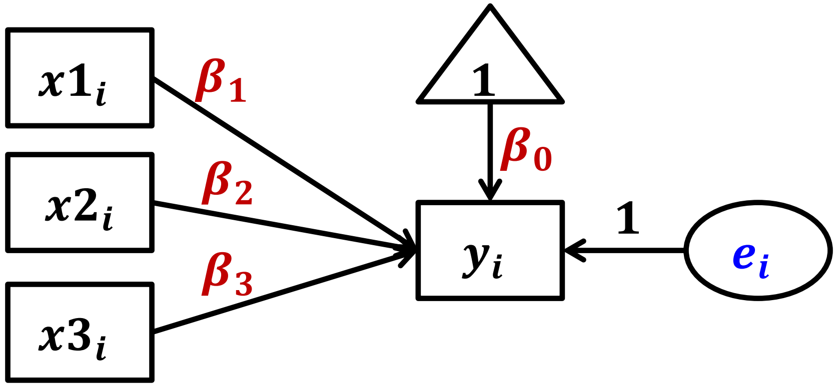
General Linear Models (GLMs) with Multiple Fixed Slopes for a Single Predictor

- Topics:
  - Reviewing empty GLMs and single predictor GLMs
  - GLM special cases: 2+ fixed slopes to describe a predictor's effect
    - "Analysis of Variance" (ANOVA) for a one categorical predictor
      - e.g., income differences across 3 categories of employment class
    - Nonlinear effects of a single quantitative predictor
      - e.g., quadratic continuous effect of years of age on income
      - e.g., piecewise discontinuous effect of years of education on income
    - Testing linear effects of a single ordinal predictor
      - e.g., linear vs. nonlinear effect of 5-category happiness on income

Where we're headed in this unit…



This figure is a **path diagram**. It illustrates a GLM with **3 x_i predictors of 1 y_i outcome**. The "1" triangle is a constant used by the fixed intercept. The picture generates this equation:

$$y_i = \beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + \beta_3(x3_i) + e_i$$

- **Synonyms for y_i outcome:** dependent variable, criterion, thing-to-be explained or predicted
- **Synonyms for each x_i predictor:** regressor, independent variable (if manipulated), covariate (if quantitative or if it must be included to show incremental contributions beyond it)
  - This unit will cover the use of multiple variables to describe the effect of a single conceptual predictor (up next will be multiple conceptual predictors)
- Ways to describe the **goal of a model**:
  - "Examine effects of (the x_i predictors) on (the y_i outcome)"
  - "Regress (outcome y_i) on (the x_i predictors)"

Review: Empty Models and Single-Predictor Models

- Predictive linear models create a **custom expected outcome** for each person through a linear combination of fixed effects that multiply predictor variables

- <u>Empty GLM:</u> **Actual** $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{e_i}$, **Predicted** $\boldsymbol{\hat{y}_i} = \boldsymbol{\beta_0}$
  - $\boldsymbol{\beta_0}$ = **intercept** = expected $\boldsymbol{y_i}$ = here is mean $\boldsymbol{\bar{y}}$ (best naïve guess if no predictors)
  - $\boldsymbol{e_i}$ = **residual** = is **always** the deviation between the actual $\boldsymbol{y_i}$ and predicted $\boldsymbol{\hat{y}_i}$
    - Because $\boldsymbol{\hat{y}_i}$ = $\boldsymbol{\bar{y}}$ for all, the $\boldsymbol{e_i}$ residual variance across persons ($\boldsymbol{\sigma_e^2}$) is **all the $\boldsymbol{y_i}$ variance**

- <u>Add a predictor:</u> **Actual** $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x_i}) + \boldsymbol{e_i}$, **Predicted** $\boldsymbol{\hat{y}_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x_i})$
  - $\boldsymbol{\beta_0}$ = **intercept** = expected $\boldsymbol{y_i}$ when $\boldsymbol{x_i} = \mathbf{0}$ (so always ensure $x_i = 0$ makes sense)
  - $\boldsymbol{\beta_1}$ = **slope of $\boldsymbol{x_i}$** = difference in $\boldsymbol{y_i}$ per one-unit difference in $\boldsymbol{x_i}$
  - $\boldsymbol{e_i}$ = **residual** = is always the deviation between the actual $\boldsymbol{y_i}$ and predicted $\boldsymbol{\hat{y}_i}$
    - Now $\boldsymbol{\hat{y}_i}$ differs by $\boldsymbol{x_i}$, so $\boldsymbol{e_i}$ residual variance across persons ($\boldsymbol{\sigma_e^2}$) is **leftover $\boldsymbol{y_i}$ variance**

1 Fixed Effect for a Single Predictor

- $\boldsymbol{\beta_1}$ for the **slope of $\boldsymbol{x_i}$** is scale-specific → is "unstandardized"
- Unstandardized results for $\boldsymbol{\beta_1}$ include:
  - **Estimate/Coefficient** = (Est) = optimal slope value for our sample
  - **Standard Error** (SE) = index of inconsistency across samples = how far away on average a sample $\boldsymbol{x_i}$ slope is from the population $\boldsymbol{x_i}$ slope
    - With only a single slope in the model, the SE for its estimate depends on the model residual variance ($\boldsymbol{\sigma_e^2}$), variance of $\boldsymbol{x_i}$ (σ_x^2), and $DF_{denominator}$: sample size minus k, the number of $\boldsymbol{\beta}$ model fixed effects ($N - k$)
  - **Test-statistic $\boldsymbol{t}$** = $(Est - H_0)/SE$ → "**Univariate Wald test**" gives p-value for slope's significance using t-distribution and $DF_{denominator} = N - k$

- Can also request a "standardized" slope to provide an $\boldsymbol{r}$ **effect size**:
  - For a GLM with a **single** quantitative or binary predictor, $\beta_{std} = Pearson\ r$

$$\beta_{std} = \boldsymbol{\beta_{unstd}} * \frac{SD_x}{SD_y}$$

GLMs with Predictors: Binary vs. 3+ Categories

- To examine a **binary predictor**, we only need **2 fixed effects** to tell us **3 things**: the outcome mean for Category=0, the outcome mean for Category=1, and the outcome mean difference
- **Actual** $y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{Category_i}) + \boldsymbol{e_i}$, **Predicted** $\hat{y}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{Category_i})$
  - Category 0 Mean: $\hat{y}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\mathbf{0}) = \boldsymbol{\beta_0}$ ← fixed effect #1
  - Difference of Category 1 from Category 0: $(\boldsymbol{\beta_0} + \boldsymbol{\beta_1}) - (\boldsymbol{\beta_0}) = \boldsymbol{\beta_1}$ ← fixed effect #2
  - Category 1 Mean: $\hat{y}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\mathbf{1}) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}$ ← linear combination of fixed effects
  - Ask for the estimate and SE for any mean created from a <u>linear combination</u> of fixed effects using `glht` in R
  - Btw, this type of GLM is also called a "two-sample" or "independent groups" *t*-test

- To examine the effect of a **predictor** with 3+ categories, the GLM needs **as many fixed effects as the number of predictor variable categories =** $\boldsymbol{C}$
  - If $\boldsymbol{C} = \mathbf{3}$, then we need the $\boldsymbol{\beta_0}$ intercept and 2 predictor slopes: $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$
  - If $\boldsymbol{C} = \mathbf{4}$, then we need the $\boldsymbol{\beta_0}$ intercept and 3 predictor slopes: $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, and $\boldsymbol{\beta_3}$
  - # pairwise mean differences = $\frac{C!}{2!(C-2)!}$ → e.g., given $\boldsymbol{C} = \mathbf{3}$, # diffs = $\frac{3*2*1}{(2*1)(1)} = 3$
  - Btw, goes by the name "**Analysis of Variance**" (**ANOVA**) in which the term "category" usually means "group"

"Indicator Coding" for a 3-Category Predictor

- Comparing outcome means across 3 categories requires creating **2 new binary predictors** to be included simultaneously along with the intercept, for example, as coded so **Low= Intercept (reference)**

workclass variable ($N = 734$)	LvM: Lower vs Middle?	LvU: Lower vs Upper?
1. Lower ($n = 436$)	0	0
2. Middle ($n = 278$)	**1**	0
3. Upper ($n = 20$)	0	**1**

Actual: $income_i = \beta_0 + \beta_1(LvM_i) + \beta_2(LvU_i) + e_i$

Predicted: $\hat{y}_i = \beta_0 + \beta_1(LvM_i) + \beta_2(LvU_i)$

- Model-implied means per category (group):
  - Lower Mean: $\hat{y}_L = \beta_0 + \beta_1(0) + \beta_2(0) = \beta_0$ ← fixed effect #1
  - Middle Mean: $\hat{y}_M = \beta_0 + \beta_1(1) + \beta_2(0) = \beta_0 + \beta_1$ ← found as linear combination
  - Upper Mean: $\hat{y}_U = \beta_0 + \beta_1(0) + \beta_2(1) = \beta_0 + \beta_2$ ← found as linear combination
- Model-implied differences between each pair of categories (groups):
  - Lower vs Middle: $(\beta_0 + \beta_1) - (\beta_0) = \beta_1$ ← fixed effect #2
  - Lower vs. Upper: $(\beta_0 + \beta_2) - (\beta_0) = \beta_2$ ← fixed effect #3
  - Middle vs Upper: $(\beta_0 + \beta_2) - (\beta_0 + \beta_1) = \beta_2 - \beta_1$ ← found as linear combination

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

R Syntax and Output for workclass → income

```
# Create indicator-binary-coded predictors
Lecture3$LvM=NA; Lecture3$LvU=NA  # Make 2 new empty variables
Lecture3$LvM[which(Lecture3$workclass==1)]=0  # Replace each for lower
Lecture3$LvU[which(Lecture3$workclass==1)]=0
Lecture3$LvM[which(Lecture3$workclass==2)]=1  # Replace each for middle
Lecture3$LvU[which(Lecture3$workclass==2)]=0
Lecture3$LvM[which(Lecture3$workclass==3)]=0  # Replace each for upper
Lecture3$LvU[which(Lecture3$workclass==3)]=1
# LvM: Lower=0 vs Middle=1 Class, LvU: Lower=0 vs Upper=1 Class

print("GLM Predicting Income from 2 New Binary Variables for workclass")
ModelClass = lm(data=Lecture3, formula=income~1+LvM+LvU)
obj=LMsummary(ModelClass, explain=TRUE)  # Custom output
```

```
Fixed Effects Table
             Est    SE      t      p    LCI    UCI
Intercept 13.650 0.626 21.795 <0.001 12.421 14.880
LvM        8.854 1.004  8.822 <0.001  6.884 10.825
LvU       10.985 2.990  3.673 <0.001  5.114 16.856
```

R Syntax and Output for workclass Linear Combinations

```
print("Get predicted income per category and category differences")
print("Values below are multipliers for each fixed effect IN ORDER")
PredClass = multcomp::glht(model=ModelClass, linfct=rbind(
  "Pred Income: Lower"    = c(1, 0, 0),  # already in model
  "Pred Income: Middle"   = c(1, 1, 0),
  "Pred Income: Upper"    = c(1, 0, 1),
  "Lower vs Middle Diff" = c(0, 1, 0),  # already in model
  "Lower vs Upper Diff"  = c(0, 0, 1),  # already in model
  "Middle vs Upper Diff" = c(0,-1, 1)))
obj=glhtSummary(glhtObject=PredClass, effectsizes=TRUE) # Custom output
```

```
Linear Combinations Table
                        Est    SE      t      p    LCI    UCI
Pred Income: Lower   13.650 0.626 21.795 <0.001 12.421 14.880
Pred Income: Middle  22.504 0.784 28.693 <0.001 20.965 24.044
Pred Income: Upper   24.635 2.924  8.425 <0.001 18.894 30.375
Lower vs Middle Diff  8.854 1.004  8.822 <0.001  6.884 10.825
Lower vs Upper Diff  10.985 2.990  3.673 <0.001  5.114 16.856
Middle vs Upper Diff  2.130 3.027  0.704  0.482 -3.813  8.074
```

GLM 3-Category Predictor: workclass Results

Empty Model: $income_i = \beta_0 + e_i$

- Model parameters:
  - Intercept β_0: $Est = 17.30, SE = 0.51$
  - Residual Variance σ_e^2: $Est = 190.21$

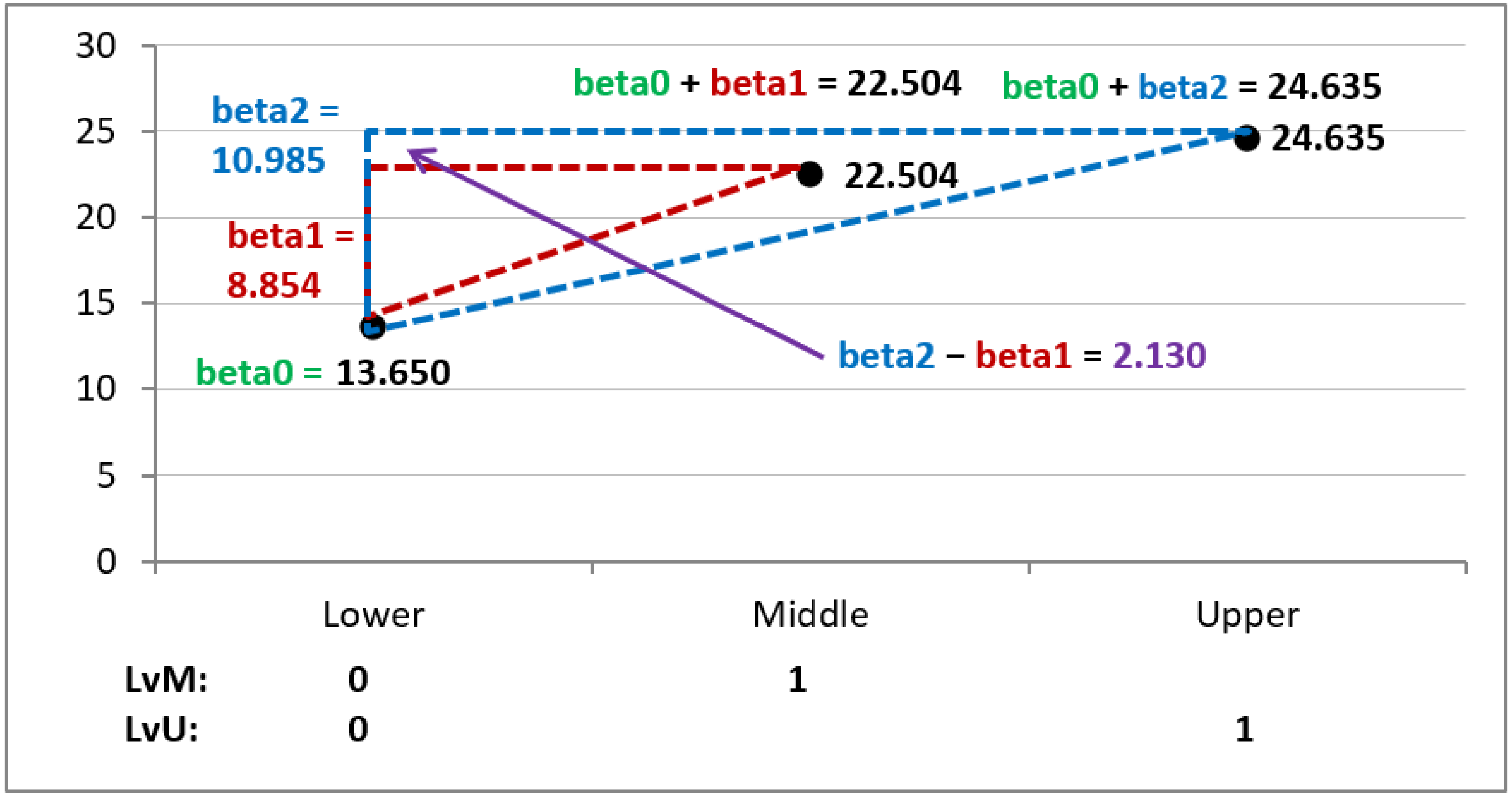
Group ($N = 734$)	LvM	LvU
1. Lower ($n = 436$)	0	0
2. Middle ($n = 278$)	**1**	0
3. Upper ($n = 20$)	0	**1**

Predictor Model: $income_i = \beta_0 + \beta_1(LvM_i) + \beta_2(LvU_i) + e_i$

- Model parameters:
  - Intercept β_0: $Est = 13.65, SE = 0.63, p < .001$ → Mean for L ($= \hat{y}_L$)
  - Slope β_1: $Est = 8.85, SE = 1.00, p < .001$ → Mean diff for L vs M
  - Slope β_2: $Est = 10.98, SE = 2.99, p < .001$ → Mean diff for L vs U
  - Residual Variance σ_e^2: $Est = 171.01$
- Linear combinations of model parameters:
  - M Mean: $\hat{y}_M = 13.65 + 8.85(1) + 10.98(0) = 22.50, SE = 0.78, p < .001$
  - U Mean: $\hat{y}_U = 13.65 + 8.85(0) + 10.98(1) = 24.63, SE = 2.92, p < .001$
  - Mean diff of M vs U = $\beta_2 - \beta_1 = 2.13, SE = 3.03, p = .482$

GLM 3-Category Predictor: workclass Results

Fixed Effects			Predictors			Category	Pred
beta0	beta1	beta2	Intercept	LvM	LvU	workclass	y-hat
13.650	8.854	10.985	1	0	0	Lower	13.650
13.650	8.854	10.985	1	1	0	Middle	22.504
13.650	8.854	10.985	1	0	1	Upper	24.635



Example of a 4-Category Predictor

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

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

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

- This model gives us **the predicted outcome mean for each category** as follows:

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

- Model provides directly 3 mean differences (control vs. each treatment), and indirectly another 3 mean differences among treatments as **linear combinations of fixed effects**…

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

Example with a 4-Category Predictor

Control (Ref) Mean = 10	Treatment 1 Mean = 12	Treatment 2 Mean = 15	Treatment 3 Mean = 19
β_0	$\beta_0 + \beta_1(d1_i)$	$\beta_0 + \beta_2(d2_i)$	$\beta_0 + \beta_3(d3_i)$

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

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

	Alt Group		Ref Group	Difference
C vs. T1 =	$(\beta_0 + \beta_1)$	$-$	(β_0)	$= \beta_1 = 2$
C vs. T2 =	$(\beta_0 + \beta_2)$	$-$	(β_0)	$= \beta_2 = 5$
C vs. T3 =	$(\beta_0 + \beta_3)$	$-$	(β_0)	$= \beta_3 = 9$
T1 vs. T2 =	$(\beta_0 + \beta_2)$	$-$	$(\beta_0 + \beta_1)$	$= \beta_2 - \beta_1 = 5 - 2 = 3$
T1 vs. T3 =	$(\beta_0 + \beta_3)$	$-$	$(\beta_0 + \beta_1)$	$= \beta_3 - \beta_1 = 9 - 2 = 7$
T2 vs. T3 =	$(\beta_0 + \beta_3)$	$-$	$(\beta_0 + \beta_2)$	$= \beta_3 - \beta_2 = 9 - 5 = 4$

Section Review: Coding for Categorical Predictors

Imagine a categorical predictor with **5 categories** (groups): A, B, C, D, and E

1. How many slopes are needed to distinguish the outcome mean across all 5 categories?

2. How many outcome mean differences between pairs of categories will the model fixed effects (intercept + slopes) provide directly?

3. How many other outcome mean differences between pairs of categories are implied (indirectly provided) by the model fixed effects?

4. How can you ensure that category E is the reference for the model?

Back to the 3-Category Predictor GLM

- The ANOVA-type question "Does group membership predict $\boldsymbol{y_i}$?" translates to "Are there significant group mean differences in $\boldsymbol{y_i}$"?
  - Can be answered **<u>specifically</u> via <u>pairwise</u> group differences** given directly by (or created from) the model fixed effects:
    For example: $\boldsymbol{income_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{LvM_i}) + \boldsymbol{\beta_2}(\boldsymbol{LvU_i}) + \boldsymbol{e_i}$,
    - Is $\boldsymbol{\beta_1} \neq 0$? If so, then $\hat{\boldsymbol{y}}_{\boldsymbol{M}} \neq \hat{\boldsymbol{y}}_{\boldsymbol{L}}$ (given directly because of our coding)
    - Is $\boldsymbol{\beta_2} \neq 0$? If so, then $\hat{\boldsymbol{y}}_{\boldsymbol{U}} \neq \hat{\boldsymbol{y}}_{\boldsymbol{L}}$ (given directly because of our coding)
    - Is $(\boldsymbol{\beta_2} - \boldsymbol{\beta_1}) \neq 0$? If so, then $\hat{\boldsymbol{y}}_{\boldsymbol{U}} \neq \hat{\boldsymbol{y}}_{\boldsymbol{M}}$ (requested as linear combination)
  - A **more <u>general</u> answer** to "**Does group matter?**" requires testing if $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ differ from 0 **<u>jointly</u>**, in other words:
    - Is the **residual variance** from this model with two grouping predictors **significantly lower** than the total variance from the empty model?
    - Does the **predicted** $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ provided by this model with two grouping predictors **correlate significantly with the actual $\boldsymbol{y_i}$?**

Prediction Gained vs. DF spent

- To get a **more general answer** to "**Does group matter?**" we need to consider the impact of our prediction <u>relative to</u> how many fixed effects we needed to predict $\hat{y}_i$ and how good they did (i.e., relative to what is left unknown)
  - This is one example of a more general "**multivariate Wald test**"
  - "Relative" is quantified using two types of **Degrees of Freedom** = $\boldsymbol{DF}$ = total number of fixed effects possible → total DF = sample size N
    - "$DF_{numerator}$" = $k - 1$ = number of fixed <u>slopes</u> in the model
    - "$DF_{denominator}$" = number of DF left over (not yet spent): $N - k$
  - In GLMs, the amount of information captured by the model's prediction and the amount of information left over are quantified using different sources of "**sums of squares**" ($\boldsymbol{SS}$)
    - Basic form of $\boldsymbol{SS}$ is the **numerator** in computing variance: $\frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}$
    - For example, "**outcome (or total)** $\boldsymbol{SS}$" = $\boldsymbol{SS_{total}} = \sum_{i=1}^{N}(y_i - \bar{y})^2$

Prediction Gained vs. DF spent

- How much information is provided by our **model prediction** is quantified by "**model sums of squares**": $SS_{model} = \sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2$
- To quantify the **relative** size of that predicted info, we need to adjust it for $DF_{numerator}$ = number of fixed slopes = $k - 1$

−1 because intercept doesn't get counted

  - Then get "**Model Mean Square**" = $MS_{model} = \frac{SS_{model}}{k-1}$
  - MS_{model} = "how much information has been captured per point spent"

- How much information is **leftover** is quantified by "**residual** (or **error**) **sums of squares**": $SS_{residual} = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
- To quantify the relative size of that leftover information, we need to adjust it for $DF_{denominator}$ = $N - k$
  - "**Residual** (or **Error**) **Mean Square**" = $MS_{residual} = \frac{SS_{residual}}{N-k}$
  - $MS_{residual}$ = "how much information left to explain per point remaining"

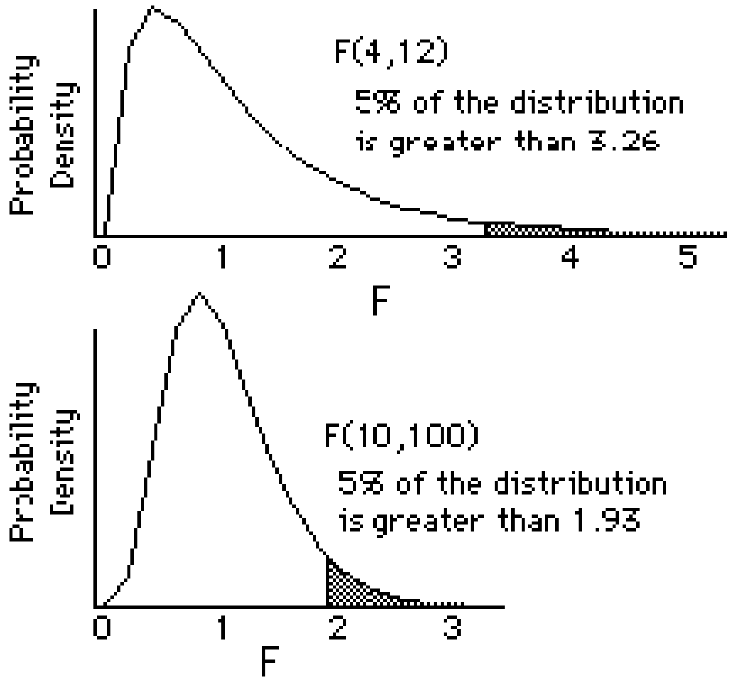
Prediction Gained vs. DF spent

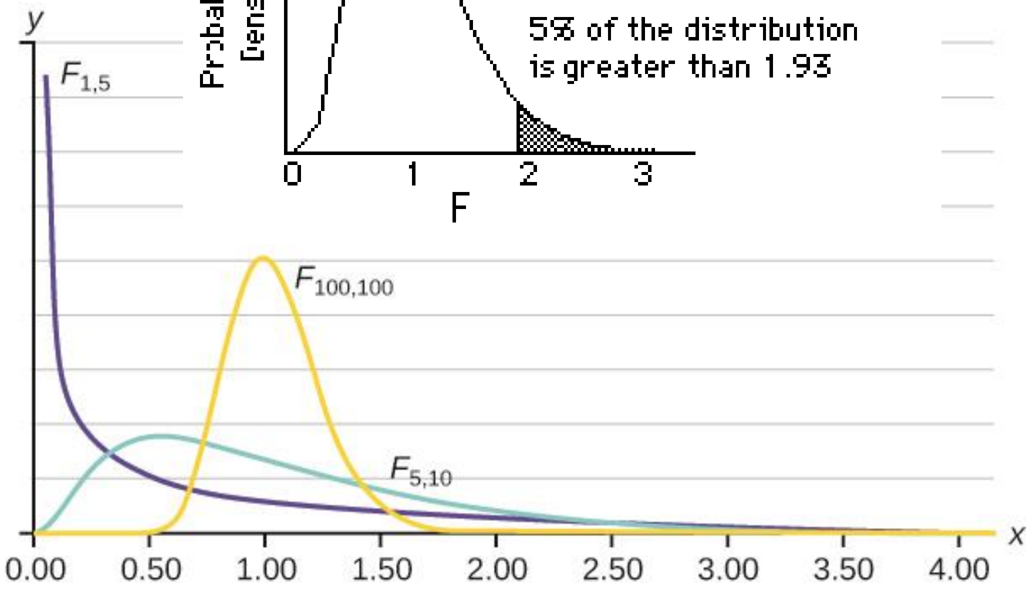
Source of Outcome Information	Sums of Squares (each summed from $i = 1$ to N)	Degrees of Freedom	Mean Square
Model (**known** because of predictor slopes)	SS_{model}: $(\hat{y}_i - \bar{y})^2$	DF_{num}: $k - 1$	MS_{model}: $\frac{SS_{model}}{k-1}$
Residual (leftover after predictors; still **unknown**)	$SS_{residual}$: $(y_i - \hat{y}_i)^2$	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)**

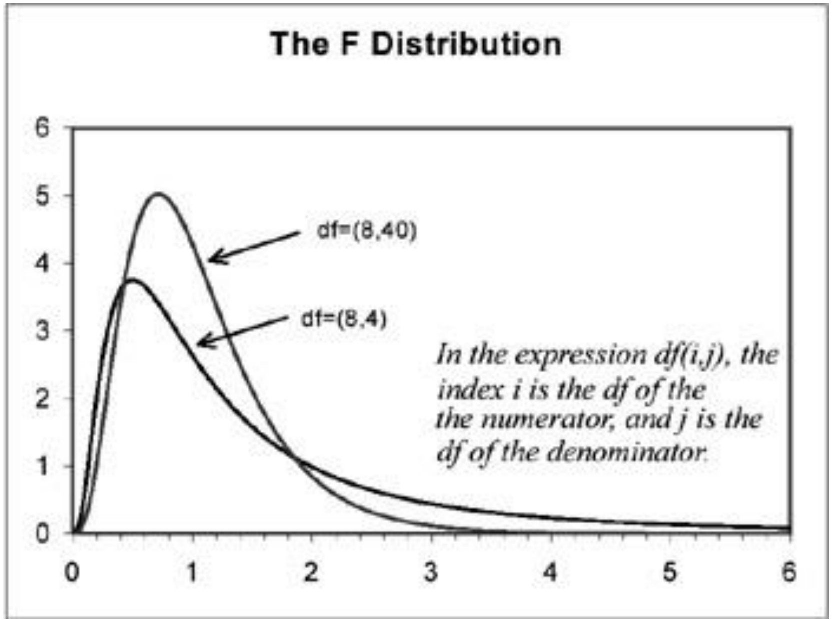
- This table now provides us with a way to answer the more general question of "Does group membership predict y_i?" → **Is our <u>model</u> significant**?
  - **Variance explained** by model fixed slopes: $R^2 = \frac{SS_{total} - SS_{residual}}{SS_{total}}$
  - R^2 = square of correlation between model-predicted $\hat{y}_i$ and actual y_i
  - F **test-statistic** for significance of $R^2 > 0$**?** is given two equivalent ways:

    $F(DF_{num}, DF_{den}) = \frac{MS_{model}}{MS_{residual}}$ or $F(DF_{num}, DF_{den}) = \frac{(N-k)R^2}{(k-1)(1-R^2)}$

Your New Friend, the F distribution







- The F test-statistic (F-value) is a ratio (in a **squared** metric) of "info explained over info unknown", so **F-values must be positive**
- Its shape (and thus critical value for the boundary of where "expected" starts) varies by DF_{num} (like χ^2) and by DF_{den} (like t, which is flatter for smaller $N - \mathrm{k}$)

Top left image borrowed from: https://www.statsdirect.com/help/distributions/f.htm
Top right image borrowed from: https://www.globalspec.com/reference/69569/203279/11-9-the-f-distribution
Bottom left image borrowed from: https://www.texasgateway.org/resource/133-facts-about-f-distribution

Summary: Steps in Significance Testing

- **Choose critical region: % alpha ("unexpected") and possible directions**
  - Both directions or just one?
  - Alpha (α) (1 −% confidence)?
  - Distribution for test-statistic will be dictated as follows:

Uses Denominator Degrees of Freedom?	Test 1 slope*	Test 2+ slopes*
No: implies infinite N	z	$\chi^2 (= z^2$ if 1)
Yes: adjusts based on N	t	$\boldsymbol{F} (= t^2$ if 1)

- If the **test-statistic exceeds** the distribution's critical value (i.e., goal posts), then the obtained **p-value is less than the chosen alpha** level:
  - You "**reject the null hypothesis**"—it is sufficiently **unexpected** to get a test-statistic that extreme *if the null hypothesis is true;* result is "**significant**"
- If the **test-statistic does NOT exceed** the distribution's critical value, then the **p-value is greater than or equal to the chosen alpha** level:
  - You "**DO NOT reject the null hypothesis**"—it is sufficiently **expected** to get a test-statistic that extreme *if the null hypothesis is true;* result is "**not significant**"

* # Fixed slopes (or associations) = **numerator degrees of freedom** = $k - 1$

Significance of the Model Prediction

- With **only 1 predictor**, we don't need a separate F test-statistic of the model R^2 significance; for example: $\boldsymbol{y_i = \beta_0 + \beta_1(x_i) + e_i}$
  - Significance of unstandardized $\boldsymbol{\beta_1}$ comes from $\boldsymbol{t = (Est - H_0)/SE}$
    - Significance of the model prediction $\boldsymbol{R^2}$ from $\boldsymbol{F = t^2}$ already
    - So if $\boldsymbol{\beta_1}$ is significant via $|\boldsymbol{t_{\beta_1}}| > \boldsymbol{t_{critical}}$, then the $\boldsymbol{F}$ **test-statistic** for the model is significant, too → sufficiently unexpected if H_0 were true
  - Standardized $\boldsymbol{\beta_1}$ = Pearson's $\boldsymbol{r}$ between predicted $\boldsymbol{\hat{y}_i}$ and actual $\boldsymbol{y_i}$
    - So model $\boldsymbol{R^2} = \frac{SS_{total} - SS_{residual}}{SS_{total}}$ is the same as **(Pearson's $\boldsymbol{r}$)^2**

- With **2+ fixed slopes**, we DO need to examine model $\boldsymbol{F}$ test-statistic and $\boldsymbol{R^2}$, for example: $\boldsymbol{income_i = \beta_0 + \beta_1(LvsM_i) + \beta_2(LvsU_i) + e_i}$
  - $\boldsymbol{F}$ test-statistic: Is the $\boldsymbol{\hat{y}_i}$ predicted from $\boldsymbol{\beta_1}$ AND $\boldsymbol{\beta_2}$ together significantly correlated with actual $\boldsymbol{y_i}$? The square of that correlation is the **model $\boldsymbol{R^2}$**
  - F test-statistic evaluates model R^2 *per DF spent to get it and DF leftover*

Significance of the Model: workclass Results

- For example, using workclass: $\boldsymbol{income_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{LvsM_i}) + \boldsymbol{\beta_2}(\boldsymbol{LvsU_i}) + \boldsymbol{e_i}$
- Group-specific means: We already know that L<M, L<U, and L≈M
- Significance test of the overall model (both slopes at once): $R^2 = .103$
- **Report as** $\boldsymbol{F(DF_{num}, DF_{den}) = Fvalue}$**,** **MSE** $\boldsymbol{= MS_{res}}$**,** $\boldsymbol{p < p - value}$

Source of Outcome Information	Sums of Squares (each summed from $i = 1$ to N)	Degrees of Freedom	Mean Square	F Value
Model (**known**)	SS_{model}: $(\hat{y}_i - \bar{y})^2$ $= 14{,}414.03$	$\boldsymbol{DF_{num}}$: $\boldsymbol{k - 1}$ $= 2$ slopes (−1 for intercept)	$\boldsymbol{MS_{model}}$: $\boldsymbol{\frac{SS_{model}}{k-1}}$ $= 7{,}207.01$	42.14
Residual (**"error"**)	$SS_{residual}$: $(y_i - \hat{y}_i)^2$ $= 125{,}009.25$	$\boldsymbol{DF_{den}}$: $\boldsymbol{N - k}$ $= 731$ leftover	$\boldsymbol{MS_{residual}}$: $\boldsymbol{\frac{SS_{residual}}{N-k}}$ $= 171.01$	
Corrected Total (after $\boldsymbol{\bar{y}}$)	SS_{total}: $(y_i - \bar{y})^2$ $= 139{,}423.23$	$\boldsymbol{N = 734 - 1}$ $= 733$ total corrected for int		

R Syntax and Output for workclass → income

```
print("GLM Predicting Income from 2 New Binary Variables for workclass")
ModelClass = lm(data=Lecture3, formula=income~1+LvM+LvU)
obj=LMsummary(ModelClass, explain=TRUE)  # Custom output

   Sums of Squares Table
                  SS  DF       MS      F      p    R2
   Model   14414.026   2 7207.013 42.144 <0.001 0.103
   Error  125009.205 731  171.011
   Total  139423.232 733  190.209

   Explanation:
   SS = Sum of Squares, MS = Mean Square, DF = Degrees of Freedom,
   F = F test-statistic, p = two-sided p-value, R2 = R-square

   Fixed Effects Table
                Est    SE      t      p    LCI    UCI
   Intercept 13.650 0.626 21.795 <0.001 12.421 14.880
   LvM        8.854 1.004  8.822 <0.001  6.884 10.825
   LvU       10.985 2.990  3.673 <0.001  5.114 16.856
```

Another version of R2: "Adjusted R2"

- Just like one may want to adjust Pearson's r for bias due to small sample size, some feel the need to **adjust the model $\boldsymbol{R}^2$**
  - $R^2 = \frac{SS_{total} - SS_{residual}}{SS_{total}}$ → Must be positive if computed this way
  - $\boldsymbol{R^2_{adj}} = 1 - \frac{(1-R^2)(N-1)}{N-k-1} = 1 - \frac{MS_{residual}}{MS_{total}}$ → Change in residual variance relative to empty model
    - R^2_{adj} can be negative! (i.e., for a really-not-useful set of fixed slopes)

- Although adjusted R^2 is considered as the only "correct" version by a few, I have never once been asked to report it…
  - But just in case Reviewer 3 wants it some day, here you go…
  - For our example: $R^2_{adj} = 1 - \frac{(1-.103)(734-1)}{734-3} = .101$ ($R^2_{unadj} = .103$)
  - Both R^2 versions are given by `lm` in R (default summary function)

Effect Size per Fixed Slope

- **Model R^2 value** (the square of the correlation between predicted $\hat{y}_i$ and actual y_i) provides a **general effect size**, but one may also want **effect size for each fixed slope**
  - Why? To standardize the effect magnitude and/or to estimate power
  - For models with one slope only, the **standardized slope** (found using z-scored variables with $M = 0$ and $SD = 1$) is the Pearson correlation → unambiguous "bivariate" effect size
  - For models with 2+ slopes, there are multiple potential measures of slope-specific effect size that you can choose from… (stay tuned)

- Although **standardized slopes** are often used to index effect size in multiple-slope models, they **have problems** in some cases:
  - Ambiguous results for quadratic or multiplicative terms (because the z-scored product of 2 variables is not equal to product of 2 z-scored variables)
  - Differences in sample size across groups create different standardized slopes for categorical predictors given the same unstandardized mean difference (see Darlington & Hayes, 2016 ch. 8 for more)

Effect Size per Fixed Slope from t

- We can use t test-statistics to compute 2 different metrics of **partial effect sizes** (for slopes or their linear combinations)
  - Here "**partial**" refers to a slope's **unique effect** in models with multiple fixed slopes *(stay tuned for "semi-partial" alternatives)*
  - Why use the t-value? To get effect sizes for fixed effect linear combinations, too!
  - **Partial correlation r** (range is ± 1): partial $r = \frac{t}{\sqrt{t^2 + DF_{den}}}$
    - Useful for quantitative predictors to convey strength of unique association for that slope
    - Can also get partial r from `pcor.test` in R package `ppcor`
  - **("Partial") Cohen's d** (range is $\pm\infty$): $d = \frac{2t}{\sqrt{DF_{den}}}$
    - Conveys difference between two groups in standard deviation units
    - ("Partial") is not used in describing Cohen's d (even though it's that kind of effect size) because there is not another kind possible (i.e., as in "semi-partial" r, stay tuned)

From r to d:

$$d \approx \frac{2r}{\sqrt{1 + r^2}}$$

$$r \approx \frac{d}{\sqrt{4 + d^2}}$$

R Output for workclass Linear Combinations

```
obj=glhtSummary(glhtObject=PredClass, effectsizes=TRUE) # Custom output
```

```
  Linear Combinations Table
                           Est    SE      t      p    LCI    UCI
  Pred Income: Lower    13.650 0.626 21.795 <0.001 12.421 14.880
  Pred Income: Middle   22.504 0.784 28.693 <0.001 20.965 24.044
  Pred Income: Upper    24.635 2.924  8.425 <0.001 18.894 30.375
  Lower vs Middle Diff   8.854 1.004  8.822 <0.001  6.884 10.825
  Lower vs Upper Diff   10.985 2.990  3.673 <0.001  5.114 16.856
  Middle vs Upper Diff   2.130 3.027  0.704  0.482 -3.813  8.074

  Effect Sizes for Linear Combinations Table
                           Est      p     d    pr   sR2
  Pred Income: Lower    13.650 <0.001 1.612 0.628 0.583
  Pred Income: Middle   22.504 <0.001 2.122 0.728 1.010
  Pred Income: Upper    24.635 <0.001 0.623 0.297 0.087
  Lower vs Middle Diff   8.854 <0.001 0.653 0.310 0.095
  Lower vs Upper Diff   10.985 <0.001 0.272 0.135 0.017
  Middle vs Upper Diff   2.130  0.482 0.052 0.026 0.001
```

Effect Sizes for the workclass Results and Sample Sizes Needed for Power = .80

- **LvM Diff as $\boldsymbol{\beta_1}$**: $Est = 8.85, SE = 1.00, t(731) = 8.82, p < .001$
  - $r = \frac{8.82}{\sqrt{8.82^2+731}} = 0.31, d = \frac{2*8.82}{\sqrt{731}} = 0.65$ → ~per-group $\boldsymbol{n > 45}$

- **LvU Diff as $\boldsymbol{\beta_2}$**: $Est = 10.98, SE = 2.99, t(731) = 3.67, p < .001$
  - $r = \frac{3.67}{\sqrt{3.67^2+731}} = 0.13, d = \frac{2*3.67}{\sqrt{731}} = 0.27$ → ~ per-group $\boldsymbol{n > 175}$

- **MvU Diff as**: $\boldsymbol{\beta_2} - \boldsymbol{\beta_1}$: $Est = 2.13, SE = 3.03, t(731) = 0.70, p = .482$
  - $r = \frac{0.70}{\sqrt{0.70^2+731}} = 0.03, d = \frac{2*0.70}{\sqrt{731}} = 0.05$ → ~ per-group $\boldsymbol{n > 2,102}$

- **Model $\boldsymbol{R^2}$** $= .103, r = .322$ → ~overall $\boldsymbol{N > 85}$

Example Results: workclcass → income

We used a general linear model (i.e., analysis of variance) to examine the extent to which annual income in thousands of dollars (M = 17.30, SD = 13.79) could be predicted from three categories of self-reported working class membership (lower = 59.40%, middle = 37.87%, and upper = 2.72%). We created two binary contrasts to distinguish the three classes, in which lower-class respondents served as the reference group to be compared separately to middle-class and upper-class respondents. Cohen's *d* standardized mean differences were then computed from the *t* test-statistics to index effect size per slope.

Class membership significantly predicted annual income, $F(2, 731) = 42.14$, MSE = 171.01, $p < .001$, $R^2 = .10$. Relative to lower-class respondents, annual income was significantly higher for both middle-class respondents (difference = 8.85, SE = 1.00, d = 0.65) and upper-class respondents (difference = 10.98, SE = 2.99, d = 0.27). However, upper-class respondents did not differ significantly from middle-class respondents (difference = 2.13, SE = 3.03, d = 0.05).

Intermediate Summary

- For GLMs with **<u>one fixed slope</u>**, the significance test for that fixed slope **is the same** as the significance test for the model
  - Slope $\boldsymbol{\beta}_{unstd}$: $t = \frac{Est - H_0}{SE}$, $\boldsymbol{\beta}_{std}$ = Pearson r
  - Model: $F = t^2$, $R^2 = r^2$ because predicted $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ only uses $\boldsymbol{\beta}_{unstd}$

- For GLMs with **<u>2+ fixed slopes</u>**, the significance tests for those fixed slopes (or any linear combinations thereof) are **NOT the same** as the significance test for the overall model
  - Single test of <u>one fixed slope</u> via $\boldsymbol{t}$ (or z) → "Univariate Wald Test"
  - Joint test of <u>2+ fixed slopes</u> via $\boldsymbol{F}$ (or χ^2) → "Multivariate Wald Test"
    - F test-statistic is used to test the significance of the model $\boldsymbol{R^2}$ (the square of the $\boldsymbol{r}$ between predicted $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ and actual $\boldsymbol{y}_{\boldsymbol{i}}$, which is necessary whenever the predicted $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ uses multiple $\boldsymbol{\beta}_{unstd}$ slopes)
    - F test-statistic evaluates model R^2 *per DF spent to get it and DF leftover*

Section Review: Model Testing

1. Conceptually, what does the F test statistic for the model fixed slopes tell us?

2. What does the model R^2 value tell us?

3. What are the two definitions of the model R^2?

4. Why does Lesa say “power for what” when asked to conduct power analyses for grant applications?

Nonlinear Trends of Quantitative Predictors

- Besides predictors with 3+ categories, another situation in which a single predictor variable may require more than one fixed slope to create its model prediction (its "effect" or "trend") is **when a quantitative predictor has a nonlinear relation with the outcome**

- We will examine <u>three types of examples</u> of this scenario:
  - **Curvilinear effect** of a quantitative predictor
    - Combine linear and quadratic slopes to create U-shape curve
    - Use natural-log transformed predictor to create an exponential curve
  - **Piecewise effects** for "sections" of a quantitative predictor
    - Also known as "linear splines" (but each slope could be nonlinear, too)
  - **Testing the assumption of <u>linearity</u>**: that equal differences between predictor values create equal outcome differences
    - Relevant for ordinal variables in which numbers are really just labels
    - Relevant for count predictors in which "more" may mean different things at different predictor values (e.g., "if and how much" predictors)

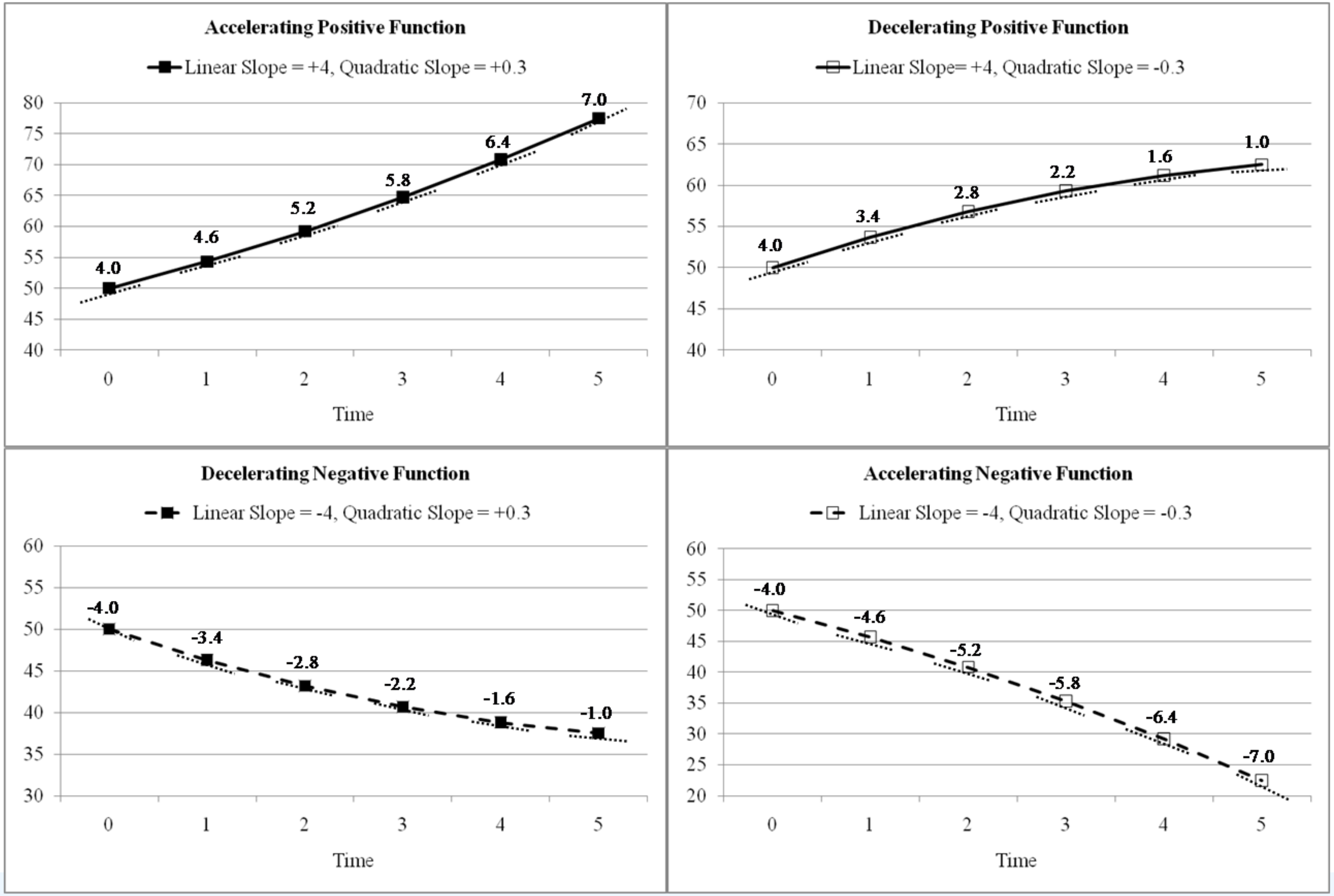
Curvilinear Trends of Quantitative Predictors

- The effect of a quantitative predictor does NOT have to be linear—curvilinear effects may be more theoretically reasonable or fit better
- There are many kinds of nonlinear trends—here are two examples:

  - **Quadratic (i.e., U-shaped)**: created by combining two predictors
    - "Linear": what it means when you enter the predictor by itself
    - "Quadratic": from also entering the predictor^2 (multiplied by itself)
    - Good to create relationships that **change directions**
    - Example for quadratic trend of x_i: $y_i = \beta_0 + \beta_1(x_i) + \beta_2(x_i)^2 + e_i$

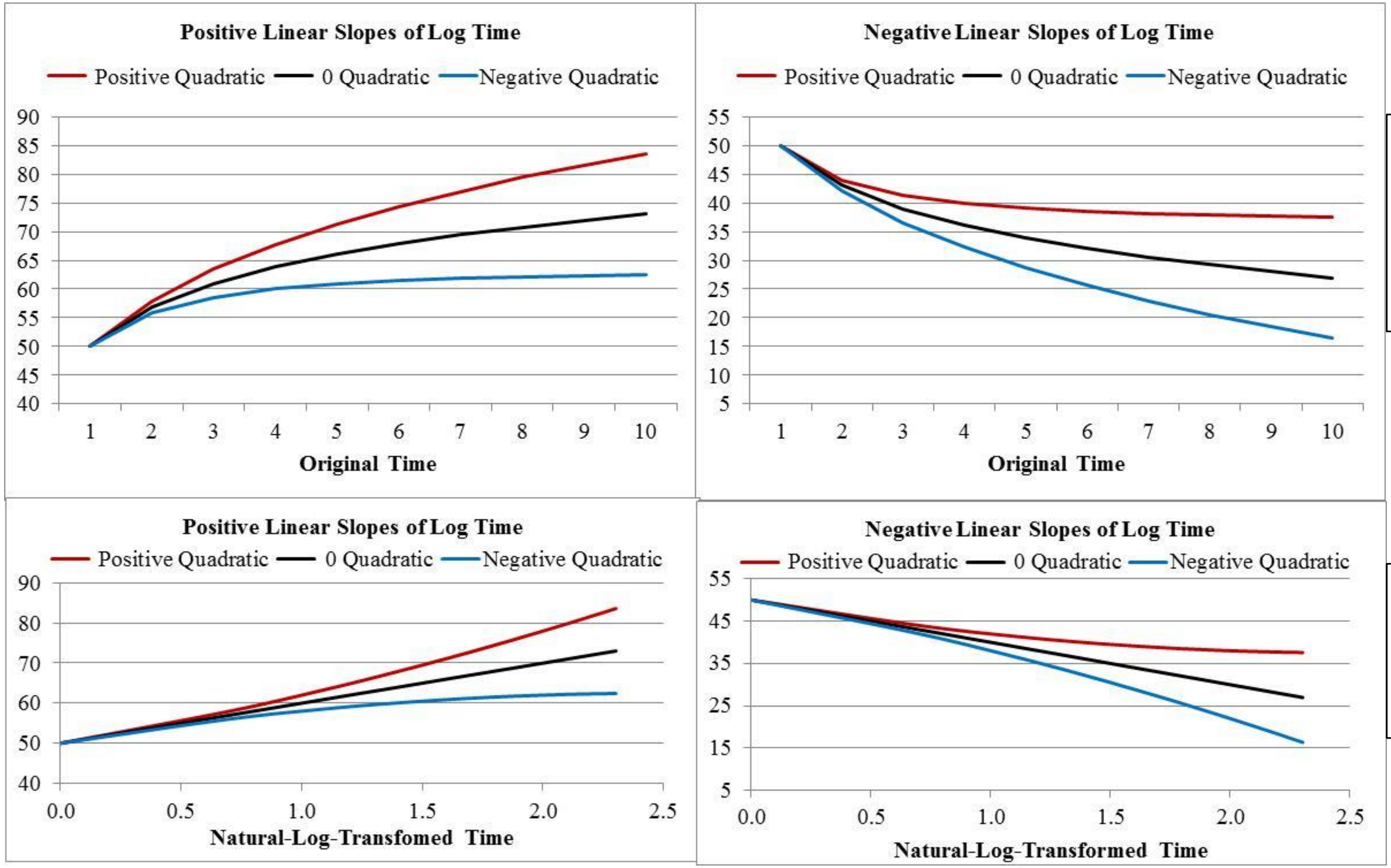
  - **Exponential(ish)**: created from one nonlinearly-transformed predictor
    - Predictor = natural-log transform of predictor (for positive values only)
    - Good to create relationships that look like **diminishing returns**
    - Example for exponential(ish) trend of x_i: $y_i = \beta_0 + \beta_1(log[x_i]) + e_i$

Quadratic Trends: Example of x_i = Time



Exponential(ish) Trends: Example of x_i = Time

- A **linear slope of log x_i** (black lines) **mimics an exponential trend across original x_i**; adding a quadratic slope of log x_i (red or blue lines) speeds it up or slows it down



There is an **exponential** effect of **original x_i** on the outcome

There is a **linear** effect of **log x_i** on the outcome

How to Interpret Quadratic Slopes

- A quadratic slope makes the effect of x_i change across itself!
  - Related to the ideas of position, velocity, and acceleration in physics
- Quadratic slope = HALF the rate of acceleration/deceleration
  - So to describe how the linear slope for x_i changes per unit difference in x_i, you must **multiply the quadratic slope (for $\boldsymbol{x_i^2}$) by 2**

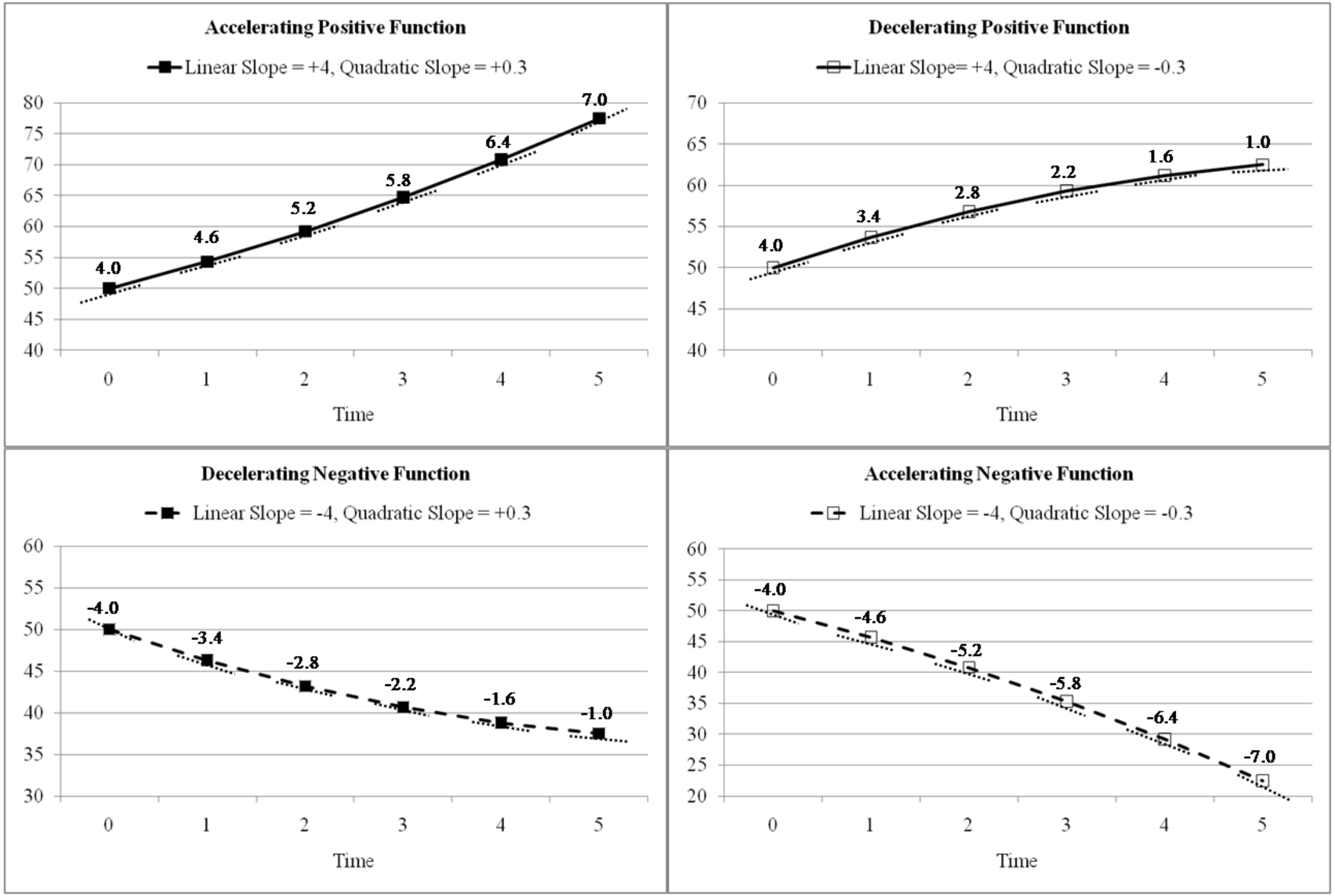
- If fixed linear slope = 4 at $x_i = 0$, with quadratic slope = 0.3?
  - "Instantaneous" linear rate of change is 4.0 at $x_i = 0$, is 4.6 at $x_i = 1$…
  - Btw: The "twice" rule comes from the second derivative of the function for y_i with respect to x_i:

Intercept (position) at $\boldsymbol{x_i = x}$: $y_x = 50 + 4(x_i) + 0.3(x_i^2)$

First derivative (velocity) at $\boldsymbol{x}$: $\frac{dy_x}{d(x)} = 4 + (2 * 0.3)(x_i)$

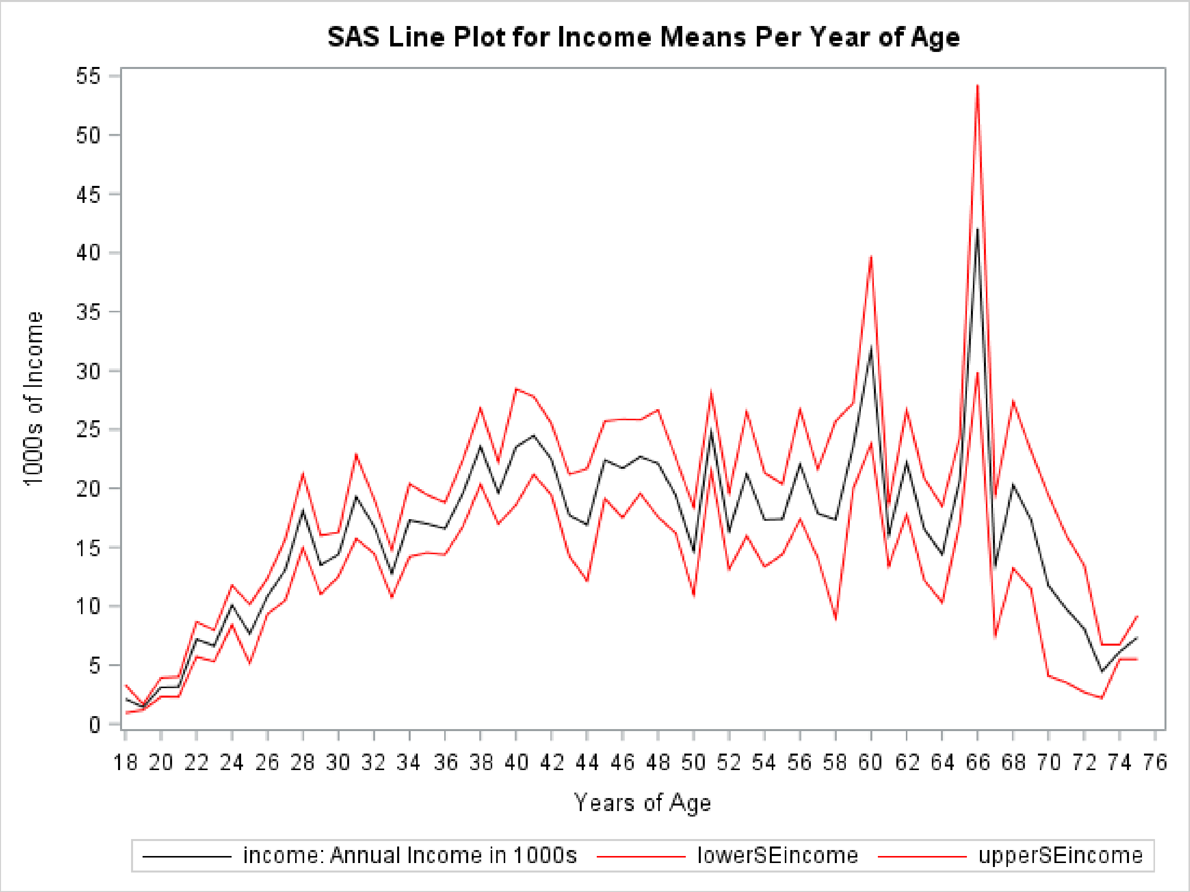
Second derivative (acceleration) at $\boldsymbol{x}$: $\frac{d^2y_x}{d(x)} = (2 * 0.3)$

Quadratic Trends: Example of x_i = Time



Quadratic Trend for age → income

- Black line = mean for each year of age; red lines = ±1 SE of mean
- Although noisy, this plot shows a clear quadratic function of age in predicting annual income (yay middle age!)



- Let's see what happens when we fit **a quadratic effect of age** (centered at 18, the minimum age) predicting annual income…

R Syntax and Output for quadratic age → income

```
# Create new age variable centered at 18 (minimum in sample)
Lecture3$age18=Lecture3$age-18  # age18: Age (0=18 years)

print("GLM Predicting Income from Linear+Quadratic Centered Age")
print("I(x^2) squares predictor to create quadratic term")
ModelQuadAge = lm(data=Lecture3, formula=income~1+age18+I(age18^2))
obj=LMsummary(ModelQuadAge)   # Custom output
```

```
 Sums of Squares Table
              SS  DF        MS      F      p    R2
 Model  15885.462   2 7942.731 46.999 <0.001 0.114
 Error 123537.770 731  168.998
 Total 139423.232 733  190.209

 Fixed Effects Table
              Est    SE      t      p    LCI    UCI
 Intercept   2.677 1.584  1.690  0.091 -0.432  5.785
 age18       1.223 0.135  9.055 <0.001  0.958  1.488
 I(age18^2) -0.020 0.003 -7.809 <0.001 -0.024 -0.015
```

R Syntax and Output for age Linear Combinations

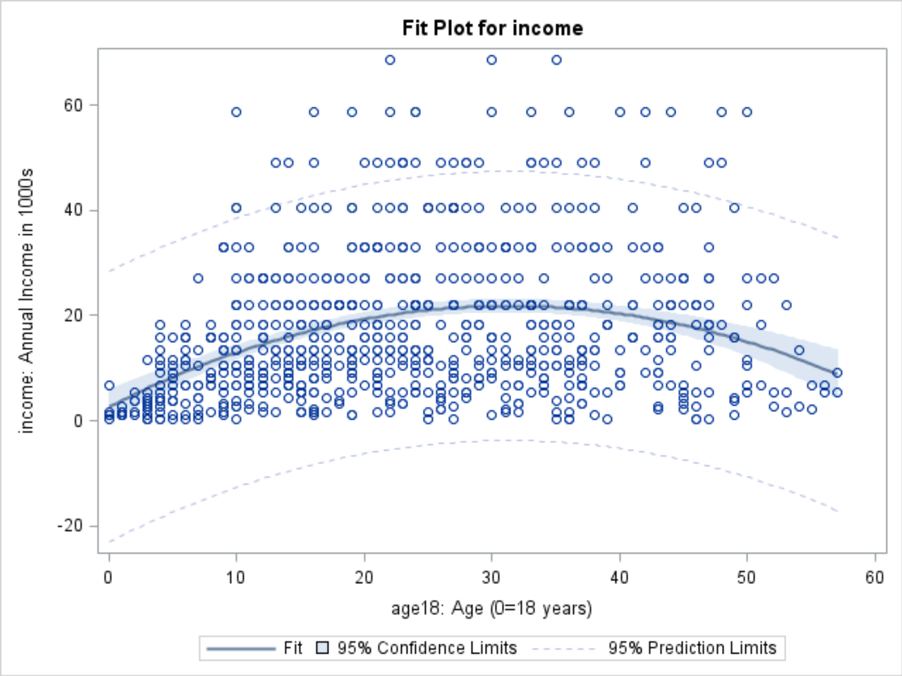
```
print("Get predicted income and predicted linear age slopes for example ages")
print("In number lists below, values are multiplier for each fixed effect IN ORDER")
print("Linear slopes = beta1 + 2*beta2*age18")
PredQuadAge = multcomp::glht(model=ModelQuadAge, linfct=rbind(
  "Pred Income at Age 30 (age18=12)" = c(1,12, 144),
  "Pred Income at Age 50 (age18=32)" = c(1,32,1024),
  "Pred Income at Age 70 (age18=52)" = c(1,52,2704),
  "Pred Linear Age Slope at Age 30"  = c(0, 1,  24),
  "Pred Linear Age Slope at Age 50"  = c(0, 1,  64),
  "Pred Linear Age Slope at Age 70"  = c(0, 1, 104)))
obj=glhtSummary(glhtObject=PredQuadAge)  # Custom output
```

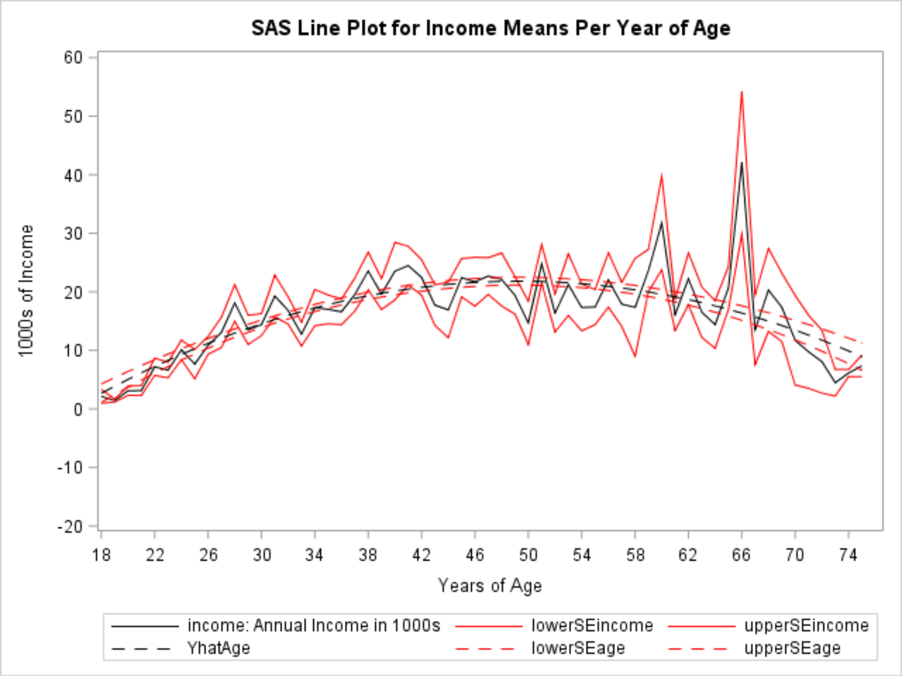
```
Linear Combinations Table
                                    Est    SE      t      p    LCI    UCI
Pred Income at Age 30 (age18=12) 14.540 0.647 22.465 <0.001 13.270 15.811
Pred Income at Age 50 (age18=32) 21.809 0.668 32.642 <0.001 20.497 23.121
Pred Income at Age 70 (age18=52) 13.448 1.659  8.106 <0.001 10.191 16.705
Pred Linear Age Slope at Age 30   0.754 0.079  9.569 <0.001  0.599  0.909
Pred Linear Age Slope at Age 50  -0.027 0.047 -0.584  0.559 -0.119  0.064
Pred Linear Age Slope at Age 70  -0.809 0.135 -5.998 <0.001 -1.074 -0.544
```

Results for quadratic age → income

- $Income_i = \beta_0 + \beta_1(Age_i - 18) + \beta_2(Age_i - 18)^2 + e_i$
  - **Intercept:** β_0 = expected income at age 18 → $Est = 2.677, SE = 1.584, p < .001$
  - **Linear Age Slope:** β_1 = instantaneous rate of change (or difference, actually) in income per year of age **at age = 18** → $Est = 1.223, SE = 0.135, p < .001$
  - **Quadratic Age Slope:** β_2 = half the rate of acceleration per year of age **at any age** → $Est = -0.020, SE = 0.003, p < .001$
- Predicted income at other ages via linear combinations of fixed effects:
  - Age 30: $\hat{y}_{x=30} = 2.677 + 1.223(12) - 0.020(12)^2 = 14.540, SE = 0.647$
  - Age 50: $\hat{y}_{x=50} = 2.677 + 1.223(32) - 0.020(32)^2 = 21.809, SE = 0.668$
  - Age 70: $\hat{y}_{x=70} = 2.677 + 1.223(52) - 0.020(52)^2 = 13.448, SE = 1.659$
- Predicted linear age slope at other ages via linear combinations:
  - Age 30: $\widehat{\beta_1}_{x=30} = 1.223 - 0.020(2 * 12) = 0.754, SE = 0.079$
  - Age 50: $\widehat{\beta_1}_{x=50} = 1.223 - 0.020(2 * 32) = -0.027, SE = 0.047$
  - Age 70: $\beta_{1_{x=70}} = 1.223 - 0.020(2 * 52) = -0.809, SE = 0.135$
- Predicted "aperture" – here, age at max income (where linear age slope = 0): $\frac{-\beta_1}{2 * \beta_2} + 18 = 48.575$

Results for quadratic age → income





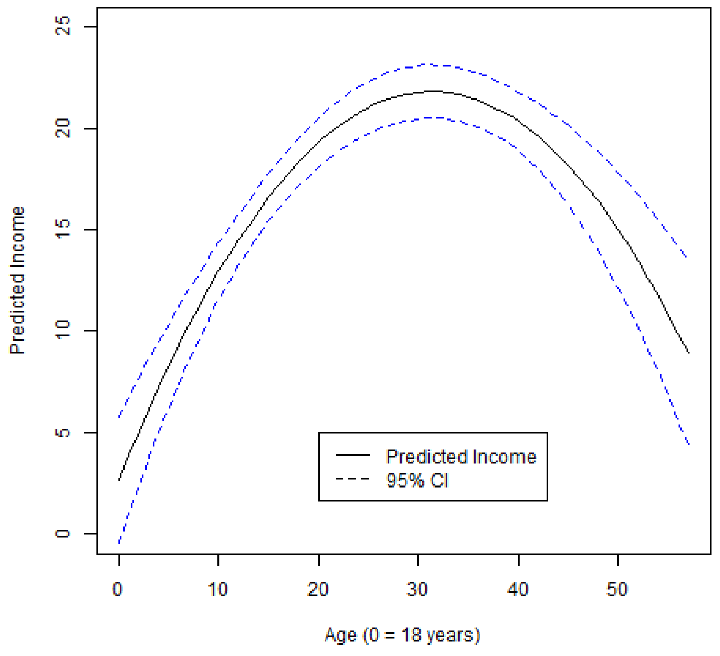
- Left: predicted regression line over individual scatterplot
  - From: $\mathbf{2.677 + 1.223(\mathit{Age_i} - 18) - 0.020(\mathit{Age_i} - 18)^2}$
- Right: predicted regression line over mean per age
  - $F(2,\ 731) = 47.00, MSE = 169.00, p < .001, R^2 = .114\ (r = .338)$
    - Since age and age^2 work together, I'd recommend model r as effect size

R Syntax and Output for Many Predicted Outcomes

```
# Generate predicted values using predict more efficiently and plot them
PredQuadAge = data.frame(age18=seq(from=0, to=57, by=1))  # Real ages 18 to 75 (min and max)
PredQuadAge = predict(object=ModelQuadAge, newdata=PredQuadAge, se.fit=TRUE, interval="confidence")
PredQuadAge = as.data.frame(PredQuadAge)  # Need to put x variable back in
PredQuadAge = cbind(PredQuadAge, data.frame(age18=seq(from=0, to=57, by=1)))

png(file = "Predicted Income by Quadratic Age Plot.png")  # open file
plot(y=PredQuadAge$fit.fit, x=PredQuadAge$age18, ylim=c(0,25), xlim=c(0,57),
     lty=1, type="l", ylab="Predicted Income", xlab="Age (0 = 18 years)")
lines(y=PredQuadAge$fit.upr, x=PredQuadAge$age18, lty=2, col="blue1")
lines(y=PredQuadAge$fit.lwr, x=PredQuadAge$age18, lty=2, col="blue1")
legend(x=20, y=5, legend=c("Predicted Income", "95% CI"), lty=1:2)
dev.off()  # close file
```

	fit.fit	fit.lwr	fit.upr	se.fit	df	residual.scale	age18
1	2.677	-0.4322	5.785	1.5835	731	13	0
2	3.880	1.0035	6.757	1.4653	731	13	1
3	5.045	2.3876	7.702	1.3534	731	13	2
4	6.170	3.7199	8.620	1.2480	731	13	3
5	7.256	4.9996	9.513	1.1495	731	13	4



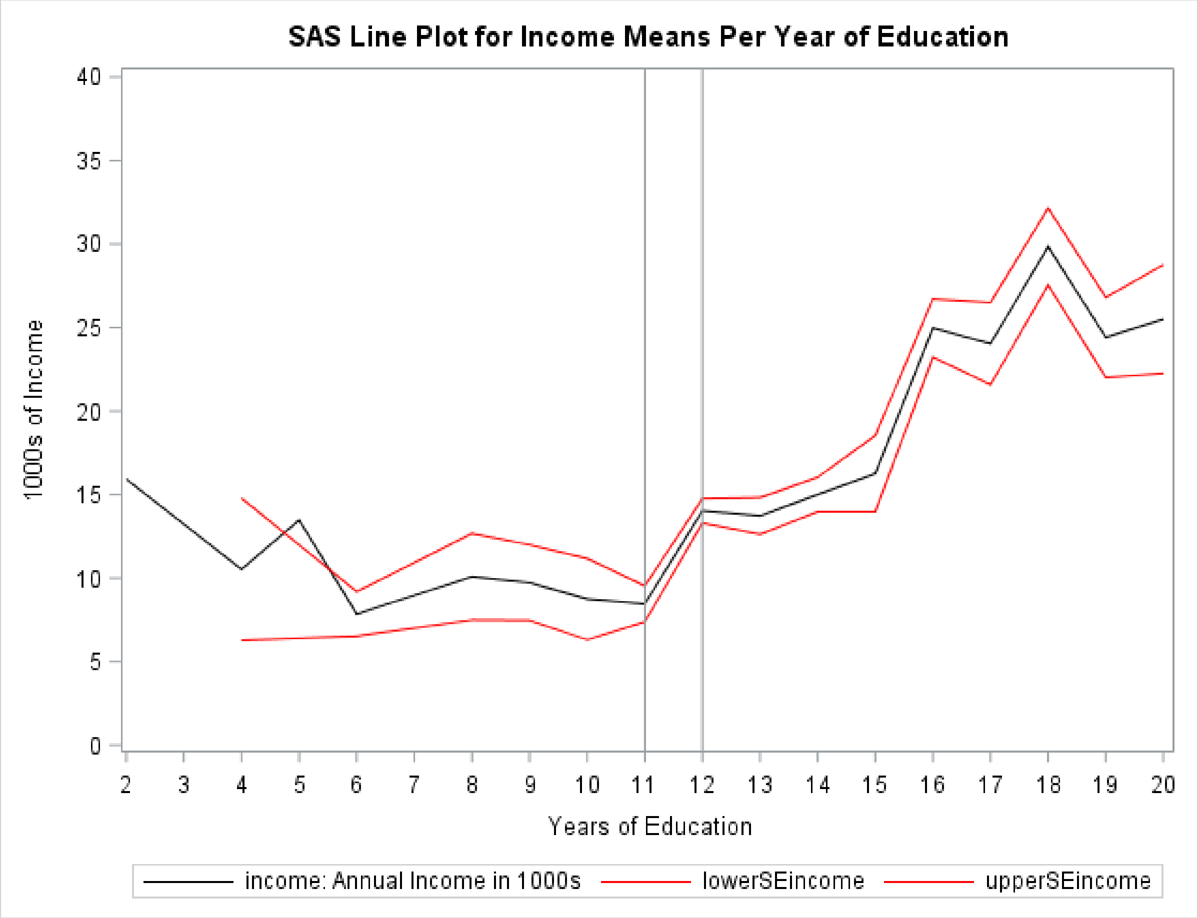
Example Results: quadratic age → income

We used a general linear model (i.e., linear regression) to examine the extent to which annual income in thousands of dollars (*M* = 17.30, *SD* = 13.79) could be predicted from years of age (*M* = 42.06, *SD* = 13.38, range = 18 to 75). We first examined the age-specific income means to identify plausible types of nonlinear associations. Given the apparent curvilinear trend (in which age appeared positively associated with income until middle age, after which it appeared negatively associated instead), we fit a model including fixed linear and quadratic slopes for age (in which age was centered such that 0 = 18 years, the minimum age in the sample). The quadratic age model captured a significant amount of variance in annual income, $F(2, 731) = 47.00$, MSE = 169.00, $p < .001$, $R^2 = .114$. The quadratic age model was also a significant improvement over a linear age model, as indicated by the significant fixed slope for the quadratic effect of age.

The model fixed effects can be interpreted as follows. The fixed intercept indicated that at age 18, annual income was predicted to be 2.676 thousand dollars (SE = 1.584) and was expected to be significantly greater by 1.223 thousand dollars per year of age (i.e., the instantaneous linear slope for age at age 18; SE = 0.135, $p < .001$). The linear age slope at age 18 was predicted to become significantly more negative per year of age by twice the quadratic coefficient of −0.020 (SE = 0.002, $p < .001$). As given by the quantity (−1*linear slope) / (2*quadratic slope) + 18, the age of maximum predicted personal income was 48.575 (i.e., the age at which the linear age slope is predicted to be 0). For example, the linear effect of age as evaluated at age 30 was significantly positive (Est = 0.754, SE = 0.079), the linear effect of age as evaluated at age 50 was nonsignificantly negative (Est = −0.027, SE = 0.047), and the linear effect of age as evaluated at age 70 was significantly negative (Est = −0.809, SE = 0.135).

Piecewise Slopes: education → income

- What if the effect of "more education" varies across education? For example, I hypothesize for predicted annual income:
  - **Less than HS degree?** No effect of educ
  - **Get HS degree?** Acute "bump" relative to less than HS degree
  - **More than HS degree?** Positive effect of more educ (likely nonlinear)

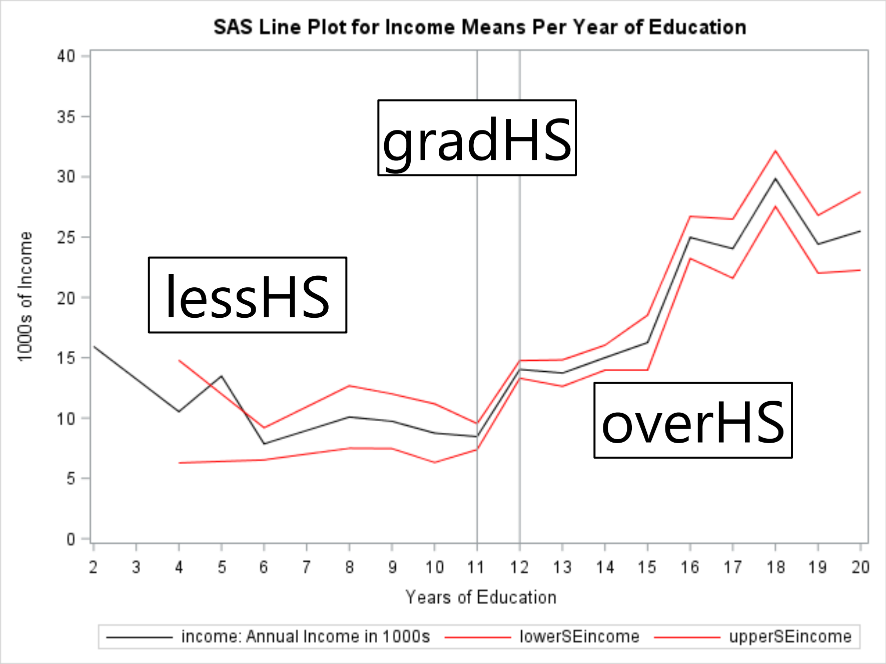


- Plot: black line shows mean per year of educ, red lines show ± 1 SE

Piecewise Slopes Coding for education

Years Educ (x)	lessHS: Slope if x <12	gradHS: HS Grad? (0=no, 1=yes)	overHS: Slope if x >12
9	-2	0	0
10	-1	0	0
11 (int)	0	0	0
12	0	1	0
13	0	1	1
14	0	1	2
15	0	1	3
16	0	1	4
17	0	1	5
18	0	1	6

- Intercept = grade 11 (when all slopes = 0)
- **3 predictors** for educ:
  - **lessHS**: from grade 2 to 11
  - **gradHS**: acute bump for 12+
  - **overHS**: after grade 12 (to 20)



R Syntax to Create Piecewise Slope Predictors

```
# Create 3 new variables for sections of education
# Make 3 new empty variables
Lecture3$lessHS=NA; Lecture3$gradHS=NA; Lecture3$overHS=NA
# Replace each for educ less than 12
Lecture3$lessHS[which(Lecture3$educ<12)]=
  Lecture3$educ[which(Lecture3$educ<12)]-11
Lecture3$gradHS[which(Lecture3$educ<12)]=0
Lecture3$overHS[which(Lecture3$educ<12)]=0
# Replace each for educ greater or equal to 12
Lecture3$lessHS[which(Lecture3$educ>=12)]=0
Lecture3$gradHS[which(Lecture3$educ>=12)]=1
Lecture3$overHS[which(Lecture3$educ>=12)]=
  Lecture3$educ[which(Lecture3$educ>=12)]-12
# lessHS: Slope for Years Ed Less Than High School
# gradHS: Acute Bump for Graduating High School
# overHS: Slope for Years Ed After High School
```

Years Educ (x)	lessHS: Slope if x <12	gradHS: HS Grad? (0=no, 1=yes)	overHS: Slope if x >12
9	-2	0	0
10	-1	0	0
11 (int)	0	0	0
12	0	1	0
13	0	1	1
14	0	1	2
15	0	1	3
16	0	1	4
17	0	1	5
18	0	1	6

R Syntax and Output for Piecewise education → income

```
print("GLM Predicting Income from 3 Piecewise Linear Slopes for Education")
ModelPieceEd = lm(data=Lecture3, formula=income~1+lessHS+gradHS+overHS)
obj=LMsummary(ModelPieceEd, effectsizes=TRUE)  # Custom output
```

```
Sums of Squares Table
              SS  DF       MS      F      p    R2
Model  22906.561   3 7635.520 47.838 <0.001 0.164
Error 116516.671 730  159.612
Total 139423.232 733  190.209

Fixed Effects Table
              Est    SE      t      p    LCI    UCI
Intercept   8.535 1.729  4.935 <0.001  5.140 11.930
lessHS     -0.269 0.599 -0.449  0.654 -1.444  0.907
gradHS      4.685 1.876  2.498  0.013  1.002  8.367
overHS      2.125 0.214  9.941 <0.001  1.705  2.544

Effect Sizes for Fixed Effects Table
          Est      p      d     pr   sR2
lessHS -0.269  0.654 -0.033 -0.017 0.000
gradHS  4.685  0.013  0.185  0.092 0.007
overHS  2.125 <0.001  0.736  0.345 0.113
```

Results for Piecewise Slopes of education → income

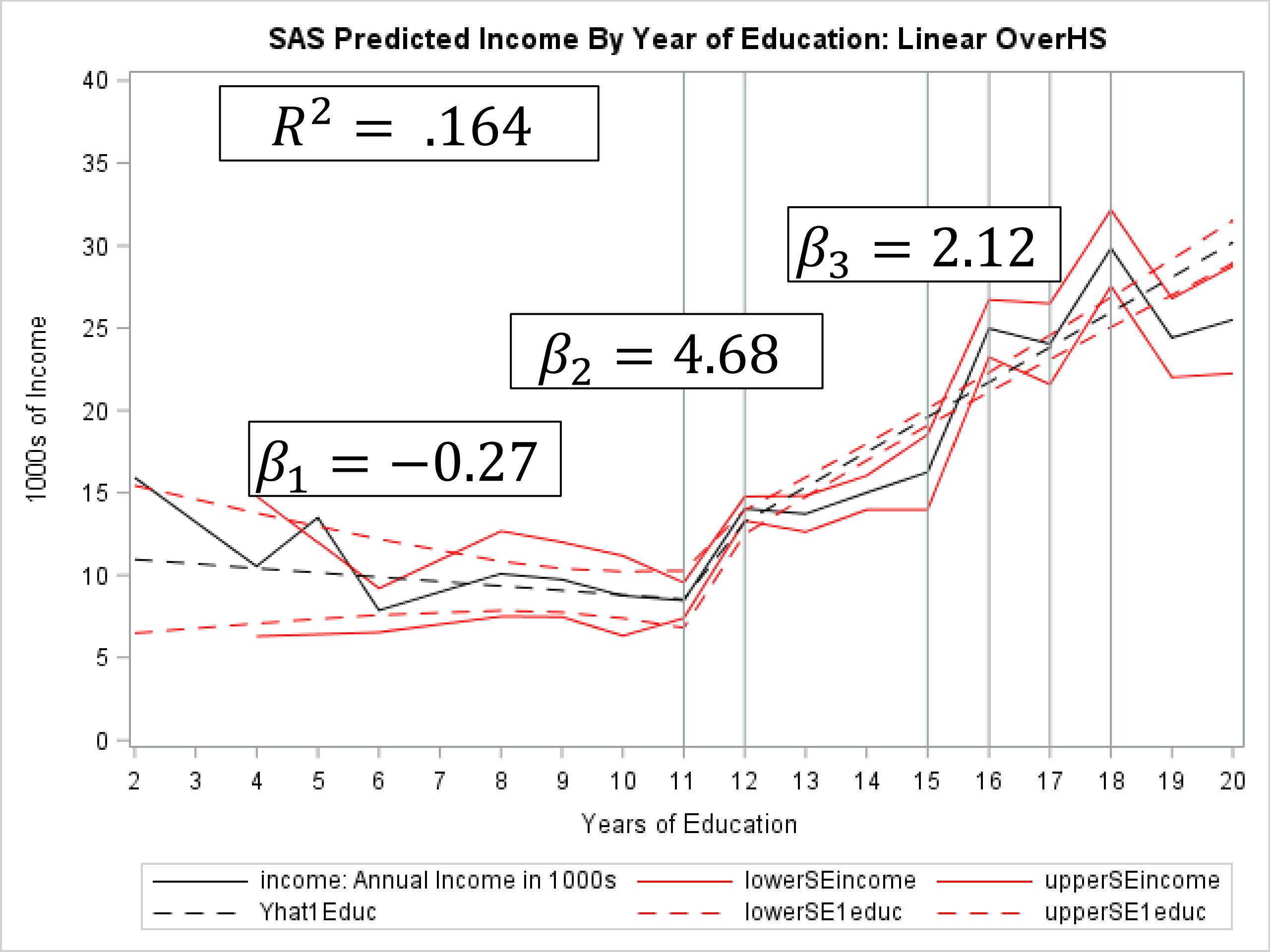
- After putting all three slopes in the model at the same time:

$$\boldsymbol{income_i = \beta_0 + \beta_1(lessHS_i) + \beta_2(gradHS_i) + \beta_3(overHS_i) + e_i}$$

- Model: $F(3,\ 730) = 47.84, MSE = 159.61, p < .001, R^2 = .164\ (r = .404)$
  - $r = .404$ is effect size for overall prediction by education (three slopes)
- $\boldsymbol{\beta_0}$ = expected income when all predictors = 0 → 11 years of ed here
  - $Est = 8.53, SE = 1.73$ *(significance and effect size not relevant)*
- $\boldsymbol{\beta_1}$ = slope for difference in income per year education from 2 to 11 years
  - $Est = -0.27, SE = 0.60, t(730) = 0.65, p = .654, pr = -.017$
- $\boldsymbol{\beta_2}$ = acute difference (jump) in income between educ=11 and educ=12
  - $Est = 4.68, SE = 1.88, t(730) = 2.05, p = .013, pr = .092$
- $\boldsymbol{\beta_3}$ = slope for difference in income per year education from 12 to 20 years
  - $Est = 2.12, SE = 0.214, t(730) = 9.94, p < .001, pr = .345$

Piecewise Slopes: Linear Past 12 Years of Ed?

- The model (dashed lines) appears to capture the mean trend (solid lines) pretty well until 12 years of education…
- I think we need even more piecewise slopes after ed=12!
  - From 12 to 15, 15 to 17–18, and 17–18 to 20
- But what if reviewer 2 questions whether 3 distinct slope are needed in the first place?



R Syntax and Output for education Linear Combinations

```
print("GLM Predicting Income from 3 Piecewise Linear Slopes for Education")
ModelPieceEd = lm(data=Lecture3, formula=income~1+lessHS+gradHS+overHS)
obj=LMsummary(ModelPieceEd, effectsizes=TRUE)  # Custom output

print("R Example of how to test differences between slopes")
print("Values below are multiplier for each fixed effect IN ORDER")
PredPieceEd = multcomp::glht(model=ModelPieceEd, linfct=rbind(
  "Diff in ed slope:  2-11 vs 11-12" = c(0,-1, 1,0),
  "Diff in ed slope: 11-12 vs 12-20" = c(0, 0,-1,1)))
obj=glhtSummary(glhtObject=PredPieceEd, effectsizes=TRUE)  # Custom output
```

```
 Linear Combinations Table
                                    Est    SE      t     p    LCI   UCI
 Diff in ed slope:  2-11 vs 11-12  4.954 2.282  2.170 0.030  0.473 9.434
 Diff in ed slope: 11-12 vs 12-20 -2.560 1.947 -1.315 0.189 -6.382 1.262

 Effect Sizes for Linear Combinations Table
                                    Est     p      d     pr   sR2
 Diff in ed slope:  2-11 vs 11-12  4.954 0.030  0.161  0.080 0.005
 Diff in ed slope: 11-12 vs 12-20 -2.560 0.189 -0.097 -0.049 0.002
```

Example Results: piecewise education → income

We used a general linear model (i.e., linear regression) to examine the extent to which annual income in thousands of dollars (*M* = 17.30, *SD* = 13.79) could be predicted from years of education (*M* = 13.81, *SD* = 2.91). We first examined the means of income by each year of education to identify plausible types of nonlinear associations. The effect of education predicting annual income appeared to differ across regions of education, suggesting a piecewise trend with the distinct region-specific slopes that could be captured by linear splines. Specifically, we fit one linear slope for the effect of education from 2 to 11 years, a second linear slope of education from 11 to 12 years, and a third linear slope of education from 12 to 20 years. Partial correlations were then computed from the *t* test-statistics to index effect size per slope. The model including these three education slopes captured a significant amount of variance in annual income, $F(3, 730) = 47.84$, MSE = 159.61, $p < .001$, $R^2 = .164$.

The model fixed slopes can be interpreted as follows. Annual income was expected to be nonsignificantly lower by 0.27 thousand dollars per year of education from 2 to 11 years ($SE = 0.60$, $p = .654$, $r = -.017$), resulting in predicted annual income of 8.53 thousand dollars (SE = 1.73) at 11 years of education (i.e., as given by the fixed intercept). Annual income was then expected to be significantly higher by 4.68 thousand dollars (SE =1.88, $p = .013$, $r = .092$) for those achieving a high school degree (i.e., a significant difference between 11 and 12 years of education). Annual income was expected to be significantly higher by 2.12 thousand dollars (SE = 0.21, $p < .001$, $r = .345$) per year of additional education past 12 years. However, examining a plot of the observed versus predicted means for annual income at each year of education suggested a linear slope was not sufficient in capturing the observed differences in income from 12 to 20 years of education. We recommend considering in future research the use of additional piecewise slopes corresponding to distinct levels of higher education (e.g., bachelors, masters, or doctoral college degrees).

Section Review: Nonlinear Trends

1. For what kinds of patterns would a quadratic slope be useful? Provide an example from your world (i.e., job, research, life).

2. For what kinds of patterns would a linear slope of $\log(x_i)$ be useful? Provide an example from your world (i.e., job, research, life).

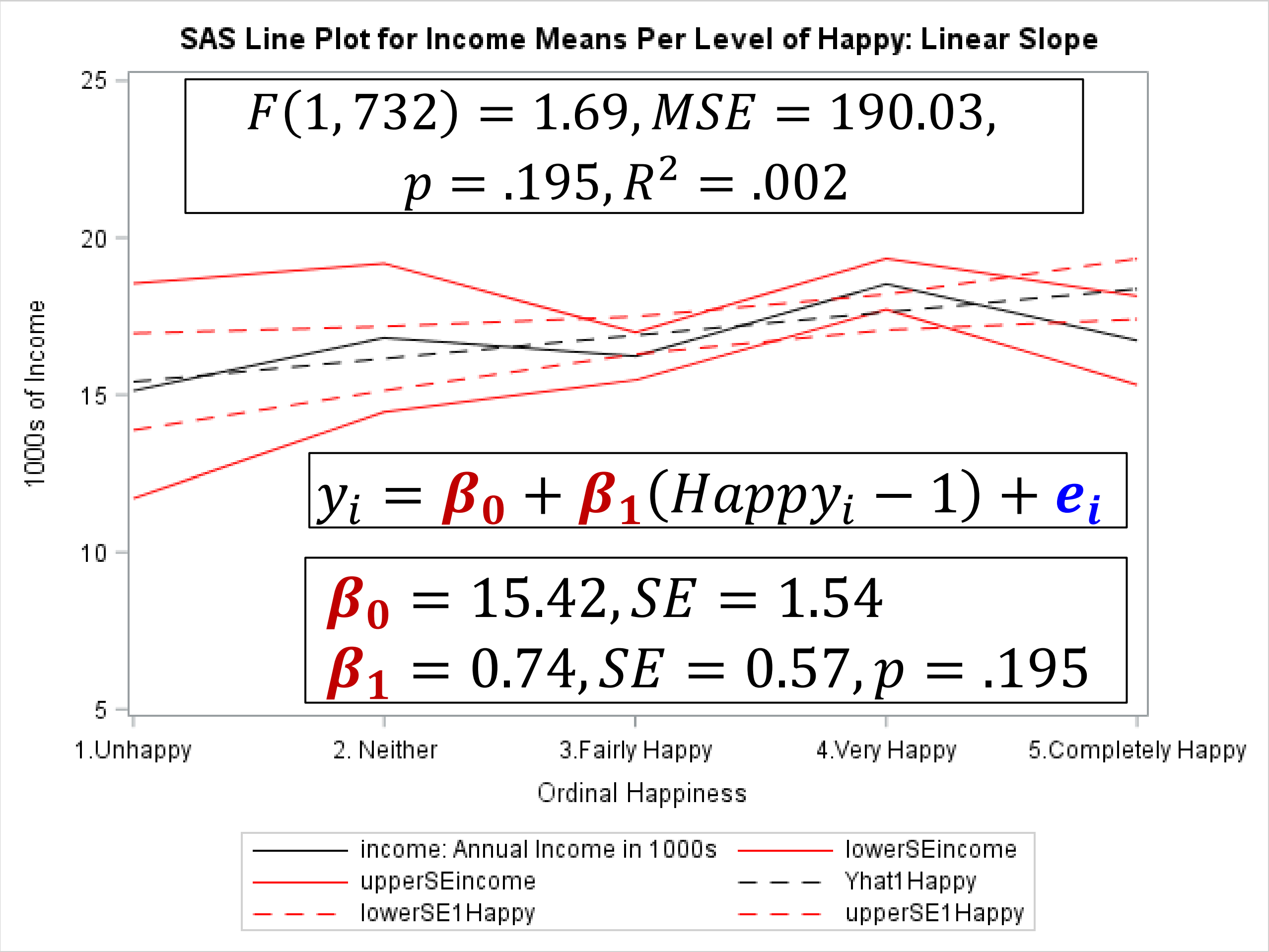
3. For what kinds of patterns would piecewise slopes (i.e., linear splines) be useful? Provide an example from your world (i.e., job, research, life).

4. What do you need to know in order to use the type of piecewise slopes model presented here?

A Linear Slope for an Ordinal Predictor???

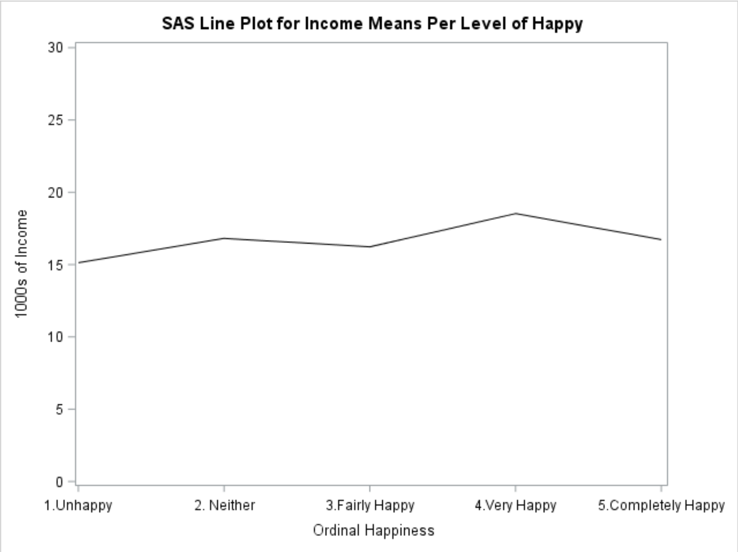
- Ordinal predictors with 5+ categories are often treated as interval by fitting a single linear slope for their overall effect ☹

- We can test this interval assumption by comparing outcome differences between adjacent predictor values
  - Here: need 4 slopes, 1 for each transition between categories
  - Use "**sequential coding**" to treat the predictor as "categorical" → 5 fixed effects used to distinguish the 5 outcome means



Sequential Slopes for Ordinal happy

Happy (x)	h1v2: Dif from 1 to 2	h2v3: Dif from 2 to 3	h3v4: Dif from 3 to 4	h4v5: Dif from 4 to 5
1 (int)	0	0	0	0
2	**1**	0	0	0
3	1	**1**	0	0
4	1	1	**1**	0
5	1	1	1	**1**



Group	LvM: Diff for Low vs Mid	LvU: Diff for Low vs Upp
Low	0	0
Mid	**1**	0
Upp	0	**1**

- Happy = 1 is where all slopes are 0, so it is the reference category (→ model intercept)
- The 4 slopes capture each **adjacent category difference** because each stays at 1 when done
  - Right: In indicator coding, the LvM slope went back to 0, so the second slope is NOT successive (i.e., it reflects LvU, not MvU)

R Syntax to Create Sequential Slopes for happy

```
# Create 4 new sequential-coded binary predictors for happy
# Make 4 new empty variables
Lecture3$h1v2=NA; Lecture3$h2v3=NA
Lecture3$h3v4=NA; Lecture3$h4v5=NA
# Replace each with 0 values
Lecture3$h1v2[which(Lecture3$happy<2)]=0
Lecture3$h2v3[which(Lecture3$happy<3)]=0
Lecture3$h3v4[which(Lecture3$happy<4)]=0
Lecture3$h4v5[which(Lecture3$happy<5)]=0
# Replace each with 1 values
Lecture3$h1v2[which(Lecture3$happy>=2)]=1
Lecture3$h2v3[which(Lecture3$happy>=3)]=1
Lecture3$h3v4[which(Lecture3$happy>=4)]=1
Lecture3$h4v5[which(Lecture3$happy>=5)]=1
# h1v2: Slope from Happy 1 to 2
# h2v3: Slope from Happy 2 to 3
# h3v4: Slope from Happy 3 to 4
# h4v5: Slope from Happy 4 to 5
```

Happy (x)	h1v2: Dif from 1 to 2	h2v3: Dif from 2 to 3	h3v4: Dif from 3 to 4	h4v5: Dif from 4 to 5
1 (int)	0	0	0	0
2	**1**	0	0	0
3	1	**1**	0	0
4	1	1	**1**	0
5	1	1	1	**1**

R Syntax and Output for Sequential happy Slopes

```
print("GLM Predicting Income from 4 Sequential Slopes for happy")
ModelHappy = lm(data=Lecture3, formula=income~1+h1v2+h2v3+h3v4+h4v5)
obj=LMsummary(ModelHappy, effectsizes=TRUE)  # Custom output
```

```
Fixed Effects Table
             Est    SE      t      p    LCI    UCI
Intercept 15.129 2.703  5.597 <0.001  9.822 20.435
h1v2       1.685 3.489  0.483  0.629 -5.165  8.536
h2v3      -0.586 2.369 -0.248  0.805 -5.238  4.065
h3v4       2.299 1.150  1.999  0.046  0.041  4.557
h4v5      -1.797 1.670 -1.076  0.282 -5.076  1.482

Effect Sizes for Fixed Effects Table
        Est     p      d     pr   sR2
h1v2  1.685 0.629  0.036  0.018 0.000
h2v3 -0.586 0.805 -0.018 -0.009 0.000
h3v4  2.299 0.046  0.148  0.074 0.005
h4v5 -1.797 0.282 -0.080 -0.040 0.002
```

Happy (x)	h1v2: Dif from 1 to 2	h2v3: Dif from 2 to 3	h3v4: Dif from 3 to 4	h4v5: Dif from 4 to 5
1 (int)	0	0	0	0
2	**1**	0	0	0
3	1	**1**	0	0
4	1	1	**1**	0
5	1	1	1	**1**

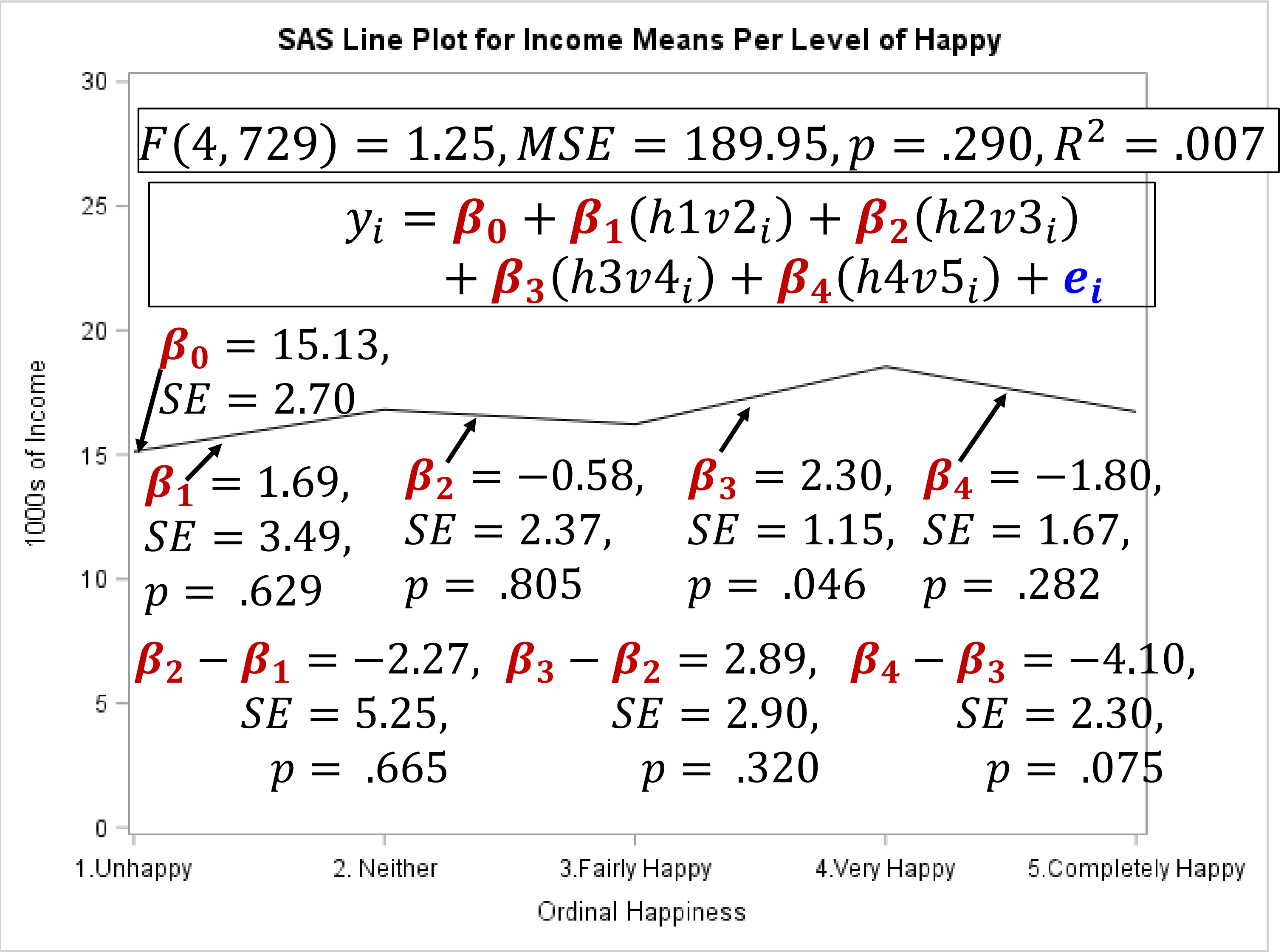
R Syntax and Output for happy Linear Combinations

```
print("Example of how to test differences between slopes")
print("Values below are multipliers for each fixed effect IN ORDER")
PredHappy = multcomp::glht(model=ModelHappy, linfct=rbind(
        "Diff in Slope 1-2 vs Slope 2-3" = c(0,-1, 1, 0, 0),
        "Diff in Slope 2-3 vs Slope 3-4" = c(0, 0,-1, 1, 0),
        "Diff in Slope 3-4 vs Slope 4-5" = c(0, 0, 0,-1, 1)))
obj=glhtSummary(glhtObject=PredHappy, effectsizes=TRUE)  # Custom output
```

```
 Linear Combinations Table
                                  Est    SE      t     p     LCI   UCI
 Diff in Slope 1-2 vs Slope 2-3 -2.272 5.247 -0.433 0.665 -12.573 8.029
 Diff in Slope 2-3 vs Slope 3-4  2.886 2.902  0.995 0.320  -2.811 8.582
 Diff in Slope 3-4 vs Slope 4-5 -4.096 2.297 -1.784 0.075  -8.605 0.413

 Effect Sizes for Linear Combinations Table
                                  Est     p      d     pr   sR2
 Diff in Slope 1-2 vs Slope 2-3 -2.272 0.665 -0.032 -0.016 0.000
 Diff in Slope 2-3 vs Slope 3-4  2.886 0.320  0.074  0.037 0.001
 Diff in Slope 3-4 vs Slope 4-5 -4.096 0.075 -0.132 -0.066 0.004
```

Results for happy Slopes and Slope Differences



Example Results: sequential happy → income

We used a general linear model (i.e., linear regression) to examine the extent to which annual income in thousands of dollars (*M* = 17.30, *SD* = 13.79) could be predicted from five-category ordinal happiness (unhappy = 3.54%, neither happy nor unhappy = 5.31%, fairly happy = 34.88%, very happy = 44.55%, completely happy = 11.72%). In first examining a linear effect of happiness (centered at unhappy = 0), the model fixed effects indicated that annual income was predicted to be 15.42 thousand dollars (SE = 1.54) for unhappy respondents (i.e., as given by the fixed intercept), and that annual income was predicted to be nonsignificantly greater by 0.74 thousand dollars (SE = 0.57, p = .195, R^2 = .002) per additional ordinal level of happiness.

However, given that a linear slope for happiness assumes interval differences with respect to predicted income, we tested this assumption by specifying a piecewise slopes model by which to estimate all sequential differences in predicted annual income by ordinal level of happiness. Partial correlations were then computed from the *t* test-statistics to index effect size per slope and slope difference. The revised model—predicting four sequential differences across the five levels of happiness—did not capture a significant amount of variance in annual income, $F(4, 729)$ = 1.25, MSE = 189.95, p = .290, R^2 = .007. The model fixed effects indicated that annual income was 15.13 thousand dollars (SE = 2.70) for unhappy respondents (i.e., as given by the fixed intercept). Annual income was nonsignificantly higher by 1.69 thousand dollars (SE = 3.49, p = .629, r = .018) for neither than unhappy respondents, nonsignificantly lower by 0.59 thousand dollars (SE = 2.37, p = .804, r = −.009) for fairly happy than neither respondents, significantly higher by 2.30 thousand dollars (SE = 1.15, p = .046, r = .073) for very happy than fairly happy respondents, and nonsignificantly lower by 1.80 thousand dollars (SE = 1.67, p = .282, r = −.040) for completely happy than very happy respondents. None of the differences between these adjacent differences were significant (as given by linear combinations of the model fixed effects, requested separately). Thus, there is little evidence that annual income can be predicted by self-rated happiness, whether treated as interval (through a linear slope) or treated as ordinal (through piecewise slopes).

Summary: Predictors with Multiple Fixed Slopes

- There are many scenarios in which a single predictor $\boldsymbol{x_i}$ needs multiple fixed slopes to describe its prediction of outcome $\boldsymbol{y_i}$:
  - Predictor variables with $\boldsymbol{C}$ categories needs $\boldsymbol{C} - 1$ fixed slopes for binary variables to distinguish its $\boldsymbol{C}$ possible different outcome means
    - "**Indicator** coding" is useful for nominal or ordinal predictors
    - "**Sequential** coding" can be more useful for ordinal predictors
  - Should report significance and effect size for each mean difference of theoretical interest (not necessarily all possible differences, though)
  - Nonlinear effects of quantitative predictor variables (via quadratic or exponential curves; piecewise slopes or curves) may require 2+ slopes
    - Predictors work together to summarize overall "trend" of $\boldsymbol{x_i}$ (so effect size for each fixed slope may be less important than overall model R^2)

- We want to know the significance of each fixed slope (via univariate Wald test of $(Est - H_0)/SE$ via t test-statistic) as well as significance of the model $\boldsymbol{R^2}$ (as multivariate Wald test via F test-statistic)
  - Model R^2 = squared Pearson $\boldsymbol{r}$ between predicted $\boldsymbol{\hat{y}_i}$ and actual $\boldsymbol{y_i}$