β_0



Intercept β_0

What if I want a different intercept?

- Choosing a location for your model-estimated intercept does not lock you into only that value for the reference location...
- **ESTIMATE/LSMEANS statements (in SAS) to the rescue!**
 - **TEST in SPSS, LINCOM/MARGINS in STATA, NEW in Mplus, GLHT/EMMEANS/LSMEANS in R (varies by package, of course)**
- These statements allow to you to request linear combinations of fixed effects given any values of your predictors
 - In other words, new intercept values = predicted outcomes
- Rules for ESTIMATE-type statements:
 - If you want a predicted outcome, you **MUST** include the intercept
 - Variable names sometimes refer to their predictor values, and sometimes to their fixed effects, depending on what is being estimated (ugh, I know)
 - The default value for quantitative predictors is 0 (i.e., if excluded, = 0)
 - The default value for categorical predictors varies by program

Categorical Predictors (3+ Groups)

- Two alternatives for how to include grouping predictors
 1. **Manually create** and include dummy-coded group contrasts
 - Need $C - 1$ contrasts for C categories, added all at once, **treated as quantitative** (WITH in SPSS, by default in SAS and R, c. in STATA)
 - Corresponds more directly to linear model representation
 - Can be easier to set own reference group and contrasts of interest
 2. **Let the program** create and include group contrasts for you
 - **Treated as categorical:** BY in SPSS, CLASS in SAS, i. in STATA, factor in R
 - SPSS and SAS: reference = highest/last group; STATA: reference = lowest/first group
 - Can be more convenient if you have many groups, want many contrasts, or have interactions among grouping predictors
 - But it marginalizes over main effects when estimating other effects ☹️

Categorical Predictors Via Manual Contrasts

- Model: $y_i = \beta_0 + \beta_1(d1_i) + \beta_2(d2_i) + \beta_3(d3_i) + e_i$
 - “group” variable: Control=0, Treat1=1, Treat2=2, Treat3=3
 - New variables to be created for the model:
 - $d1 = 0, 1, 0, 0 \rightarrow$ difference between Control and Treat1
 - $d2 = 0, 0, 1, 0 \rightarrow$ difference between Control and Treat2
 - $d3 = 0, 0, 0, 1 \rightarrow$ 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 the difference...

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

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

Categorical Predictors Via Manual Contrasts

- 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
β_0	$\beta_0 + \beta_1(d1_i)$	$\beta_0 + \beta_2(d2_i)$	$\beta_0 + \beta_3(d3_i)$

- | | <u>Alt Group</u> | <u>Ref Group</u> | <u>Difference</u> |
|------------------|-----------------------|-----------------------|-----------------------|
| • Control vs. T1 | $(\beta_0 + \beta_1)$ | (β_0) | $= \beta_1$ |
| • Control vs. T2 | $(\beta_0 + \beta_2)$ | (β_0) | $= \beta_2$ |
| • Control vs. T3 | $(\beta_0 + \beta_3)$ | (β_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 | $(\beta_0 + \beta_3)$ | $(\beta_0 + \beta_2)$ | $= \beta_3 - \beta_2$ |

Program-Created Contrast 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 contrast variables, for example “group” with $C = 4$:

New Predictors Created Internally Mean this:	Control	Treat1	Treat2	Treat3
IsControl	1	0	0	0
IsTreat1	0	1	0	0
IsTreat2	0	0	1	0
IsTreat3	0	0	0	1

- It then figures out how many of these internal contrast 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 contrast variables

Program-Created Contrast Predictors

- If you let **SAS**/SPSS do the dummy coding via **CLASS**/BY, then the **highest/last group is default reference**
 - In SAS 9.4 you can change reference group: REF='level' | FIRST | LAST but it changes that group to be last in the data (→ confusing, so don't do it)
 - "Type III test of fixed effects" provide multivariate Wald tests by default
 - LSMEANS/EMMEANS can be used to get all cell means and comparisons without specifying each individual contrast, but you still have to ask for interaction contrasts (add / E to end of ESTIMATE to see the order of category values)
- If you let STATA/R do the dummy coding via i.group/factor , then the **lowest/first group is default reference**
 - In STATA, can change reference group, e.g., last = ref → ib(last).group
 - In R, you can use RELEVEL to change the reference group (says The Google)
 - STATA CONTRAST (R ANOVA) used to get omnibus tests (not provided by default)
 - STATA MARGINS (R EMMEANS) can be used to get all means and comparisons with much less code than describing each individual contrast
- Btw, no such thing as "categorical" predictors in Mplus ☹
 - You must create contrasts manually for all grouping variables

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 unit difference in 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)
 - 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: $y_i = \beta_0 + \beta_1(w_i) + e_i$
 - β_1 is “bivariate marginal”: difference in y_i per unit w_i
- Model: $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i) + e_i$
 - β_1 is “unique marginal”: diff in y_i per unit w_i , controlling for x_i and z_i
 - β_2 is “unique marginal”: diff in y_i per unit x_i , controlling for w_i and z_i
 - β_3 is “unique marginal”: diff in y_i per unit z_i , controlling for w_i and x_i
- Model: $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i) + \beta_4(x_i)(z_i) + e_i$
 - β_1 is “unique marginal”: diff in y_i per unit w_i , controlling for x_i and z_i
 - β_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., β_2 is a “simple” main effect slope)
 - β_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., β_3 is a “simple” main effect slope)
 - β_4 is “unique marginal” (unconditional), but how do we interpret it???

Interpreting Interaction Terms

- Model: $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i) + \beta_4(x_i)(z_i) + e_i$
 - β_4 is “unique marginal” → slope controlling for other slopes in model
 - Rather than talk about what happens to the predicted outcome y_i , interaction slopes are described by **what they do to their main effects**
- Two-way interaction has two equally correct interpretations:
 - How slope of x_i is moderated by z_i : $\beta_4 =$ difference in β_2 per unit z_i
 - How slope of z_i is moderated by x_i : $\beta_4 =$ difference in β_3 per unit x_i
- This means that effects of x_i and z_i are **linear combinations** (find common terms, factor out predictor the slope is for, and then the term in brackets is the equation for the simple effect)
 - Model-implied effect of x_i : $\beta_2(x_i) + \beta_4(x_i)(z_i) \rightarrow [\beta_2 + \beta_4(z_i)](x_i)$
 - Model-implied effect 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!

The 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 β_4 makes effect of x_i ??? per unit higher z_i
10	2	more positive
10	-2	less positive
-10	-2	more negative
-10	2	less negative

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):
 - β_1 = diff in y_i per unit w_i specifically when $z_i = 0$
 - β_2 = diff in y_i per unit x_i specifically when $z_i = 0$
 - β_3 = diff in y_i per unit z_i specifically when $w_i = 0$ and $x_i = 0$
- Interaction slopes (β_4 and β_5) are “unique marginal”
 - β_4 is now controlling for β_5 , but interpretation of β_4 is unchanged:
 - How slope of x_i is moderated by z_i : β_4 = difference in β_2 per unit z_i
 - How slope of z_i is moderated by x_i : β_4 = difference in β_3 per unit x_i
 - New β_5 has two equally correct interpretations:
 - How slope of w_i is moderated by z_i : β_5 = difference in β_1 per unit z_i
 - How slope of z_i is moderated by w_i : β_5 = difference in β_3 per 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 effects of w_i , x_i and z_i are **linear combinations** (find common terms, factor out predictor the slope is for, and then the term in brackets is the equation for the simple effect)
 - 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 examples in Hoffman 2015 ch. 2) can be used to 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 separate dummy-coded contrast**
- If you let the program create the dummy-coded contrasts for you, all needed interaction contrasts will be automatically included

- e.g., **adding an interaction of 4-category group with age (0=85):**

- New variables to be created for the model:
 - $d1 = 0, 1, 0, 0 \rightarrow$ difference between Control and Treat1
 - $d2 = 0, 0, 1, 0 \rightarrow$ difference between Control and Treat2
 - $d3 = 0, 0, 0, 1 \rightarrow$ 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 (β_5 , β_6 , and β_7) in order to test the "group*age" interaction
- Group difference slopes (β_1 , β_2 , and β_3) are each conditional on age = 85
- Age slope (β_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 variables to be created for the model:
 - $d1 = 0, 1, 0, 0 \rightarrow$ difference between Control and Treat1
 - $d2 = 0, 0, 1, 0 \rightarrow$ difference between Control and Treat2
 - $d3 = 0, 0, 0, 1 \rightarrow$ 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 simple effects**

- 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 two variables = 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 variable = 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)

Yes, we will do 3-way interactions!

- Model: $y_i = \beta_0 + \beta_1(w_i) + \beta_2(x_i) + \beta_3(z_i) + \beta_4(x_i)(z_i) + \beta_5(w_i)(z_i) + \beta_6(x_i)(w_i) + \beta_7(w_i)(x_i)(z_i) + e_i$
- **Equations for each simple (conditional) main effect:**
 - 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 : $[\beta_3 + \beta_4(x_i) + \beta_5(w_i) + \beta_7(w_i)(x_i)](z_i)$
- **Equations for each simple (conditional) 2-way interaction:**
 - 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

Summary

- Today was all about fixed effects in the model for the means
 - Within the context of GLM as a unifying starting point, but these same concepts will readily apply to longitudinal MLM
 - Output will result from REML or ML instead of ordinary least squares
- Key points to take with you:
 - Fixed slopes are tested for significance using:
 - 1 slope: Univariate Wald tests (t -value or z -value from ratio of estimate / SE)
 - 2+ slopes: Multivariate Wald tests (F or χ^2 analog for “lumps” of slopes)
 - All predictors should always have a meaningful 0 value (adjusting predictor scales is called “centering” or “recoding”)
 - You can represent one categorical predictor variable through separate dummy-coded variables, which are then “quantitative”
 - Code and some results will look different if you let the program create the dummy-coded variables for you internally (SAS CLASS, STATA i., R factor variable)
 - Model-implied fixed effects can be created through linear combinations
 - e.g., for predicted outcomes, differences among non-reference groups, or simple slopes within interaction terms (SAS ESTIMATE, STATA LINCOM, R GLHT)