General Linear Models with More than One Conceptual Predictor

- Topics:
  - Review: specific and general model results
  - Unique effect sizes: standardized slopes; semi-partial (part) and partial versions of correlation and squared versions
  - Special cases of GLM (and corresponding effect sizes):
    - "Multiple (Linear) Regression" with 2+ quantitative predictors
    - "Analysis of Covariance" (ANCOVA) with both categorical and quantitative predictors—requires joint significance tests and effect sizes
  - Some examples of 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 $\boldsymbol{x_i}$ 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 spline 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)
  - Effect size can be given by converting $\boldsymbol{t}$**-value** into **partial** $\boldsymbol{r}$ 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 $x1_i$ ($\beta_{1std} = r_{y,x1}$)
  - $\boldsymbol{e_i}$ = discrepancy from $y_i - \hat{y}_i$ 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 $x2_i$ ($\beta_{2std} = r_{y,x2}$)
  - $\boldsymbol{e_i}$ = discrepancy from $y_i - \hat{y}_i$ 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 describing the relation of $\boldsymbol{x_i}$** with y_i significantly explains y_i variance (i.e., if $R^2 > 0$) is tested via a "**Multivariate Wald Test**" (usually with F using denominator DF, or χ^2 otherwise)
  - $F(DF_{num}, DF_{den}) = \frac{SS_{model}/(k-1)}{SS_{residual}/(N-k)} = \frac{(N-k)R^2}{(k-1)(1-R^2)} = \frac{known}{unknown}$
  - $\boldsymbol{F}$ **test-statistic** ("F-test") evaluates model R^2 *per slope spent to get to it AND per slope leftover* (is weighted 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 $\boldsymbol{y_i}$; also the proportion reduction in residual variance relative to empty model
  - $\boldsymbol{R^2_{adj}} = 1 - \frac{(1-R^2)(N-1)}{N-k-1} = 1 - \frac{MS_{residual}}{MS_{total}}$ = correction used for small N

- For GLMs with **only 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 $\boldsymbol{\beta}_{unstandardized}$: $t = \frac{Est(-H_0)}{SE}$, $\boldsymbol{\beta}_{standardized}$ = Pearson r
  - Model: $F = t^2$, $R^2 = r^2$ because predicted $\boldsymbol{\hat{y}_i}$ only uses $\boldsymbol{\beta}_{unstd}$

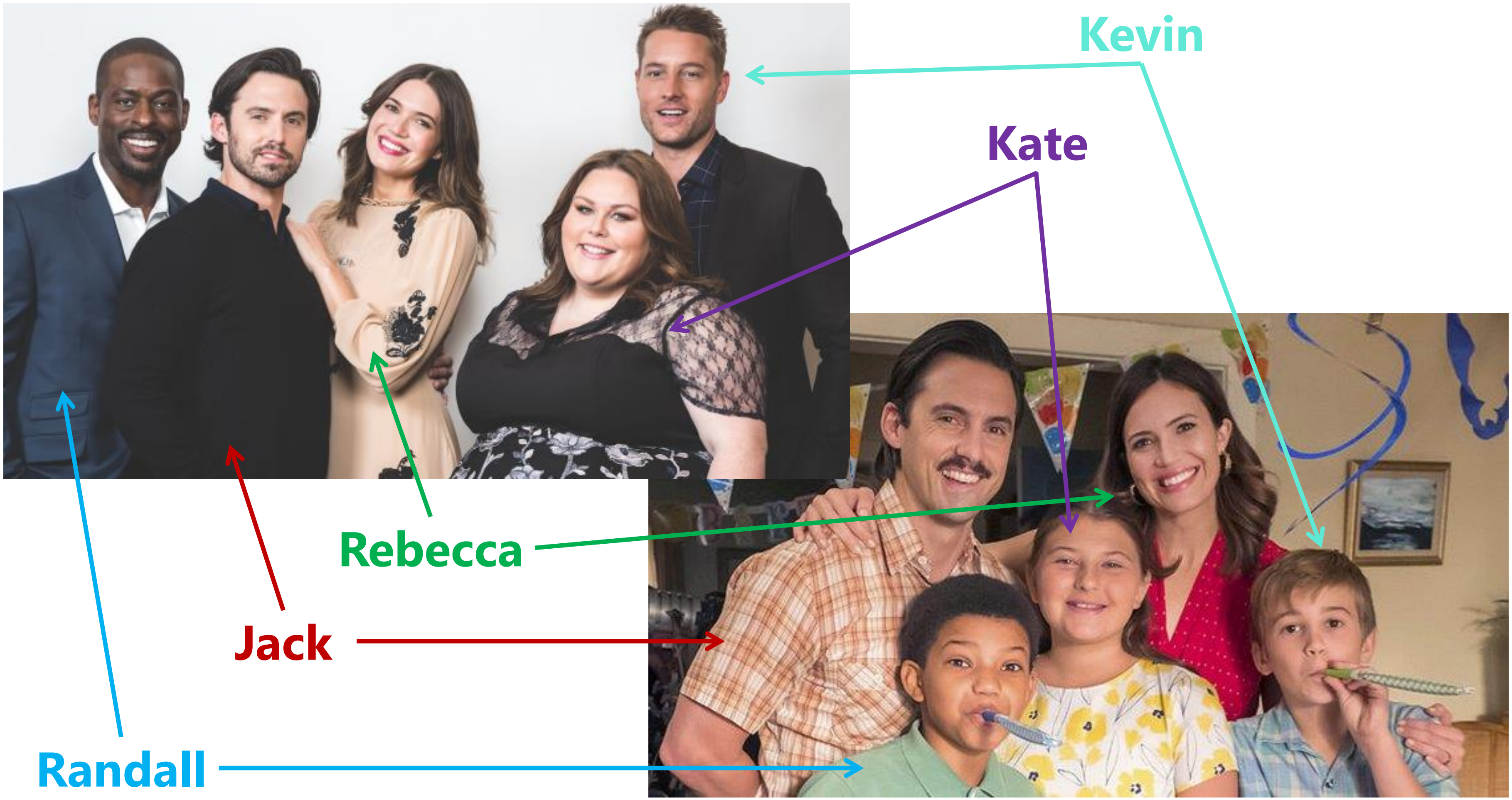
Moving On: GLMs with Multiple Predictors

- So far each **set of fixed slopes** within a separate model have **worked together** to describe the effect of a **single variable**
  - Thus, 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 relationship to y_i

- 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
    - Multiple possible metrics by which to quantify "unique" effect size

A Real-World Analog 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 (didn't get cleaned)				

- 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 (didn't get cleaned)				

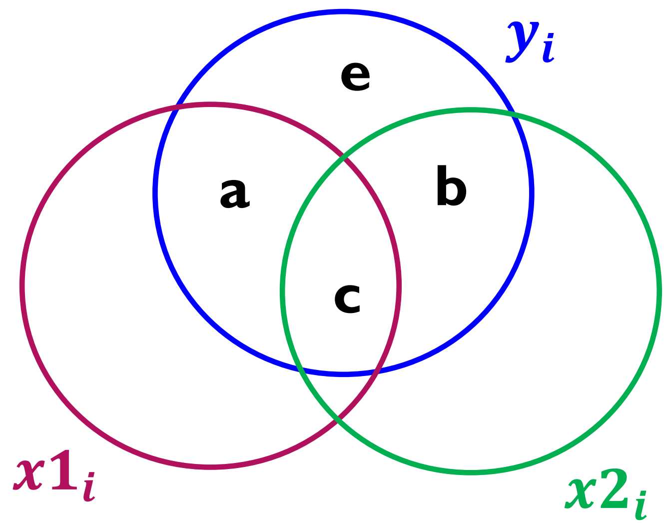
- Jack, Kevin, and Kate: only one Nintendo game each for cleaning one **unique** room (can't assign rewards for overlapping rooms)
- Randall: three Nintendo games for three **unique** rooms
- No one gets credit for overlapping rooms (but the family gets a TV)

From Cleaning to Modeling: 2 Goals

1. **General Utility:** Do the model predictors explain a significant amount of variance?
   - Is the model R^2 (the $\boldsymbol{r^2}$ of $\hat{\boldsymbol{y}}_i$ with $\boldsymbol{y}_i$) significantly > 0 (is F-test significant)?
   - **Model R^2 is includes shared AND unique effects of predictor variables: for diagram on right, $R^2 = \frac{a+b+c}{a+b+c+e}$**

2. **Specific Utility:** What is each predictor's **unique** contribution to the model R^2 after discounting (i.e., controlling for) its redundancy with the other predictors?
   - **No predictors get credit for what they have in common** (area c on the right) in predicting $\boldsymbol{y}_i$, even though that shared variance still increases the R^2

Areas below describe partitions of y_i variance:
a = y_i unique to $x1_i$
b = y_i unique to $x2_i$
c = y_i shared by $x1_i$ and $x2_i$
e = y_i leftover (residual)



GLMs with Multiple Predictors: New Interpretation of Fixed Effects

- GLM with 2 predictor variables: $\boldsymbol{y_i = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + e_i}$
  - $\boldsymbol{\beta_0}$ = **intercept** = expected $\boldsymbol{y_i}$ when $\boldsymbol{x1_i = 0}$ AND when $\boldsymbol{x2_i = 0}$
  - $\boldsymbol{\beta_1}$ = **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 $r_{x1,x2} \neq 0$)
    - But $\boldsymbol{\beta_1}$ is still assumed to be constant over all values of $\boldsymbol{x2_i}$ (and $\boldsymbol{x1_i}$)
  - $\boldsymbol{\beta_2}$ = **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 $r_{x1,x2} \neq 0$)
    - But $\boldsymbol{\beta_2}$ is still assumed to be constant over all values of $\boldsymbol{x1_i}$ (and $\boldsymbol{x2_i}$)
  - Here $x1_i$ and $x2_i$ have "additive effects" (effect = slope in this context)… *stay tuned for "multiplicative effects" via interaction terms in unit 5!*

Btw: From Pearson Correlations and Covariances to Standardized Slopes

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

- 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: $\beta_1 = \beta_{1std}\frac{SD_y}{SD_{x1}}$, $\beta_2 = \beta_{2std}\frac{SD_y}{SD_{x2}}$

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 $\hat{y}_i$ and y_i
  - Old R^2 said differently: $R^2 = \frac{Var_{\hat{y}_i}}{Var_{y_i}} = \frac{explained\ variance}{total\ variance}$
  - New: $R^2 = \frac{r_{y,x1}^2 + r_{y,x2}^2 - (2 * r_{y,x1} * r_{y,x2} * r_{x1,x2})}{1 - R_{x1,x2}^2}$
  - New: $R^2 = \beta_{1std}^2 + \beta_{2std}^2 + (2 * \beta_{1std} * \beta_{2std} * r_{x1,x2})$
- In general: $\boldsymbol{R^2}$ **= unique effects + function of common effects**
  - General effect size for magnitude of prediction by the model
- The standard errors of each "unique" slope also must be adjusted to reflect the unique variance of its predictor variable relative to other predictor variables…

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):

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

N = sample size
k = number of fixed effects

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

$$\text{SE}_{\beta_x} = \sqrt{\frac{\text{residual variance of Y}}{\text{Var}(x_i) * \left(\mathbf{1 - R_x^2}\right) * (N-k)}}$$

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 (index of inconsistency) is smaller when:
  - More of the outcome variance has been reduced (better predictive model)
    - So fixed slopes can become significant if added later (if R^2 is higher than before)
  - The predictor has less correlation 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

Recommended Model-Building Strategies

- **Step 0:** Create new variables out of each conceptual predictor
  - Quantitative: center (subtract a constant) so that 0 is meaningful
  - Categorical: represent differences using dummy-coded (0/1) predictors

- **Step 1:** Examine **bivariate** relations of each conceptual predictor with y_i
  - "Bivariate" = "zero-order" relation for two variables (x_i and y_i)
  - For a quantitative or binary predictor that has a linear relation with y_i, its bivariate relation is given by Pearson correlation r (use matrix for many)
    - Square of Pearson r = "shared variance" for x_i and y_i
  - Otherwise, you need a GLM for each conceptual predictor in order to include multiple fixed slopes (e.g., 3+ categories; linear+quadratic slopes)
    - Model R^2 = "shared variance" for x_i and y_i

- **Step 2:** Examine bivariate relations of each conceptual predictor with the other predictors—useful to get a sense of how they will compete with each other when combined into the same model predicting y_i
  - Via correlation matrices when possible, using models otherwise
  - Quantify shared variance using same process as in step 1

Recommended Model-Building Strategies

- **Step 3**: Combine conceptual predictors into the same model in whatever way corresponds to your **research questions**… here are two examples:

- **Simultaneous**: How does y_i relate to $x1_i$, $x2_i$, and $x3_i$?
  - Put all slopes into same model—report model test (F for R^2), as well as direction, significance, and effect size per predictor (stay tuned for options)

- **Stepwise using $\boldsymbol{R^2}$ change:** (a) After controlling for $x1_i$, how does $x2_i$ predict y_i? (b) After controlling for $x1_i$ AND $x2_i$, how does $x3_i$ predict y_i?
  - (a) **Put $\boldsymbol{x1_i}$** into model and report its direction, significance, and effect size. **Add $\boldsymbol{x2_i}$** into model—report model test (F for R^2), change in model test (F for R^2 change), as well as $x2_i$ direction (also significance and effect size per slope if not redundant with change in model test). Comment on how the slope(s) for $x1_i$ changed after $x2_i$.
  - (b) **Add $\boldsymbol{x3_i}$** into model—report model test (F for R^2), change in model test (F for R^2 change), as well as $x3_i$ direction (also significance and effect size if not redundant with change in model test). Comment on how $x1_i$ and $x2_i$ slopes changed after $x3_i$.
- Stepwise strategy is useful if there is a clear hierarchy for the inclusion of predictors, but if not, a simultaneous strategy is more defensible!
  - I will show you how to get unique contributions for a set of slopes from same model!
  - Btw, atheoretical automated routines can also find optimal combos of predictors…

What about "Multicolinearity"? Meh.

- A frequently worried-about problem is "**multicolinearity**" *(see also "multicollinearity" or just "colinearity" or "collinearity")*
- 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 overhyped 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 truly a problem**—when a predictor is a perfect linear combination of the others (redundant)
  - e.g., when including two subscale scores AND their total as predictors
  - e.g., when including intercept + 3 dummy-coded predictors for 3 groups
  - You will get a row of dots instead of results for redundant predictors

Addressing (Multi)Collinearity

- Use the bivariate relationships among your to-be-considered predictors to guide the possibility of "equivalent" models
  - e.g., invasive biological measure vs. highly related but non-invasive alternative measure—can one sufficiently replace the other?

- Such questions require comparing non-nested models
  - **Nested** = one model **is a subset** of other (model A vs. model A+B+C)
    - Btw, I will show you how to test nested models using just one model
  - **Non-nested** = models are **not subsets** (model A+B vs. model A+C)
    - "Hotelling's t" can be used for significance test of R from each model (must save $\hat{y}_i$ for each model and compute their correlation first)
  - See also "dominance analysis" (see Darlington & Hayes 2016, sec. 8.3)

- Or just try to reduce the slope SEs by adding predictors that are related to y_i but that are (mostly) unrelated to other predictors
  - Less residual variance → smaller SE for each predictor → more power

Metrics of Effect Size per Fixed Slope

- Unstandardized fixed slopes **cannot be used to ascertain the relative importance** of each predictor because **they are scale-dependent** (so differences in "one unit" matter)

- So we also need to report **some kind of "unique" effect size**
  - Could be relevant **per fixed slope** (for predictors whose effect on y_i is described by a single slope) or **per conceptual predictor** (for predictors whose effect on y_i require multiple slopes to describe)
  - Why? Beyond putting the slope magnitudes on same scale, specific effect sizes are also used in **meta-analyses** and to **predict power**
  - Choices in $\boldsymbol{r}$ **metric**: standardized slopes (which are not really correlations, see next slide), semi-partial r, or partial r → r gets called $\boldsymbol{\eta}$ ("**eta**" when using R^2) or $\boldsymbol{\omega}$ ("**omega**" when adjusted by N, to be used with adjusted R^2)
    - Btw, also Cohen's d in standardized mean difference metric—is "partial" version
  - Fewer useful in $\boldsymbol{R^2}$ **metric**: semi-partial $\boldsymbol{\eta^2}$ or $\boldsymbol{\omega^2}$; see also Cohen's $\boldsymbol{f^2}$

- *Let's examine more closely how these differ from each other…*

Standardized Slopes: Confusing and Limited

- **Standardized slopes** (solution using z-scored variables, each with $M = 0$ and $SD = 1$) are supposed to describe the **change in y_i per "SD" of x_i**
  - Provided in SAS PROC REG or in STATA REGRESS with BETA option
  - Can also get by z-scoring all variables, then doing usual GLM (i.e., as implemented in R's lm function by putting scale( ) around each variable)

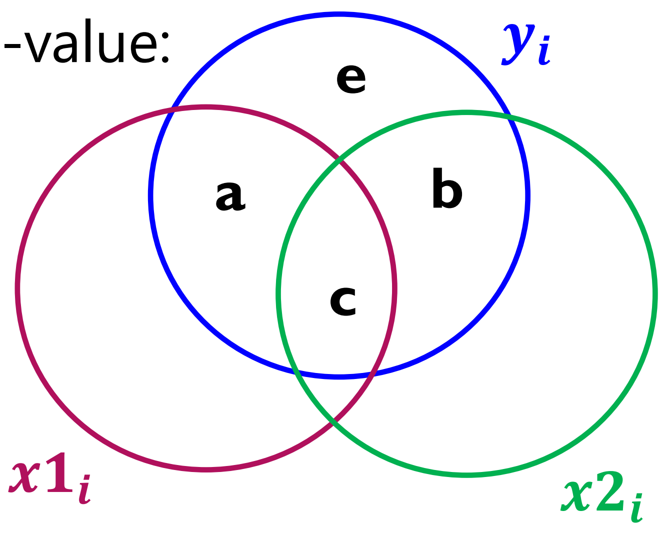
- Although **standardized slopes** (β_{std}) are often used to index effect size in GLMs and path models, they **are confusing and limited in scope**:
  - They range from $\pm\infty$, not -1 to 1 (so are not correlations), because the SD of original x_i is almost always larger than the SD for "unique" x_i variance
    - Btw, multiplying β_{std} by unique SD of x_i (as $\sqrt{Tolerance}$) = semi-partial r
  - Yield ambiguous results for quadratic or multiplicative terms (z-scored product of 2 variables is not equal to product of 2 z-scored variables)
  - Differences in sample size across subgroups create different standardized slopes for categorical predictors given the same unstandardized mean difference (see Darlington & Hayes, 2016, sec. 5.1.5 and ch. 8)
  - Do not readily extend to more complex types of prediction models (e.g., generalized linear models, multilevel or "mixed-effects" models)

For a helpful blog post on this topic, see: http://www.daviddisabato.com/blog/2016/4/8/on-effect-sizes-in-multiple-regression

Semi-partial (aka, "Part") Eta-Squared

- Given this GLM: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x1_i}) + \boldsymbol{\beta_2}(\boldsymbol{x2_i}) + \boldsymbol{e_i}$
- **For $\boldsymbol{x1_i}$, semi-partial $\boldsymbol{\eta^2} = \boldsymbol{sr^2} = \frac{\boldsymbol{SS_{x1}}}{\boldsymbol{SS_{total}}} = \frac{\boldsymbol{a}}{\boldsymbol{a+b+c+e}}$**
  - "Unique" sums of squares / total sums of squares: amount of model R^2 that is due to $x1_i$ → directly intuitive ☺
  - Will NOT be influenced by adding extra predictors to the model to explain residual variance → comparability across studies ☺
  - Btw, $\boldsymbol{\eta}$ version can also be found from $\boldsymbol{t}$-value:
    - $\boldsymbol{sr = t_{x1}\sqrt{\frac{1-R^2}{DF_{den}}}}$ — SQRT part → prop. *unexplained* variance
    - Btw, there is no analog to Cohen's d (b/c group is needed in the model)

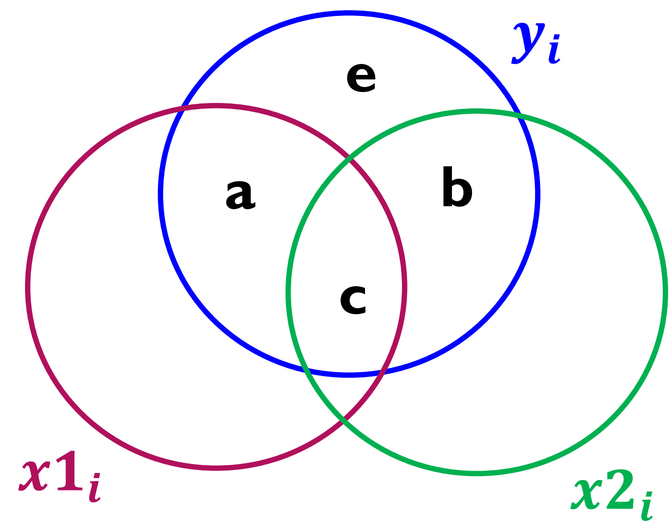
Overall model $\boldsymbol{R^2} = \frac{\boldsymbol{a+b+c}}{\boldsymbol{a+b+c+e}}$



Partial Eta and Eta-Squared

- Given this GLM: $y_i = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + e_i$
- **For $x1_i$, partial $\eta^2 = pr^2 = \frac{SS_{x1}}{SS_{x1}+SS_{residual}} = \frac{a}{a+e}$**
  - Unique SS / (unique SS + residual SS) → **R^2 for what's left**
  - WILL BE influenced by adding extra predictors to explain residual variance → lack of comparability across models/studies ☹
  - More useful η version can also be found from t-value:
    - **Partial η** $= pr = \frac{t}{\sqrt{t^2+DF_{den}}}$
    - **Btw, Partial Cohen's d** for mean differences in SD units: $pd = \frac{2t}{\sqrt{DF_{den}}}$
  - The word "partial" is used as a synonym for "unique" effects

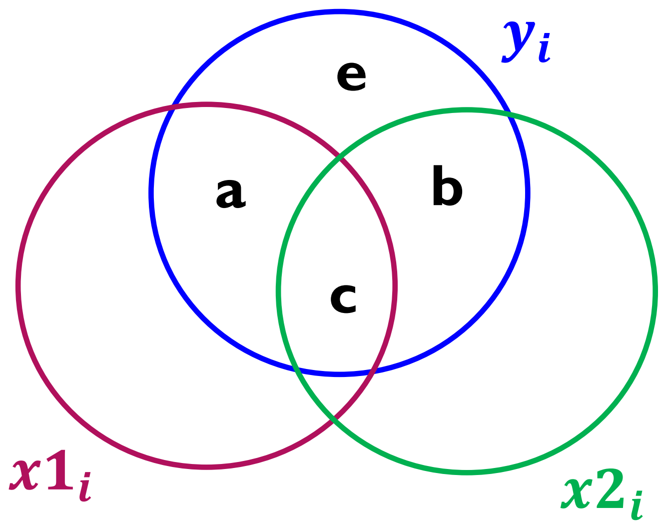
Overall model $R^2 = \frac{a+b+c}{a+b+c+e}$



Summarizing Effect Sizes (for $x1_i$ here)

- **Semi-partial** $\boldsymbol{\eta^2 = sr^2 = \frac{a}{a+b+c+e}}$
  - Unique / total: amount of model R^2 due to $x1_i$ (directly useful)
- **Partial** $\boldsymbol{\eta^2 = pr^2 = \frac{a}{a+e}}$
  - Unique / (unique+residual): $x1_i$ contribution setting aside $x2_i$
  - Given that it describes a subset of model R^2, η (or d) version can be less prone to misinterpretation
- **Cohen's** $\boldsymbol{f^2 = \frac{a}{e}} = ?????$
  - But is often used in power analysis!

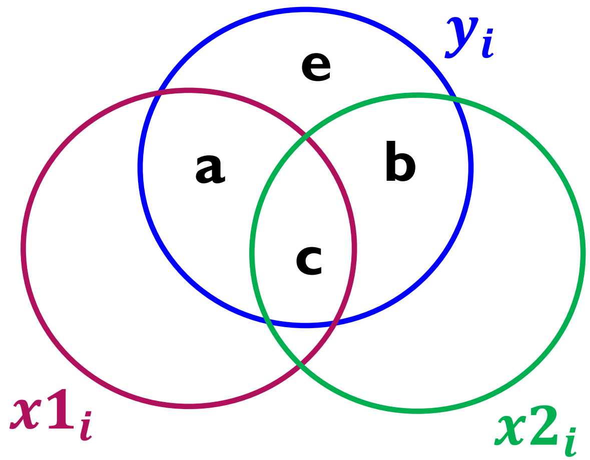
Areas below describe partitions of $\boldsymbol{y_i}$ variance:
a = $\boldsymbol{y_i}$ unique to $\boldsymbol{x1_i}$
b = $\boldsymbol{y_i}$ unique to $\boldsymbol{x2_i}$
c = $\boldsymbol{y_i}$ shared by $\boldsymbol{x1_i}$ and $\boldsymbol{x2_i}$
e = $\boldsymbol{y_i}$ leftover (residual)



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

Interpreting Effect Sizes with the Pearsons

- Effect sizes for $x1_i$
  - **Semi-partial** $\eta^2 = sr^2 = \frac{a}{a+b+c+e} = \frac{unique}{total}$
  - **Partial** $\eta^2 = pr^2 = \frac{a}{a+e} = \frac{unique}{unique+residual}$
    - Should not be compared across studies whose models differ in predictor content—**here's why**:



- Using the Pearsons—of 10 rooms, Randall cleaned 4 rooms, Kevin cleaned 1 room, and Randall and Kevin cleaned 2 common rooms
  - Randall: a = 4, Kevin: b =1, common: c = 2, residual: e = 3 (for this)
  - Randall: $sr^2 = \frac{4}{4+1+2+3} = .40$, $pr^2 = \frac{4}{4+3} = .57$
    - Randall cleaned 40% of the house, and 57% of the house *that Kevin didn't*
  - Kevin: $sr^2 = \frac{1}{4+1+2+3} = .10$, $pr^2 = \frac{1}{1+3} = .25$
    - Kevin cleaned 10% of the house, and 25% of the house *that Randall didn't*

Example of "Multiple Linear Regression"

- Models from example 2 (here, $R^2 = sr^2 = pr^2$)

  Sum of separate $R^2 = .1986$

  - Empty: $\boldsymbol{income_i} = \boldsymbol{\beta_0}$, $R^2 = 0$
  - Education: $\boldsymbol{income_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{educ_i} - \mathbf{12}) + \boldsymbol{e_i}$, $R^2 = .1480$
  - Marital Status: $\boldsymbol{income_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_2}(\boldsymbol{marry01_i}) + \boldsymbol{e_i}$, $R^2 = .0506$

- Combined: $\boldsymbol{income_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{educ_i} - \mathbf{12}) + \boldsymbol{\beta_2}(\boldsymbol{marry01_i}) + \boldsymbol{e_i}$
  - $R^2 = .1903$ for both < sum of separate $R^2 = .1986$ b/c of common
  - Education $\boldsymbol{\beta_1}$: semi-partial $sr^2 = .1396$, partial $pr^2 = .1471$ (t → sig*)
    - Explained 13.96% of income variance (14.71% of unexplained by marital)
  - Marital $\boldsymbol{\beta_2}$: semi-partial $sr^2 = .0423$, partial $pr^2 = .0496$ (t → sig*)
    - Explained 4.23% of income variance (4.96% of unexplained by educ)

- Significance of effect sizes given directly *per conceptual predictor* (linear education and binary marital status required 1 slope each)

More Complex "Multiple Linear Regression"

- Separate models from example 3 (here, $R^2 = sr^2 = pr^2$)
  - 3-category Workclass (2 slopes): $R^2 = .1034$
  - Linear +Quadratic Age (2 slopes): $R^2 = .1139$
  - Piecewise Education (3 slopes): $R^2 = .1643$

Sum of separate $R^2 = .3816$

- Combined: $$\boldsymbol{Income_i = \beta_0 + \beta_1 (LvsM_i) + \beta_2 (LvsU_i) + \beta_3 (Age_i - 18) + \beta_4 (Age_i - 18)^2 + \beta_5 (LessHS_i) + \beta_6 (GradHS_i) + \beta_7 (OverHS_i) + e_i}$$
  - $R^2 = .2887$ for all < sum of separate $R^2 = .3816$ b/c of common
  - **Workclass** $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$: semi-partial $sr^2 = .0428$, partial $pr^2 = .0567$
    - Explained 4.28% of income variance (5.67% of unexplained by others)
  - **Age** $\boldsymbol{\beta_3}$, $\boldsymbol{\beta_4}$: semi-partial $sr^2 = .0805$, partial $\eta^2 := .1017$
    - Explained 8.05% of income variance (10.17% of unexplained by others)
  - **Education** $\boldsymbol{\beta_5}$, $\boldsymbol{\beta_6}$, $\boldsymbol{\beta_7}$: semi-partial $sr^2 = .0807$, partial $\eta^2 := .1019$
    - Explained 8.07% of income variance (10.19% of unexplained by others)

More Complex "Multiple Linear Regression"

- Combined: $\boldsymbol{Income_i = \beta_0 + \beta_1 (LvsM_i) + \beta_2 (LvsU_i)}$
  $\boldsymbol{+ \beta_3 (Age_i - 18) + \beta_4 (Age_i - 18)^2}$
  $\boldsymbol{+ \beta_5 (LessHS_i) + \beta_6 (GradHS_i) + \beta_7 (OverHS_i) + e_i}$
  - *Btw, this model might also be called "Analysis of Covariance" (or ANCOVA)*

- Effect size per slope is problematic for two conceptual predictors:
  - **Working Class:** slopes $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ share a common reference (low group) and imply 3 pairwise group differences (2 in model; 1 given as linear combination; other types of differences could be requested as needed)
    - So the unique $\boldsymbol{sr^2}$ values across three possible group differences will sum to more than they should (given a single 3-category predictor)
  - **Age**: Linear age slope $\boldsymbol{\beta_3}$ is specific to centered age = 0, so its unique $\boldsymbol{sr^2}$ would change if age were centered differently; also, the unique $\boldsymbol{sr^2}$ values for linear and quadratic age cannot be summed directly to create total $\boldsymbol{sr^2}$ for age because of the correlation among the two predictors
  - **Education**: although the unique $\boldsymbol{sr^2}$ values for $\boldsymbol{\beta_5}$, $\boldsymbol{\beta_6}$, and $\boldsymbol{\beta_7}$ are ok to use in this case, they also cannot be summed directly to create total $\boldsymbol{sr^2}$ for education because of the correlation among the three predictors

How to Get Significance Tests and Effect Sizes for a Set of Slopes in Software

- **In SAS GLM**, semi-partial and partial η^2 (or ω^2 to use with adjusted R^2 instead) given by adding **EFFECTSIZE** to MODEL statement options
  - Then effect sizes provided directly for each fixed slope by default
  - Effect size and F-test also provided for a set of slopes via **CONTRAST** statements (e.g., for "omnibus" group effects, for linear+quadratic slopes)
  - Can choose hierarchical (Type I SS) or not (Type II, III, or IV SS), but hierarchical (in which order of predictors matters) is rarely appropriate (**Type III** most common)

- **In STATA**, **PCORR** provides semi-partial and partial η and η^2
  - Only works for single slopes—for a set of slopes, you have to compute semi-partial and partial η^2 using sums of squares relative to a model without them
  - **TEST after REGRESS** will provide F-tests for a set of slopes, though

- **R package ppcor** has pcor.test for partial η and spcor.test for semi-partial η
  - Only works for single slopes—for a set of slopes, you have to compute semi-partial and partial η^2 using sums of squares relative to a model without them
  - **glht after lm** will provide F-tests for a set of slopes, though

Effect Sizes for a Set of Slopes

- How to compute effect sizes for a set of slopes manually using unique sums of squares (SS)—see Example 4a for illustration
  - Step 1: From the full model, get SS for the model: SS_{Full}
    From the full model, get SS for the 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 b/t models: $SS_{Test} = SS_{Full} - SS_{Reduced}$
  - Step 4: Compute effect sizes: $sr^2 = \frac{SS_{Test}}{SS_{Total}}$, $pr^2 = \frac{SS_{Test}}{SS_{Total} - SS_{Test}}$
  - Step 5: Repeat steps 1–4 per set of slopes to be tested

- Given that this extra work is not needed in SAS, for fairness, your homework for this unit will instead use sequential models
  - Then the change in the model R^2 after adding new slopes will directly provide sr^2 for the new slopes (at each step, so these contributions will differ from what they would be in a full simultaneous model)

Example: Testing R^2 vs. Change in R^2

Example Model Fixed Effects	MSE residual variance (leftover)	Model R2 (relative to empty model)	Change in R2 from new slopes = Semipartial r2
1. intercept	200	0.00	
2. intercept + A	180	0.10	0.10
3. intercept + A + B	140	0.30	**0.20**
4. intercept + A + B + C + D	80	0.60	**0.30**

- F-tests assess the significance of a set of multiple slopes
  - F-test for **model R^2** is given by default (for **all slopes** in model)
- To assess the **change in the R^2** after adding **new slopes**:
  - **1 slope**? Its **p-value** tests R^2 change directly (e.g., model 2 to 3)
  - **2+ slopes**? Must request a **separate F-test** for new slopes added
    - e.g., for R^2 change from model 3 to 4—list slopes C and D only in SAS CONTRAST, STATA TEST, or R glht (see Example 4a and 4b)

Unexpected Results: Suppression

- *In general*, the semi-partial r for each predictor (and its unique standardized slope) will be smaller in magnitude than the bivariate r (and its standardized slope 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	t	Sig.
		B	Std. Error	Beta		
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	t	Sig.
		B	Std. Error	Beta		
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 average predictor contribution

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

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	t	Sig.
		B	Std. Error	Beta		
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 unique effect—this is because of their strong(ish) correlations with each other (and "common" still contributes to R^2)

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:
  - Do the slopes join to create a model $R^2 > 0$? Check p-value for model F
  - What is the model's effect size? Check $R^2 = (r$ of $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ with $\boldsymbol{y}_{\boldsymbol{i}})^2$
  - Is each slope significantly ≠ 0? Check p-value for $t = (Est - H_0)/SE$
  - What is each slope's effect size? Compute partial r or d from t

- 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 $t = (Est - H_0)/SE$
    - >1 slopes: check p-value for F-test of joint effect (requested separately)
  - What is the effect size for each conceptual 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