Example 3: General Linear Models with Multiple Fixed Effects of a Single Conceptual Predictor
(complete syntax, data, and output available for SAS, STATA, and R electronically)

The data for this example were selected from the 2012 General Social Survey dataset (and were also used for examples 1 and 2). The current example will use general linear models to predict a single quantitative outcome (annual income) when multiple fixed effects are needed to describe a predictor's relationship to the outcome: for categorical predictors with more than two categories (3-category working class), for quantitative predictors with nonlinear effects (quadratic years of age or piecewise years of education), or for testing the assumption of a single linear slope for ordinal predictors (5-category happiness).

SAS Syntax for Importing and Preparing Data for Analysis:

```
* Paste in the folder address where "GSS_Example.xlsx" is saved after = before ;
%LET filesave= \\Client\C:\Dropbox\22SP_PSQF6243\PSQF6243_Example3;

* IMPORT GSS_Example.xlsx data using filesave reference and exact file name;
* from the Excel workbook in DATAFILE= location from SHEET= ;
* New SAS file is in "work" library place with name "Example3";
* "GETNAMES" reads in the first row as variable names;
* DBMS=XLSX (can also use EXCEL or XLS for .xls files);
PROC IMPORT DATAFILE="&filesave.\GSS_Example.xlsx"
            OUT=work.Example3 DBMS=XLSX REPLACE;
     SHEET="GSS_Example";
     GETNAMES=YES;
RUN;
* All data transformations must happen inside a DATA+SET combo to know where to use them;
* Here is how to make a new variable: new = old;
DATA work.Example3; SET work.Example3;
* Label variables and apply value formats for variables used below;
* LABEL name=      "name: Descriptive Variable Label";
  LABEL workclass= "workclass: 1=Lower, 2=Middle, 3=Upper"
        age=       "age: Years of Age"
        educ=      "educ: Years of Education"
        happy=     "happy: 5-Category Happy Rating"
        income=    "income: Annual Income in 1000s";
* Select cases complete on all variables to be used;
  WHERE NMISS(income,workclass,age,educ,happy)=0;
RUN;
* Now dataset work.Example3 is ready to use;
```

Note: All SAS commands and comments end in a semi-colon!

STATA Syntax for Importing and Preparing Data for Analysis:

```
// Paste in the folder address where "GSS_Example.xlsx" is saved between " "
cd "\\Client\C:\Dropbox\22SP_PSQF6243\PSQF6243_Example3"

// IMPORT GSS_Example.xlsx data from working directory using exact file name
// To change all variable names to lowercase, remove "case(preserve")
clear // Clear before means close any open data
import excel "GSS_Example.xlsx", case(preserve) firstrow clear
// Clear after means re-import if it already exists (if need to start over)

// Label variables and apply value formats for variables used below
// label variable name   "name: Descriptive Variable Label"
label variable workclass "workclass: 1=Lower, 2=Middle, 3=Upper"
label variable age       "age: Years of Age"
label variable educ      "educ: Years of Education"
label variable happy     "happy: 5-Category Happy Rating"
label variable income    "income: Annual Income in 1000s"

// Select cases complete on variables to be used
egen nmiss = rowmiss(income workclass age educ happy)
drop if nmiss>0
// Now dataset is ready to use
```

R Syntax for Importing and Preparing Data for Analysis:

```
# Set working directory (to import and export files to)
# Paste in the folder address where "GSS_Example.xlsx" is saved in quotes
setwd("C:/Dropbox/22SP_PSQF6243/PSQF6243_Example3")

# Import GSS_Example.xlsx data from working directory -- path = file name
Example3 = read_excel(path="GSS_Example.xlsx", sheet="GSS_Example")
# Convert to data frame to use for analysis
Example3 = as.data.frame(Example3)

# Label variables used below (add descriptive titles) using comments instead
```

Syntax and SAS Output for Data Description:

```
TITLE "SAS Descriptive Statistics for Quantitative and Ordinal Variables";
PROC MEANS NDEC=3 DATA=work.Example3;
     VAR income age educ happy;
RUN; TITLE;

display "STATA Descriptive Statistics for Quantitative and Ordinal Variables"
summarize income age educ happy

print("R Descriptive Statistics for Quantitative or Ordinal Variables")
describe(x=Example3[ , c("income","age","educ","happy")])
```

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
income	income: Annual Income in 1000s	734	17.303	13.792	0.245	68.600
age	age: Years of Age	734	42.063	13.378	18.000	75.000
educ	educ: Years of Education	734	13.812	2.909	2.000	20.000
happy	happy: 5-Category Happy Rating	734	3.556	0.895	1.000	5.000

```
TITLE "SAS Descriptive Statistics for Categorical Variables";
PROC FREQ DATA=work.Example3;
     TABLE workclass happy;
RUN; TITLE;

display "SAS Descriptive Statistics for Categorical Variables"
tabulate workclass
tabulate happy

print("R Descriptive Statistics for Categorical Variables")
prop.table(table(x=Example3$workclass,useNA="ifany"))
prop.table(table(x=Example3$happy,useNA="ifany"))
```

workclass: 1=Lower, 2=Middle, 3=Upper				
workclass	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	436	59.40	436	59.40
2	278	37.87	714	97.28
3	20	2.72	734	100.00

We will need 2 slopes to represent the differences across the 3 categories.

happy: 5-Category Happy Rating				
happy	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	26	3.54	26	3.54
2	39	5.31	65	8.86
3	256	34.88	321	43.73
4	327	44.55	648	88.28
5	86	11.72	734	100.00

We will need 4 slopes to represent the differences across the 5 categories.

Syntax to Create Indicator-Dummy-Coded Predictors—2 needed for 3 workclass categories:

Categorical variables with 3+ categories cannot be included directly as predictors in the model, or else a single linear slope will be estimated to differentiate the total C categories—this doesn't make any sense, especially for nominal predictor variables. Instead, **we need to create $C-1$ new predictors to distinguish the predicted outcome for each of the C categories**. The coding scheme we are using is "indicator-dummy-coding" where each category has a 1 for only a single predictor (that "activates" the predictor for that category).

```
* SAS code to create 2 new indicator-dummy-coded binary predictors;
* DATA + SET means "save as itself" after adding new predictors;
DATA work.Example3; SET work.Example3;
  LvM=.; LvU=.; * Make 2 new empty variables;
  IF workclass=1 THEN DO; LvM=0; LvU=0; END; * Replace each for lower;
  IF workclass=2 THEN DO; LvM=1; LvU=0; END; * Replace each for middle;
  IF workclass=3 THEN DO; LvM=0; LvU=1; END; * Replace each for upper;
  LABEL LvM="LvM: Low=0 vs Mid=1 Class"
        LvU="LvU: Low=0 vs Upp=1 Class";
RUN;
```

```
// STATA code to create 2 new indicator-dummy-coded binary predictors
gen LvM=. // Make two new empty variables
gen LvU=.
replace LvM=0 if workclass==1 // Replace each for lower
replace LvU=0 if workclass==1
replace LvM=1 if workclass==2 // Replace each for middle
replace LvU=0 if workclass==2
replace LvM=0 if workclass==3 // Replace each for upper
replace LvU=1 if workclass==3
label variable LvM "LvM: Low=0 v Mid=1 Class"
label variable LvU "LvU: Low=0 v Upp=1 Class"
```

Group (N = 734)	LvM	LvU
1. Low (n = 436)	0	0
2. Mid (n = 278)	**1**	0
3. Upp (n = 20)	0	**1**

```
# R code to create indicator-dummy-coded binary predictors
Example3$LvM=NA; Example3$LvU=NA  # Make 2 new empty variables
Example3$LvM[which(Example3$workclass==1)]=0  # Replace each for lower
Example3$LvU[which(Example3$workclass==1)]=0
Example3$LvM[which(Example3$workclass==2)]=1  # Replace each for middle
Example3$LvU[which(Example3$workclass==2)]=0
Example3$LvM[which(Example3$workclass==3)]=0  # Replace each for upper
Example3$LvU[which(Example3$workclass==3)]=1
# LvM: Low=0 vs Mid=1 Class
# LvU: Low=0 vs Upp=1 Class
```

Syntax and SAS Output for 3-Category Working Class Predicting Income:

Model with workclass via two indicator-dummy-coded predictors:

$$Income_i = \beta_0 + \beta_1(LvM_i) + \beta_2(LvU_i) + e_i$$

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

The syntax below will also request the predicted outcome for each category and all possible pairwise differences.

Low Mean: $\hat{y}_L = \beta_0 + \beta_1(0) + \beta_2(0) = \beta_0$ ← fixed effect #1
Mid Mean: $\hat{y}_M = \beta_0 + \beta_1(1) + \beta_2(0) = \beta_0 + \beta_1$ ← linear combination
Upp Mean: $\hat{y}_U = \beta_0 + \beta_1(0) + \beta_2(1) = \beta_0 + \beta_2$ ← linear combination

Diff of Low vs Mid: $(\beta_0 + \beta_1) - (\beta_0) = \beta_1$ ← fixed effect #2
Diff of Low vs. Upp: $(\beta_0 + \beta_2) - (\beta_0) = \beta_2$ ← fixed effect #3
Diff of Mid vs Upp: $(\beta_0 + \beta_2) - (\beta_0 + \beta_1) = \beta_2 - \beta_1$ ← linear combination

```
TITLE "SAS GLM Predicting Income from 2 New Variables for workclass";
PROC GLM DATA=work.Example3 NAMELEN=100;
  MODEL income = LvM LvU / SOLUTION ALPHA=.05 CLPARM SS3;
* Ask for predicted income per group and group differences;
  ESTIMATE "Pred Income: Low"  intercept 1 LvM  0 LvU 0;
  ESTIMATE "Pred Income: Mid"  intercept 1 LvM  1 LvU 0;
  ESTIMATE "Pred Income: Upp"  intercept 1 LvM  0 LvU 1;
  ESTIMATE "Low vs Mid Diff"               LvM  1 LvU 0;
  ESTIMATE "Low vs Upp Diff"               LvM  0 LvU 1;
  ESTIMATE "Mid vs Upp Diff"               LvM -1 LvU 1;
* Save requested estimates as SAS dataset to do math on them;
  ODS OUTPUT Estimates=work.ClassEstimates;
RUN; QUIT; TITLE;
```

```
display "STATA GLM Predicting Income from 2 New Variables for workclass"
regress income c.LvM c.LvU, level(95)
// Ask for predicted income per group and group differences
   lincom _cons*1 + c.LvM*0  + c.LvU*0  // Pred Income: Low
   lincom _cons*1 + c.LvM*1  + c.LvU*0  // Pred Income: Mid
   lincom _cons*1 + c.LvM*0  + c.LvU*1  // Pred Income: Upp
   lincom           c.LvM*1  + c.LvU*0  // Low vs Mid Diff
   lincom           c.LvM*0  + c.LvU*1  // Low vs Upp Diff
   lincom           c.LvM*-1 + c.LvU*1  // Mid vs Upp Diff
```

```
print("R GLM Predicting Income from 2 New Variables for workclass")
ModelClass = lm(data=Example3, formula=income~1+LvM+LvU)
anova(ModelClass)    # anova to print residual variance
summary(ModelClass) # summary to print fixed effects solution
confint(ModelClass, level=.95) # confint for level% CI for fixed effects
```

```
print("R Ask for predicted income per group and group differences")
print("In number lists below, values are multiplier for each fixed effect IN ORDER")
PredClass = glht(model=ModelClass, linfct=rbind(
  "Pred Income: Low"  = c(1, 0, 0),
  "Pred Income: Mid"  = c(1, 1, 0),
  "Pred Income: Upp"  = c(1, 0, 1),
  "Low vs Mid Diff"   = c(0, 1, 0),
  "Low vs Upp Diff"   = c(0, 0, 1),
  "Mid vs Upp Diff"   = c(0,-1, 1)))
print("Save glht linear combination results with unadjusted p-values and 95% CIs")
SavePredClass = summary(PredClass, test=adjusted("none"))
print(SavePredClass); confint(PredClass, level=.95, calpha=univariate_calpha())
```

SAS GLM Predicting Income from 2 New Variables for workclass

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	**2**	14414.0265	7207.0132	**42.14**	**<.0001**
Error	**731**	125009.2054	**171.0112**		
Corrected Total	733	139423.2319			

Mean Square Error, the residual variance, is 171.01 after including 2 slopes for workclass as a predictor (which accounted for 10.34% of the variance in income as the model R^2). The *F*-test tells us this R^2 is significantly > 0, *F*(2, 731) = 42.14, MSE = 171.01, *p* < .001.

R-Square	Coeff Var	Root MSE	income Mean
0.103383	75.57777	13.07713	17.30287

Table of Model-Estimated Fixed Effects (normally is last)

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits		
Intercept	13.65004014	0.62628075	21.80	<.0001	12.42051668	14.87956360	**Beta0**
LvM	8.85426742	1.00368116	8.82	<.0001	6.88382600	10.82470884	**Beta1**
LvU	10.98470986	2.99044960	3.67	0.0003	5.11381580	16.85560393	**Beta2**

Interpret β_0 = Intercept:

Interpret β_1 = slope of Low vs Mid:

Interpret β_2 = slope of Low vs Upp:

Table of Extra Requested Linear Combinations of Model-Estimated Fixed Effects

```
                                    Standard
Parameter              Estimate        Error    t Value    Pr > |t|    95% Confidence Limits

Pred Income: Low     13.6500401   0.62628075      21.80      <.0001    12.4205167   14.8795636 Beta0
Pred Income: Mid     22.5043076   0.78431390      28.69      <.0001    20.9645311   24.0440840 Beta0+Beta1
Pred Income: Upp     24.6347500   2.92413427       8.42      <.0001    18.8940472   30.3754528 Beta0+Beta2

Low vs Mid Diff       8.8542674   1.00368116       8.82      <.0001     6.8838260   10.8247088 Beta1
Low vs Upp Diff      10.9847099   2.99044960       3.67      0.0003     5.1138158   16.8556039 Beta2
Mid vs Upp Diff       2.1304424   3.02749229       0.70      0.4818    -3.8131743    8.0740592 Beta2-Beta1
```

Syntax and SAS Output to Compute Partial Effect Sizes from Requested Category Differences:

```
* SAS code to compute effect sizes from stored ESTIMATE results;
DATA work.ClassEstimates; SET work.ClassEstimates;
* Cohen d is partial standardized mean difference;
  PartialD=(2*tValue)/SQRT(731); * SQRT(number) = DF denominator;
* PartialR is partial correlation;
  PartialR = tvalue/(SQRT(tvalue**2 +731)); * +number = DF denominator;
RUN;
* Print estimates table with effect sizes added;
TITLE "PartialD and PartialR Effect Sizes for 3-Category workclass";
PROC PRINT NOOBS DATA=work.ClassEstimates;
     VAR Parameter--PartialR; * Print all contiguous columns;
RUN; TITLE;
```

Btw, effect sizes for predicted outcomes are not meaningful (but these rows were included in the dataset of saved estimates).

Parameter	Estimate	StdErr	tValue	Probt	LowerCL	UpperCL	PartialD	PartialR
Pred Income: Low	13.6500401	0.62628075	21.80	<.0001	12.4205167	14.8795636	1.61226	0.62760
Pred Income: Mid	22.5043076	0.78431390	28.69	<.0001	20.9645311	24.0440840	2.12250	0.72780
Pred Income: Upp	24.6347500	2.92413427	8.42	<.0001	18.8940472	30.3754528	0.62319	0.29749
Low vs Mid Diff	8.8542674	1.00368116	8.82	<.0001	6.8838260	10.8247088	0.65257	0.31019
Low vs Upp Diff	10.9847099	2.99044960	3.67	0.0003	5.1138158	16.8556039	0.27172	0.13462
Mid vs Upp Diff	2.1304424	3.02749229	0.70	0.4818	-3.8131743	8.0740592	0.05205	0.02602

```
// STATA code to compute effect sizes from stored results per lincom
lincom c.LvM*1  + c.LvU*0 // Low vs Mid Diff
   display "PartialD= " (2*(r(estimate)/r(se)))/sqrt(r(df))
   display "PartialR= " (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.LvM*0  + c.LvU*1  // Low vs Upp Diff
   display "PartialD= " (2*(r(estimate)/r(se)))/sqrt(r(df))
   display "PartialR= " (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.LvM*-1 + c.LvU*1  // Mid vs Upp Diff
   display "PartialD= " (2*(r(estimate)/r(se)))/sqrt(r(df))
   display "PartialR= " (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
```

```
# R code to compute effect sizes from stored model and GLHT results
PredClassPartialD=(2*SavePredClass$test$tstat)/sqrt(ModelClass$df.residual)
PredClassPartialR=SavePredClass$test$tstat/
             sqrt(SavePredClass$test$tstat^2+ModelClass$df.residual)
# Concatenate effect sizes to results table for mean differences
data.frame(Estimate=SavePredClass$test$coefficients, pvalue=SavePredClass$test$pvalues,
           PartialD=PredlassPartialD, PartialR=PredClassPartialR)
```

<u>Example Results Section for Income Mean Differences by Working Class:</u>

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 contrasts to distinguish the three classes, in which lower-class respondents served as the reference group to be compared separately to middle-class and to upper-class respondents. We found that 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).

Syntax to Center Age at 18 years (minimum of sample):

```
* SAS code to create 1 new age variable centered at 18 (minimum in sample);
DATA work.Example3; SET work.Example3;
  age18=age-18; LABEL age18= "age18: Age (0=18 years)";
RUN;

// STATA code to create 1 new age variable centered at 18 (minimum in sample)
gen age18=age-18
label variable age18 "age18: Age (0=18 years)"

# R code to make new age variable centered at 18 (minimum in sample)
Example3$age18=Example3$age-18 # age18: Age (0=18 years)
```

Syntax and SAS Output for Age Predicting Income:

First Testing a <u>Linear</u> Effect of Age (0=18): $Income_i = \beta_0 + \beta_1(Age_i - 18) + e_i$

The syntax below will also request the predicted outcome for example ages 30, 50, and 70.

```
TITLE "SAS GLM Predicting Income from Linear Centered Age (0=18)";
PROC GLM DATA=work.Example3 NAMELEN=100;
  MODEL income = age18 / SOLUTION ALPHA=.05 CLPARM SS3;
* Ask for predicted income for example ages;
  ESTIMATE "Pred Income: Age 30 (age18=12)"  intercept 1 age18 12;
  ESTIMATE "Pred Income: Age 50 (age18=32)"  intercept 1 age18 32;
  ESTIMATE "Pred Income: Age 70 (age18=52)"  intercept 1 age18 52;
RUN; QUIT; TITLE;

display "STATA GLM Predicting Income from Linear Centered Age (0=18)"
regress income c.age18, level(95)
// Ask for predicted income for example ages
   lincom _cons*1 + c.age18*12  // Pred Income: Age 30 (age18=12)
   lincom _cons*1 + c.age18*32  // Pred Income: Age 50 (age18=32)
   lincom _cons*1 + c.age18*52  // Pred Income: Age 70 (age18=52)

print("R GLM Predicting Income from Linear Centered Age")
ModelLinAge = lm(data=Example3, formula=income~1+age18)
anova(ModelLinAge)    # anova to print residual variance
summary(ModelLinAge) # summary to print fixed effects solution
confint(ModelLinAge, level=.95) # confint to print level% CI for fixed effects

print("R Ask for predicted income for example ages")
print("In number lists below, values are multiplier for each fixed effect IN ORDER")
PredLinAge = glht(model=ModelLinAge, linfct=rbind(
  "Pred Income: Age 30 (age18=12)" = c(1,12),
  "Pred Income: Age 50 (age18=32)" = c(1,32),
  "Pred Income: Age 70 (age18=52)" = c(1,52)))
print("Print glht linear combination results with unadjusted p-values and 95% CIs")
summary(PredLinAge, test=adjusted("none"))
confint(PredLinAge, level=.95, calpha=univariate_calpha())
```

```
SAS GLM Predicting Income from Linear Centered Age (0=18)
                                          Sum of
Source                      DF           Squares     Mean Square    F Value    Pr > F
Model                        1         5580.7424       5580.7424      30.52    <.0001
Error                      732       133842.4895        182.8449
Corrected Total            733       139423.2319

R-Square      Coeff Var      Root MSE      income Mean
0.040027       78.14896      13.52202         17.30287
```

Mean Square Error, the residual variance, is 182.84 after including a linear effect of age (which accounted for 4.00% of the variance in income as the model R^2). The *F*-test tells us this R^2 is significantly > 0, $F(1, 732) = 30.52$, MSE = 182.84, $p < .001$.

Table of Model-Estimated Fixed Effects (normally is last)

```
                                  Standard
Parameter          Estimate          Error    t Value    Pr > |t|      95% Confidence Limits
Intercept       12.33998883     1.02765825      12.01      <.0001     10.32247980  14.35749786
age18            0.20624834     0.03733240       5.52      <.0001      0.13295699   0.27953969
```

Interpret β_0 = Intercept:

Interpret β_1 = slope of age18:

Table of Extra Requested Linear Combinations of Model-Estimated Fixed Effects

```
                                                Standard
Parameter                        Estimate          Error  t Value  Pr > |t|    95% Confidence Limits
Pred Income Age 30 (age18=12)  14.8149689     0.67223750    22.04    <.0001    13.4952255    16.1347124
Pred Income Age 50 (age18=32)  18.9399357     0.58044193    32.63    <.0001    17.8004063    20.0794652
Pred Income Age 70 (age18=52)  23.0649026     1.15623917    19.95    <.0001    20.7949622    25.3348429
```

Second, Keeping a Linear Slope for Age and Adding a Quadratic Slope for Age (0=18):

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

The syntax below will also request not only the predicted outcome for example ages 30, 50, and 70, but also the predicted instantaneous linear slopes at those ages too: $Linear\ Age\ Slope = \beta_1 + 2\beta_2(Age_i - 18)$

```
TITLE "SAS GLM Predicting Income from Linear+Quadratic Centered Age";
PROC GLM DATA=work.Example3 NAMELEN=100;
* Asterisk creates multiplied predictor variable;
  MODEL income = age18 age18*age18 / SOLUTION ALPHA=.05 CLPARM SS3;
* Ask for predicted income for example ages;
  ESTIMATE "Pred Income: Age 30 (age18=12)"  intercept 1 age18 12 age18*age18  144;
  ESTIMATE "Pred Income: Age 50 (age18=32)"  intercept 1 age18 32 age18*age18 1024;
  ESTIMATE "Pred Income: Age 70 (age18=52)"  intercept 1 age18 52 age18*age18 2704;
* Linear age slope changes by 2*quadratic coefficient per year, so multiply age*2;
  ESTIMATE "Pred Linear Age Slope: Age 30 (age18=12)"    age18 1  age18*age18  24;
  ESTIMATE "Pred Linear Age Slope: Age 50 (age18=32)"    age18 1  age18*age18  64;
  ESTIMATE "Pred Linear Age Slope: Age 70 (age18=52)"    age18 1  age18*age18 104;
* Save predicted income and SE to new dataset to make pictures;
  OUTPUT OUT=work.PredIncomebyAge PREDICTED=YhatAge STDP=SEyhatAge;
RUN; QUIT; TITLE;

display as result "STATA GLM Predicting Income from Linear+Quadratic Centered Age (0=18)"
regress income c.age18 c.age18#c.age18, level(95) // Hashtag multiplies predictors
// Ask for predicted income for example ages
   lincom _cons*1 + c.age18*12 + c.age18#c.age18*144  // Pred Income: Age 30 (age18=12)
   lincom _cons*1 + c.age18*32 + c.age18#c.age18*1024 // Pred Income: Age 50 (age18=32)
   lincom _cons*1 + c.age18*52 + c.age18#c.age18*2704 // Pred Income: Age 70 (age18=52)
// Linear age slope changes by 2*quadratic coefficient, so multiply age*2
   lincom  c.age18*1 + c.age18#c.age18*24  // Pred Linear Age Slope: Age 30 (age18=12)
   lincom  c.age18*1 + c.age18#c.age18*64  // Pred Linear Age Slope: Age 50 (age18=32)
   lincom  c.age18*1 + c.age18#c.age18*104 // Pred Linear Age Slope: Age 70 (age18=52)
```

```
print("R GLM Predicting Income from Linear+Quadratic Centered Age")
ModelQuadAge = lm(data=Example3, formula=income~1+age18+I(age18^2)) # I(x^2) squares x
anova(ModelQuadAge)    # anova to print residual variance
summary(ModelQuadAge) # summary to print fixed effects solution
confint(ModelQuadAge, level=.95) # confint to print level% CI for fixed effects

print("R Ask for predicted income and predicted linear age slopes for example ages")
print("In number lists below, values are multiplier for each fixed effect IN ORDER")
PredQuadAge = glht(model=ModelQuadAge, linfct=rbind(
  "Pred Income: Age 30 (age18=12)" = c(1,12, 144),
  "Pred Income: Age 50 (age18=32)" = c(1,32,1024),
  "Pred Income: Age 70 (age18=52)" = c(1,52,2704),
  "Pred Linear Age Slope: Age 30 (age18=12)" = c(0,1, 24),
  "Pred Linear Age Slope: Age 50 (age18=32)" = c(0,1, 64),
  "Pred Linear Age Slope: Age 70 (age18=52)" = c(0,1,104)))
print("Print glht linear combination results with unadjusted p-values and 95% CIs")
summary(PredQuadAge, test=adjusted("none"))
confint(PredQuadAge, level=.95, calpha=univariate_calpha())
```

SAS GLM Predicting Income from <u>Linear+Quadratic</u> Centered Age (0=18)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	**2**	15885.4618	7942.7309	**47.00**	**<.0001**
Error	**731**	123537.7701	**168.9983**		
Corrected Total	733	139423.2319			

R-Square	Coeff Var	Root MSE	income Mean
0.113937	75.13165	12.99994	17.30287

Mean Square Error, the residual variance, is now 169.00 from the two effects of age (which accounted for 11.39% of the variance in income as the model R^2). The *F*-test says this R^2 is significantly > 0, $F(2, 731) = 47.00$, MSE = 169.00, $p < .001$.

Table of Model-Estimated Fixed Effects (normally is last)

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits		
Intercept	2.676597431	1.58352919	1.69	0.0914	-0.432210062	5.785404923	**Beta0**
age18	1.223080607	0.13507406	9.05	<.0001	0.957901252	1.488259961	**Beta1**
age18*age18	-0.019537211	0.00250199	-7.81	<.0001	-0.024449155	-0.014625267	**Beta2**

Interpret β_0 = Intercept:

Interpret β_1 = slope of age18:

Interpret β_2 = slope of age18^2:

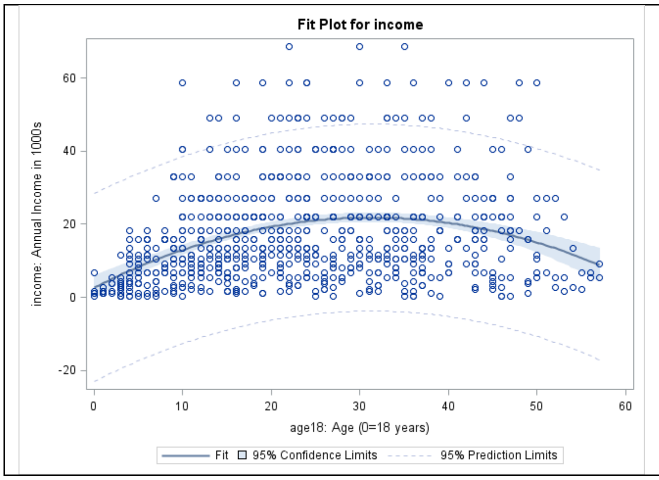
Interpret R^2 two different ways:

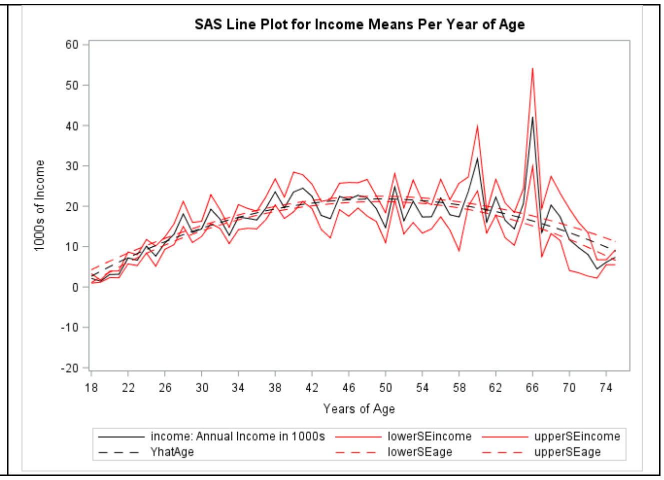
The R^2 went from .040 to .114, an increase of .074. Do we know if the R^2 increased <u>significantly</u> relative to the linear age model?

Table of Extra Requested Linear Combinations of Model-Estimated Fixed Effects

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Pred Income: Age 30 (age18=12)	14.5402064	0.64723977	22.46	<.0001	13.2695359	15.8108769
Pred Income: Age 50 (age18=32)	21.8090730	0.66813438	32.64	<.0001	20.4973819	23.1207641
Pred Income: Age 70 (age18=52)	13.4481710	1.65902182	8.11	<.0001	10.1911553	16.7051867
Pred Linear Age Slope: Age 30 (age18=12)	0.7541875	0.07881678	9.57	<.0001	0.5994533	0.9089218
Pred Linear Age Slope: Age 50 (age18=32)	-0.0273009	0.04671950	-0.58	0.5592	-0.1190213	0.0644195
Pred Linear Age Slope: Age 70 (age18=52)	-0.8087893	0.13485251	-6.00	<.0001	-1.0735337	-0.5440449

Left: model-predicted regression line through scatterplot (provided automatically)
Right: model-predicted regression line through means for age (see extra code online)





We forgo requesting standardized slopes for this model given the ambiguity of how to interpret them for models with interactions… R^2 is a sufficiently useful effect size to describe the overall effect (trend) of age here.

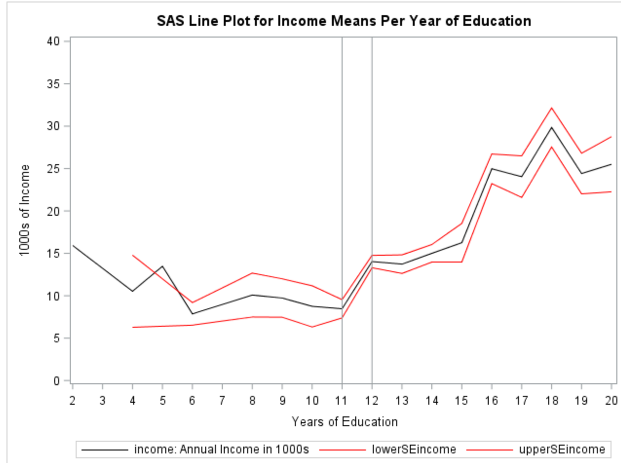
<u>Example Results Section for the Linear and Quadratic Effects of Age:</u>

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 means of income by age to identify plausible types of nonlinear associations. Given the apparent curvilinear trend (in which age appeared positively associated with income until middle age, upon which it appeared negatively associated instead), we fit a model including 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 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 = 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).

Syntax to Create 3 Predictors for Piecewise Slopes for Education:

The idea is to represent the 3 different sections of education using 3 different predictors, that way the slope for each section is captured separately.

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




```
* SAS code to create 3 new predictor variables for sections of education;
DATA work.Example3; SET work.Example3;
  lessHS=.; gradHS=.; overHS=.; * Make 3 new empty variables;
* Replace each for educ less than 12;
  IF educ LT 12 THEN DO; lessHS=educ-11; gradHS=0; overHS=0;        END;
* Replace each for educ greater or equal to 12;
  IF educ GE 12 THEN DO; lessHS=0;        gradHS=1; overHS=educ-12; END;
  LABEL lessHS= "lessHS: Slope for Years Ed Less Than High School"
        gradHS= "gradHS: Bump for Graduating High School"
        overHS= "overHS: Slope for Years Ed After High School";
RUN;
```

```
// STATA code to create 3 new predictor variables for sections of education
gen lessHS=. // Make 3 new empty variables
gen gradHS=.
gen overHS=.
// Replace each 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 each 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
// Label variables
label variable lessHS "lessHS: Slope for Years Ed Less Than High School"
label variable gradHS "gradHS: Acute Bump for Graduating High School"
label variable overHS "overHS: Slope for Years Ed After High School"
```

```
# R code to make to make 3 new variables for sections of education
# Make 3 new empty variables
Example3$lessHS=NA; Example3$gradHS=NA; Example3$overHS=NA
# Replace each for educ less than 12
Example3$lessHS[which(Example3$educ<12)]=Example3$educ[which(Example3$educ<12)]-11
Example3$gradHS[which(Example3$educ<12)]=0
Example3$overHS[which(Example3$educ<12)]=0
# Replace each for educ greater or equal to 12
Example3$lessHS[which(Example3$educ>=12)]=0
Example3$gradHS[which(Example3$educ>=12)]=1
Example3$overHS[which(Example3$educ>=12)]=Example3$educ[which(Example3$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
```

Syntax and SAS Output for Piecewise Linear Slopes of Education Predicting Income:

$$Income_i = \beta_0 + \beta_1(LessHS_i) + \beta_2(GradHS_i) + \beta_3(OverHS_i) + e_i$$

The syntax below will also show how to test slope differences.

```
TITLE "SAS GLM Predicting Income from 3 Piecewise Linear Slopes for Education";
PROC GLM DATA=work.Example3 NAMELEN=100;
  MODEL income = lessHS gradHS overHS / SOLUTION ALPHA=.05 CLPARM SS3;
* Example of how to compare slopes;
  ESTIMATE "Diff in ed slope:  2-11 vs 11-12" lessHS -1 gradHS 1;
  ESTIMATE "Diff in ed slope: 11-12 vs 12-20" gradHS -1 overHS 1;
* Save predicted income and SE to new dataset to make pictures;
  OUTPUT OUT=work.PredIncomebyEduc PREDICTED=YhatEduc STDP=SEyhatEduc;
* Save fixed effect estimates and requested estimates as SAS datasets to do math on them;
  ODS OUTPUT ParameterEstimates=work.EducSolution ParameterEstimates=work.EducEstimates;
RUN; QUIT; TITLE;

display "STATA GLM Predicting Income from 3 Piecewise Linear Slopes for Education"
regress income c.lessHS c.gradHS c.overHS, level(95)
// Example of how to compare slopes
   lincom c.lessHS*-1 + c.gradHS*1 // Diff in ed slope:  2-11 vs 11-12
   lincom c.gradHS*-1 + c.overHS*1 // Diff in ed slope: 11-12 vs 12-20

print("R GLM Predicting Income from 3 Piecewise Linear Slopes for Education ")
ModelEd3 = lm(data=Example3, formula=income~1+lessHS+gradHS+overHS)
anova(ModelEd3)    # anova to print residual variance
SaveModelEd3 = summary(ModelEd3) # summary to print fixed effects solution
print(SaveModelEd3); confint(ModelEd3, level=.95) # confint for level% CI for fixed effects

print("R Example of how to compare slopes")
print("In number lists below, values are multiplier for each fixed effect IN ORDER")
PredEd3 = glht(model=ModelEd3, 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)))
print("Print glht linear combination results with unadjusted p-values and 95% CIs")
SavePredEd3 = summary(PredEd3, test=adjusted("none"))
Print(SavePredEd3); confint(PredEd3, level=.95, calpha=univariate_calpha())
```

SAS GLM Predicting Income from 3 Piecewise Linear Slopes for Education

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	**3**	22906.5605	7635.5202	**47.84**	**<.0001**
Error	**730**	116516.6714	**159.6119**		
Corrected Total	733	139423.2319			

R-Square	Coeff Var	Root MSE	income Mean
0.164295	73.01538	12.63376	17.30287

Mean Square Error, the residual variance, is 159.61 given the piecewise education slopes (which accounted for 16.43% of the variance in income as the model R^2). The *F*-test says this R^2 is significantly > 0, $F(3, 730) = 47.84$, MSE = 159.61, $p < .001$.

Table of Model-Estimated Fixed Effects (normally is last)

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits		
Intercept	8.534867248	1.72935077	4.94	<.0001	5.139773001	11.929961495	**Beta0**
lessHS	-0.268784499	0.59880153	-0.45	0.6537	-1.444363022	0.906794023	**Beta1**
gradHS	4.684746178	1.87568395	2.50	0.0127	1.002367857	8.367124499	**Beta2**
overHS	2.124528973	0.21372442	9.94	<.0001	1.704941139	2.544116806	**Beta3**

Interpret β_0 = Intercept:

Interpret β_1 = slope of lessHS:

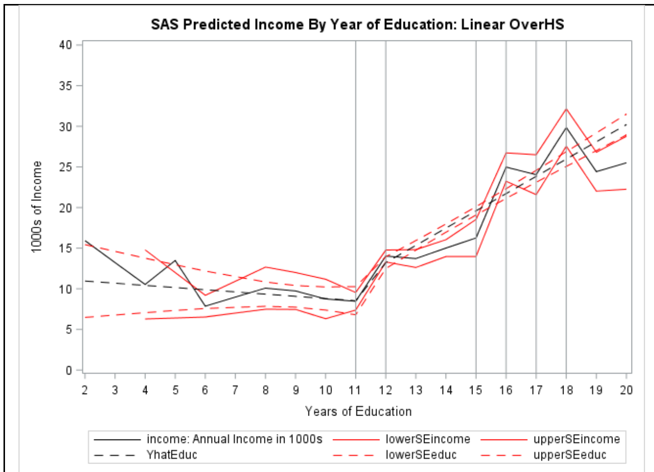
Interpret β_2 = slope of gradHS:

Interpret β_3 = slope of overHS:

Table of Extra Requested Linear Combinations of Model-Estimated Fixed Effects

```
                                                         Standard
Parameter                              Estimate             Error    t Value    Pr > |t|      95% Confidence Limits
Diff in ed slope:  2-11 vs 11-12     4.95353068        2.28222698       2.17      0.0303      0.47301937    9.43404199
Diff in ed slope: 11-12 vs 12-20    -2.56021721        1.94673385      -1.32      0.1889     -6.38208203    1.26164762
```

Comparisons of Slopes Above: The slope for gradHS is significantly more positive than the slope for lessHS by 4.95 per year (indicating that they should not be constrained to be the same). The slope for overHS is nonsignificantly less positive than the slope for gradHS (by −2.56 per year, indicating that they *could* be constrained to be the same). However, it's important to note that the slope for overHS—implying a linear effect of each additional year of education—does not appear to fit the means well. So efforts to refine the model should focus on better capturing differences after 12 years first…



Left: model-predicted regression line through means for education (see extra code online)

As shown by the misfit of the data to the model (dashed line), it looks like the effect of education after 12 years should have additional piecewise slopes (i.e., 12–15, 15–17, 17–18, 18–20)… if you are feeling brave, give it a try and let me know what happens!

Syntax and SAS Output to Compute Partial Effect Sizes from Requested Piecewise Slopes:

```
* SAS code to compute effect sizes from stored fixed effect results;
DATA work.EducEffectSizes; LENGTH Parameter $50;
  SET work.EducSolution work.EducEstimates; * Combine tables;
* PartialR is partial correlation (using +DFden);
  PartialR = tvalue/(SQRT(tvalue**2 +730)); * +number = DF denominator;
RUN;
* Print estimates table with effect sizes added;
TITLE "PartialR Effect Sizes for Piecewise Slopes for Education";
PROC PRINT NOOBS DATA=work.EducEffectSizes;
     VAR Parameter--PartialR; * Print all contiguous columns;
RUN; TITLE;
```

Parameter	Estimate	StdErr	tValue	Probt	LowerCL	UpperCL	PartialR
Intercept	8.534867248	1.72935077	4.94	<.0001	5.139773001	11.929961495	0.17969
lessHS	-0.268784499	0.59880153	-0.45	0.6537	-1.444363022	0.906794023	-0.01661
gradHS	4.684746178	1.87568395	2.50	0.0127	1.002367857	8.367124499	0.09205
overHS	2.124528973	0.21372442	9.94	<.0001	1.704941139	2.544116806	0.34529
Diff in ed slope: 2-11 vs 11-12	4.953530678	2.28222698	2.17	0.0303	0.473019369	9.434041986	0.08008
Diff in ed slope: 11-12 vs 12-20	-2.560217205	1.94673385	-1.32	0.1889	-6.382082034	1.261647623	-0.04862

```
* SAS alternative method to compute partial correlations for fixed slopes;
TITLE "SAS Partial Correlation of income with lessHS";
PROC CORR NOSIMPLE DATA=work.Example3; VAR income lessHS; PARTIAL gradHS overHS; RUN;
TITLE "SAS Partial Correlation of income with gradHS";
PROC CORR NOSIMPLE DATA=work.Example3; VAR income gradHS; PARTIAL lessHS overHS; RUN;
TITLE "SAS Partial Correlation of income with overHS";
PROC CORR NOSIMPLE DATA=work.Example3; VAR income overHS; PARTIAL lessHS gradHS; RUN;
TITLE;
```

```
// STATA code to compute effect sizes from stored results per lincom
lincom c.lessHS*1 // Slope for 2-11 years
   display "PartialR= " (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.gradHS*1 // Slope for 11-12 years
   display "PartialR= " (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.overHS*1 // Slope for 12+ years
   display "PartialR= " (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.lessHS*-1 + c.gradHS*1 // Diff in ed slope:  2-11 vs 11-12
   display "PartialR= " (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.gradHS*-1 + c.overHS*1 // Diff in ed slope: 11-12 vs 12-20
   display "PartialR= " (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))

// STATA alternative method to compute partial correlations from fixed slopes
display "STATA Partial Correlations of Income with Education Slopes"
pcorr income lessHS gradHS overHS


# R code to compute effect sizes from stored model fixed effects
ModelEd3PartialR=SaveModelEd3$coefficients[,"t value"]/
            sqrt(SaveModelEd3$coefficients[,"t value"]^2+ModelEd3$df.residual)
# Concatenate effect sizes to results table for fixed effects
data.frame(SaveModelEd3$coefficients, PartialR=ModelEd3PartialR)

# R code to compute effect sizes from stored glht results
PredEd3PartialR=SavePredEd3$test$tstat/sqrt(SavePredEd3$test$tstat^2+ModelEd3$df.residual)
# Concatenate effect sizes to results table for mean differences
data.frame(Estimate=SavePredEd3$test$coefficients, pvalue=SavePredEd3$test$pvalues,
           PartialR=PredEd3PartialR)

# R alternative method to compute partial correlations for fixed slopes
print("R Partial Correlation of income with lessHS")
pcor.test(Example3$income,Example3$lessHS, Example3[,c("gradHS","overHS")])
print("R Partial Correlation of income with gradHS")
pcor.test(Example3$income,Example3$gradHS, Example3[,c("lessHS","overHS")])
print("R Partial Correlation of income with overHS")
pcor.test(Example3$income,Example3$overHS, Example3[,c("lessHS","gradHS")])
```

Example Results Section for 3 Piecewise Linear Slopes for the Effect of Education:

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 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 slopes to 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 effects 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). Although 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, 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).

Syntax to Center 5-category Ordinal Happiness at 1 (minimum):

```
* SAS code to create 1 new happy variable centered at lowest value;
DATA work.Example3; SET work.Example3;
  happy1=happy-1; LABEL happy1= "happy1: Happy Category (0=1)";
RUN;

// STATA code to create 1 new happy variable centered at lowest value
   gen happy1=happy-1
   label variable happy1 "happy1: Happy Category (0=1)"

# R code to make a single happy variable centered at lowest value
Example3$happy1=Example3$happy-1 # happy1: Happy Category (0=1)
```

Syntax and SAS Output for 5-Category Ordinal Happiness Predicting Income:

First Testing a <u>Linear</u> Effect of Happy (0=1): $Income_i = \beta_0 + \beta_1(Happy_i - 1) + e_i$

```
TITLE "SAS GLM Predicting Income from Linear Centered Happy (0=1)";
PROC GLM DATA=work.Example3 NAMELEN=100;
  MODEL income = happy1 / SOLUTION ALPHA=.05 CLPARM SS3;
* Save predicted income and SE to new dataset to make pictures;
  OUTPUT OUT=work.PredIncomebyHappy1 PREDICTED=Yhat1Happy STDP=SEyhat1Happy;
RUN; QUIT; TITLE;

display "STATA GLM Predicting Income from Linear Centered Happy (0=1)"
regress income c.happyc1, level(95)

print("R GLM Predicting Income from Linear Centered Age")
ModelHappy1 = lm(data=Example3, formula=income~1+happy1)
anova(ModelHappy1)    # anova to print residual variance
summary(ModelHappy1) # summary to print fixed effects solution
confint(ModelHappy1, level=.95) # confint for level% CI for fixed effects
```

SAS GLM Predicting Income from <u>Linear</u> Centered Happy (0=1)

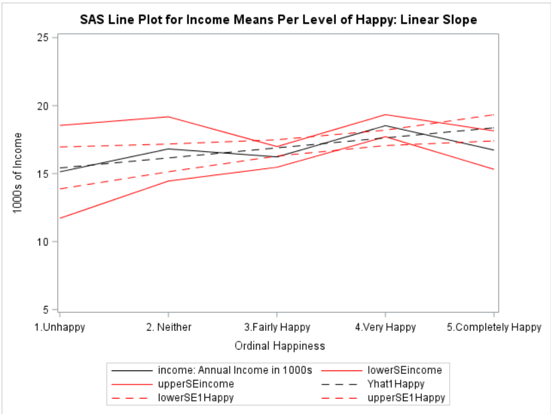
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	**1**	320.3981	320.3981	**1.69**	**0.1945**
Error	**732**	139102.8338	**190.0312**		
Corrected Total	733	139423.2319			

R-Square	Coeff Var	Root MSE	income Mean
0.002298	79.66988	13.78518	17.30287

Mean Square Error, the residual variance, is 190.03 after a linear effect of happy (which accounted for 0.23% of the variance in income as the model R^2). The *F*-test tells us this R^2 is **not** significantly > 0, $F(1, 732) = 1.69$, MSE = 190.03, $p = .195$.

Table of Model-Estimated Fixed Effects (normally is last)

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits		
Intercept	15.41494808	1.54042165	10.01	<.0001	12.39077678	18.43911937	**Beta0**
happy1	0.73866637	0.56887362	1.30	0.1945	-0.37815205	1.85548479	**Beta1**



Interpret β_0 = Intercept:

Interpret β_1 = slope of happy1:

Syntax to Create Sequential-Dummy-Coded Predictors—4 needed for 5 happy categories:

In addition to not really making sense (i.e., these values are ordinal, so they aren't really numbers), a single linear slope predicting the same difference between each pair of happiness categories doesn't seem to fit the pattern of means. So let's fit a piecewise slopes model created through sequential-dummy-coding, in which the slopes capture each shift between adjacent categories.

```
* SAS code to create 4 new sequential-dummy-coded binary predictors for happy;
DATA work.Example3; SET work.Example3;
  h1v2=.; h2v3=.; h3v4=.; h4v5=.; * Make 4 new empty variables;
  IF happy=1 THEN DO; h1v2=0; h2v3=0; h3v4=0; h4v5=0; END; * Replace each for happy=1;
  IF happy=2 THEN DO; h1v2=1; h2v3=0; h3v4=0; h4v5=0; END; * Replace each for happy=2;
  IF happy=3 THEN DO; h1v2=1; h2v3=1; h3v4=0; h4v5=0; END; * Replace each for happy=3;
  IF happy=4 THEN DO; h1v2=1; h2v3=1; h3v4=1; h4v5=0; END; * Replace each for happy=4;
  IF happy=5 THEN DO; h1v2=1; h2v3=1; h3v4=1; h4v5=1; END; * Replace each for happy=5;
  LABEL 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";
RUN;
```

```
// STATA code to make 4 new sequential-dummy-coded variables for happy
// Make 4 new empty variables
   gen h1v2=.
   gen h2v3=.
   gen h3v4=.
   gen h4v5=.
// Replace each with 0 values
   replace h1v2=0 if happy < 2
   replace h2v3=0 if happy < 3
   replace h3v4=0 if happy < 4
   replace h4v5=0 if happy < 5
// Replace each with 1 values
   replace h1v2=1 if happy >= 2
   replace h2v3=1 if happy >= 3
   replace h3v4=1 if happy >= 4
   replace h4v5=1 if happy == 5
// Label variables
   label variable h1v2 "Slope from Happy 1 to 2"
   label variable h2v3 "Slope from Happy 2 to 3"
   label variable h3v4 "Slope from Happy 3 to 4"
   label variable 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 code to create 4 new sequential-dummy-coded predictors for happy
# Make 4 new empty variables
Example3$h1v2=NA; Example3$h2v3=NA; Example3$h3v4=NA; Example3$h4v5=NA;
# Replace each with 0 values
Example3$h1v2[which(Example3$happy<2)]=0
Example3$h2v3[which(Example3$happy<3)]=0
Example3$h3v4[which(Example3$happy<4)]=0
Example3$h4v5[which(Example3$happy<5)]=0
# Replace each with 1 values
Example3$h1v2[which(Example3$happy>=2)]=1
Example3$h2v3[which(Example3$happy>=3)]=1
Example3$h3v4[which(Example3$happy>=4)]=1
Example3$h4v5[which(Example3$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
```

Second, Testing 4 Sequential Adjacent Slopes for Happy:

$$Income_i = \beta_0 + \beta_1(h1v2_i) + \beta_2(h2v3_i) + \beta_3(h3v4_i) + \beta_3(h4v5_i) + e_i$$

```
TITLE "SAS GLM Predicting Income from Sequential Slopes for Happy";
PROC GLM DATA=work.Example4 NAMELEN=100;
     MODEL income = h1v2 h2v3 h3v4 h4v5 / SOLUTION ALPHA=.05 CLPARM SS3;
     * Example of how to compare slopes;
     ESTIMATE "Diff in Slope 1-2 vs 2-3" h1v2 -1 h2v3 1;
     ESTIMATE "Diff in Slope 2-3 vs 3-4" h2v3 -1 h3v4 1;
     ESTIMATE "Diff in Slope 3-4 vs 4-5" h3v4 -1 h4v5 1;
* Save fixed effect estimates and requested estimates as SAS datasets to do math on them;
  ODS OUTPUT ParameterEstimates=work.HappySolution Estimates=HappyEstimates;
RUN; QUIT; TITLE;
```

```
display "STATA GLM Predicting Income from Sequential Slopes for Happy"
regress income c.h1v2 c.h2v3 c.h3v4 c.h4v5, level(95)
// Example of how to compare slopes
   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
```

```
print("R GLM Predicting Income from Sequential Slopes for Happy")
ModelHappy5 = lm(data=Example3, formula=income~1+h1v2+h2v3+h3v4+h4v5)
anova(ModelHappy5)    # anova to print residual variance
SaveModelHappy5 = summary(ModelHappy5) # summary to print fixed effects solution
print(SaveModelHappy5); confint(ModelHappy5, level=.95) # confint for level% CI

print("R Example of how to compare slopes")
print("In number lists below, values are multiplier for each fixed effect IN ORDER")
PredHappy5 = glht(model=ModelHappy5, 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)))
print("Print glht linear combination results with unadjusted p-values and 95% CIs")
SavePredHappy5 = summary(PredHappy5, test=adjusted("none"))
print(SavePredHappy5); confint(PredHappy5, level=.95, calpha=univariate_calpha())
```

SAS GLM Predicting Income from Sequential Slopes for Happy

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	**4**	946.3348	236.5837	**1.25**	**0.2902**
Error	**729**	138476.8971	**189.9546**		
Corrected Total	733	139423.2319			

R-Square	Coeff Var	Root MSE	income Mean
0.006787	79.65383	13.78240	17.30287

Mean Square Error, the residual variance, is 189.95 after adding the 4 slopes of happy (which accounted for 0.68% of the variance in income as the model R^2). The *F*-test tells us this R^2 is **not** significantly > 0, $F(4, 729) = 1.25$, MSE = 189.95, $p = .290$.

Table of Model-Estimated Fixed Effects (normally is last)

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits		
Intercept	15.12875000	2.70295132	5.60	<.0001	9.82225260	20.43524740	**Beta0**
h1v2	1.68516026	3.48949515	0.48	0.6293	-5.16549843	8.53581894	**Beta1**
h2v3	-0.58648838	2.36910124	-0.25	0.8045	-5.23756348	4.06458671	**Beta2**
h3v4	2.29929831	1.15017869	2.00	0.0460	0.04124054	4.55735608	**Beta3**
h4v5	-1.79692367	1.67023208	-1.08	0.2823	-5.07596246	1.48211511	**Beta4**

The fixed intercept gives the mean for happy=1, and each slope gives the difference to the next category.

Table of Extra Requested Linear Combinations of Model-Estimated Fixed Effects

```
                                                   Standard
Parameter                          Estimate           Error  t Value  Pr > |t|    95% Confidence Limits
Diff in Slope 1-2 vs 2-3        -2.27164864      5.24694941    -0.43    0.6652  -12.57258275   8.02928547
Diff in Slope 2-3 vs 3-4         2.88578669      2.90164986     0.99    0.3203   -2.81080036   8.58237373
Diff in Slope 3-4 vs 4-5        -4.09622198      2.29660358    -1.78    0.0749   -8.60496798   0.41252402
```

Comparisons of Slopes Above: No pairwise differences between slopes are significant, which means we would not lose anything predictive informative by constraining the slopes to be equal in these data.

Syntax and SAS Output to Compute Partial Effect Sizes from Requested Piecewise Slopes:

```
* SAS code to compute effect sizes from stored fixed effect results;
DATA work.HappyEffectSizes; LENGTH Parameter $50;
  SET work.HappySolution work.HappyEstimates; * Combine tables;
  IF INDEX(Parameter, "Intercept")>0 THEN DELETE; * Remove intercept;
* PartialR is partial correlation using +DFden);
  PartialR = tvalue/(SQRT(tvalue**2 +729)); * +number = DF denominator;
RUN;
* Print estimates table with effect sizes added;
TITLE "PartialR Effect Sizes for Sequential Slopes for Happy";
PROC PRINT NOOBS DATA=work.HappyEffectSizes;
     VAR Parameter--PartialR; * Print all contiguous columns;
RUN; TITLE;
```

Parameter	Estimate	StdErr	tValue	Probt	LowerCL	UpperCL	PartialR
Intercept	15.12875000	2.70295132	5.60	<.0001	9.82225260	20.43524740	0.20299
h1v2	1.68516026	3.48949515	0.48	0.6293	-5.16549843	8.53581894	0.01788
h2v3	-0.58648838	2.36910124	-0.25	0.8045	-5.23756348	4.06458671	-0.00917
h3v4	2.29929831	1.15017869	2.00	0.0460	0.04124054	4.55735608	0.07384
h4v5	-1.79692367	1.67023208	-1.08	0.2823	-5.07596246	1.48211511	-0.03981
Diff in Slope 1-2 vs 2-3	-2.27164864	5.24694941	-0.43	0.6652	-12.57258275	8.02928547	-0.01603
Diff in Slope 2-3 vs 3-4	2.88578669	2.90164986	0.99	0.3203	-2.81080036	8.58237373	0.03681
Diff in Slope 3-4 vs 4-5	-4.09622198	2.29660358	-1.78	0.0749	-8.60496798	0.41252402	-0.06592

```
* SAS alternative method to compute partial correlations for fixed slopes;
TITLE "SAS Partial Correlation of income with h1v2";
PROC CORR NOSIMPLE DATA=work.Example3; VAR income h1v2; PARTIAL h2v3 h3v4 h4v5; RUN;
TITLE "SAS Partial Correlation of income with h2v3";
PROC CORR NOSIMPLE DATA=work.Example3; VAR income h2v3; PARTIAL h1v2 h3v4 h4v5; RUN;
TITLE "SAS Partial Correlation of income with h3v4";
PROC CORR NOSIMPLE DATA=work.Example3; VAR income h3v4; PARTIAL h1v2 h2v3 h4v5; RUN;
TITLE "SAS Partial Correlation of income with h4v5";
PROC CORR NOSIMPLE DATA=work.Example3; VAR income h4v5; PARTIAL h1v2 h2v3 h3v4; RUN;
TITLE;
```

```
// STATA code to compute effect sizes from stored results per lincom
lincom c.h1v2*1 // Slope for 1-2 happy
   display "PartialR=" (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.h2v3*1 // Slope for 2-3 happy
   display "PartialR=" (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.h3v4*1 // Slope for 3-4 happy
   display "PartialR=" (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.h4v5*1 // Slope for 4-5 happy
   display "PartialR=" (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.h1v2*-1 + c.h2v3*1 // Diff in Slope 1-2 vs Slope 2-3
   display "PartialR=" (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.h2v3*-1 + c.h3v4*1 // Diff in Slope 2-3 vs Slope 3-4
   display "PartialR=" (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
lincom c.h3v4*-1 + c.h4v5*1 // Diff in Slope 3-4 vs Slope 4-5
   display "PartialR=" (r(estimate)/r(se))/sqrt((r(estimate)/r(se))^2+r(df))
```

```
// STATA alternative method to compute partial correlations for fixed slopes
display "STATA Partial Correlations of Income with Happy Slopes"
pcorr income h1v2 h2v3 h3v4 h4v5


# R code to compute effect sizes from stored model fixed effects
ModelHappy5PartialR=SaveModelHappy5$coefficients[,"t value"]/
               sqrt(SaveModelHappy5$coefficients[,"t value"]^2+ModelHappy5$df.residual)
# Concatenate effect sizes to results table for fixed effects
data.frame(SaveModelHappy5$coefficients, PartialR=ModelHappy5PartialR)

# R code to compute effect sizes from stored glht results
PredHappy5PartialR=SavePredHappy5$test$tstat/
              sqrt(SavePredHappy5$test$tstat^2+ModelHappy5$df.residual)
# Concatenate effect sizes to results table for mean differences
data.frame(Estimate=SavePredHappy5$test$coefficients, pvalue=SavePredHappy5$test$pvalues,
           PartialR=PredHappy5PartialR)

# R alternative method to compute partial correlations for fixed slopes
print("R Partial Correlation of income with h1v2")
pcor.test(Example3$income,Example3$h1v2, Example3[,c("h2v3","h3v4","h4v5")])
print("R Partial Correlation of income with h2v3")
pcor.test(Example3$income,Example3$h2v3, Example3[,c("h1v2","h3v4","h4v5")])
print("R Partial Correlation of income with h3v4")
pcor.test(Example3$income,Example3$h3v4, Example3[,c("h1v2","h2v3","h4v5")])
print("R Partial Correlation of income with h4v5")
pcor.test(Example3$income,Example3$h4v5, Example3[,c("h1v2","h2v3","h3v4")])
```

Example Results Section for the Linear and Piecewise Sequential Slopes for Happy:

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 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 adjacent differences in predicted annual income by ordinal level of happiness. The revised model—predicting four adjacent 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).