General Linear Models with More than One Predictor

- Topics:
  - Review: specific and general model results
  - Special case of GLM 1:
    "Multiple (Linear) Regression" with 2+ quantitative predictors
  - Special case of GLM 2:
    "Analysis of Covariance" (ANCOVA) with both categorical and quantitative predictors
  - Non-problems and unexpected results

Review: Specific Info for Fixed Effects

- The role of **each predictor variable** $\boldsymbol{x_i}$ in creating a custom expected **outcome** $\boldsymbol{y_i}$ is described using <u>one or more fixed slopes:</u>
  - **One slope** is sufficient to capture the mean difference between two categories for a **binary** $\boldsymbol{x_i}$ or to capture a **linear effect of a quantitative** $\boldsymbol{x_i}$ (or an exponential-ish curve if x_i is log-transformed)
  - **More than one slope** is needed to capture other **nonlinear effects of a quantitative** $\boldsymbol{x_i}$ (e.g., quadratic curves or piecewise slopes)
  - $\boldsymbol{C-1}$ **slopes** are needed to capture the mean differences in the outcome across a **categorical predictor** with C categories
    - # pairwise mean differences = $\frac{C!}{2!(C-2)!}$, but only $C-1$ are given directly

- For each fixed slope, we obtain an **unstandardized** solution:
  - **Estimate, SE, $\boldsymbol{t}$-value, $\boldsymbol{p}$-value** (in which [Est−0]/SE = t, in which $DF_{num} = 1$ and $DF_{den} = N-k$ are used to find the p-value; this is a “Univariate Wald Test” (or a “modified” test given use of t, not z)
  - **Standardized** effect size can be given by converting $\boldsymbol{t}$ into $\boldsymbol{r}$ or $\boldsymbol{d}$

GLMs with Single Predictors: Review of Fixed Effects

- Predictor $\boldsymbol{x1_i}$ alone: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x1_i}) + \boldsymbol{e_i}$
  - $\boldsymbol{\beta_0}$ = **intercept** = expected $\boldsymbol{y_i}$ when $\boldsymbol{x1_i = 0}$
  - $\boldsymbol{\beta_1}$ = **slope of** $\boldsymbol{x1_i}$ = difference in $\boldsymbol{y_i}$ per one-unit difference in $\boldsymbol{x1_i}$
    - Standardized slope for $\boldsymbol{\beta_1}$ = Pearson's r for y_i with $x1_i$ ($\beta_{1std} = r_{y,x1}$)
  - $\boldsymbol{e_i}$ = difference of $y_i - \hat{y}_i$ where $\boldsymbol{\hat{y}_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x1_i})$

- Predictor $\boldsymbol{x2_i}$ alone : $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_2}(\boldsymbol{x2_i}) + \boldsymbol{e_i}$
  - $\boldsymbol{\beta_0}$ = **intercept** = expected $\boldsymbol{y_i}$ when $\boldsymbol{x2_i = 0}$
  - $\boldsymbol{\beta_2}$ = **slope of** $\boldsymbol{x2_i}$ = difference in $\boldsymbol{y_i}$ per one-unit difference in $\boldsymbol{x2_i}$
    - Standardized slope for $\boldsymbol{\beta_2}$ = Pearson's r for y_i with $x2_i$ ($\beta_{2std} = r_{y,x2}$)
  - $\boldsymbol{e_i}$ = difference of $y_i - \hat{y}_i$ where $\boldsymbol{\hat{y}_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_2}(\boldsymbol{x2_i})$

Review: General Test of Fixed Effects

- Whether the **set of fixed slopes for $\boldsymbol{x_i}$** significantly explains y_i variance (i.e., if $R^2 > 0$) is tested via "**Multivariate Wald Test**"
  - $F(DF_{num}, DF_{den}) = \frac{SS_{model}/(k-1)}{SS_{residual}/(N-k)} = \frac{(N-k)R^2}{(k-1)(1-R^2)} = \frac{known}{unknown}$
  - ***F* test-statistic** ("F-test") evaluates model R^2 *per DF spent to get to it and DF leftover* (is a ratio of info known to info unknown)
  - $\boldsymbol{R^2} = \frac{SS_{total} - SS_{residual}}{SS_{total}}$ = square of $\boldsymbol{r}$ of predicted $\boldsymbol{\hat{y}_i}$ with $\boldsymbol{y_i}$

- For GLMs with **one fixed slope**, the Univariate Wald (t) test for that slope is the same as the Multivariate Wald (F) Test for the model R^2
  - Slope $\boldsymbol{\beta_{unstd}}$: $t = \frac{Est(-H_0)}{SE}$, $\boldsymbol{\beta_{unstd}}$ = Pearson r
  - Model: $F = t^2$, $R^2 = r^2$ because predicted $\boldsymbol{\hat{y}_i}$ only uses $\boldsymbol{\beta_{unstd}}$

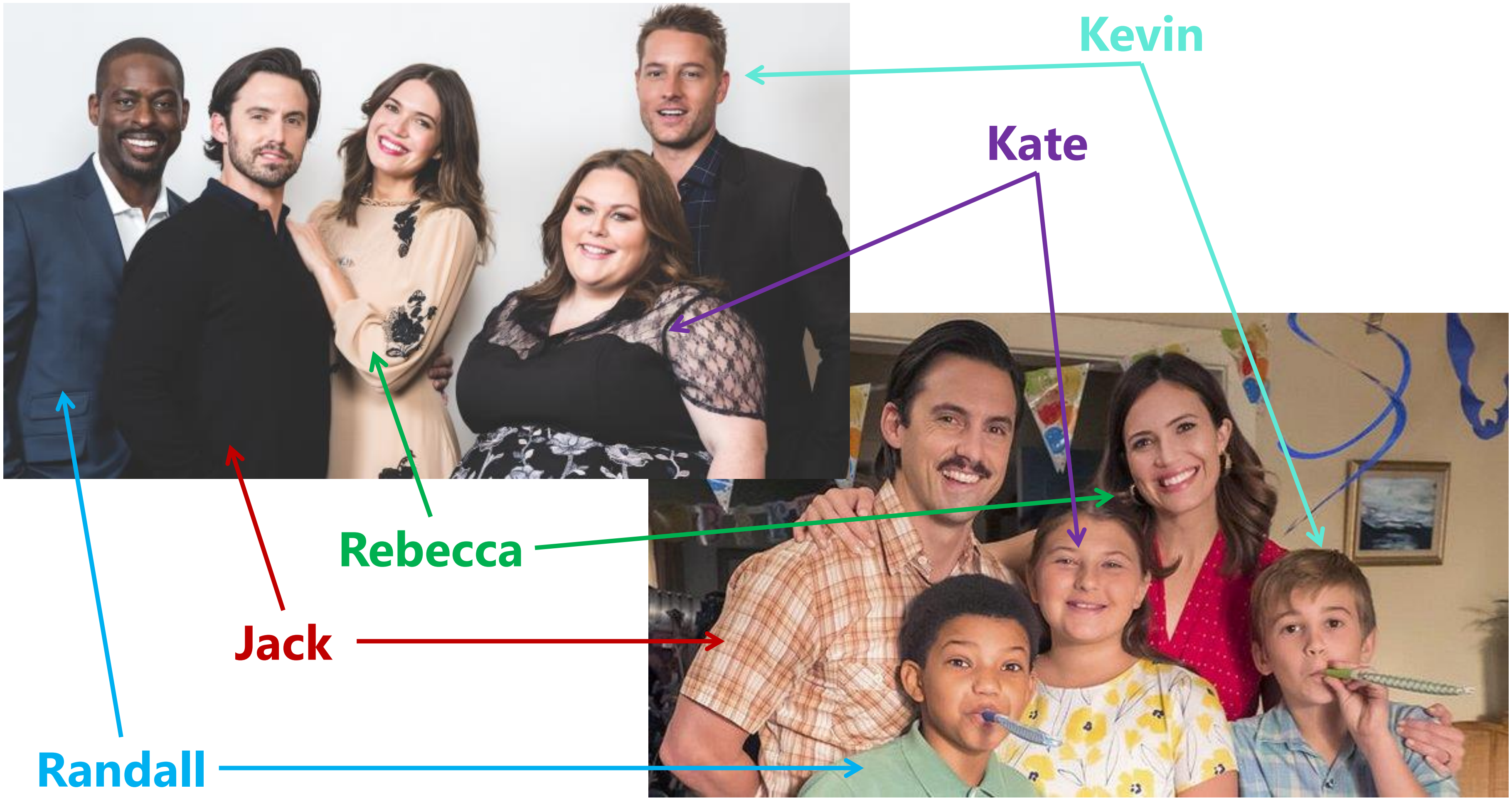
Moving On: GLMs with Multiple Predictors

- So far each **set of fixed slopes** within a separate model have **worked together** to describe the effect of a **single variable**
  - So the F-test of the model R^2 has reflected the contribution of **one predictor variable conceptually** in forming $\hat{y}_i$, albeit with one or more fixed slopes to capture its effect

- Now we will see what happens to the fixed slopes for each variable when combined into a single model that includes **multiple predictor variables**, each with its own fixed slope(s)
  - Short answer: fixed slopes go from representing "bivariate" to "**unique**" relationships (i.e., controlling for the other predictors), and $\hat{y}_i$ is created from all predictors' fixed slopes simultaneously
    - Standardized slopes are no longer equal to bivariate Pearson's r, and are instead related to a "semipartial" (or "part") correlation

A Real-World Example of "Unique" Effects

- House-cleaning with the Pearsons—the cast from "This is Us"



A Real-World Example of "Unique" Effects

- Scenario: Rebecca Has. Had. It. with 3 messy tween-agers and decides to provide an incentive for them to clean the house
  - Let's say the Pearson house has 10 cleanable rooms: 4 bedrooms, 2 bathrooms, 1 living area, 1 kitchen area, 1 dining area, 1 garage
- Incentive system for each cleaner (3 children and spouse Jack):
  - Individual: one Nintendo game per room cleaned by yourself
  - Family Bonus: if ≥ 8 rooms are clean, the family gets a new TV! (8 = average of 2 rooms per person)
- Rebecca decides to let the family decide what rooms they will each be responsible for while she is shopping for necessities
  - She returns home to a cleaner house, and asks who did what…

Pearson House: Who Cleaned What?

Room	Jack	Kevin	Kate	Randall
Master bedroom	x			
Kevin bedroom		x		
Kate bedroom			x	
Randall bedroom				x
Bathroom 1				x
Bathroom 2				x
Living area		x	x	x
Kitchen area	x			x
Dining area	x			x
Garage				

- 9/10 rooms are cleaned, so the family gets a new TV—hooray!
- But what should each person get for their <u>individual</u> effort?

Pearson House: Who Cleaned What?

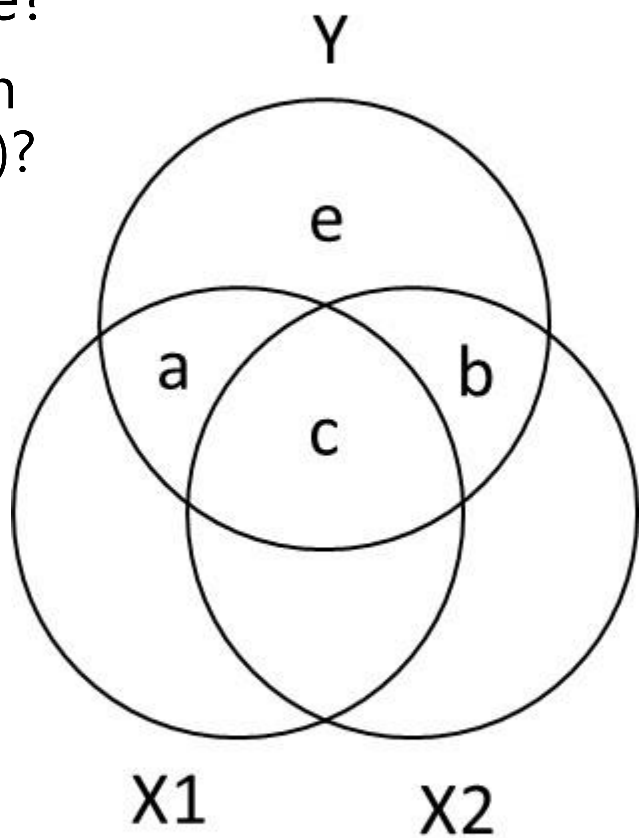
Room	Jack	Kevin	Kate	Randall
Master bedroom	x			
Kevin bedroom		x		
Kate bedroom			x	
Randall bedroom				x
Bathroom 1				x
Bathroom 2				x
Living area		x	x	x
Kitchen area	x			x
Dining area	x			x
Garage				

- Jack, Kevin, and Kate: only one Nintendo game each for cleaning one **unique** room (can't assign rewards for overlapping rooms)
- Randall: three Nintendo games for three **unique** rooms
- No one gets credit for overlapping rooms (but the family gets a TV)

From Cleaning to Modeling: 2 Goals

- General Utility: Do the model predictors explain a significant amount of variance?
  - Is the model R^2 (the squared $\boldsymbol{r}$ of $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ with $\boldsymbol{y}_{\boldsymbol{i}}$) significantly > 0 (is F-test significant)?
  - R^2 is a function of the common AND unique effects of predictor variables

- Specific Utility: What is the **unique** contribution of each predictor to the model R^2 after discounting its redundancy with the other predictors?
  - This is tricky because no predictor gets individual credit for what they have in common in predicting $\boldsymbol{y}_{\boldsymbol{i}}$, even though their common variance still increases R^2



"Ballantine" image borrowed from: http://alexanderdemos.org/Mixed1.html

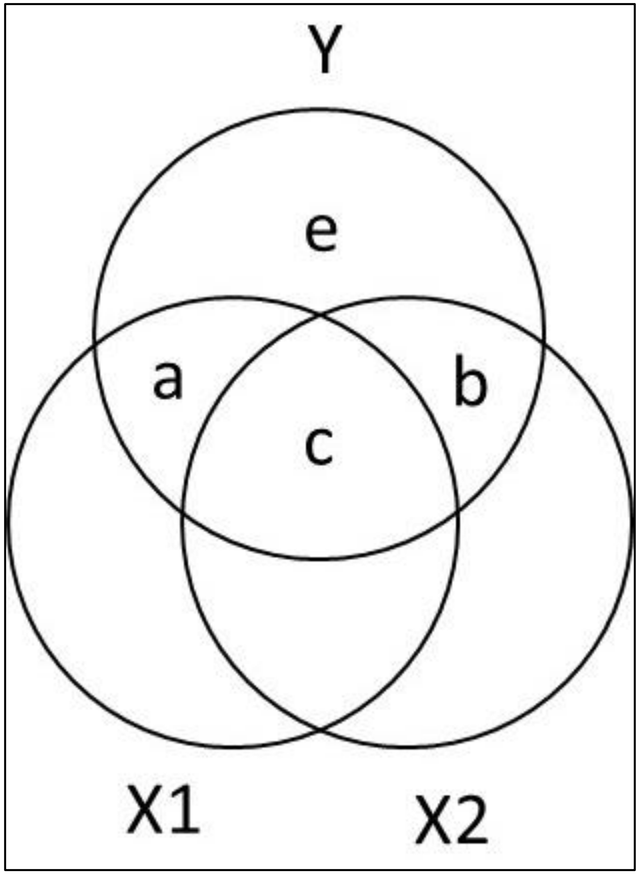
From Correlations to Standardized Slopes

- Recall for a one-predictor model: $y_i = \beta_0 + \beta_1(x1_i) + e_i$
  - Unstandardized: $\beta_0 = M_y - (\beta_1 M_{x1}), \beta_1 = r_{y,x1}\frac{SD_y}{SD_{x1}}, \beta_1 = \frac{Cov_{x1,y}}{SD^2_{x1}}$
  - Standardized: $\beta_0 = 0, \beta_{1std} = \beta_1 \frac{SD_{x1}}{SD_y}$ **(so $\beta_{1std} = r_{y,x1}$ here)**

- For a two-predictor model: $y_i = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + e_i$
  - Unstandardized: $\beta_0 = M_y - (\beta_1 M_{x1}) - (\beta_2 M_{x2})$
  - Standardized: $\beta_{1std} = \frac{r_{y,x1} - (r_{y,x2} * r_{x1,x2})}{1 - R^2_{x1,x2}}, \beta_{2std} = \frac{r_{y,x2} - (r_{y,x1} * r_{x1,x2})}{1 - R^2_{x1,x2}}$
  - Standardized to unstandardized: $\beta_1 = \beta_{1std}\frac{SD_y}{SD_{x1}}, \beta_2 = \beta_{2std}\frac{SD_y}{SD_{x2}}$

Semipartial (Part)* Correlations

- Just Pearson's r is an effect size for a bivariate (or "zero-order") relationship between two variables, **semipartial correlations provide an effect size for the relationship of $\boldsymbol{y_i}$ with the <u>unique</u> part of each $\boldsymbol{x_i}$** (i.e., after controlling each x_i for the other predictors)
  - When r is squared → R^2 → areas to the right represent "proportions of variance"
  - **Area $\boldsymbol{a}$** = $sr^2_{y,x1.x2}$ = semipartial R^2 for y_i with $x1_i$, controlling $x1_i$ for $x2_i$
  - **Area $\boldsymbol{b}$** = $sr^2_{y,x2.x1}$ = semipartial R^2 for y_i with $x2_i$, controlling $x2_i$ for $x1_i$
  - $\boldsymbol{R^2 = \frac{a+b+c}{a+b+c+e}}$ → c also contributes to R^2 but neither predictor gets "credit" for it



* Part correlations are not the same as "partial" correlations (which do not correspond to GLMs so we won't use them)

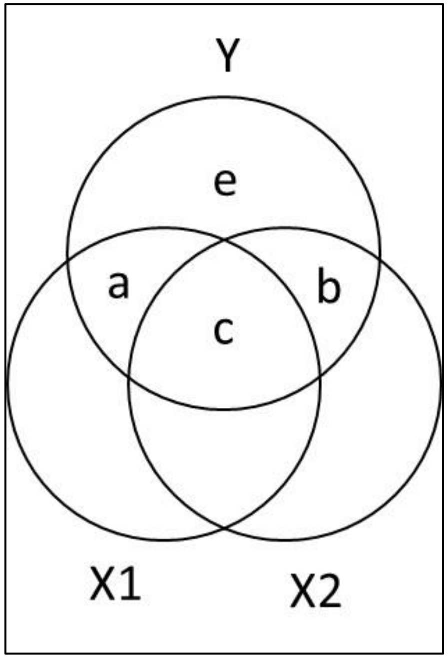
"Ballantine" image borrowed from: http://alexanderdemos.org/Mixed1.html

Standardized Slopes and Semi-Partial r

One-predictor model: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x1_i}) + \boldsymbol{e_i}$, $\boldsymbol{\beta_1} = \dfrac{\boldsymbol{Cov_{x1,y}}}{\boldsymbol{Var_{x1}}}$

For a two-predictor model: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x1_i}) + \boldsymbol{\beta_2}(\boldsymbol{x2_i}) + \boldsymbol{e_i}$

- Standardized "unique" slopes (repeated):
  - $\boldsymbol{\beta_{1std}} = \dfrac{r_{y,x1} - (r_{y,x2} * r_{x1,x2})}{1 - R^2_{x1,x2}} = \dfrac{\text{unique Cov of } y,x1}{\text{unique Var of } x1}$
  - $\boldsymbol{\beta_{2std}} = \dfrac{r_{y,x2} - (r_{y,x1} * r_{x1,x2})}{1 - R^2_{x1,x2}} = \dfrac{\text{unique Cov of } y,x2}{\text{unique Var of } x2}$
- Semipartial correlations (**.** indicates controlled):
  - $\boldsymbol{sr_{y,x1.x2}} = \sqrt{a} = \dfrac{r_{y,x1} - (r_{y,x2} * r_{x1,x2})}{\sqrt{1 - R^2_{x1,x2}}} = \dfrac{\text{unique Cov of } y,x1}{\text{unique SD of } x1}$
  - $\boldsymbol{sr_{y,x2.x1}} = \sqrt{b} = \dfrac{r_{y,x2} - (r_{y,x1} * r_{x1,x2})}{\sqrt{1 - R^2_{x1,x2}}} = \dfrac{\text{unique Cov of } y,x2}{\text{unique SD of } x2}$



$$R^2 = \frac{a+b+c}{a+b+c+e}$$

Where the "Common" Area c Goes

- Model R^2 can be understood in many ways—here, for two slopes:
  - Old: R^2 is the square of the r between predicted $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ and $\boldsymbol{y}_{\boldsymbol{i}}$
  - Old said differently: $R^2 = \frac{Var_{\hat{y}_i}}{Var_{y_i}}$
  - New: $R^2 = \frac{r^2_{y,x1} + r^2_{y,x2} - (2 * r_{y,x1} * r_{y,x2} * r_{x1,x2})}{1 - R^2_{x1,x2}}$
  - New: $R^2 = \beta_{1std * r_{y,x1}} + \beta_{2std * r_{y,x2}}$
  - New: $R^2 = \beta^2_{1std} + \beta^2_{2std} + (2 * \beta_{1std} * \beta_{2std} * r_{x1,x2})$

Note: The version of this slide given in class had errors in the first two "new" R^2 formulas (now corrected)

- In general: R^2 = unique effects + function of common effects

GLMs with Multiple Predictors: New Interpretation of Fixed Effects

- Two predictor variables: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x1_i}) + \boldsymbol{\beta_2}(\boldsymbol{x2_i}) + \boldsymbol{e_i}$
  - $\boldsymbol{\beta_0}$ = **intercept** = expected $\boldsymbol{y_i}$ when $\boldsymbol{x1_i} = \boldsymbol{0}$ AND when $\boldsymbol{x2_i} = \boldsymbol{0}$
  - $\boldsymbol{\beta_1}$ = **slope of $\boldsymbol{x1_i}$** = unique difference in $\boldsymbol{y_i}$ per one-unit difference in $\boldsymbol{x1_i}$ "controlling for" or "partialling out" or "holding constant" $\boldsymbol{x2_i}$ (so $\beta_{1std} \neq$ Pearson's bivariate $r_{y,x1}$ whenever $r_{x1,x2} \neq 0$)
  - $\boldsymbol{\beta_2}$ = **slope of $\boldsymbol{x2_i}$** = unique difference in $\boldsymbol{y_i}$ per one-unit difference in $\boldsymbol{x2_i}$ "controlling for" or "partialling out" or "holding constant" $\boldsymbol{x1_i}$ (so $\beta_{2std} \neq$ Pearson's bivariate $r_{y,x2}$ whenever $r_{x1,x2} \neq 0$)

- These unstandardized fixed effects (intercept and slopes) do create predicted $\boldsymbol{\hat{y}_i}$ in the original scale of y_i, but they cannot be used to ascertain the relative importance of each predictor to the model because unstandardized fixed effects are scale-dependent (units matter)

Standard Errors of Each Fixed Slope

- Standard Error (SE) for fixed effect estimate β_X in a one-predictor model (SE is like the SD of the estimated slope across samples):

$$\mathrm{SE}_{\beta_X} = \sqrt{\frac{\text{residual variance of Y}}{\mathrm{Var(X)}*(N-k)}}$$

N = sample size
k = number of fixed effects

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

$$\mathrm{SE}_{\beta_X} = \sqrt{\frac{\text{residual variance of Y}}{\mathrm{Var(X)}*\left(\mathbf{1-R_X^2}\right)*(N-k)}}$$

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

- So all things being equal, SE (index of inconsistency) 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 than before
  - The predictor has less covariance with other predictors
    - Best case scenario: x_i is uncorrelated with all other predictors

- If SE is smaller → t-value (or z-value) is bigger→ p-value is smaller

Effect Sizes of Single Fixed Slopes

- When a predictor's effect is captured by a single fixed slope, these unstandardized slopes have **two potential (related) effect sizes** <u>that indicate their importance relative to other predictors' slopes</u>:

  - **Standardized slope ($\boldsymbol{\beta_{std}}$)** is from a "standardized" solution in which all variables have $M = 0, SD = 1$ (often labeled as "beta" in output)
    - Provided in SAS PROC REG or in STATA REGRESS with "beta" option
    - Can also get by z-scoring all variables, then doing usual GLM

  - **Squared semipartial $\boldsymbol{r}$ ($\boldsymbol{sr_x^2}$)**, also known as **eta-squared ($\boldsymbol{\eta^2}$)**, gives the contribution to the model R^2 of a single slope or set of slopes
    - SAS: Provided in PROC REG (as SCORR2 option on MODEL)
    - STATA: Provided in separate PCORR routine (NOT same as ESTAT ESIZE)
    - Btw, semipartial r (sr_x) is related to β_{std} as $\beta_{std} = \frac{sr_x}{\sqrt{1-R_x^2}} = \frac{sr_x}{unique\ SD_x}$
    - Btw, adjacent omega-squared (ω^2) version is analogous to adjusted-R^2
    - Can be confused with "partial" eta-squared, which partials the other predictors out of outcome, too (and which is thus less helpful to know)

Effect Sizes for Multiple Fixed Slopes

- When a **single predictor's effect** is captured by **multiple fixed slopes**, a standardized slope can still provide an effect size for each
- But **slope-specific squared semipartial $\boldsymbol{r}$ ($\boldsymbol{sr^2}$) values should not be used** because of the dependency between the slopes
  - e.g., $\boldsymbol{Income_i = \beta_0 + \beta_1(Age_i - 18) + \beta_2(Age_i - 18)^2 + e_i}$
    - Linear age slope $\boldsymbol{\beta_1}$ is specific to when centered age = 0, so its unique $\boldsymbol{sr^2_{\beta_1}}$ would change if age were centered differently (even though the model R^2 and the F-test of its significance would be the same)
  - e.g., $\boldsymbol{Income_i = \beta_0 + \beta_1(LvsM_i) + \beta_2(LvsU_i) + e_i}$
    - Slopes $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ share a common reference (low group) and imply (at least) 3 possible group differences (2 in model; 1 as linear combination)
    - So the squared semipartial $\boldsymbol{r}$ values across these group differences will sum to more than they should (given a single 3-category predictor)

- What to do instead? Request a multivariate squared semipartial $\boldsymbol{r}$ **($\boldsymbol{sr^2}$)** that is estimated jointly across the fixed slopes… stay tuned!

Special Cases of the GLM, continued

- So far we've seen what many would call "**multiple (linear) regression**"—this term generally refers to GLMs with quantitative predictors (and "linear" differentiates normal residuals specifically)
  - However, all of these concepts hold for categorical predictors (i.e., as called "**analysis of covariance**" when they are paired with quantitative predictors) or for quantitative predictors that use 2+ fixed slopes, for which one extra piece of info is often of interest
- For example, consider a quantitative predictor $x1_i$ paired with a three-category predictor of low vs. middle vs. upper class:

$$\boldsymbol{y_i = \beta_0 + \beta_1(x1_i) + \beta_2(LvsM_i) + \beta_3(LvsU_i) + e_i}$$

  - **Multivariate Wald $\boldsymbol{F}$-Test** of model R^2 includes all 3 fixed slopes
  - To see if **3-category group contributes significantly** to the R^2, we need to request another Multivariate Wald F-test **for just $\boldsymbol{\beta_2}$ and $\boldsymbol{\beta_3}$**
  - The same request is needed when considering multiple slopes for any quantitative predictor variable—is their **joint effect** significant?

Multivariate Wald Tests for Joint Effects

- The F-test for the model R^2 was your first instance of a **Multivariate Wald (F) Test**, a general way to jointly test the significance of **2 or more fixed slopes simultaneously**
  - Remember, for one fixed slope, its $t^2 = F$, so separate F is not needed

- These F-tests can be requested for any combo of fixed slopes:
  - SAS: CONTRAST within PROC GLM (which also provides effect sizes) or TEST within PROC REG (which does not provide effect sizes)
  - STATA: TEST within REGRESS (which does not provide effect sizes)
  - They take into account the covariance among the predictors as part of the test of their joint contribution to the model R^2
  - Can choose hierarchical (Type I SS) or not (Type II, III, or IV SS), but hierarchical (in which order of predictors matters) is rarely needed
  - Can replace “hierarchical regression” (entering predictors in sequence) as a way to test the change in the model R^2 for new fixed slopes

Joint squared semi-partial correlations

- In SAS PROC GLM, CONTRAST can be used to get Multivariate Wald F-tests for 2+ slopes simultaneously AND get their joint sr^2 value
  - Takes into account any covariance among predictors (i.e., dependency between slopes) due to a common reference group
  - sr^2 value is labeled as "Semipartial Eta-Square" on SAS output

- In STATA, TEST can be used to get F-tests but not joint sr^2 values for 2+ slopes, but you can compute them using unique sums of squares (SS)
  - Step 1: From the full model, get SS for the model: SS_{Full}
    From the full model, get SS for corrected total: SS_{Total}
  - Step 2: Get the model SS from <u>a reduced model</u> without
    the slopes for which you want a joint test: $SS_{Reduced}$
  - Step 3: Compute SS difference between models: $SS_{Test} = SS_{Full} - SS_{Reduced}$
  - Step 4: Compute squared semipartial r: $sr^2 = \frac{SS_{Test}}{SS_{Total}}$
  - Tip: STATA is a calculator if you type "display" and then the math you want it to do (the result of the calculation shows in output)

General Linear Model Assumptions

- For us to believe in the accuracy of our model results, the following assumptions must be plausible (some are testable, some are not)
- *Because we have selected the GLM specifically:*
  - Individual residual $\boldsymbol{e_i}$ values (i.e., as formed from $\boldsymbol{y_i} - \boldsymbol{\hat{y}_i}$) are independent and normally distributed with constant variance (i.e., with homoscedasticity) across predicted outcomes and predictor variables
    - If not independent, need multilevel (mixed-effects) models
    - If not normal/constant variance, may need generalized linear models (or general linear models that allow heterogeneous variance by predictors)
    - If not both, need generalized multilevel (mixed-effects) models

- *Also applicable to any linear model using observed variables:*
  - All variables are measured without error (what "structural equation models" with regressions among "latent" variables try to solve; see PSQF 6249)
  - All predictors have their effects specified in the correct form (e.g., no missing nonlinearity or non-additivity in their effects)
  - Any predictors not included would have had a fixed slope ≈ 0 (only testable with the predictors you have measured, unfortunately)

Dealing with Problematic Reviewers

- A frequently claimed non-problem is "**multicolinearity**" (see also "multicollinearity" or just "colinearity" or "collinearity")
- As shown before, the SE for a predictor's slope will be greater to the extent that the predictor has in common (more correlation) with the other predictors—that makes it harder to determine its unique effect
- Diagnostics for this supposed danger are given in many forms
  - "**tolerance**" = unique predictor variance = $1 - R_X^2$ (<.10 = "bad")
  - "**variance inflation factor**" (VIF) = 1/tolerance (> 10 = "bad")
  - Computers used to have numerical stability problems with high collinearity, but these problems are largely nonexistent nowadays
- Only when you have "**singularity**" is it actually a problem—when one predictor is a perfect linear combination of the others (such as when including two subscale scores AND their total as predictors)
  - It's always a good idea to examine the bivariate relationships among your to-be-modeled predictors to see to what extent they are redundant for conceptual reasons to consider the possibility of "equivalent" models

Problematic Participants

- Fixed effects estimates are the values that collectively minimize the sum of the squared residuals for the sample (→ smallest residual variance; "OLS")
- Extreme values can have undue influence on these estimates, though
  - "**Distance**" = extreme on e_i (e.g., A, C)
  - "**Leverage**" = extreme on x_i (e.g., B)
  - "**Influence**" = impact on slope (e.g., C)
    - Measured by absolute value of change in slope if each case were removed: Cook's D = $(\beta_{1new} - \beta_1)^2$
- Btw, "quantile regression" can avoid bias in results due to cases with influence
  - Predict median (or any percentile) instead of mean; in my generalized models class

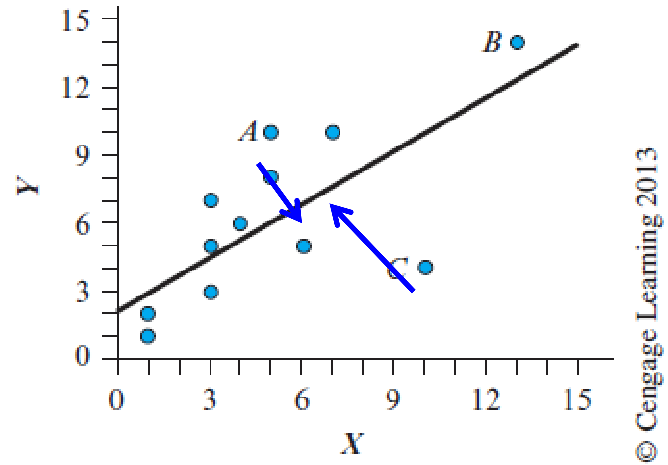


Figure 15.5 Scatterplot of Y on X

What to do with any high influence cases? There are no good uniform solutions… it depends on how much you believe the aberrant cases are representative…

Image borrowed from: Howell, D. C. (201). *Statistical methods for psychology* (8th ed). Belmont, CA: Cengage Wadsworth.

Unexpected Results: Suppression

- *In general*, the semipartial r for each predictor (and its unique standardized effect) will be smaller in magnitude than the bivariate r (and its standardized effect when by itself) with y_i
- However, this will not always be the case given **suppression**: when the relationship between the predictors is hiding (suppressing) their "real" relationship with the outcome
  - Occurs given $r_{y,x1} > 0$ and $r_{y,x2} > 0$ in three conditions: (a) $r_{y,x1} < r_{y,x2} * r_{x1,x2}$, (b) $r_{y,x2} < r_{y,x1} * r_{x1,x2}$, or (c) $\boldsymbol{r_{x1,x2} < 0}$
  - For example: Consider y_i = sales success as predicted by $x1_i$= assertiveness and $x2_i$= record-keeping diligence
    - $r_{y,x1} = .403, r_{y,x2} = .127$, and $r_{x1,x2} = -.305$ (so is condition c)
    - Standardized: $\hat{y}_i = 0 + 0.487(x1_i) + 0.275(x2_i)$
    - So these standardized slopes (for the predictors' unique effects) are greater than their bivariate correlations with the outcome!
- This is one of the reasons why you cannot anticipate just from bivariate correlations what will happen in a model with multiple predictors…

Example taken from Cohen, Cohen, Aiken, & West (2002)

Unexpected Results: Multivariate Power

Correlations

		Y	X1	X2	X3	X4	X5
Y	Pearson Correlation	1	.191	.192	.237	.174	.110
	Sig. (2-tailed)	.	.119	.117	.081	.155	.371
	N	68	68	68	68	68	68
X1	Pearson Correlation	.191	1	-.250*	-.077	-.079	-.110
	Sig. (2-tailed)	.119	.	.039	.535	.521	.371
	N	68	68	68	68	68	68
X2	Pearson Correlation	.192	-.250*	1	-.077	.361**	.013
	Sig. (2-tailed)	.117	.039	.	.532	.003	.917
		68	68	68	68	68	68
X3	Pearson Correlation	.237	-.077	-.077	1	.203	.219
	Sig. (2-tailed)	.081	.535	.532	.	.098	.073
	N	68	68	68	68	68	68
X4	Pearson Correlation	.174	-.079	.361**	.203	1	.162
	Sig. (2-tailed)	.155	.521	.003	.098	.	.187
	N	68	68	68	68	68	68
X5	Pearson Correlation	.110	-.110	.013	.219	.162	1
	Sig. (2-tailed)	.371	.371	.917	.073	.187	.
	N	68	68	68	68	68	68

*. Correlation is significant at the 0.05 level (2-tailed).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-350.742	195.472		-1.794	.078
	X1	3.327	1.376	.290	2.418	.019
	X2	2.485	1.185	.271	2.098	.040
	X3	3.125	1.479	.257	2.112	.039
	X4	.366	1.342	.035	.273	.786
	X5	.844	1.309	.077	.644	.522

Even though none of these five predictors has a significant bivariate correlation with y_i, they still combined to create a significant model R^2

$$F(5,62) = 2.77,$$
$$MSE = 272631.57,$$
$$p = .025, R^2 = .183$$

This is most likely when the predictors have little correlation amongst themselves (and thus can contribute uniquely)

Example borrowed from: https://psych.unl.edu/psycrs/statpage/mr_rem.pdf

Unexpected Results: Null Washout

Correlations

		P1	P2	P3	P4	P5	P6	P7	P8	P9
Y	Pearson Correlation	.230	.059	.004	.079	-.100	-.028	-.040	-.007	.013
	Sig. (2-tailed)	.002	.432	.953	.294	.186	.709	.595	.927	.863
	N	177	177	177	177	177	177	177	177	177

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	100.454	17.866		5.623	.000
	P1	.115	.038	.233	3.047	.003
	P2	4.511E-02	.077	.044	.583	.561
	P3	-1.93E-02	.076	-.019	-.254	.800
	P4	7.511E-02	.076	.075	.988	.325
	P5	-9.22E-02	.070	-.099	-1.320	.189
	P6	6.555E-04	.077	.001	.009	.993
	P7	-4.86E-02	.076	-.048	-.640	.523
	P8	-4.13E-02	.073	-.044	-.568	.571
	P9	6.592E-03	.076	.007	.087	.931

Even though P1 has a significant bivariate correlation with y_i and a significant unique effect, the model R^2 is not significant—because it measures the <u>average</u> predictor contribution

$$F(9,167) = 1.49,$$
$$MSE = 93.76,$$
$$p = .155, R^2 = .074$$

Example borrowed from: https://psych.unl.edu/psycrs/statpage/mr_rem.pdf

Unexpected Results: A Significant Model R^2 with No Significant Predictors???

		Y	P1	P2	P3	P4	P5
Y	Pearson Correlation	1	.298**	.198**	.221**	.221**	.251**
	Sig. (2-tailed)	.	.000	.008	.003	.003	.001
	N	177	177	177	177	177	177
P1	Pearson Correlation	.298**	1	.689**	.712**	.742**	.728**
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000
	N	177	177	177	177	177	177
P2	Pearson Correlation	.198**	.689**	1	.499**	.500**	.520**
	Sig. (2-tailed)	.008	.000	.	.000	.000	.000
	N	177	177	177	177	177	177
P3	Pearson Correlation	.221**	.712**	.499**	1	.471**	.494**
	Sig. (2-tailed)	.003	.000	.000	.	.000	.000
	N	177	177	177	177	177	177
P4	Pearson Correlation	.221**	.742**	.500**	.471**	1	.593**
	Sig. (2-tailed)	.003	.000	.000	.000	.	.000
	N	177	177	177	177	177	177
P5	Pearson Correlation	.251**	.728**	.520**	.494**	.593**	1
	Sig. (2-tailed)	.001	.000	.000	.000	.000	.
	N	177	177	177	177	177	177

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	93.378	1.899		49.184	.000
	P1	.115	.080	.244	1.441	.151
	P2	-1.23E-02	.073	-.017	-.169	.866
	P3	1.555E-02	.076	.022	.206	.837
	P4	-4.41E-03	.077	-.006	-.057	.954
	P5	5.211E-02	.074	.076	.707	.481

This model R^2 is definitely significant:

$F(5{,}171) = 3.455$,
$MSE = 89.85$,
$p = .005, R^2 = .190$

Yet no predictor has a significant <u>unique</u> effect—this is because of their strong(ish) correlations with each other (and "common" still contributes to R^2)

Example borrowed from: https://psych.unl.edu/psycrs/statpage/mr_rem.pdf

GLM with Multiple Predictors: Summary

- For any GLM with multiple fixed slopes, we want to know:
  - Is each slope significantly ≠ 0? Check p-value for t = (Est−0)/SE
  - What is each slope's effect size? Check r (=β_{std}) or find d from t
  - Do the slopes join to create a model R^2 > 0? Check p-value for F
  - What is the model's effect size? Check R^2 (r of $\boldsymbol{\hat{y}_i}$ with $\boldsymbol{y_i}$)^2

- When combining the fixed slopes from different conceptual predictor variables into the same model, we also want to know:
  - Is each slope *still* significantly ≠ 0? If yes, has a "unique" effect
    - Unique effect is *usually* smaller than bivariate effect (but not necessarily)
    - 1 slope: check p-value for t = (Est−0)/SE
    - >1 slopes: check p-value for F-test of joint effect (requested separately)
  - What is the effect size for each predictor's unique effect?
    - 1 slope: check sr^2 (or β_{std}) or find "adjusted" d or r from t
    - >1 slopes: check joint sr^2 for predictor's overall contribution to R^2