# 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 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…
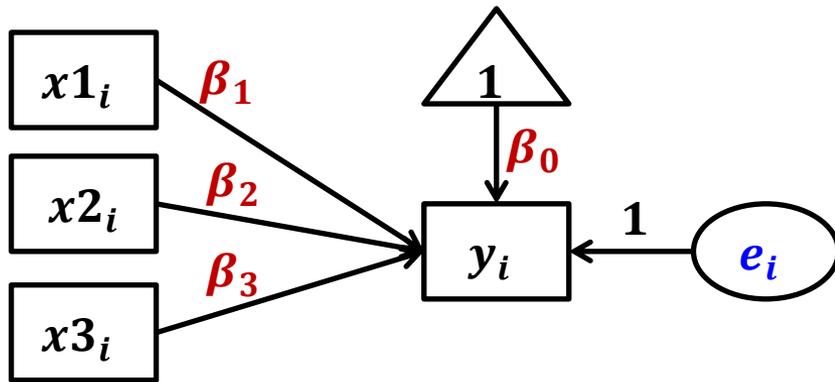


This figure is a **path diagram**. It illustrates a GLM with 3 $x_i$ predictors of 1 $y_i$ outcome. The "1" triangle is a constant used by the fixed intercept. Here is the equation the picture generates:

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

- **Synonyms for $y_i$ outcome:**
  dependent variable, criterion, thing-to-be explained/predicted/accounted for

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

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

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

# GLMs with Predictors: Binary vs. 3+ Categories

- To examine a **binary predictor** of a quantitative outcome, 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** $y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(Category_i) + \boldsymbol{e_i}$, **Predicted** $\hat{y}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(Category_i)$

  - Category 0 Mean: $\hat{y}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{0}) = \boldsymbol{\beta_0}$ ← 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: $\hat{y}_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{1}) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}$ ← linear combination of fixed effects

  - To get the estimate, SE, and *p*-value for any mean created from a <u>linear combination</u> of fixed effects, you need to ask for it via SAS ESTIMATE, STATA LINCOM, or R GLHT

  - 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 = $C$**

  - If $C = 3$, then we need the $\boldsymbol{\beta_0}$ intercept and 2 predictor slopes: $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$

  - If $C = 4$, 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 $C = 3$, \# 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

# "Indicator Coding" for a 3-Category Predictor

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

| 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:** $Income_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(LvM_i) + \boldsymbol{\beta_2}(LvU_i) + e_i$

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

- Model-implied means per category (group):

  - Lower Mean: $\hat{y}_L = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(0) + \boldsymbol{\beta_2}(0) = \boldsymbol{\beta_0}$ ← fixed effect #1

  - Middle Mean: $\hat{y}_M = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(1) + \boldsymbol{\beta_2}(0) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}$ ← found as linear combination

  - Upper Mean: $\hat{y}_U = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(0) + \boldsymbol{\beta_2}(1) = \boldsymbol{\beta_0} + \boldsymbol{\beta_2}$ ← found as linear combination

- Model-implied differences between each pair of categories (groups):

  - Lower vs Middle: $(\boldsymbol{\beta_0} + \boldsymbol{\beta_1}) - (\boldsymbol{\beta_0}) = \boldsymbol{\beta_1}$ ← fixed effect #2

  - Lower vs. Upper: $(\boldsymbol{\beta_0} + \boldsymbol{\beta_2}) - (\boldsymbol{\beta_0}) = \boldsymbol{\beta_2}$ ← fixed effect #3

  - Middle vs Upper: $(\boldsymbol{\beta_0} + \boldsymbol{\beta_2}) - (\boldsymbol{\beta_0} + \boldsymbol{\beta_1}) = \boldsymbol{\beta_2} - \boldsymbol{\beta_1}$ ← 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.

# GLM 3-Category Predictor: GSS Results

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

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

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

**Predictor Model:** $y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{LvM_i}) + \boldsymbol{\beta_2}(\boldsymbol{LvU_i}) + \boldsymbol{e_i}$

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

# GLM 3-Category Predictor: GSS 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 |
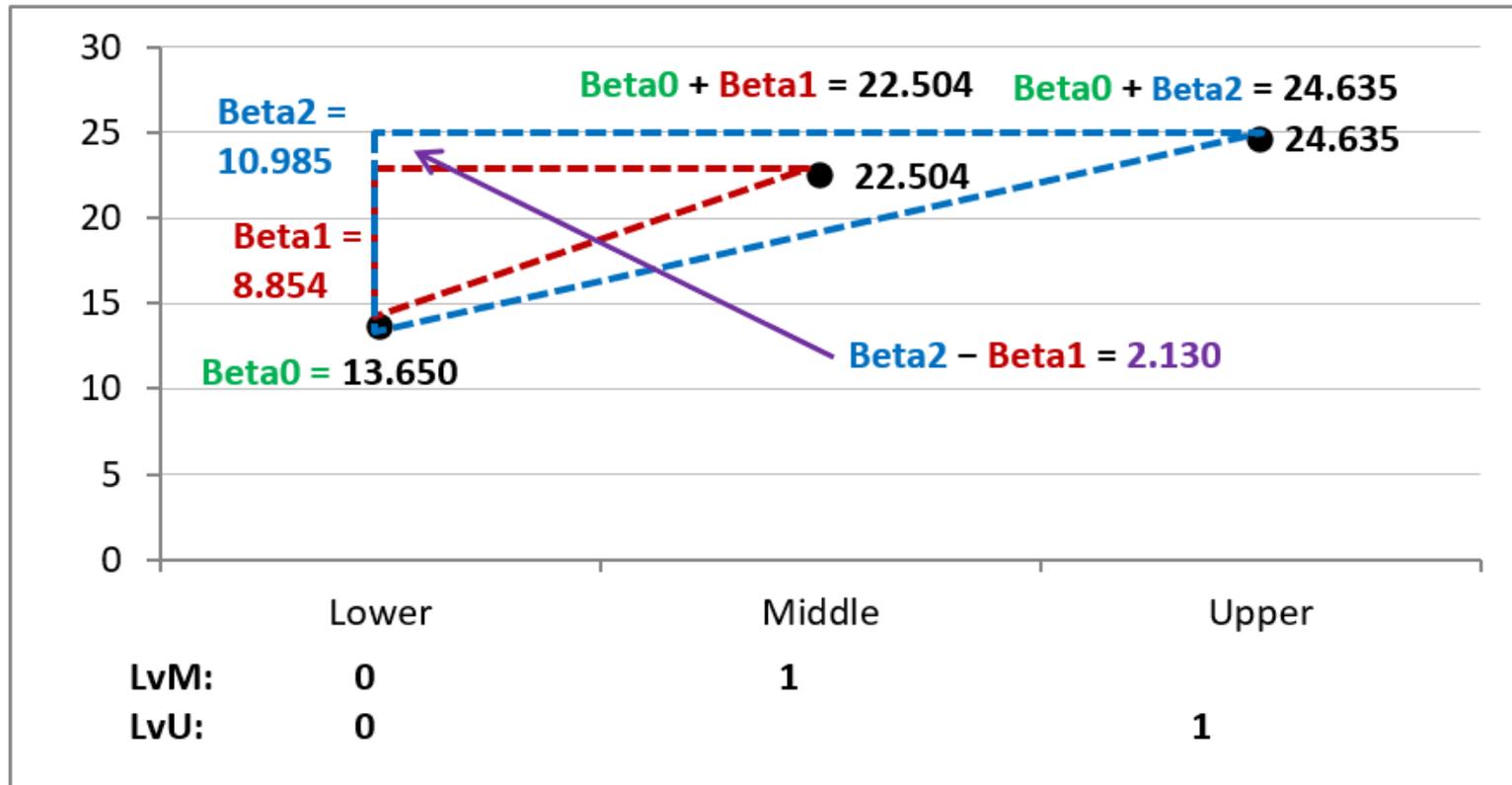
# Example of a 4-Category Predictor

Comparing outcome means across **4 groups** requires creating **3 new binary predictors** to be included **simultaneously** along with the intercept—for example, using "**indicator dummy-coded**" 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: $y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(d1_i) + \boldsymbol{\beta_2}(d2_i) + \boldsymbol{\beta_3}(d3_i) + \boldsymbol{e_i}$
  - The model gives us **the predicted outcome mean for each category** as follows:

| Control (Ref) Mean | Treatment 1 Mean | Treatment 2 Mean | Treatment 3 Mean |
|---|---|---|---|
| $\boldsymbol{\beta_0}$ | $\boldsymbol{\beta_0} + \boldsymbol{\beta_1}(d1_i)$ | $\boldsymbol{\beta_0} + \boldsymbol{\beta_2}(d2_i)$ | $\boldsymbol{\beta_0} + \boldsymbol{\beta_3}(d3_i)$ |

  - Model directly provides 3 mean differences (control vs. each treatment), and indirectly provides another 3 mean differences (differences between treatments) as **linear combinations of the fixed effects**... let's see how this works

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** |
|:---:|:---:|:---:|:---:|
| $\boldsymbol{\beta_0}$ | $\boldsymbol{\beta_0 + \beta_1(d1_i)}$ | $\boldsymbol{\beta_0 + \beta_2(d2_i)}$ | $\boldsymbol{\beta_0 + \beta_3(d3_i)}$ |

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

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

|  | Alt Group | Ref Group | Difference |
|---|---|---|---|
| • C vs. T1 = | $(\beta_0 + \boldsymbol{\beta_1})$ − | $(\beta_0)$ | $= \boldsymbol{\beta_1} = 2$ |
| • C vs. T2 = | $(\beta_0 + \boldsymbol{\beta_2})$ − | $(\beta_0)$ | $= \boldsymbol{\beta_2} = 5$ |
| • C vs. T3 = | $(\beta_0 + \boldsymbol{\beta_3})$ − | $(\beta_0)$ | $= \boldsymbol{\beta_3} = 9$ |
| • T1 vs. T2 = | $(\beta_0 + \boldsymbol{\beta_2})$ − | $(\beta_0 + \boldsymbol{\beta_1})$ | $= \boldsymbol{\beta_2} - \boldsymbol{\beta_1} = 5 - 2 = 3$ |
| • T1 vs. T3 = | $(\beta_0 + \boldsymbol{\beta_3})$ − | $(\beta_0 + \boldsymbol{\beta_1})$ | $= \boldsymbol{\beta_3} - \boldsymbol{\beta_1} = 9 - 2 = 7$ |
| • T2 vs. T3 = | $(\beta_0 + \boldsymbol{\beta_3})$ − | $(\beta_0 + \boldsymbol{\beta_2})$ | $= \boldsymbol{\beta_3} - \boldsymbol{\beta_2} = 9 - 5 = 4$ |

# Back to the 3-Category Predictor GLM

- 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: $y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(LvM_i) + \boldsymbol{\beta_2}(LvU_i) + \boldsymbol{e_i}$,

    - Is $\boldsymbol{\beta_1} \neq 0$? If so, then $\widehat{y}_M \neq \widehat{y}_L$ (given directly because of our coding)
    - Is $\boldsymbol{\beta_2} \neq 0$? If so, then $\widehat{y}_U \neq \widehat{y}_L$ (given directly because of our coding)
    - Is $(\boldsymbol{\beta_2} - \boldsymbol{\beta_1}) \neq 0$? If so, then $\widehat{y}_U \neq \widehat{y}_M$ (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** $\widehat{y}_i$ provided by this model with two grouping 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 <u>relative to</u> how many fixed effects we needed to generate predicted $\widehat{\boldsymbol{y}}_{\boldsymbol{i}}$ and how good they did (i.e., relative to what is left unknown)

  - This is an example of a "**multivariate Wald test**" (stay tuned for others)

  - "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**": $SS_{model} = \sum_{i=1}^{N}(\widehat{y}_i - \overline{y})^2$

- To quantify the **relative** size of that predicted info, we need to adjust it for $DF_{numerator}$ = number of fixed <u>slopes</u> = $k - 1$

  ➢ Then get "**Model Mean Square**" = $MS_{model} = \frac{SS_{model}}{k-1}$  | $-1$ because intercept doesn't get counted |

  ➢ $MS_{model}$ = "how much information has been captured per point spent"

- How much information is **leftover** is quantified by "**residual** (or **error**) **sums of squares**": $SS_{residual} = \sum_{i=1}^{N}(y_i - \widehat{y}_i)^2$

- 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



F(4,12)

5% of the distribution is greater than 3.26

F(10,100)

5% of the distribution is greater than 1.93

**The F Distribution**

df=(8,40)

df=(8,4)

*In the expression df(i,j), the index i is the df of the the numerator, and j is the df of the denominator.*

$F_{1,5}$

$F_{100,100}$

$F_{5,10}$
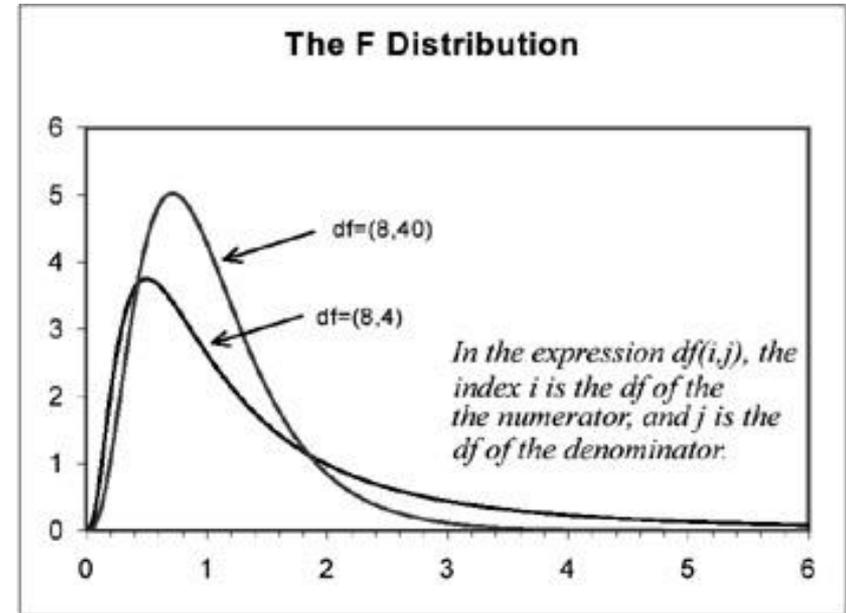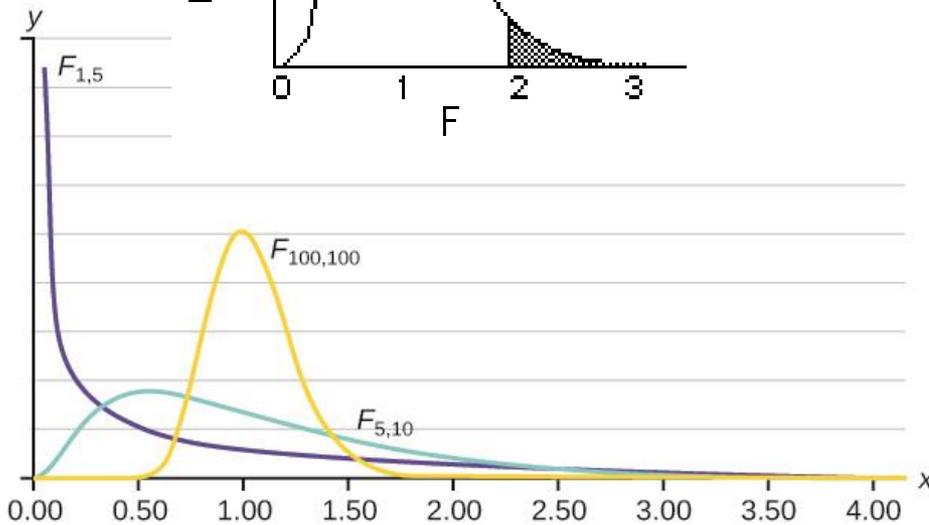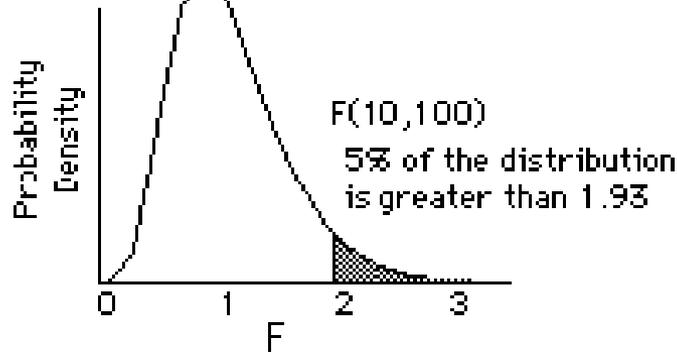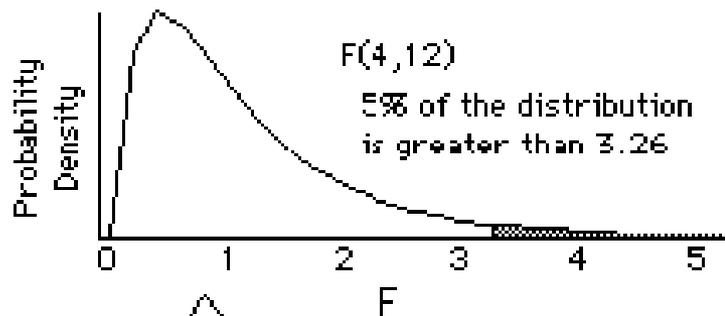
- The $F$ test-statistic ($F$-value) is a ratio (in a **squared** metric) of "info explained over info unknown", so ***F*-values must be positive**

- Its shape (and thus the critical value for the boundary of where "expected" starts) varies by $DF_{num}$ (like $\chi^2$) and by $DF_{den}$ (like $t$, which is flatter for smaller $N - $ k)

# Summary: Steps in Significance Testing

- **Choose critical region: % alpha ("unexpected") and possible directions**

  - Both directions or just one?

  - Alpha ($\alpha$) (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 (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 **<u>only 1 predictor</u>**, we don't need a separate $F$ test-statistic of the model $R^2$ significance; for example: $\boldsymbol{y_i = \beta_0 + \beta_1(x_i) + e_i}$

  - 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 $|t_{\beta_1}| > t_{critical}$, 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 $\boldsymbol{\widehat{y}_i}$ and actual $\boldsymbol{y_i}$

    - So model $\boldsymbol{R^2 = \dfrac{SS_{total} - SS_{residual}}{SS_{total}}}$ is the same as **(Pearson's $r$)$^2$**

- With **<u>2+ fixed slopes</u>**, we DO need to examine model $\boldsymbol{F}$ test-statistic and $\boldsymbol{R^2}$, for example: $\boldsymbol{y_i = \beta_0 + \beta_1(LvsM_i) + \beta_2(LvsU_i) + e_i}$

  - $\boldsymbol{F}$ test-statistic: Is the $\boldsymbol{\widehat{y}_i}$ 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 in GSS data: $y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(LvsM_i) + \boldsymbol{\beta_2}(LvsU_i) + \boldsymbol{e_i}$
- 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** $F(DF_{num}, DF_{den}) = Fvalue, \ \text{MSE} = MS_{res}, \ p < pvalue$

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

# Another version of R²:"Adjusted R²"

- 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 $R^2$**

  ➢ $R^2 = \frac{SS_{total} - SS_{residual}}{SS_{total}}$ → Must be positive if computed this way

  ➢ $\boldsymbol{R^2_{adj}} = 1 - \frac{(1 - R^2)(N-1)}{N-k-1} = 1 - \frac{MS_{residual}}{MS_{total}}$ $\boxed{\text{→ Change in residual variance relative to empty model}}$

    ▪ $R^2_{adj}$ 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: $R^2_{adj} = 1 - \frac{(1 - .103)(734 - 1)}{734 - 3} = .101 \ (R^2_{unadj} = .103)$

  ➢ Btw, we need to use SAS PROC REG instead of SAS PROC GLM to get $R^2_{adj}$ (both $R^2$ versions are given by STATA REGRESS and R LM)

# Effect Size per Fixed Slope

- The **model $R^2$ value** (the square of the correlation between predicted $\widehat{y}_i$ and actual $y_i$) provides a **general effect size**, but you may also want an **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 same as Pearson's 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…

- 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 (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 $t$-value? Effect sizes for fixed effect linear combinations, too

  - **Partial correlation $r$** (range is $\pm 1$): partial $r = \dfrac{t}{\sqrt{t^2 + DF_{den}}}$

    - Useful for quantitative predictors to convey strength of unique association for that slope
    - Can also get partial $r$ from SAS PROC CORR, STATA PCORR, and pcor.test in R package ppcor

  - **(Partial) Cohen's $d$** (range is $\pm\infty$): $d = \dfrac{2t}{\sqrt{DF_{den}}}$

    - Conveys difference between two groups in standard deviation units
    - (Partial) is not used in describing Cohen's $d$, because there is not another kind possible (i.e., as in "semi-partial" $r$, stay tuned)

> **From $r$ to $d$:**
>
> $$d \approx \frac{2r}{\sqrt{1 + r^2}}$$
>
> $$r \approx \frac{d}{\sqrt{4 + d^2}}$$

# Effect Sizes for Our GSS Results and Sample Sizes Needed for Power = .80

- **LvM Diff as $\boldsymbol{\beta_1}$**: $Est = 8.85, SE = 1.00, t(731) = 8.82, p < .001$

  ➢ $r = \dfrac{8.82}{\sqrt{8.82^2 + 731}} = 0.31, d = \dfrac{2*8.82}{\sqrt{731}} = 0.65 \rightarrow$ ~per-group $\boldsymbol{n > 45}$

- **LvU Diff as $\boldsymbol{\beta_2}$**: $Est = 10.98, SE = 2.99, t(731) = 3.67, p < .001$

  ➢ $r = \dfrac{3.67}{\sqrt{3.67^2 + 731}} = 0.13, d = \dfrac{2*3.67}{\sqrt{731}} = 0.27 \rightarrow$ ~ per-group $\boldsymbol{n > 175}$

- **MvU Diff as**: $\boldsymbol{\beta_2 - \beta_1}$: $Est = 2.13, SE = 3.03, t(731) = 0.70, p = .482$

  - $r = \dfrac{0.70}{\sqrt{0.70^2 + 731}} = 0.03, d = \dfrac{2*0.70}{\sqrt{731}} = 0.05 \rightarrow$ ~ per-group $\boldsymbol{n > 2,102}$

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

# Intermediate Summary

- For GLMs with **one fixed slope**, 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 $\hat{\boldsymbol{y}}_i$ only uses $\boldsymbol{\beta}_{unstd}$

- For GLMs with **2+ fixed slopes**, 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 $\chi^2$) → "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 $\hat{\boldsymbol{y}}_i$ and actual $\boldsymbol{y}_i$, which is necessary whenever the predicted $\hat{\boldsymbol{y}}_i$ 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 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 $x_i$: $y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(x_i) + \boldsymbol{\beta_2}(x_i)^2 + \boldsymbol{e_i}$

  - **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 $x_i$: $y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(\boldsymbol{Log}[x_i]) + \boldsymbol{e_i}$

# 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 $x_i$ 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 for $y_i$ with respect to $x_i$:

> **Intercept (position) at $x_i = x$:** $y_x = 50 + 4(x_i) + 0.3(x_i^2)$
>
> **First derivative (velocity) at $x$:** $\frac{dy_x}{d(x)} = 4 + (2 * 0.3)(x_i)$
>
> **Second derivative (acceleration) at $x$:** $\frac{d^2 y_x}{d(x)} = (2 * 0.3)$

# Quadratic Trends: Example of $x_i$ = Time
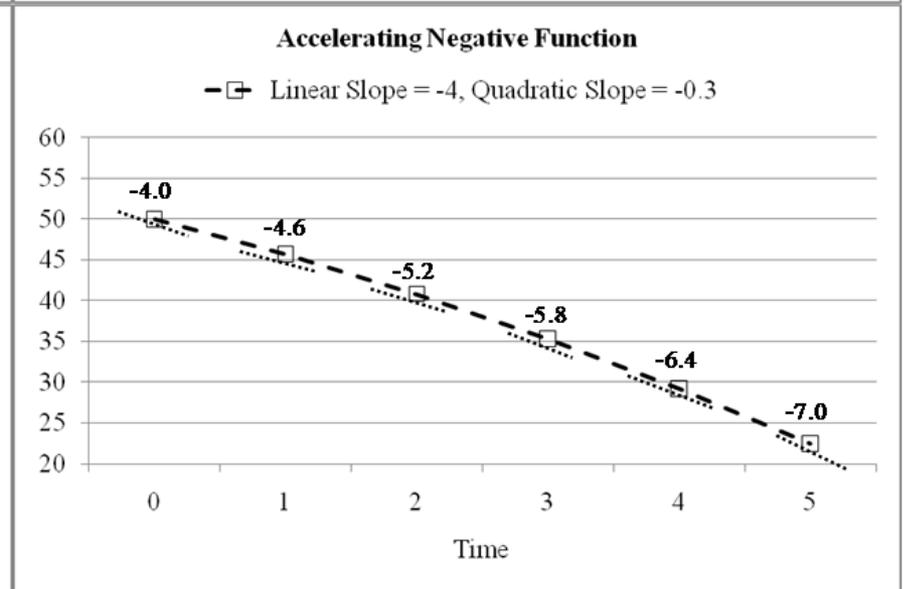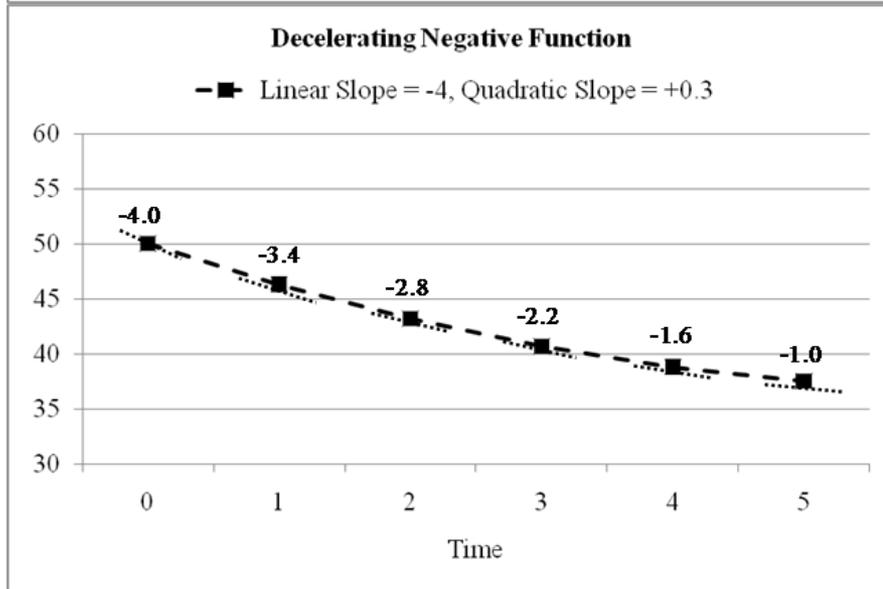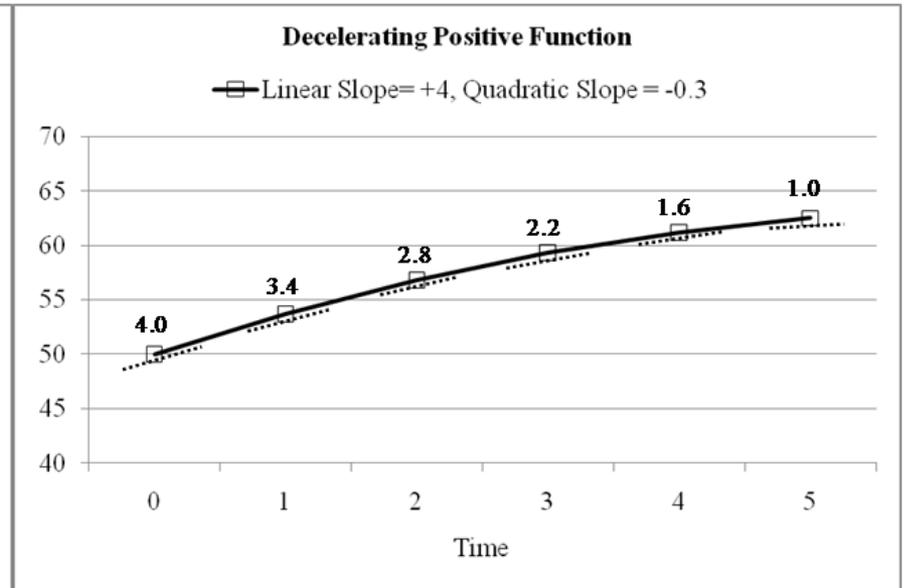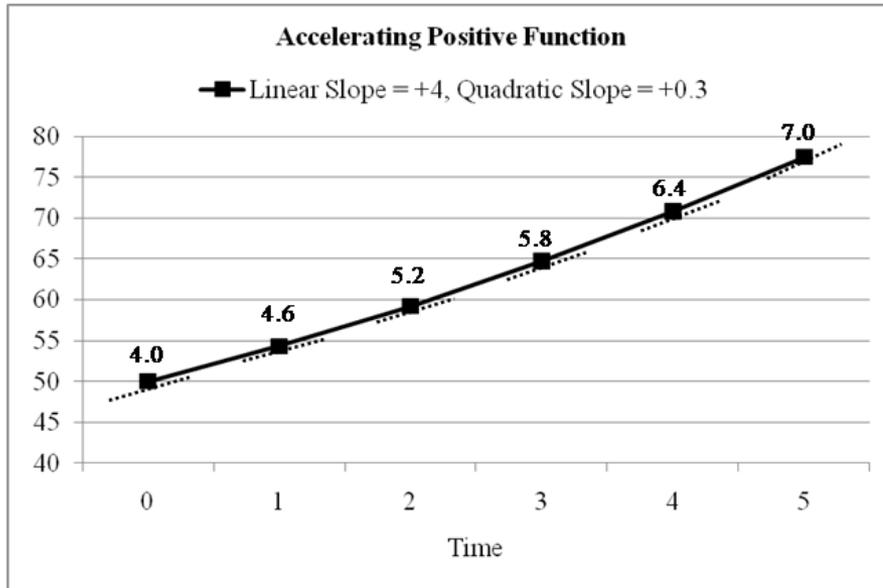
# Quadratic Trend for Age: GSS Example

- 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 annual income (yay middle age!)



**SAS Line Plot for Income Means Per Year of Age**
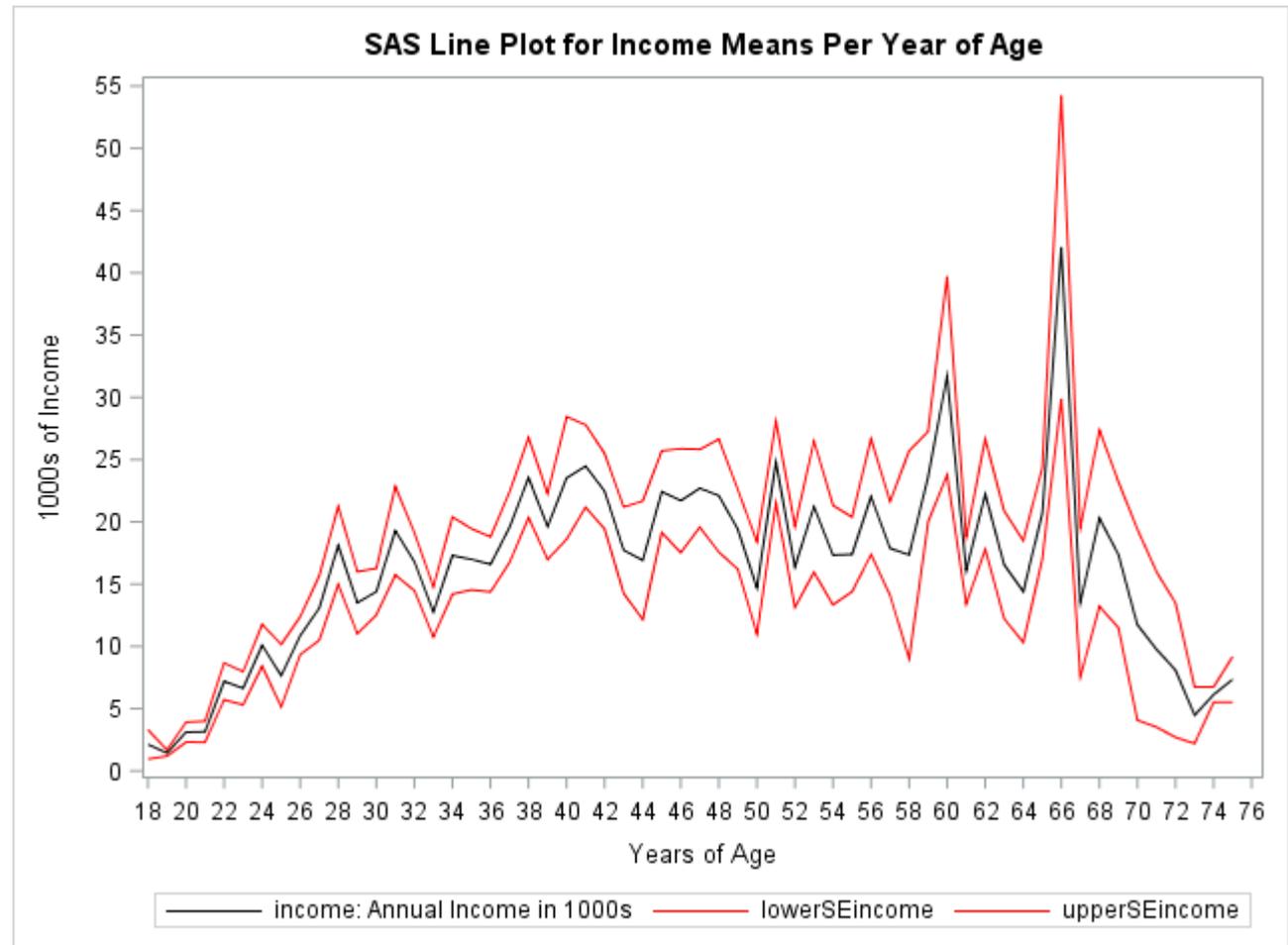
- Let's see what happens when we fit **a quadratic effect of age** (centered at 18, the minimum age) predicting annual income…

# Quadratic Trend for Age: GSS Example

- $Income_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(Age_i - \mathbf{18}) + \boldsymbol{\beta_2}(Age_i - \mathbf{18})^{\mathbf{2}} + \boldsymbol{e_i}$

  - **Intercept:** $\boldsymbol{\beta_0}$ = expected income at age 18 → $Est = 2.677, SE = 1.584, p < .001$

  - **Linear Age Slope:** $\boldsymbol{\beta_1}$ = 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:** $\boldsymbol{\beta_2}$ = <u>half</u> the rate of acceleration (or deceleration here) per year of age **at any age** → $Est = -0.020, SE = 0.003, p < .001$

- <u>Predicted income</u> at other ages via linear combinations of fixed effects:

  - Age 30: $\hat{y}_{x=30} = \mathbf{2.677} + \mathbf{1.223}(\mathbf{12}) - \mathbf{0.020}(\mathbf{12})^{\mathbf{2}} = 14.540, SE = 0.647$

  - Age 50: $\hat{y}_{x=50} = \mathbf{2.677} + \mathbf{1.223}(\mathbf{32}) - \mathbf{0.020}(\mathbf{32})^{\mathbf{2}} = 21.809, SE = 0.668$

  - Age 70: $\hat{y}_{x=70} = \mathbf{2.677} + \mathbf{1.223}(\mathbf{52}) - \mathbf{0.020}(\mathbf{52})^{\mathbf{2}} = 13.448, SE = 1.659$

- <u>Predicted linear age slope</u> at other ages via linear combinations:

  - Age 30: $\widehat{\boldsymbol{\beta_1}}_{x=30} = \mathbf{1.223} - \mathbf{0.020}(\mathbf{2}*\mathbf{12}) = 0.754, SE = 0.079$

  - Age 50: $\widehat{\boldsymbol{\beta_1}}_{x=50} = \mathbf{1.223} - \mathbf{0.020}(\mathbf{2}*\mathbf{32}) = -0.027, SE = 0.047$

  - Age 70: $\boldsymbol{\beta_1}_{x=70} = \mathbf{1.223} - \mathbf{0.020}(\mathbf{2}*\mathbf{52}) = -0.809, SE = 0.135$

- Predicted age at max income (where linear age slope = 0): $\frac{-\beta_1}{2*\beta_2} + 18 = 48.575$

# Quadratic Trend for Age: GSS Example
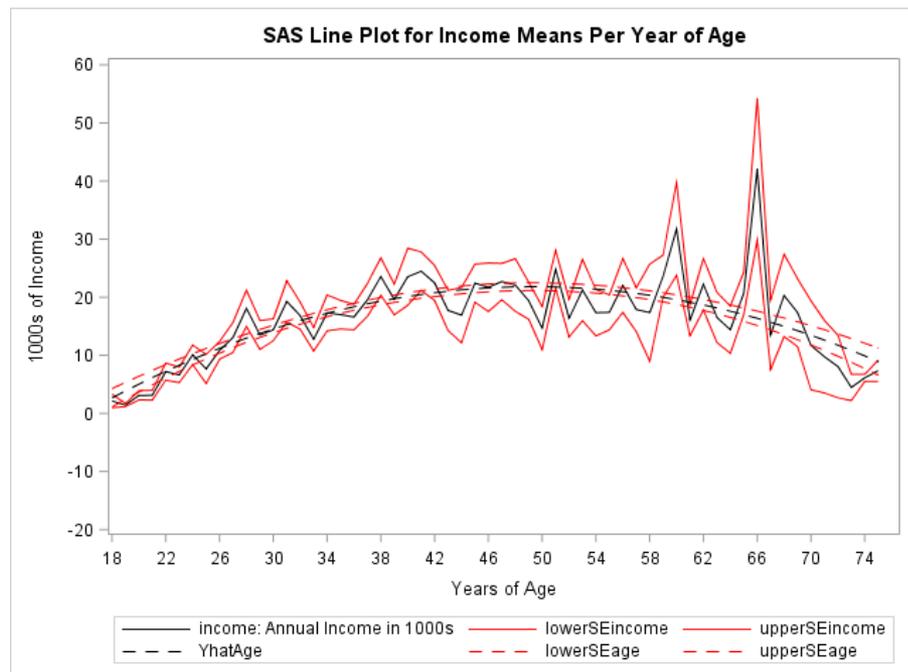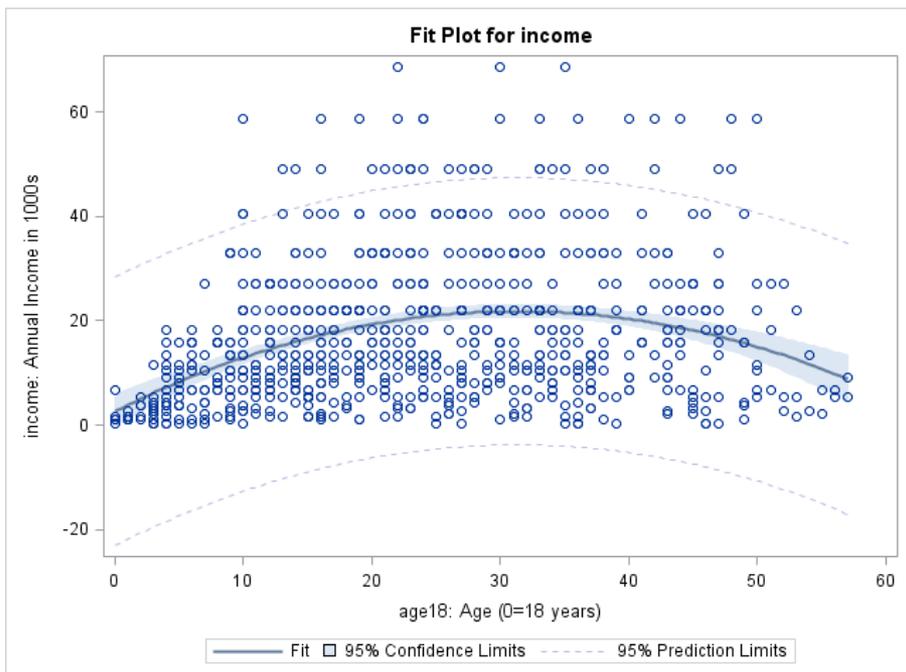


- Left: predicted regression line over individual scatterplot
  - From: $\mathbf{2.677 + 1.223(Age_i - 18) - 0.020(Age_i - 18)^2}$
- Right: predicted regression line over mean per age
  - $F(2, 731) = 47.00, MSE = 169.00, p < .001, R^2 = .114 \ (r = .338)$
    - Since age and age² work together, I'd use model $r$ as effect size

# 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
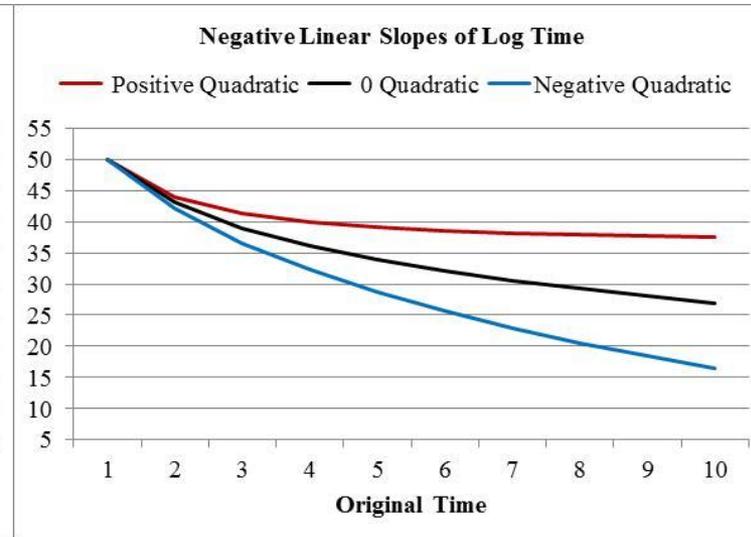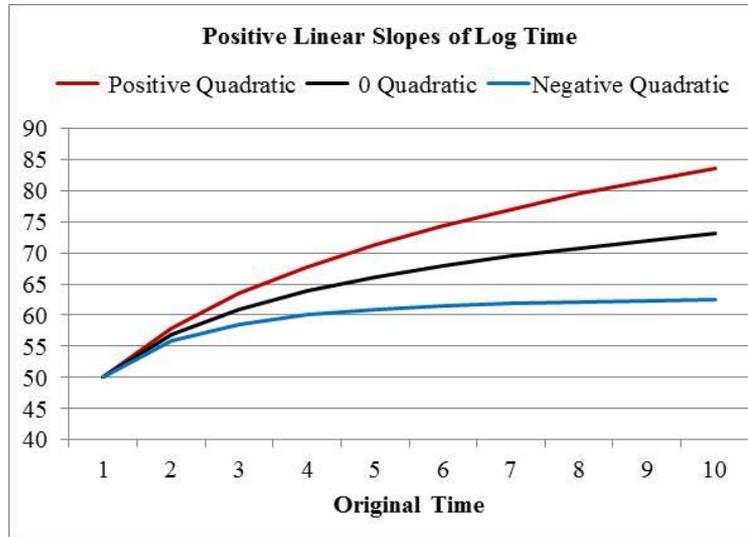


There is an **exponential** effect of **original $x_i$** on the outcome
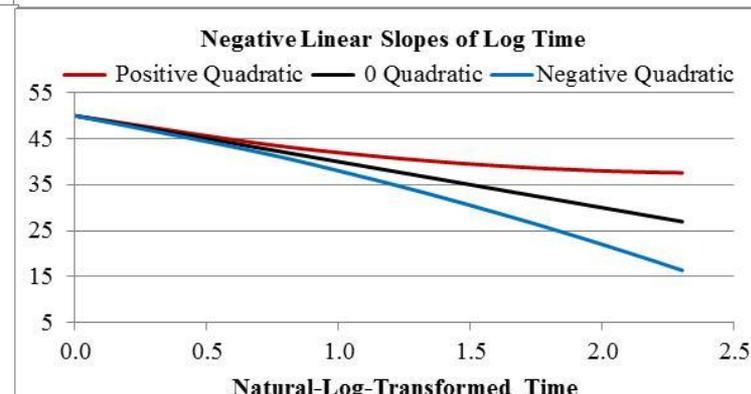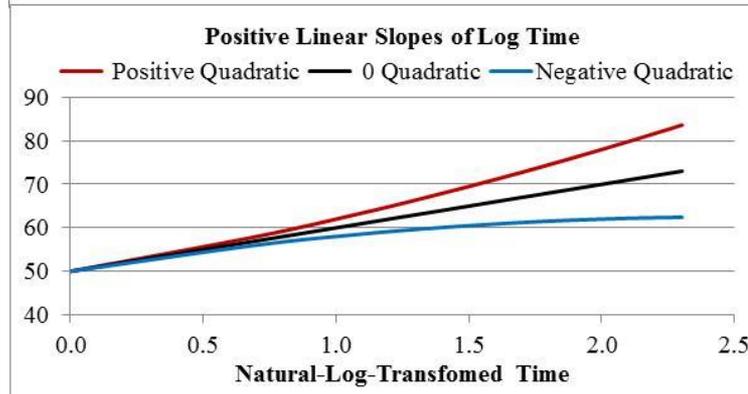
There is a **linear** effect of **log $x_i$** on the outcome

# Piecewise Slopes: GSS Example

- 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



SAS Line Plot for Income Means Per Year of Education
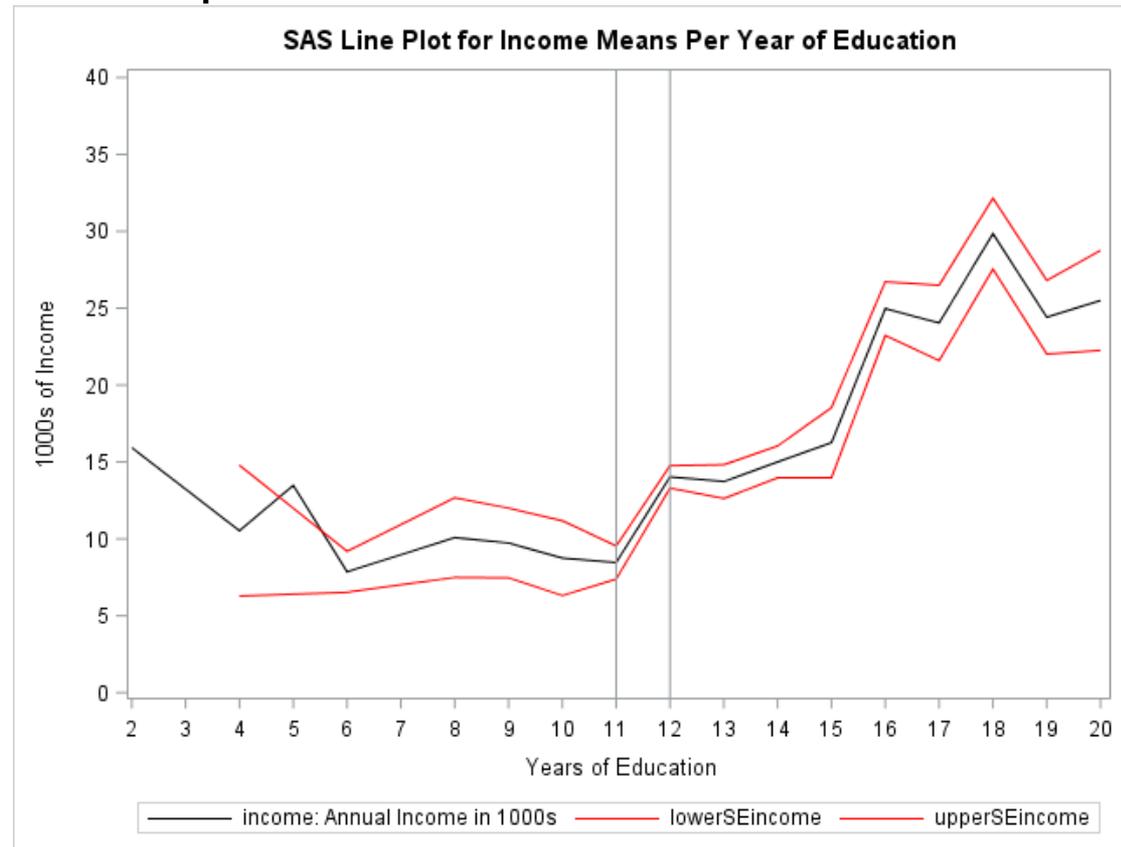
# Piecewise Slopes Coding: GSS Example

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



SAS Line Plot for Income Means Per Year of Education

gradHS

lessHS

overHS

1000s of Income

Years of Education

income: Annual Income in 1000s    lowerSEincome    upperSEincome
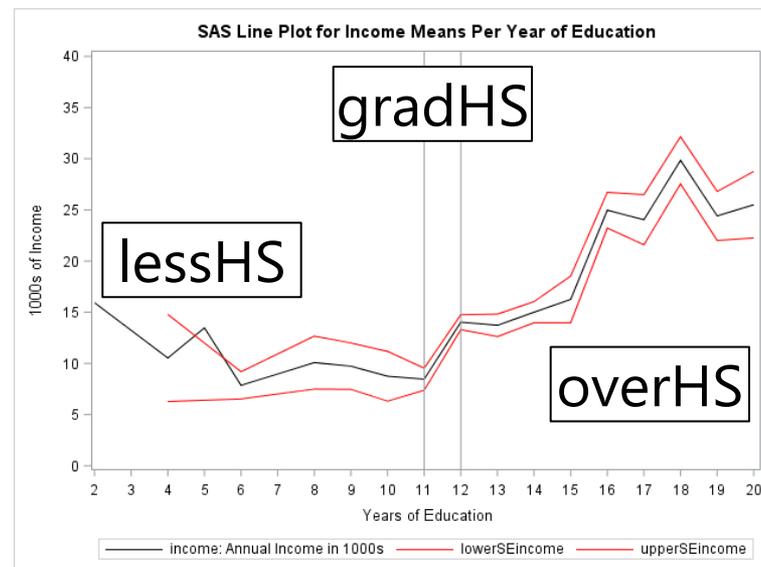
# Piecewise Slopes: GSS Results

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

  - $r = .404$ 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 <u>per year education</u> 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 <u>per year education</u> 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 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

**SAS Predicted Income By Year of Education: Linear OverHS**

$R^2 = .164$

$\beta_3 = 2.12$

$\beta_2 = 4.68$

$\beta_1 = -0.27$

1000s of Income

Years of Education

— income: Annual Income in 1000s  — lowerSEincome  — upperSEincome
- - - Yhat1Educ  - - - lowerSE1educ  - - - upperSE1educ
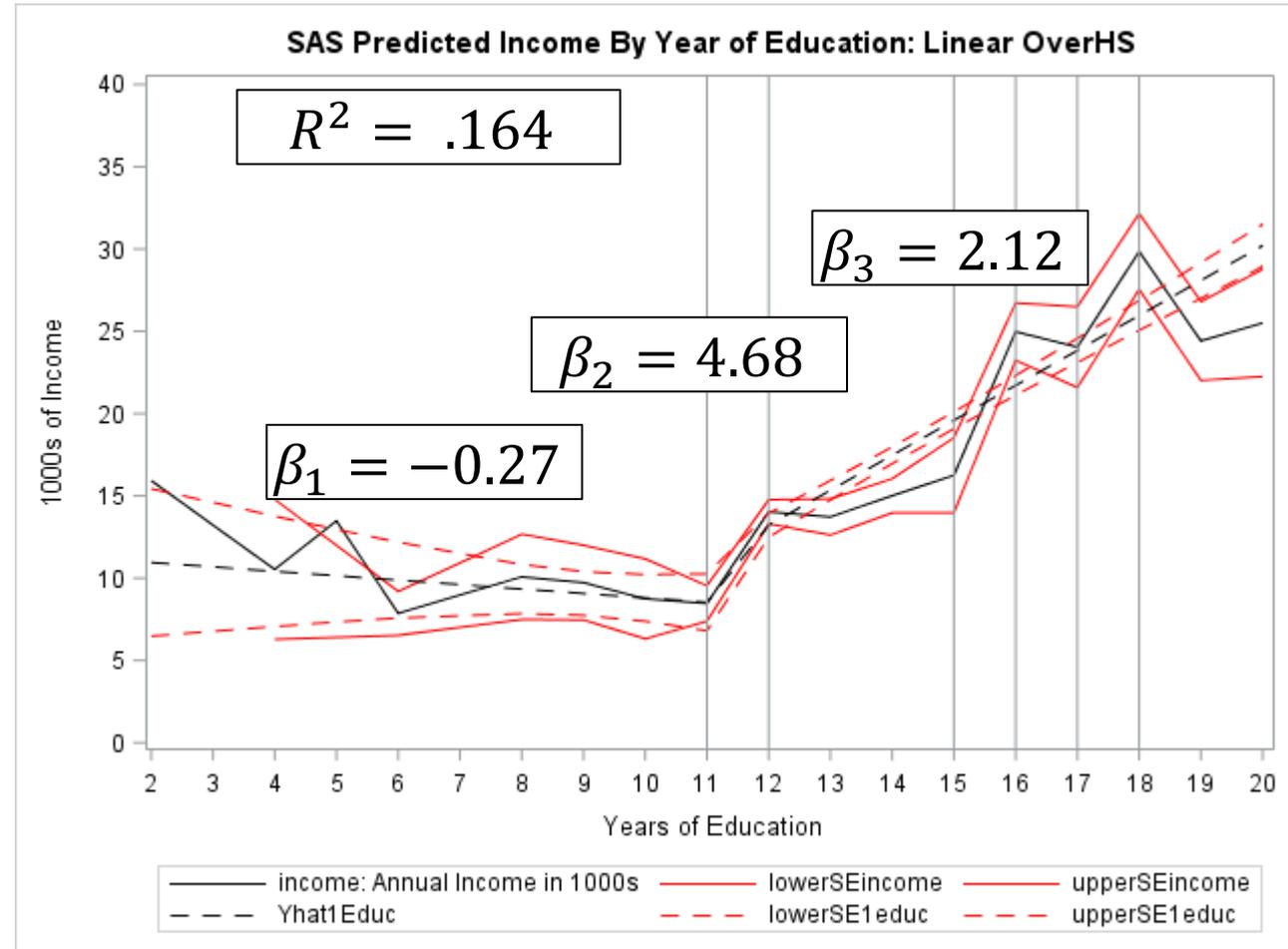
# A Linear Slope for an Ordinal Predictor???

- 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 the outcome differences between adjacent predictor values

  - Here: need 4 slopes, 1 for each transition between categories

  - Use "**sequential dummy coding**" to treat the predictor as "categorical" → 5 fixed effects used to distinguish each of 5 categories

### SAS Line Plot for Income Means Per Level of Happy: Linear Slope

$$F(1, 732) = 1.69, MSE = 190.03,$$
$$p = .195, R^2 = .002$$

$$y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(Happy_i - 1) + \boldsymbol{e_i}$$

$$\boldsymbol{\beta_0} = 15.42, SE = 1.54$$
$$\boldsymbol{\beta_1} = 0.74, SE = 0.57, p = .195$$

1000s of Income

1.Unhappy   2. Neither   3.Fairly Happy   4.Very Happy   5.Completely Happy

Ordinal Happiness

— income: Annual Income in 1000s
— upperSEincome
- - lowerSE1Happy
— lowerSEincome
– – Yhat1Happy
- - upperSE1Happy
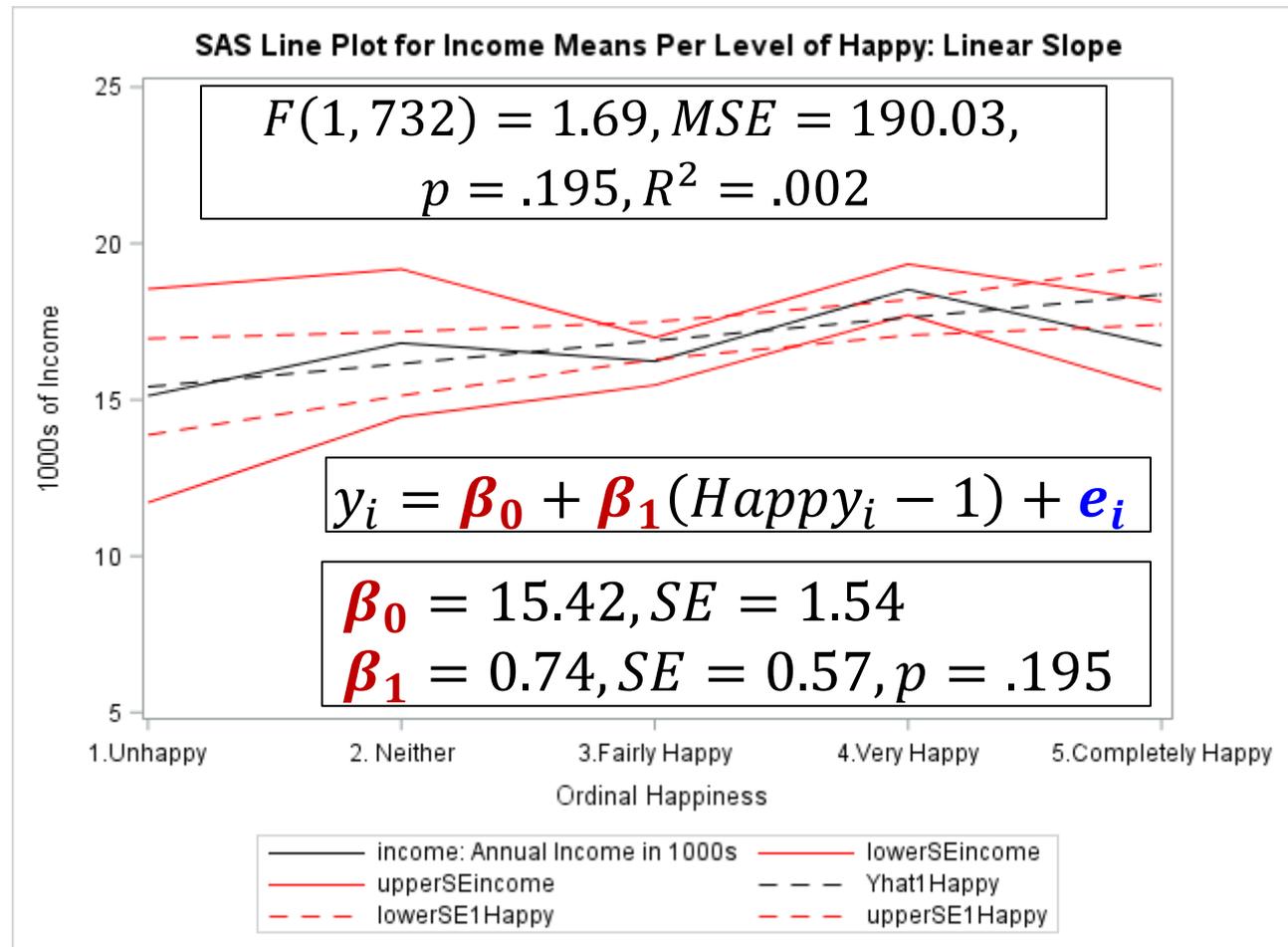
# Sequential Slopes for an Ordinal Predictor

| 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** |



SAS Line Plot for Income Means Per Level of Happy
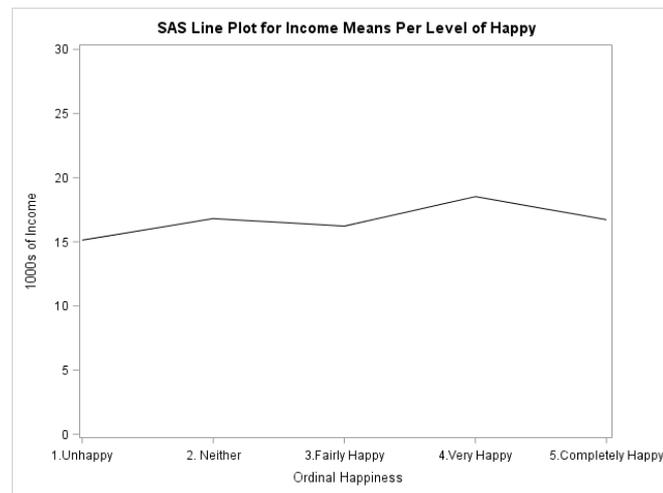
- Happy = 1 is where all slopes are 0, so it is the reference category (→ model intercept)

- The 4 slopes capture each <u>adjacent category difference</u> 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)

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

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

# GSS Results for Slopes and Slope Differences



**SAS Line Plot for Income Means Per Level of Happy**

$$F(4, 729) = 1.25, MSE = 189.95, p = .290, R^2 = .007$$

$$y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1}(h1v2_i) + \boldsymbol{\beta_2}(h2v3_i) + \boldsymbol{\beta_3}(h3v4_i) + \boldsymbol{\beta_4}(h4v5_i) + \boldsymbol{e_i}$$

$\boldsymbol{\beta_0} = 15.13,$
$SE = 2.70$

$\boldsymbol{\beta_1} = 1.69,$
$SE = 3.49,$
$p = .629$

$\boldsymbol{\beta_2} = -0.58,$
$SE = 2.37,$
$p = .805$

$\boldsymbol{\beta_3} = 2.30,$
$SE = 1.15,$
$p = .046$

$\boldsymbol{\beta_4} = -1.80,$
$SE = 1.67,$
$p = .282$

$\boldsymbol{\beta_2} - \boldsymbol{\beta_1} = -2.27,$
$SE = 5.25,$
$p = .665$

$\boldsymbol{\beta_3} - \boldsymbol{\beta_2} = 2.89,$
$SE = 2.90,$
$p = .320$

$\boldsymbol{\beta_4} - \boldsymbol{\beta_3} = -4.10,$
$SE = 2.30,$
$p = .075$

Y-axis: 1000s of Income

X-axis: Ordinal Happiness — 1.Unhappy, 2. Neither, 3.Fairly Happy, 4.Very Happy, 5.Completely Happy
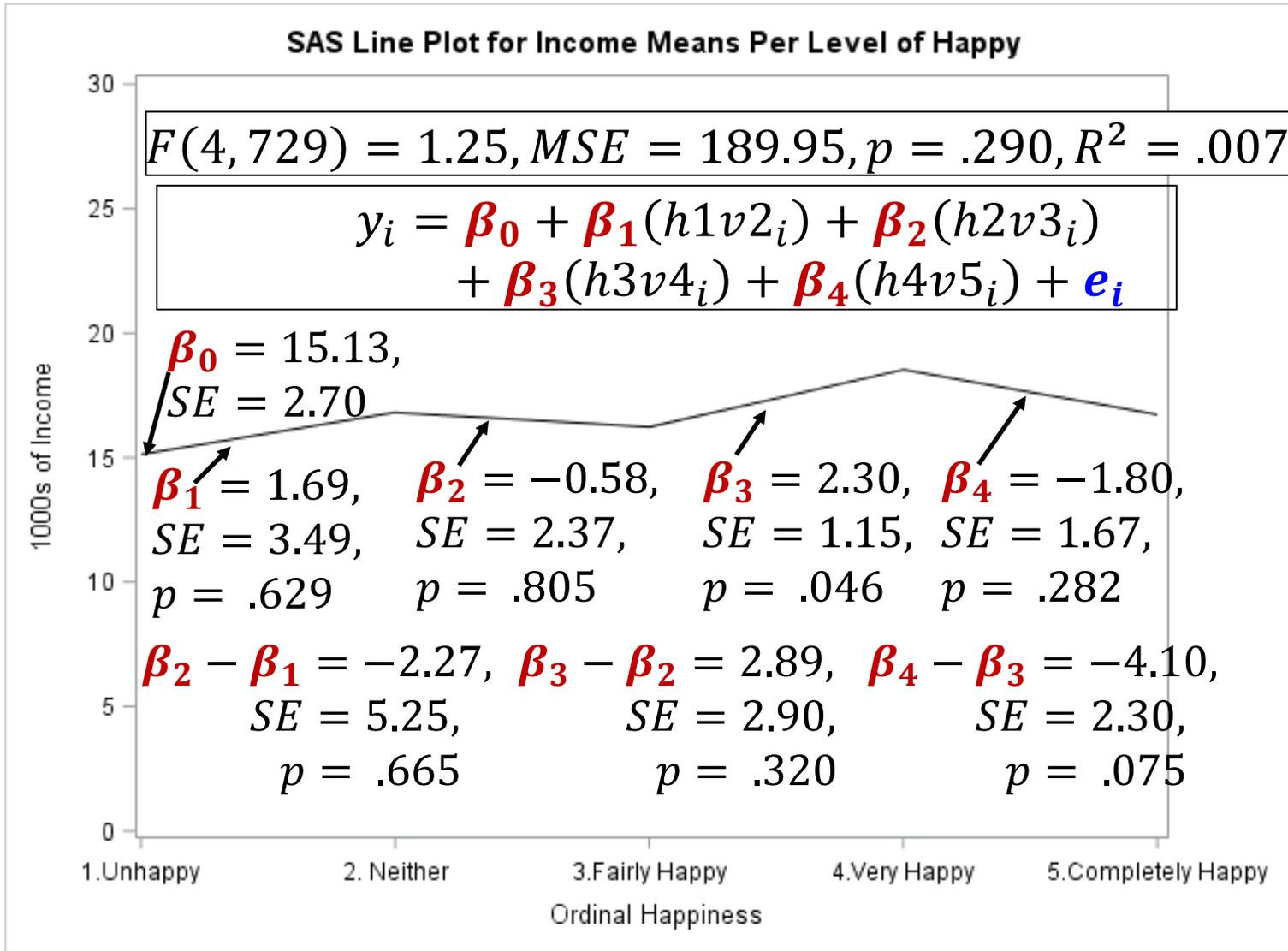
# Summary: Predictors with Multiple Fixed Slopes

- There are many scenarios in which a single predictor $x_i$ needs **multiple fixed slopes** to describe its prediction of outcome $y_i$:

  - <u>Predictor variables</u> with $C$ categories needs $C - 1$ fixed slopes to distinguish its $C$ possible different outcome means

    - "Indicator dummy coding" is useful for nominal or ordinal predictors
    - "Sequential dummy 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)

  - <u>Nonlinear effects of quantitative predictor variables</u> (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 $(Est - H_0)/SE$ via $t$ test-statistic) as well as significance of the **model $R^2$** (as multivariate Wald test via $F$ test-statistic)

  - Model $R^2$ = squared Pearson $r$ between predicted $\hat{y}_i$ and actual $y_i$