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

- Topics:
  - Reviewing empty GLMs and single predictor GLMs
  - GLM special cases: 2+ fixed effects to describe one variable
    - "Analysis of Variance" (ANOVA) for a one categorical predictor
      - e.g., income differences across 3 categories of working class
    - Nonlinear effects of a single quantitative predictor
      - e.g., quadratic continuous effect of years of age on income
      - e.g., piecewise discontinues 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

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
  - $y_i = (\text{constant} * 1) + (\text{constant} * \text{Xpred1}_i) + (\text{constant} * \text{Xpred2}_i)\ldots$
  - **General Linear Models** (**GLMs**) predict quantitative outcomes (continuous or continu-ish) using the normal distribution for the model residuals (the $\boldsymbol{e_i}$ values): $\boldsymbol{y_i} \sim \mathbf{N}(\boldsymbol{\hat{y}_i}, \boldsymbol{\sigma_e^2})$

- Empty GLM: **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**

- Add a predictor: **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** = (Est) = most likely value for the sample's slope
  - **Standard Error** (SE) = index of imprecision 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-0)/SE → "**Univariate Wald test**" provides 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, $\beta_{std} = Pearson\ r$

$$\boxed{\beta_{std} = \boldsymbol{\beta_{unstd}} * \frac{SD_x}{SD_{\boldsymbol{y}}}}$$

GLMs with Predictors: Binary vs. 3+ Categories

- To examine a **binary predictor** of a quantitative outcome, we only need **two fixed effects** to tell us **three things**: the outcome mean for Group 0, the outcome mean for Group 1, and the outcome mean difference
- **Actual** $y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{Group_i}) + \boldsymbol{e_i}$, **Predicted** $\hat{y}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{Group_i})$
  - Group 0 Mean: $\hat{y}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\mathbf{0}) = \boldsymbol{\beta_0}$ ← fixed effect #1
  - Difference of Group 1 from Group 0: $(\boldsymbol{\beta_0} + \boldsymbol{\beta_1}) - (\boldsymbol{\beta_0}) = \boldsymbol{\beta_1}$ ← fixed effect #2
  - Group 1 Mean: $\hat{y}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\mathbf{1}) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}$ ← linear combination of fixed effects
  - To get the estimate and SE for any term comprised of a <u>linear combination</u> of fixed effects, you need to ask for it via SAS ESTIMATE or STATA LINCOM
  - Btw, this specification is 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 two slopes: $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$
  - If $\boldsymbol{C} = \mathbf{4}$, then we need the $\boldsymbol{\beta_0}$ intercept and three 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 = $\frac{3*2*1}{(2*1)(1)} = 3$
  - This type of GLM goes by the name "**Analysis of Variance**" (**ANOVA**) in which the term "category" is usually replaced with "group" as a synonym

A GLM with a 3-Category Predictor

- Comparing the means of a quantitative outcome across **three groups** requires creating **two binary predictors** to be included **simultaneously** along with the intercept, for example, as coded so Low= Intercept (ref)

workclass variable ($N = 734$)	LvsM: Low vs Mid?	LvsU: Low vs Upp?
1. Low ($n = 436$)	0	0
2. Mid ($n = 278$)	1	0
3. Upp ($n = 20$)	0	1

SAS code:

```
DATA work.Example4; SET work.Example4;
LvsM=.; LvsU=.; * Make two new empty variables;
IF workclass=1 THEN DO; LvsM=0; LvsU=0; END; * Replace each for lower;
IF workclass=2 THEN DO; LvsM=1; LvsU=0; END; * Replace each for middle;
IF workclass=3 THEN DO; LvsM=0; LvsU=1; END; * Replace each for upper;
RUN;
```

STATA code (where = is used to change a value and == evaluates a condition):

```
gen LvsM=. // Make two new empty variables
gen LvsU=.
replace LvsM=0 if workclass==1 // Replace each for lower
replace LvsU=0 if workclass==1
replace LvsM=1 if workclass==2 // Replace each for middle
replace LvsU=0 if workclass==2
replace LvsM=0 if workclass==3 // Replace each for upper
replace LvsU=1 if workclass==3
```

A GLM with a 3-Category Predictor

- Comparing the means of a quantitative outcome across **three groups** requires creating **two binary predictors** to be included ***simultaneously*** along with the intercept, for example, as coded so Low= Intercept (ref)

workclass variable ($N = 734$)	LvsM: Low vs Mid?	LvsU: Low vs Upp?
1. Low ($n = 436$)	0	0
2. Mid ($n = 278$)	1	0
3. Upp ($n = 20$)	0	1

Actual: $y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{LvsM_i}) + \boldsymbol{\beta_2}(\boldsymbol{LvsU_i}) + \boldsymbol{e_i}$
Predicted: $\hat{y}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{LvsM_i}) + \boldsymbol{\beta_2}(\boldsymbol{LvsU_i})$

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

GLM 3-Category Predictor: SAS

Low Mean: $\hat{y}_L = \beta_0 + \beta_1(0) + \beta_2(0) = \beta_0$

Mid Mean: $\hat{y}_M = \beta_0 + \beta_1(1) + \beta_2(0) = \beta_0 + \beta_1$

Upp Mean: $\hat{y}_U = \beta_0 + \beta_1(0) + \beta_2(1) = \beta_0 + \beta_2$

	Alt Group	Ref Group	Difference
Diff of L vs M =	$(\beta_0 + \beta_1)$	$-\ (\beta_0)$	$= \beta_1$
Diff of L vs U =	$(\beta_0 + \beta_2)$	$-\ (\beta_0)$	$= \beta_2$
Diff of M vs U =	$(\beta_0 + \beta_2)$	$-\ (\beta_0 + \beta_1)$	$= \beta_2 - \beta_1$

For differences, note the order in the equations: the reference group mean *is subtracted from* the alternative group mean.

In SAS ESTIMATE statements, the variables refer to their fixed effects; the numbers refer to the values for the associated predictor variables.

```
TITLE "SAS GLM Predicting Income from workclass";
PROC GLM DATA=work.Example4 NAMELEN=100;
MODEL income = LvsM LvsU / SOLUTION ALPHA=.05 CLPARM;
* Ask for predicted income per group and group diffs;
ESTIMATE "Low Mean" intercept 1 LvsM  0 LvsU  0;
ESTIMATE "Mid Mean" intercept 1 LvsM  1 LvsU  0;
ESTIMATE "Upp Mean" intercept 1 LvsM  0 LvsU  1;
ESTIMATE "Low vs. Mid"          LvsM  1 LvsU  0;
ESTIMATE "Low vs. Upp"          LvsM  0 LvsU  1;
ESTIMATE "Mid vs. Upp"          LvsM -1 LvsU  1;
RUN; QUIT;
```

Intercepts are used <u>only</u> in creating predicted outcomes.

Positive values indicate addition; negative values indicate subtraction.

GLM 3-Category Predictor: STATA

Low Mean: $\hat{y}_L = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\mathbf{0}) + \boldsymbol{\beta_2}(\mathbf{0}) = \boldsymbol{\beta_0}$

Mid Mean: $\hat{y}_M = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\mathbf{1}) + \boldsymbol{\beta_2}(\mathbf{0}) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}$

Upp Mean: $\hat{y}_U = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\mathbf{0}) + \boldsymbol{\beta_2}(\mathbf{1}) = \boldsymbol{\beta_0} + \boldsymbol{\beta_2}$

	Alt Group	Ref Group	Difference
Diff of L vs M =	$(\boldsymbol{\beta_0} + \boldsymbol{\beta_1})$	$-\ (\boldsymbol{\beta_0})$	$= \boldsymbol{\beta_1}$
Diff of L vs U =	$(\boldsymbol{\beta_0} + \boldsymbol{\beta_2})$	$-\ (\boldsymbol{\beta_0})$	$= \boldsymbol{\beta_2}$
Diff of M vs U =	$(\boldsymbol{\beta_0} + \boldsymbol{\beta_2})$	$-\ (\boldsymbol{\beta_0} + \boldsymbol{\beta_1})$	$= \boldsymbol{\beta_2} - \boldsymbol{\beta_1}$

For differences, note the order of the equations: the reference group mean *is subtracted from* the alternative group mean.

In STATA LINCOM statements, the variables refer to their fixed effects; the numbers refer to the values of the predictor variables.

c. → a quantitative predictor (needed to do math in LINCOM)

```
display "STATA GLM Predicting Income from workclass"
regress income c.LvsM c.LvsU, level(95)
// Ask for predicted income per group and group differences
lincom _cons*1 + c.LvsM*0  + c.LvsU*0  // Low Mean
lincom _cons*1 + c.LvsM*1  + c.LvsU*0  // Mid Mean
lincom _cons*1 + c.LvsM*0  + c.LvsU*1  // Upp Mean
lincom           c.LvsM*1  + c.LvsU*0  // Low vs Mid
lincom           c.LvsM*0  + c.LvsU*1  // Low vs Upp
lincom           c.LvsM*-1 + c.LvsU*1  // Mid vs Upp
```

GLM 3-Category Predictor: Results

Group ($N = 734$)	LvsM	LvsU
1. Low ($n = 436$)	0	0
2. Mid ($n = 278$)	1	0
3. Upp ($n = 20$)	0	1

Empty Model: $\boldsymbol{y_i = \beta_0 + e_i}$

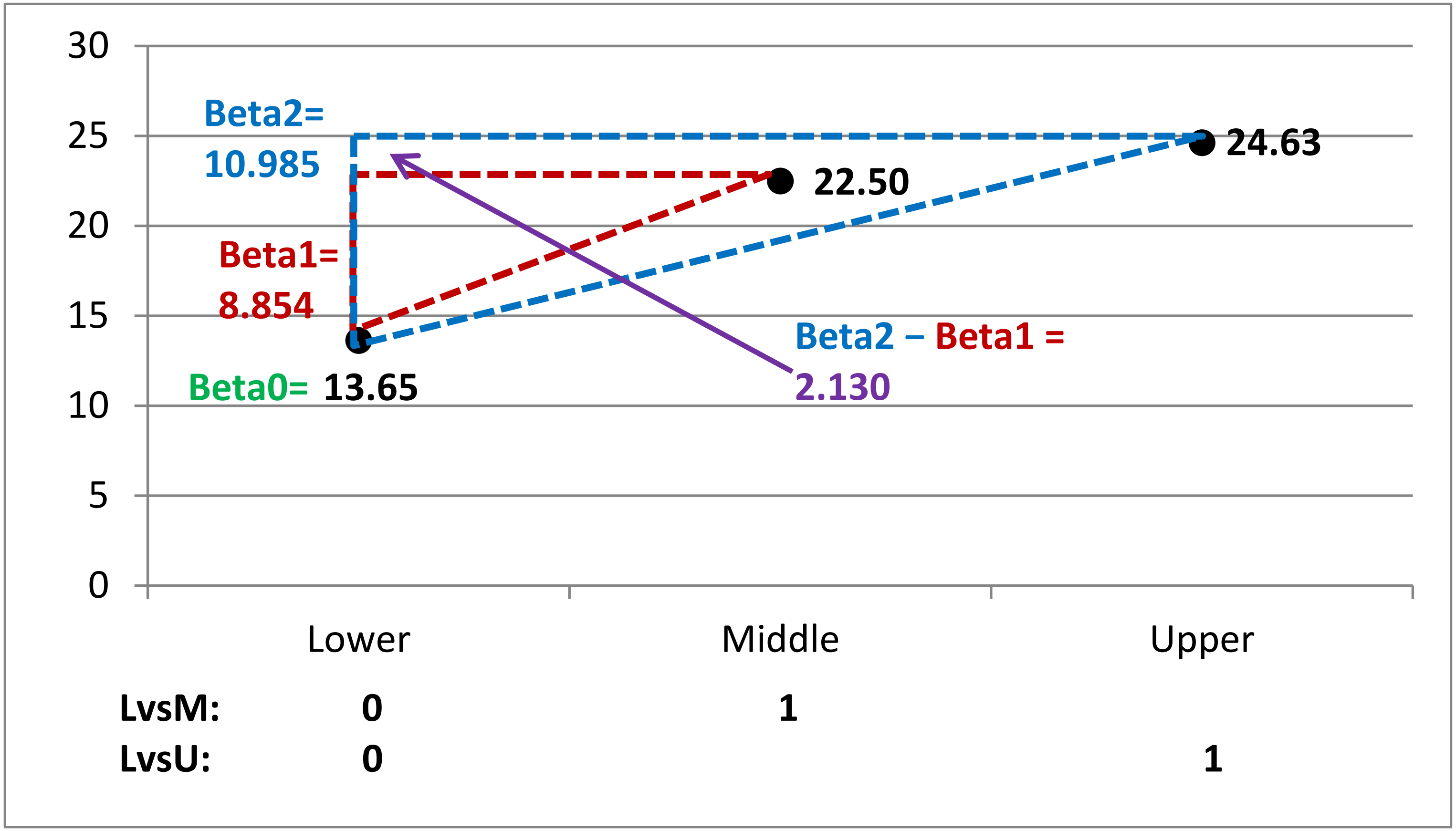
- Model parameters:
  - Intercept $\boldsymbol{\beta_0}$: $Est = 17.30$ $SE = 0.51$
  - Residual Variance $\boldsymbol{\sigma_e^2}$: $Est = 190.21$

Predictor Model: $\boldsymbol{y_i = \beta_0 + \beta_1(LvsM_i) + \beta_2(LvsU_i) + e_i}$

- Model parameters:
  - Intercept $\boldsymbol{\beta_0}$: $Est = 13.65, SE = 0.63,\ p < .001$ → Mean for Low ($= \hat{y}_L$)
  - Slope $\boldsymbol{\beta_1}$: $Est = 8.85, SE = 1.00, p < .001$ → Mean diff for Low vs Mid
  - Slope $\boldsymbol{\beta_2}$: $Est = 10.98, SE = 2.99, p < .001$ → Mean diff for Low vs Upp
  - Residual Variance $\boldsymbol{\sigma_e^2}$: $Est = 171.01$
- Linear combinations of model parameters:
  - Mid Mean: $\hat{y}_M = 13.65 + 8.85(1) + 10.98(0) = 22.50, SE = 0.78,\ p < .001$
  - Upp Mean: $\hat{y}_U = 13.65 + 8.85(0) + 10.98(1) = 24.63, SE = 2.92,\ p < .001$
  - Mean diff of Mid vs Upp $= \boldsymbol{\beta_2 - \beta_1} = 2.13, SE = 3.03, p = .482$

GLM 3-Category Predictor: Results

Fixed Effects				Predictors				Category	Pred
Beta0	Beta1	Beta2		Intercept	LvsM	LvsU		workclass	Y Hat
13.650	8.854	10.985		1	0	0		Lower	13.65
13.650	8.854	10.985		1	1	0		Middle	22.50
13.650	8.854	10.985		1	0	1		Upper	24.63



A GLM with a 3-Category Predictor

- The ANOVA-type question "Does group membership predict y_i?" translates to "Are there significant group mean differences in y_i"?

  - Can be answered **specifically via pairwise group differences** given directly by (or created from) the model fixed effects: For example: $\boldsymbol{y_i = \beta_0 + \beta_1(LvsM_i) + \beta_2(LvsU_i) + e_i}$,
    - Is $\boldsymbol{\beta_1} \neq 0$? If so, then $\hat{y}_M \neq \hat{y}_L$ (given directly because of our coding)
    - Is $\boldsymbol{\beta_2} \neq 0$? If so, then $\hat{y}_U \neq \hat{y}_L$ (given directly because of our coding)
    - Is $(\boldsymbol{\beta_2 - \beta_1}) \neq 0$? If so, then $\hat{y}_U \neq \hat{y}_M$ (requested as linear combination)

  - A **more general answer** to "Does group matter?" requires testing if $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ differ from 0 **jointly**, 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** $\hat{y}_i$ provided by this model with two group predictors **correlate significantly with the actual y_i?**

Prediction Gained vs. DF spent

- To provide a **more general answer** to "Does group matter?" we need to consider the impact of our prediction relative to how many fixed effects we needed to generate predicted $\hat{y}_i$ and how good they did (relative to what is left unknown)
  - This is an example of a "**multivariate Wald test**" (stay tuned for others)
  - "Relative" is qualified 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 slopes 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**" ($\boldsymbol{SS}$)
    - The basic form of $\boldsymbol{SS}$ is the **numerator** in 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**": $\boldsymbol{SS_{model}} = \sum_{i=1}^{N}(\boldsymbol{\hat{y}_i} - \boldsymbol{\bar{y}})^2$
- To quantify the **relative** size of that predicted info, we need to adjust it for $DF_{numerator}$ = number of fixed slopes = $k-1$
  - Then get "**Model Mean Square**" = $\boldsymbol{MS_{model}} = \frac{SS_{model}}{k-1}$ $\boxed{-1 \text{ because intercept doesn't get counted}}$
  - $\boldsymbol{MS_{model}}$ = "how much information has been captured per point spent"

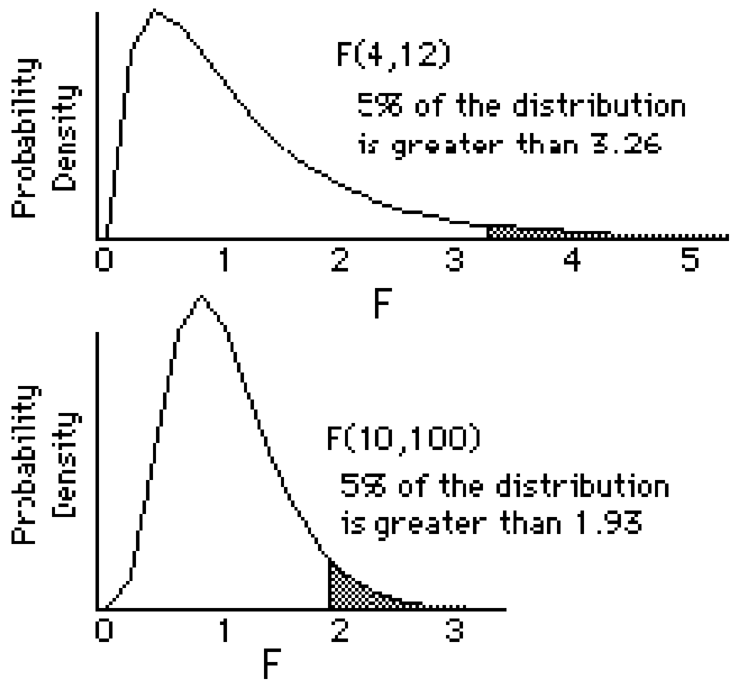
- How much information is **leftover** is quantified by "**residual** (or **error**) **sums of squares**": $\boldsymbol{SS_{residual}} = \sum_{i=1}^{N}(\boldsymbol{y_i} - \boldsymbol{\hat{y}_i})^2$
- To quantify the relative size of that leftover information, we need to adjust it for $DF_{denominator}$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

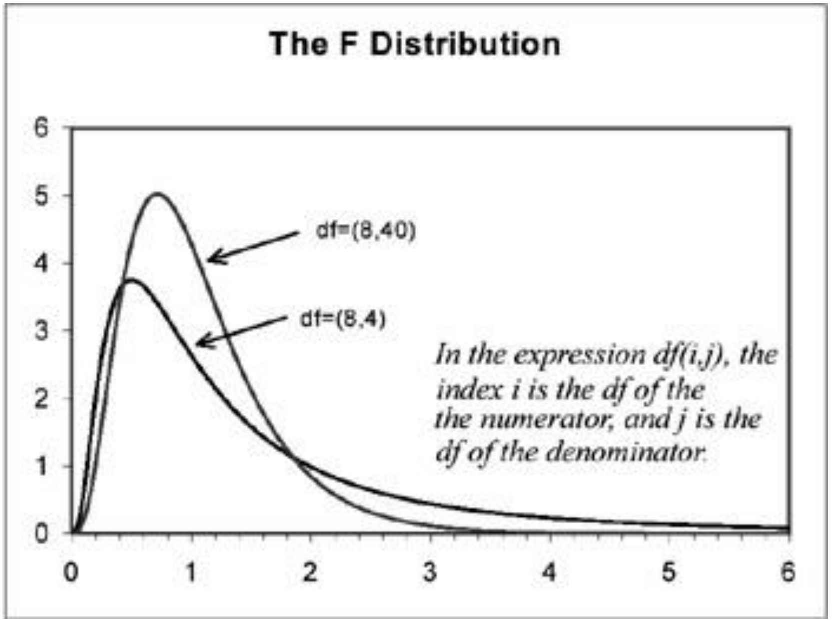
Prediction Gained vs. DF spent

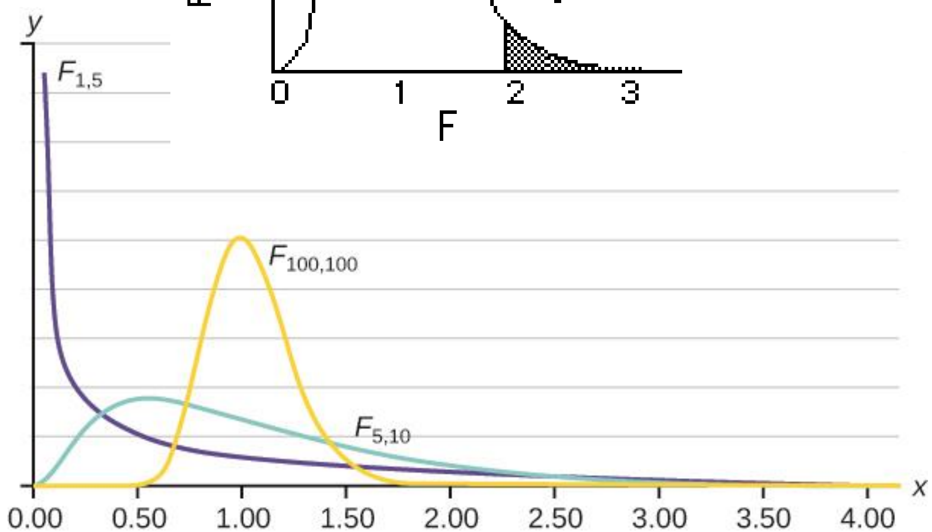
Source of Outcome Information	Sums of Squares (each summed from $i = 1$ to N)	Degrees of Freedom	Mean Square
Model (**known** because of predictor slopes)	SS_{model}: $(\hat{y}_i - \bar{y})^2$	DF_{num}: $k - 1$	MS_{model}: $\frac{SS_{model}}{k-1}$
Residual (leftover after predictors; still **unknown**)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

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 it must be positive
- Its shape (and thus the critical value for the boundary of where "unlikely" 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

Review: Steps in Significance Testing

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

Uses Denominator Degrees of Freedom?	Test 1 slope*	Test >1 slope*
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(s), then the obtained $\boldsymbol{p}$**-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(s), then the $\boldsymbol{p}$**-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 **<u>only 1 predictor</u>**, we don't need a separate F test-statistic of the model R^2 significance; for example: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x_i}) + \boldsymbol{e_i}$
  - Significance of unstandardized $\boldsymbol{\beta_1}$ comes from $\boldsymbol{t} = (\boldsymbol{Est} - \mathbf{0})/\boldsymbol{SE}$
    - Significance of the model prediction $\boldsymbol{R^2}$ from $\boldsymbol{F} = \boldsymbol{t^2}$ 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 **<u>2+ fixed slopes</u>**, we DO need to examine model $\boldsymbol{F}$ test-statistic and $\boldsymbol{R^2}$, for example: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{LvsM_i}) + \boldsymbol{\beta_2}(\boldsymbol{LvsU_i}) + \boldsymbol{e_i}$
  - $\boldsymbol{F}$ 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: Example

- For example: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{LvsM_i}) + \boldsymbol{\beta_2}(\boldsymbol{LvsU_i}) + \boldsymbol{e_i}$
- Group-specific results: We already know that L<M, L<U, and L=M
- Significance test of the model: $R^2 = .103$
- Report as $F(DF_{num}, DF_{den}) = Fvalue,\ \mathrm{MSE} = MS_{res},\ p < pvalue$

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}: $(\hat{y}_i - \bar{y})^2$ $= 14{,}414.03$	$\boldsymbol{DF_{num}: k-1}$ $= 2$ slopes (-1 for int)	$\boldsymbol{MS_{model}: \frac{SS_{model}}{k-1}}$ $= 7{,}207.01$	42.14
Residual (**"error"**)	$SS_{residual}$: $(y_i - \hat{y}_i)^2$ $= 125{,}009.25$	$\boldsymbol{DF_{den}: N-k}$ $= 731$ leftover	$\boldsymbol{MS_{residual}: \frac{SS_{residual}}{N-k}}$ $= 171.01$	
Corrected Total (after $\bar{y}$)	SS_{total}: $(y_i - \bar{y})^2$ $= 139{,}423.23$	$\boldsymbol{N = 734 - 1}$ $= 733$ total corrected for int		

Another version of R^2: "Adjusted R^2"

- Just like we may want to adjust Pearson's r for bias due to small sample size, some feel the need to adjust the model R^2
  - Pearson $r_{adj} = \sqrt{1 - \frac{(1-r^2)(N-1)}{N-2}}$
  - Model $R^2_{adj} = 1 - \frac{(1-R^2)(N-1)}{N-k-1}$

The difference for R^2_{adj} relative to r_{adj} is removing the square root, as well as generalizing to k model fixed effects to make $\hat{y}_i$

- Although adjusted R^2 is provided by many statistical programs, 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: $R^2_{adj} = 1 - \frac{(1-.103)(734-1)}{734-3} = .101$ ($R^2_{unadj} = .103$)
  - Btw, I had to use SAS PROC REG instead of SAS PROC GLM to get R^2_{adj} (both versions are provided by STATA REGRESS)

Effect Size: General vs Specific

- The model R^2 value (the square of the correlation between predicted $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ and actual $\boldsymbol{y}_{\boldsymbol{i}}$) provides a general effect size, but you may also need an **effect size for each fixed slope**
- For categorical predictors, an $\boldsymbol{r}$ effect size (standardized slope) is less helpful than an alternative effect size: **Cohen's $\boldsymbol{d}$**, a **standardized mean difference** between two groups (1 and 2)
  - $d = \frac{\bar{y}_1 - \bar{y}_2}{SD_{pooled}}$, where $SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}} = \sqrt{\boldsymbol{\sigma_e^2}}$ (group-only model)
    - Other variants you might see: Glass' delta uses SD for only 1 group; Hedges' g that weights by the relative sample size in each group
  - Can compute d from t test-statistic for a fixed effect: $d = \frac{2t}{\sqrt{DF_{den}}}$
    - Btw, d and r can be converted as: $d = \sqrt{\frac{4r^2}{1-r^2}}$, $r = \sqrt{\frac{d^2}{4+d^2}}$

Effect Sizes for Our Example and Sample Sizes Needed for Power = .80

- L vs M Diff as slope $\boldsymbol{\beta_1}$: $Est = 8.85, SE = 1.00, p < .001$
  - $t = \frac{8.85-0}{1.00} = 8.82, d = \frac{2*8.82}{\sqrt{731}} = 0.65$ → ~power > .80 for $\boldsymbol{n} > \mathbf{45}$

- L vs U Diff as slope $\boldsymbol{\beta_2}$: $Est = 10.98, SE = 2.99, p < .001$
  - $t = \frac{10.98-0}{2.99} = 3.67, d = \frac{2*3.67}{\sqrt{731}} = 0.27$ → ~power > .80 for $\boldsymbol{n} > \mathbf{175}$

- M vs U Diff as: $\boldsymbol{\beta_2} - \boldsymbol{\beta_1} = 2.13, SE = 3.03, p = .482$
  - $t = \frac{2.13-0}{3.03} = 0.70, d = \frac{2*0.70}{\sqrt{731}} = 0.05$ → ~power > .80 for $\boldsymbol{n} > \mathbf{2,102}$

- Model $R^2 = .103, r = .322$ → ~power > .80 for $\boldsymbol{N} > \mathbf{85}$

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 $\boldsymbol{\hat{y}_i}$ 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 model-predicted $\boldsymbol{\hat{y}_i}$ and actual $\boldsymbol{y_i}$, which is necessary whenever the predicted $\boldsymbol{\hat{y}_i}$ uses multiple $\boldsymbol{\beta_{unstd}}$ slopes)
    - F test-statistic evaluates model R^2 *per DF spent to get it and DF leftover*

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 describe its model prediction (its "effect" or "trend") is **when a quantitative predictor has a nonlinear relation with the outcome**
- We will examine three types of examples 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 can be nonlinear, too
  - **Testing the assumption of linearity**: 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 entering the predictor squared (multiplied by itself)
    - Good to create relationships that **change directions**
    - Example for quadratic trend of $\boldsymbol{x_i}$: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{x_i}) + \boldsymbol{\beta_2}(\boldsymbol{x_i})^2 + \boldsymbol{e_i}$
  - **Exponential(ish)**: created from one nonlinearly-transformed predictor
    - Predictor = natural-log transform of predictor (positive values only)
    - Good to create relationships that look like **diminishing returns**
    - Example for exponential(ish) trend of $\boldsymbol{x_i}$: $\boldsymbol{y_i} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{Log}[\boldsymbol{x_i}]) + \boldsymbol{e_i}$

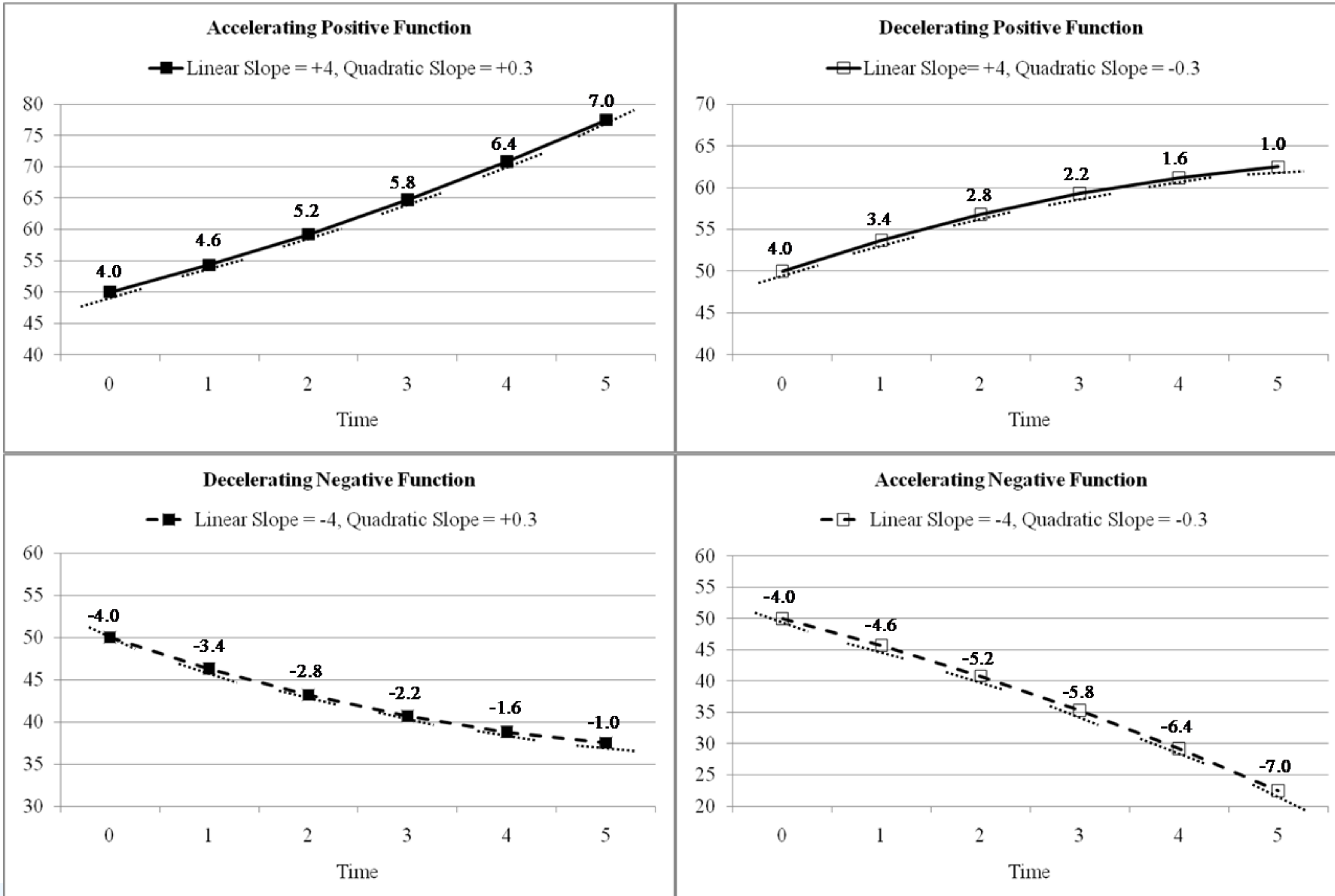
How to Interpret Quadratic Slopes

- 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 $\boldsymbol{x_i}$ 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 derivatives of the function with respect to x_i:

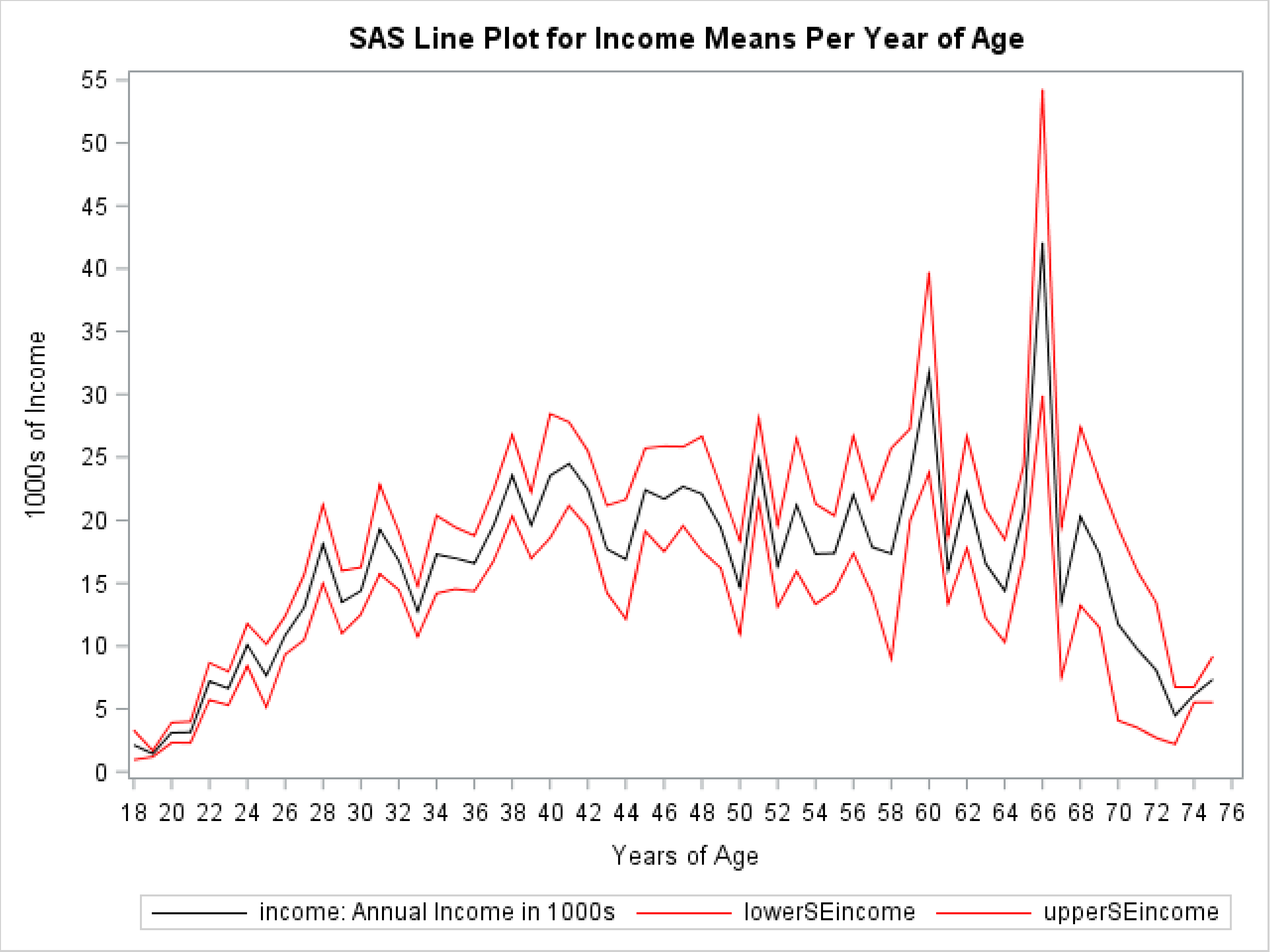
Intercept (Position) at $x_i = X$: $\hat{y}_X = 50.0 + 4.0x_i + 0.3x_i^2$
First Derivative (Velocity) at X: $\dfrac{d\hat{y}_X}{d(X)} = 4.0 + 0.6x_i$
Second Derivative (Acceleration) at X: $\dfrac{d^2\hat{y}_X}{d(X)} = 0.6$

Quadratic Trends: Example of x_i = Time



Quadratic Trend for Age: GSS Example

- 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 personal income (yay middle age!)



- Let's see how to fit a quadratic effect of age predicting income in SAS and STATA…

A GLM with a Quadratic Predictor Trend

SAS code:

```
* Make new age variable centered at 18;
DATA work.Example4; SET work.Example4;
age18=age-18;
RUN;

TITLE "SAS GLM Predicting Income from Quadratic Centered Age";
PROC GLM DATA=work.Example4 NAMELEN=100;
* Asterisk creates multiplied predictor variable;
MODEL income = age18 age18*age18 / SOLUTION ALPHA=.05 CLPARM;
RUN; QUIT; TITLE;
```

STATA code:

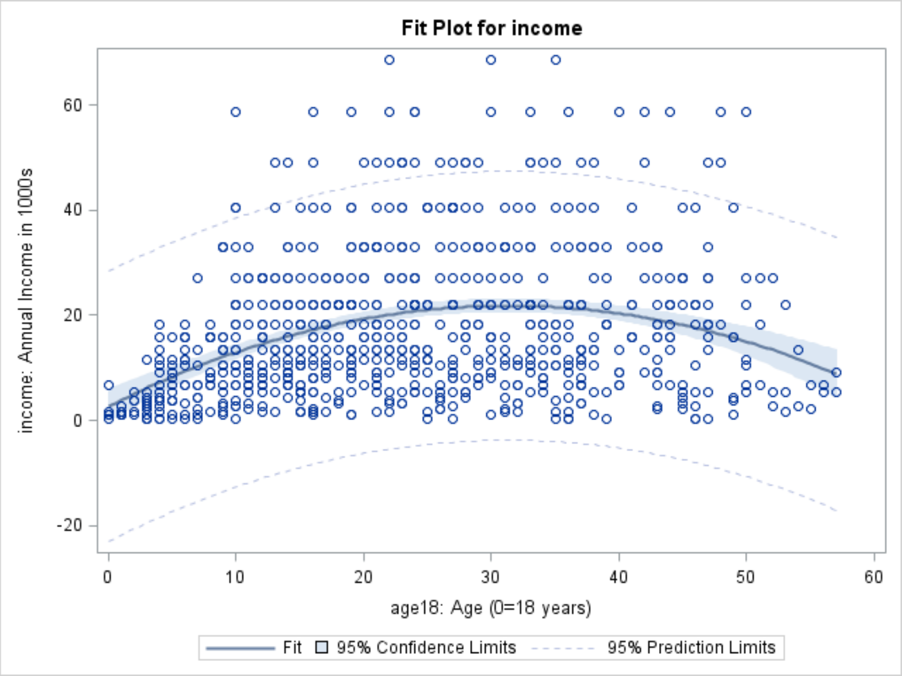
```
gen age18=age-18 // Make new age variable centered at 18

display "STATA GLM Predicting Income from Centered Quadratic Age"
// Pound sign (hashtag) creates multiplied predictor variable
regress income c.age18 c.age18#c.age18, level(95)
```

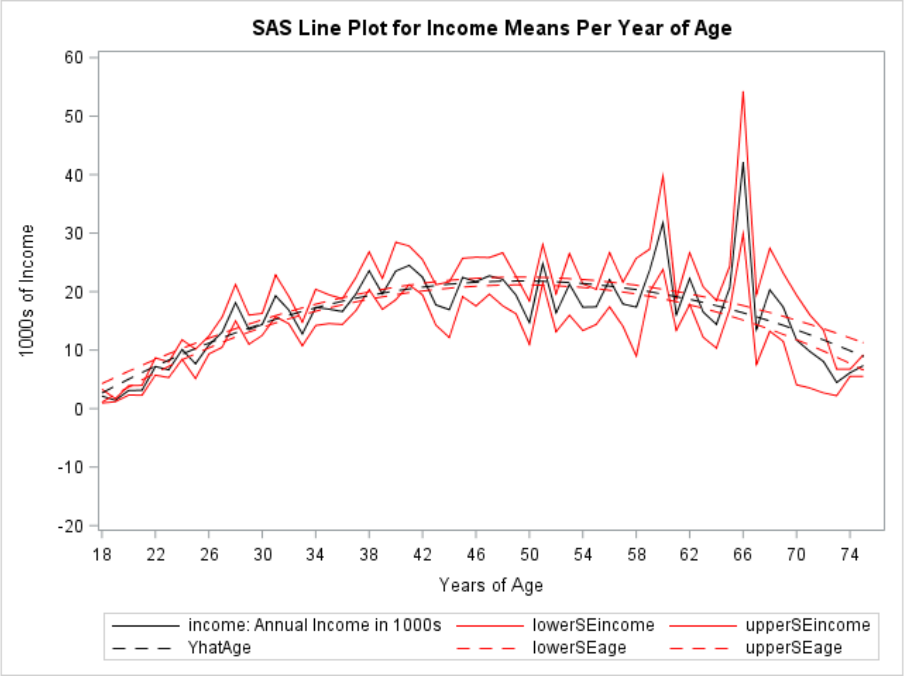
See Example 4 for SAS ESTIMATE and STATA LINCOM statements to generate illustrative predicted outcomes and age-specific instantaneous linear slopes…

Quadratic Trend for Age: GSS Example

- $Income_i = \beta_0 + \beta_1(Age_i - 18) + \beta_2(Age_i - 18)^2 + e_i$
  - **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 (or deceleration here) 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: $\hat{y}_{x=30} = \mathbf{2.677} + \mathbf{1.223}(\mathbf{12}) - \mathbf{0.020}(\mathbf{12})^2 = 14.540, SE = 0.647$
  - Age 50: $\hat{y}_{x=50} = \mathbf{2.677} + \mathbf{1.223}(\mathbf{32}) - \mathbf{0.020}(\mathbf{32})^2 = 21.809, SE = 0.668$
  - Age 70: $\hat{y}_{x=70} = \mathbf{2.677} + \mathbf{1.223}(\mathbf{52}) - \mathbf{0.020}(\mathbf{52})^2 = 13.448, SE = 1.659$
- Predicted linear age slope at other ages via linear combinations:
  - Age 30: $\widehat{\beta_1}_{x=30} = \mathbf{1.223} - \mathbf{0.020}(\mathbf{2} * \mathbf{12}) = 0.754, SE = 0.079$
  - Age 50: $\widehat{\beta_1}_{x=50} = \mathbf{1.223} - \mathbf{0.020}(\mathbf{2} * \mathbf{32}) = -0.027, SE = 0.047$
  - Age 70: $\beta_{1_{x=70}} = \mathbf{1.223} - \mathbf{0.020}(\mathbf{2} * \mathbf{52}) = -0.809, SE = 0.135$
- Predicted age with maximum income (linear age slope = 0): $\frac{-\beta_1}{2*\beta_2} + 18 = 48.575$

Quadratic Trend for Age: GSS Example

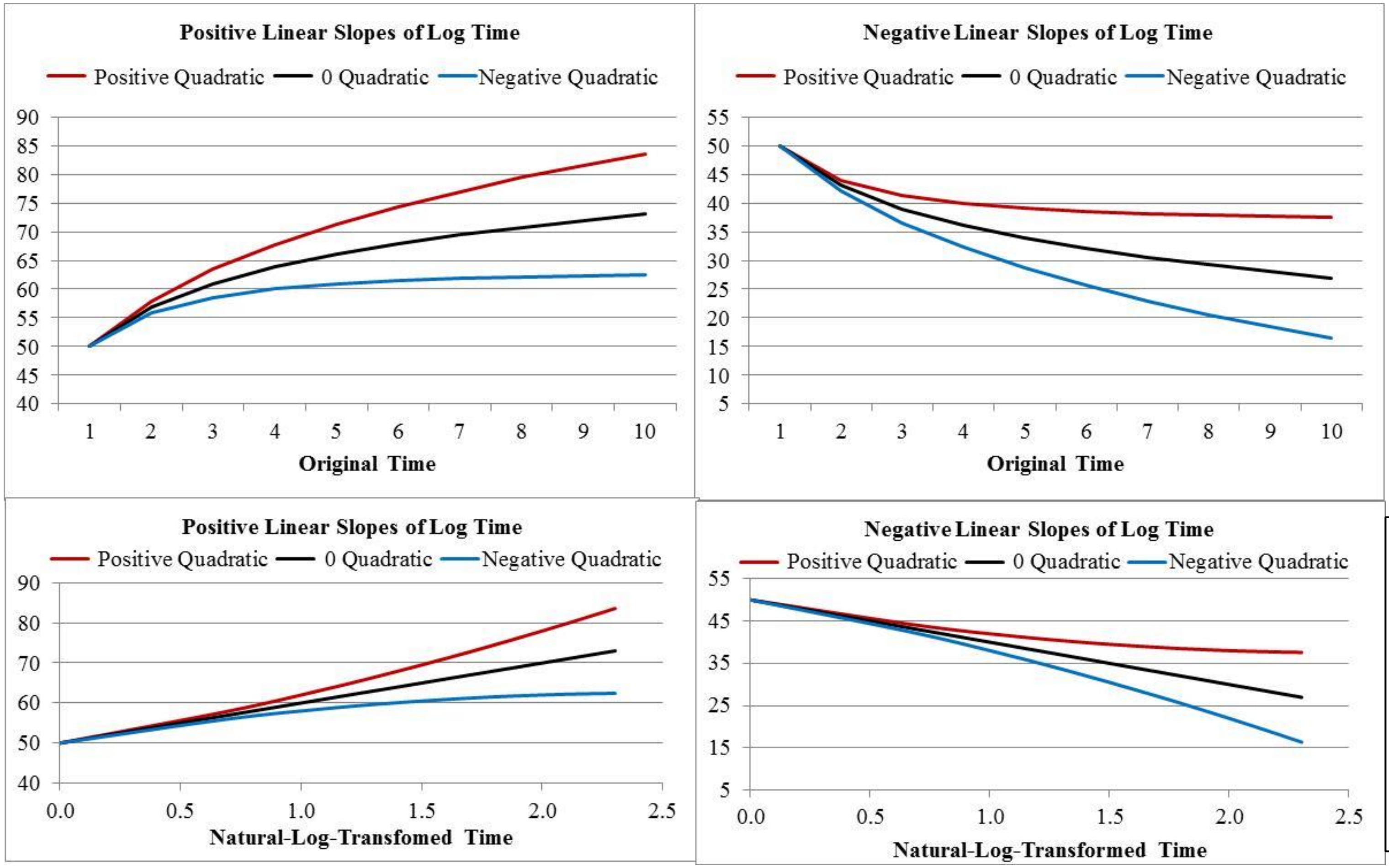




- Left: predicted regression line over individual scatterplot
- Right: predicted regression line over mean per age
  - $F(2,\ 731) = 47.00, MSE = 169.00, p < .001, R^2 = .114\ (r = .338)$

Exponential Trends: Example of x_i = Time

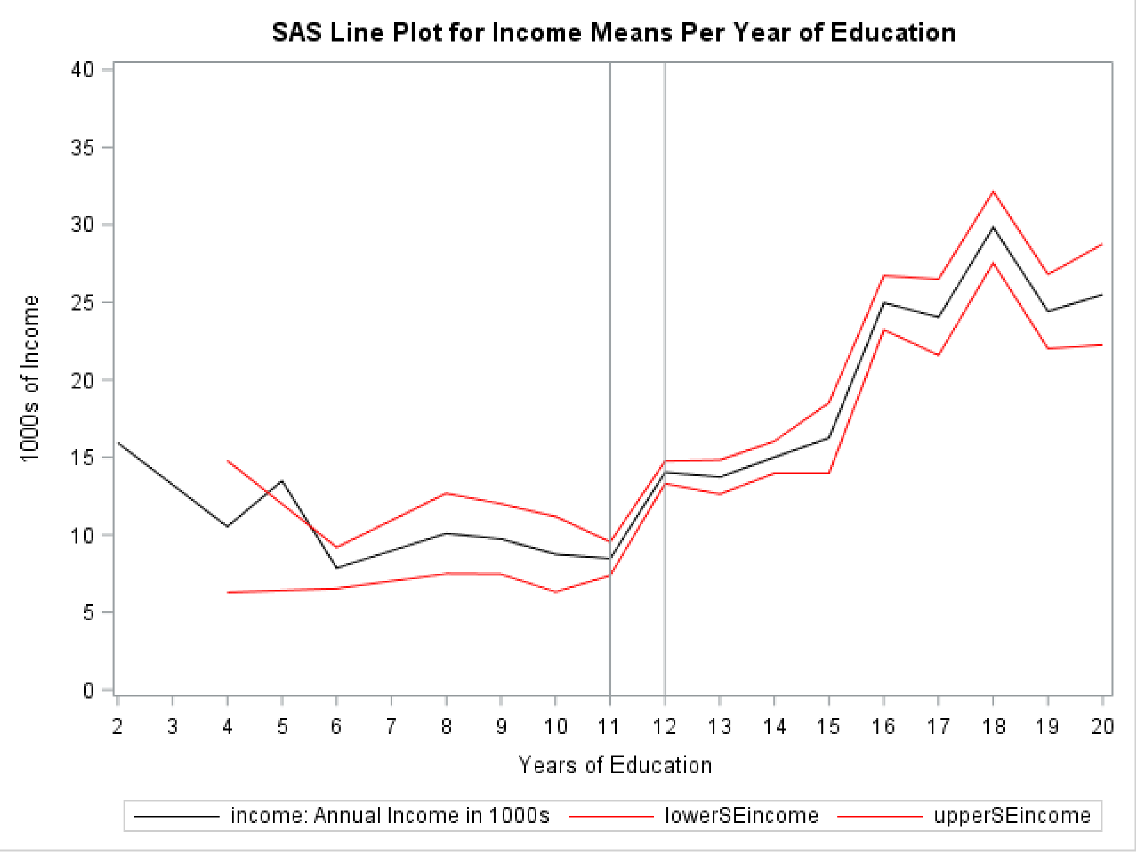
- A <u>linear</u> slope of log x_i (black lines) mimics an <u>exponential</u> trend across *original* x_i; adding a quadratic slope of log x_i (red or blue lines) can speed up or slow down the exponential(ish) trend



Bottom: There is a linear relationship between log x_i and the outcome.

Piecewise Slopes: GSS Example

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

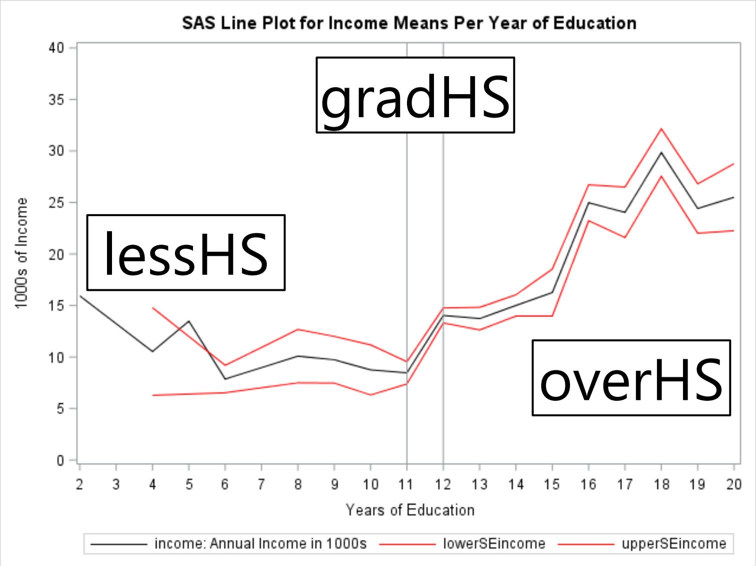


- Plot: black line shows mean per year of educ, red lines show ± 1 SE

Piecewise Slopes Coding: GSS Example

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 linear slopes for educ:
  - lessHS: from grade 2 to 11
  - gradHS: acute bump for 12+
  - overHS: after grade 12 (to 20)



Piecewise Slopes for Education: SAS

```
* Make 3 new variables for sections of education;
DATA work.Example4; SET work.Example4;
lessHS=.; gradHS=.; overHS=.; * Make three new empty variables;
* Replace for educ less than 12;
IF educ LT 12 THEN DO; lessHS=educ-11; gradHS=0; overHS=0;        END;
* Replace for educ greater or equal to 12;
IF educ GE 12 THEN DO; lessHS=0;          gradHS=1; overHS=educ-12; END;
RUN;

TITLE "SAS GLM Predicting Income
       from Piecewise Education";
PROC GLM DATA=work.Example4 NAMELEN=100;
MODEL income = lessHS gradHS overHS
             / SOLUTION ALPHA=.05 CLPARM;
RUN; QUIT; TITLE;
```

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

Piecewise Slopes for Education: STATA

```
// Make 3 new vars for sections of educ
gen lessHS=. // Make 3 new empty variables
gen gradHS=.
gen overHS=.
// Replace for educ less than 12
replace lessHS=educ-11 if educ <  12
replace gradHS=0       if educ <  12
replace overHS=0       if educ <  12
// Replace for educ greater or equal to 12
replace lessHS=0       if educ >= 12
replace gradHS=1       if educ >= 12
replace overHS=educ-12 if educ >= 12

display "STATA GLM Predicting Income from Piecewise Education"
regress income c.lessHS c.gradHS c.overHS, level(95)
```

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

Piecewise Slopes: GSS Results

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

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

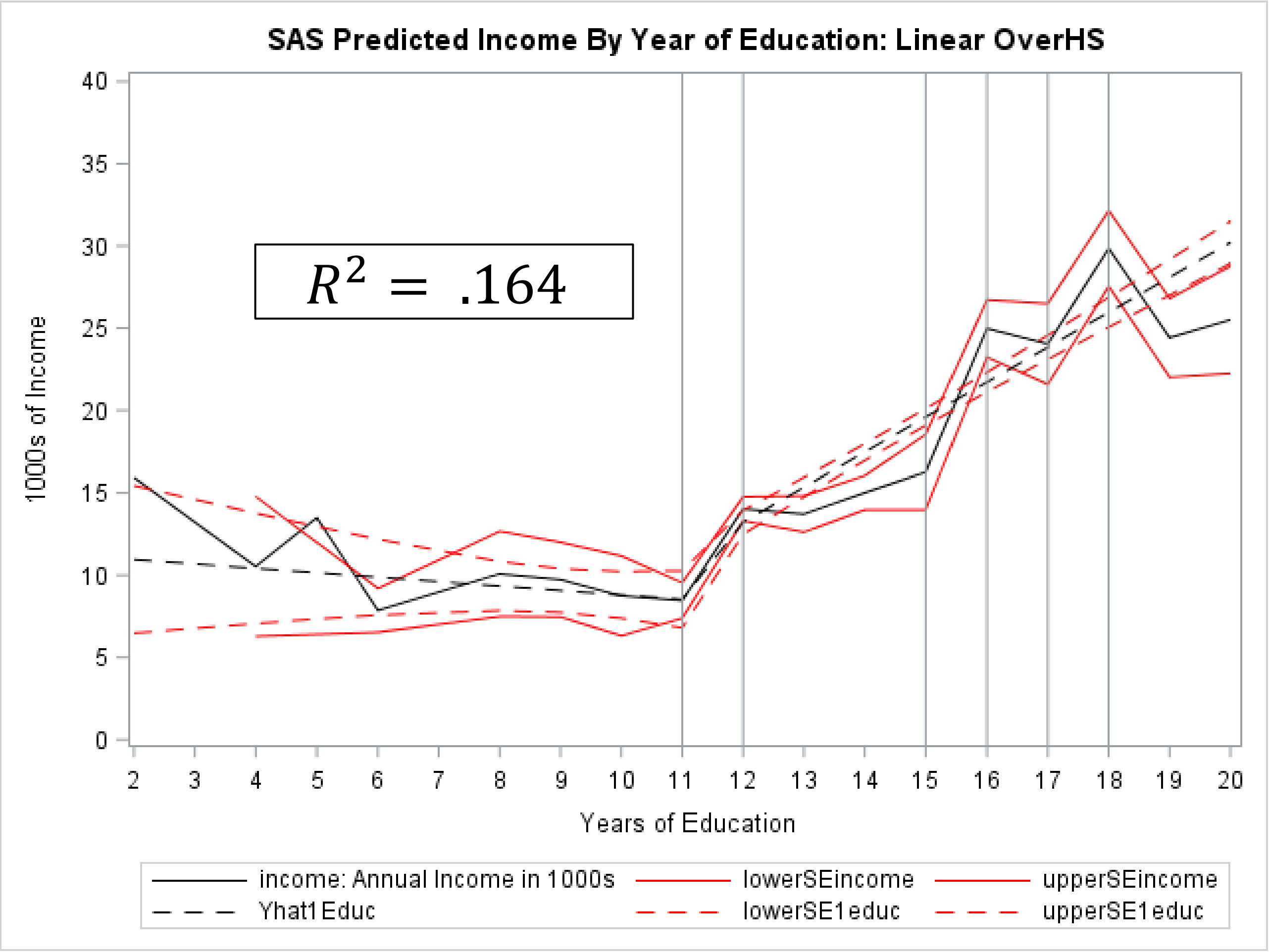
- Model: $F(3,\ 730) = 47.84, MSE = 159.61, p < .001, R^2 = .164\ (r = .404)$

- $\boldsymbol{\beta_0}$ = expected income when all predictors = 0 → 11 years of ed here
  - $Est = 8.53, SE = 1.73, p < .001$
- $\boldsymbol{\beta_1}$ = slope for difference in income <u>per year education</u> from 2 to 11 years
  - $Est = -0.27, SE = 0.60, p = .654, \beta_{1std} = -0.019$
- $\boldsymbol{\beta_2}$ = acute difference (jump) in income between educ=11 and educ=12
  - $Est = 4.68, SE = 1.88, p = .013, \beta_{1std} = 0.112, d = 0.185$
- $\boldsymbol{\beta_3}$ = slope for difference in income <u>per year education</u> from 12 to 20 years
  - $Est = 2.12, SE = 0.214, p < .001, \beta_{1std} = 0.359$

Piecewise Slopes: Linear Past 12 Years Ed?

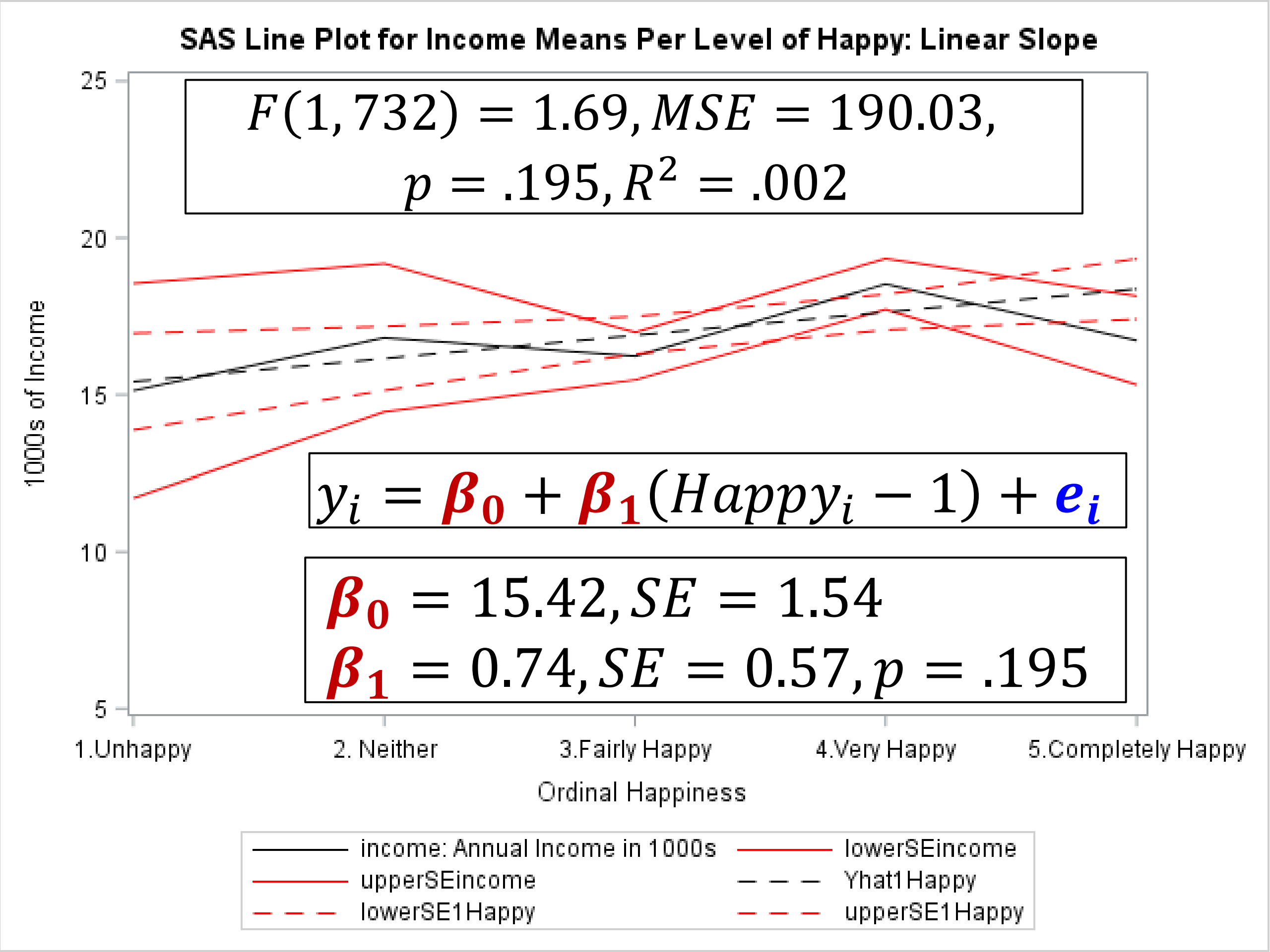
- The model (dashed lines) appears to capture the mean trend (solid lines) pretty well until 12 years of education…

- I think we need even more piecewise slopes after ed=12!
  - From 12 to 15
  - From 15 to 17-18
  - From 17-18 to 20



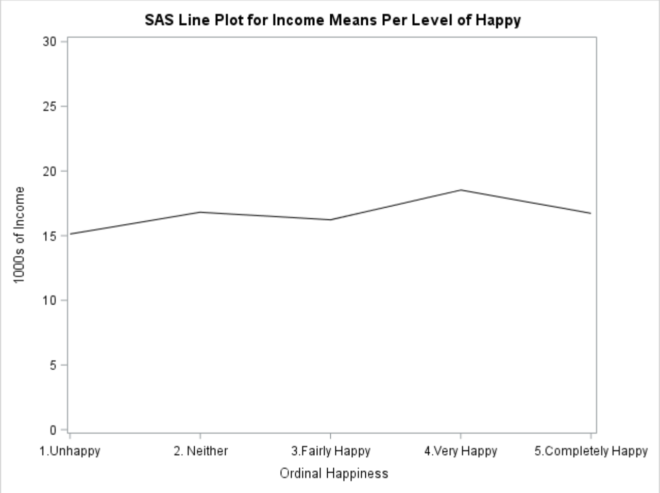
A Linear Trend for an Ordinal Predictor???

- 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 the outcome differences between adjacent predictor values
  - Here: need 4 slopes, 1 for each transition between categories
  - Is an alternative coding scheme that treats the predictor as "categorical" → 5 fixed effects used to distinguish each of 5 categories



Logic: Adjacent Slopes for an Ordinal Predictor

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



- Happy = 1 is where all slopes are 0, so it is the reference category (→ model intercept)
- The 4 slopes capture each <u>successive category difference</u> because each stays at **1** when done
  - ➢ In previous example (see right table), the LvsM slope went back to 0, so the second slope is NOT successive (i.e., it reflects LvsU, not MvsU)

Group	LvsM: Diff for Low vs Mid	LvsU: Diff for Low vs Upp
Low	0	0
Mid	1	0
Upp	0	1

SAS: Adjacent Slopes for Ordinal "Happy" Predictor

```
* Make 4 new variables for sections of happy;
DATA work.Example4; SET work.Example4;
h1v2=.; h2v3=.; h3v4=.; h4v5=.; * Make 4 new empty variables;
IF happy=1 THEN DO; h1v2=0; h2v3=0; h3v4=0; h4v5=0; END;
IF happy=2 THEN DO; h1v2=1; h2v3=0; h3v4=0; h4v5=0; END;
IF happy=3 THEN DO; h1v2=1; h2v3=1; h3v4=0; h4v5=0; END;
IF happy=4 THEN DO; h1v2=1; h2v3=1; h3v4=1; h4v5=0; END;
IF happy=5 THEN DO; h1v2=1; h2v3=1; h3v4=1; h4v5=1; END;
RUN;


TITLE "SAS GLM Predicting Income from Piecewise Happy";
PROC GLM DATA=work.Example4 NAMELEN=100;
MODEL income = h1v2 h2v3 h3v4 h4v5 / SOLUTION ALPHA=.05 CLPARM;
ESTIMATE "Diff in Slope 1-2 vs Slope 2-3" h1v2 -1 h2v3 1;
ESTIMATE "Diff in Slope 2-3 vs Slope 3-4" h2v3 -1 h3v4 1;
ESTIMATE "Diff in Slope 3-4 vs Slope 4-5" h3v4 -1 h4v5 1;
RUN; QUIT; TITLE;
```

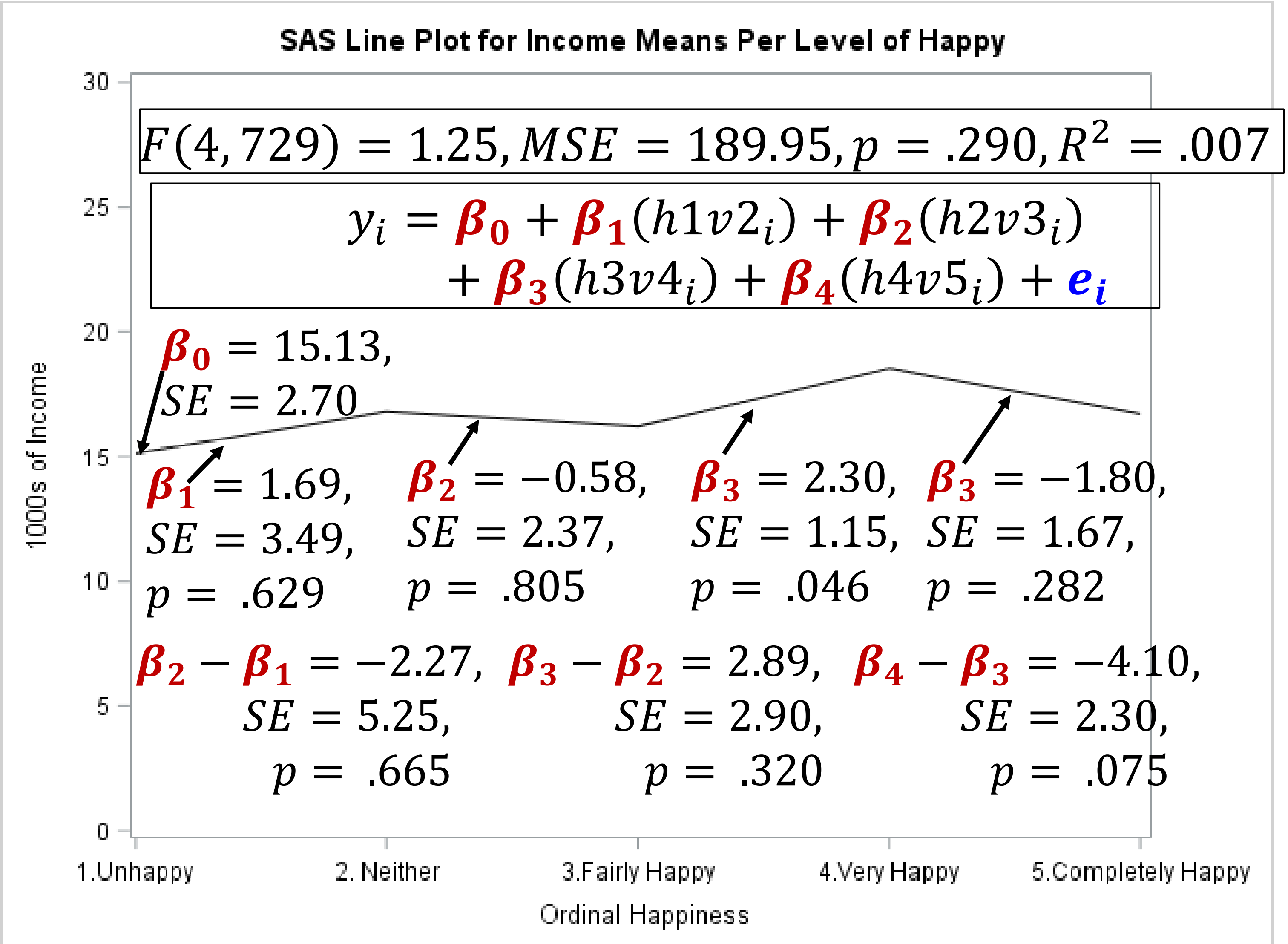
STATA: Adjacent Slopes for Ordinal "Happy" Predictor

```
// Make 4 new empty variables
gen h1v2=.
gen h2v3=.
gen h3v4=.
gen h4v5=.

// Replace with 0s
replace h1v2=0 if happy < 2
replace h2v3=0 if happy < 3
replace h3v4=0 if happy < 4
replace h4v5=0 if happy < 5
```

```
// Code continues
// Replace with 1s
replace h1v2=1 if happy >= 2
replace h2v3=1 if happy >= 3
replace h3v4=1 if happy >= 4
replace h4v5=1 if happy == 5
```

```
display "STATA GLM Predicting Income from Piecewise Happy"
regress income c.h1v2 c.h2v3 c.h3v4 c.h4v5, level(95)
lincom c.h1v2*-1 + c.h2v3*1 // Diff in Slope 1-2 vs Slope 2-3
lincom c.h2v3*-1 + c.h3v4*1 // Diff in Slope 2-3 vs Slope 3-4
lincom c.h3v4*-1 + c.h4v5*1 // Diff in Slope 3-4 vs Slope 4-5
```

Results from Testing Slope Differences



Summary: Predictors with Multiple Fixed Slopes

- There are many situations in which a single predictor x_i needs **multiple fixed slopes** to describe its prediction of outcome y_i:
  - Nominal predictor variables with $\boldsymbol{C}$ categories needs $\boldsymbol{C}-1$ fixed slopes to distinguish its $\boldsymbol{C}$ possible different outcome means
    - "Dummy coding": Means and mean differences related to the reference group (as the intercept) will be given by the model fixed effects directly, and others can be requested as linear combinations of fixed effects
    - Should report significance and effect size for all mean differences
  - 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 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-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 $\boldsymbol{y_i}$