Example 4: General Linear Models with Multiple Conceptual Predictors
(complete data, syntax, and output in R available online)

These example data were selected from the High School and Beyond 1980 dataset (in the `candisc` R package). Like Example 3, the current example will use general linear models (estimated within the base R `lm` function) to examine associations of math with nominal program type (academic, general, or vocational), ordinal four-category motivation, and a quadratic trend for writing. It will also obtain linear combinations of fixed effects to create predicted outcomes using the `glht` function from the `multcomp` R package. Additionally, this example will demonstrate how to conduct nested model comparisons when adding predictors sequentially.

Syntax for importing and preparing data for analysis (after loading packages `TeachingDemos`, `readxl`, `multcomp`, and `supernova`):

```
# Set working directory (to import and export files to)
# Paste in the folder address where your data file is saved in quotes
# Note the slashes are backwards relative to Windows file paths
setwd("C:/Dropbox/26_EDF9770/Example4/")

# Import "HSB_Example.xlsx" from sheet "Sheet1" with first row as variable names
Example4 = read_excel(path="HSB_Example.xlsx", sheet="Sheet1", col_names=TRUE)
# Convert to data frame to use for analysis
Example4 = as.data.frame(Example4)

# Load R functions for this class from R file in working directory
source("EDF9770_Functions.R")

# Create indicator-coded binary predictors for program type
Example4$AcavGen=NA; Example4$AcavVoc=NA      # Make 2 new empty variables
Example4$AcavGen[which(Example4$prog==1)]=1  # general
Example4$AcavVoc[which(Example4$prog==1)]=0
Example4$AcavGen[which(Example4$prog==2)]=0  # academic
Example4$AcavVoc[which(Example4$prog==2)]=0
Example4$AcavGen[which(Example4$prog==3)]=0  # vocational
Example4$AcavVoc[which(Example4$prog==3)]=1
# AcavGen: Academic=0 vs General=1, AcavVoc: Academic=0 vs Vocational=1
```

program	AcavGen	AcavVoc
general	1	0
academic	0	0
vocational	0	1

```
# Create sequential-coded binary predictors for motiv
Example4$m1v2=NA; Example4$m2v3=NA; Example4$m3v4=NA  # Make 3 new empty variables
# Replace each with 0 values
Example4$m1v2[which(Example4$motiv<2)]=0
Example4$m2v3[which(Example4$motiv<3)]=0
Example4$m3v4[which(Example4$motiv<4)]=0
# Replace each with 1 values
Example4$m1v2[which(Example4$motiv>=2)]=1
Example4$m2v3[which(Example4$motiv>=3)]=1
Example4$m3v4[which(Example4$motiv>=4)]=1
# m1v2: From motiv 1 to 2, m2v3: From motiv 2 to 3, m3v4: From motiv 3 to 4
```

motiv	m1v2	m2v3	m3v4
1	0	0	0
2	1	0	0
3	1	1	0
4	1	1	1

```
# Center quantitative write variable to be used as a predictor
Example4$write26 = Example4$write-26   # write26: Writing Score (0=26 as min)
```

Model 1: Do math outcomes differ across the three types of programs?

Because program type is a nominal variable (stored as 1, 2, 3), we need to recode it into two new binary variables (also known as "contrasts") so that 0 will be a meaningful value for the β_0 fixed intercept as one of the programs.

$$math_i = \beta_0 + \beta_1(General_i) + \beta_2(Vocational_i) + e_i$$

```
print("GLM Predicting Math from Program")
Model1 = lm(data=Example4, formula=math~1+AcavGen+AcavVoc)
obj=LMsummary(Model1, explain=TRUE, effectsizes=TRUE)  # Custom output
```

```
Sums of Squares Table
             SS  DF       MS      F      p    R2
Model 10515.135   2 5257.567 73.717 <0.001 0.198
Error 42578.583 597   71.321
Total 53093.718 599   88.637

Fixed Effects Table
             Est    SE       t      p     LCI    UCI
Intercept 55.811 0.481 115.980 <0.001  54.866 56.756
AcavGen   -6.719 0.851  -7.900 <0.001  -8.389 -5.049
AcavVoc   -9.543 0.847 -11.272 <0.001 -11.205 -7.880

Effect Sizes for Fixed Effects Table
           Est      p      d     pr   sR2
AcavGen -6.719 <0.001 -0.647 -0.308 0.084
AcavVoc -9.543 <0.001 -0.923 -0.419 0.171

Explanation:
Est = Estimate, p = two-sided p-value, d = Cohen's d,
pr = Partial r, sR2 = Semi-Partial R-square
```

Model 2: Does motivation predict math after controlling for program type?

Previously in Example 3 we saw that motivation (a four-category ordinal variable) appeared to have a nonlinear relation with math. Thus, we add to the previous model including nominal program type three sequential slopes to capture the difference in math for each additional category of motivation.

$$math_i = \beta_0 + \beta_1(General_i) + \beta_2(Vocational_i) + \beta_3(m1v2_i) + \beta_4(m2v3_i) + \beta_5(m3v4_i) + e_i$$

```
print("GLM Predicting Math from Program + Motivation")
Model2 = lm(data=Example4, formula=math~1+AcavGen+AcavVoc+m1v2+m2v3+m3v4)
obj=LMsummary(Model2, effectsizes=TRUE)  # Custom output
```

```
Sums of Squares Table
             SS  DF       MS      F      p    R2
Model 11085.486   5 2217.097 31.350 <0.001 0.209
Error 42008.232 594   70.721
Total 53093.718 599   88.637

Fixed Effects Table
             Est    SE       t      p     LCI    UCI
Intercept 54.252 1.133  47.884 <0.001  52.027 56.477
AcavGen   -6.354 0.858  -7.404 <0.001  -8.040 -4.669
AcavVoc   -9.067 0.861 -10.529 <0.001 -10.759 -7.376
m1v2      -0.076 1.285  -0.059  0.953  -2.599  2.448
m2v3       1.597 1.006   1.588  0.113  -0.378  3.572
m3v4       0.795 0.846   0.939  0.348  -0.867  2.457

Effect Sizes for Fixed Effects Table
           Est      p      d     pr   sR2
AcavGen -6.354 <0.001 -0.608 -0.291 0.073
AcavVoc -9.067 <0.001 -0.864 -0.397 0.148
m1v2    -0.076  0.953 -0.005 -0.002 0.000
m2v3     1.597  0.113  0.130  0.065 0.003
m3v4     0.795  0.348  0.077  0.039 0.001
```

```
Previous Model 1 results:
Effect Sizes for Fixed Effects Table
           Est      p      d     pr   sR2
AcavGen -6.719 <0.001 -0.647 -0.308 0.084
AcavVoc -9.543 <0.001 -0.923 -0.419 0.171
```

Note how the effect sizes for the program type contrasts are smaller after adding the contrasts for motivation because they have contributed less unique prediction (i.e., due to the overlap with motivation). However, it is possible that the partial correlation and Cohen's *d* effect sizes could increase after adding predictors that reduce the residual variance (leaving less "leftover" for the denominator). However, the semi-partial eta-square effect sizes will never increase after adding additional predictors because their denominator is total variance.

```
# Get F-test and effect sizes for change in R2
obj=R2compare(ReducedModel=Model1, FullModel=Model2, PredName="Motivation", explain=TRUE)

F-Test and R2 Change for Motivation Slopes
 R2reduced R2full R2diff DFnum DFden     F     p   pR2   sR2
     0.198  0.209  0.011     3   594 2.688 0.046 0.013 0.011
```

Even though each of the sequential contrasts for motivation are not significant, their overall contribution to the model is, as indicated by the F-test for the change in model R^2 after adding them. This indicates that other pairwise comparisons must be significant, so let's find out which ones:

$$\widehat{math}_i = \beta_0 + \beta_1(General_i) + \beta_2(Vocational_i) + \beta_3(m1v2_i) + \beta_4(m2v3_i) + \beta_5(m3v4_i) + e_i$$

```
print("Get motivation means and all differences")
PredModel2 = multcomp::glht(model=Model2, linfct=rbind(
  "Pred Math: motiv=1"  = c(1,0,0, 0,0,0),
  "Pred Math: motiv=2"  = c(1,0,0, 1,0,0),
  "Pred Math: motiv=3"  = c(1,0,0, 1,1,0),
  "Pred Math: motiv=4"  = c(1,0,0, 1,1,1),
  "Motiv 1 vs 3 Diff"   = c(0,0,0, 1,1,0),
  "Motiv 1 vs 4 Diff"   = c(0,0,0, 1,1,1),
  "Motiv 2 vs 4 Diff"   = c(0,0,0, 0,1,1)))
obj=glhtSummary(glhtObject=PredModel2, effectsizes=TRUE) # Custom output
```

motiv	m1v2	m2v3	m3v4
1	0	0	0
2	1	0	0
3	1	1	0
4	1	1	1

```
Linear Combinations Table
                      Est    SE      t      p    LCI    UCI
Pred Math: motiv=1 54.252 1.133 47.884 <0.001 52.027 56.477
Pred Math: motiv=2 54.177 0.876 61.828 <0.001 52.456 55.897
Pred Math: motiv=3 55.774 0.735 75.864 <0.001 54.330 57.217
Pred Math: motiv=4 56.568 0.599 94.412 <0.001 55.391 57.745
Motiv 1 vs 3 Diff   1.521 1.223  1.244  0.214 -0.880  3.923
Motiv 1 vs 4 Diff   2.316 1.182  1.960  0.050 -0.004  4.637
Motiv 2 vs 4 Diff   2.392 0.952  2.512  0.012  0.522  4.262

Effect Sizes for Linear Combinations Table
                      Est      p     d    pr    sR2
Pred Math: motiv=1 54.252 <0.001 3.929 0.891  3.054
Pred Math: motiv=2 54.177 <0.001 5.074 0.930  5.092
Pred Math: motiv=3 55.774 <0.001 6.225 0.952  7.666
Pred Math: motiv=4 56.568 <0.001 7.748 0.968 11.873
Motiv 1 vs 3 Diff   1.521  0.214 0.102 0.051  0.002
Motiv 1 vs 4 Diff   2.316  0.050 0.161 0.080  0.005
Motiv 2 vs 4 Diff   2.392  0.012 0.206 0.103  0.008
```

Model 3: Does writing predict math after controlling for program type and motivation?

Previously in Example 3 we saw that writing had a significant quadratic relation with math. Thus, we add both linear and quadratic slopes for writing to the previous model including program type and motivation.

$$math_i = \beta_0 + \beta_1(General_i) + \beta_2(Vocational_i) + \beta_3(m1v2_i) + \beta_4(m2v3_i) + \beta_5(m3v4_i) + \beta_6(write_i - 26) + \beta_7(write_i - 26)^2 + e_i$$

```
print("GLM Predicting Math from Program + Motivation + Writing")
Model3 = lm(data=Example4,
            formula=math~1+AcavGen+AcavVoc+m1v2+m2v3+m3v4+write26+I(write26^2))
obj=LMsummary(Model3, effectsizes=TRUE)  # Custom output
```

```
Sums of Squares Table
             SS  DF       MS      F      p    R2
Model 24252.048   7 3464.578 71.113 <0.001 0.457
Error 28841.670 592   48.719
Total 53093.718 599   88.637

Fixed Effects Table
                Est    SE      t      p    LCI    UCI
Intercept    45.266 1.919 23.583 <0.001 41.496 49.035
AcavGen      -3.539 0.733 -4.830 <0.001 -4.978 -2.100
AcavVoc      -4.805 0.763 -6.297 <0.001 -6.304 -3.307
m1v2         -0.546 1.068 -0.511  0.610 -2.644  1.552
m2v3          0.622 0.839  0.740  0.459 -1.027  2.270
m3v4         -0.113 0.705 -0.161  0.872 -1.497  1.270
write26       0.015 0.145  0.106  0.916 -0.269  0.300
I(write26^2)  0.011 0.003  3.591 <0.001  0.005  0.016

Effect Sizes for Fixed Effects Table
                Est      p      d     pr   sR2
AcavGen      -3.539 <0.001 -0.397 -0.195 0.021
AcavVoc      -4.805 <0.001 -0.518 -0.251 0.036
m1v2         -0.546  0.610 -0.042 -0.021 0.000
m2v3          0.622  0.459  0.061  0.030 0.001
m3v4         -0.113  0.872 -0.013 -0.007 0.000
write26       0.015  0.916  0.009  0.004 0.000
I(write26^2)  0.011 <0.001  0.295  0.146 0.012
```

```
# Get F-test and effect sizes for change in R2
obj=R2compare(ReducedModel=Model2, FullModel=Model3, PredName="Writing")

F-Test and R2 Change for Writing Slopes
 R2reduced R2full R2diff DFnum DFden       F      p   pR2   sR2
     0.209  0.457  0.248     2   592 135.127 <0.001 0.313 0.248
```

For a full report, we request the pairwise differences indirectly given for program type and motivation as linear combinations of their fixed slopes—to do so, we need to refer back to their coding schemes:

program	AcavGen	AcavVoc
general	1	0
academic	0	0
vocational	0	1

motiv	m1v2	m2v3	m3v4
1	0	0	0
2	1	0	0
3	1	1	0
4	1	1	1

```
print("Get missing program and motivation differences")
PredModel3 = multcomp::glht(model=Model3, linfct=rbind(
  "Gen vs Voc Diff"    = c(0,-1,1, 0,0,0, 0,0),
  "Motiv 1 vs 3 Diff" = c(0, 0,0, 1,1,0, 0,0),
  "Motiv 1 vs 4 Diff" = c(0, 0,0, 1,1,1, 0,0),
  "Motiv 2 vs 4 Diff" = c(0, 0,0, 0,1,1, 0,0)))
obj=glhtSummary(glhtObject=PredModel3, effectsizes=TRUE) # Custom output

Linear Combinations Table
                      Est    SE      t     p    LCI   UCI
Gen vs Voc Diff    -1.267 0.827 -1.532 0.126 -2.891 0.357
Motiv 1 vs 3 Diff   0.076 1.019  0.074 0.941 -1.925 2.076
Motiv 1 vs 4 Diff  -0.038 0.991 -0.038 0.970 -1.984 1.909
Motiv 2 vs 4 Diff   0.508 0.802  0.634 0.527 -1.067 2.083

Effect Sizes for Linear Combinations Table
                      Est     p      d     pr   sR2
Gen vs Voc Diff    -1.267 0.126 -0.126 -0.063 0.002
Motiv 1 vs 3 Diff   0.076 0.941  0.006  0.003 0.000
Motiv 1 vs 4 Diff  -0.038 0.970 -0.003 -0.002 0.000
Motiv 2 vs 4 Diff   0.508 0.527  0.052  0.026 0.000
```

The contributions to the R^2 reported above are relative to what predictors are in the model at that step. Let's check the contribution to the full model of program (from step 1) and motivation (from step 2).

```
# Get F-test and effect sizes for change in R2 from full model for program
Model4NoProg = lm(data=Example4, formula=math~1+m1v2+m2v3+m3v4+write26+I(write26^2))
obj=R2compare(ReducedModel=Model4NoProg, FullModel=Model3, PredName="Program")

F-Test and R2 Change for Program Slopes
 R2reduced R2full R2diff DFnum DFden      F      p   pR2   sR2
     0.414  0.457  0.043     2   592 23.160 <0.001 0.073 0.043 → still contributes!

# Get F-test and effect sizes for change in R2 from full model for motivation
Model4NoMotiv = lm(data=Example4, formula=math~1+AcavGen+AcavVoc+write26+I(write26^2))
obj=R2compare(ReducedModel=Model4NoMotiv, FullModel=Model3, PredName="Motivation")

F-Test and R2 Change for Motivation Slopes
 R2reduced R2full R2diff DFnum DFden     F     p   pR2   sR2
     0.456  0.457  0.001     3   592 0.205 0.893 0.001 0.001 → no longer contributes!
```

Example Results Section (in which the "bivariate results" section refers to the models in Example 3) ***(although it's more verbose than would be typical for the sake of completeness here):***

Analytic Strategy

The extent to which math could be predicted from program type, motivation, and writing examined in general linear models. All analyses were conducted using the lm function in R v. 4.5.2. Predicted outcomes and linear combinations of the fixed effects were generated using the glht function within the multcomp package v. 1.4-29. Program (a nominal predictor) was represented by two binary contrasts to distinguish the three types, in which academic programs served as the reference to be compared to general and vocational programs. Motivation (an ordinal predictor) was represented by three binary contrasts for sequential differences across categories. Writing (a quantitative predictor) was centered such that 0 = 26, near the sample minimum.

Bivariate Results

The predictors were first examined in separate models to obtain their bivariate relations (i.e., before controlling for the other predictors). Program type significantly predicted math, $F(2, 597) = 73.72$, MSE = 71.32, $p < .001$, $R^2 = .198$, such that students in academic programs had significantly greater math on average than students in general programs, who in turn had significantly higher math than students in vocational programs. Motivation also significantly predicted math, $F(3, 596) = 8.10$, MSE = 85.59, $p < .001$, $R^2 = .039$, but the only significant sequential difference was between 3 and 2, such that students with motivation = 3 had greater math than motivation = 2. Given that descriptive statistics indicated a potential curvilinear trend, a quadratic model for writing was estimated. Linear and quadratic writing also significantly predicted math, $F(2, 597) = 209.54$, MSE = 52.25, $p < .001$, $R^2 = .412$, such that the linear slope became significantly more positive as writing increased.

Sequential Results

We then examined to what extent motivation and writing could still predict math after controlling for program type by adding them in sequential models (i.e., a stepwise linear regression). Adding ordinal motivation to the model with only program type resulted in a significant model, $F(5, 594) = 31.35$, MSE = 70.72, $p < .001$, $R^2 = .209$, as well as a significant increase in the model R^2 relative to program alone, $F(3, 594) = 2.69$, $p < .001$, R^2 change = .011. Math remained significantly higher for students in academic than general or vocational programs, but none of the sequential differences in motivation were uniquely significant. Additional pairwise comparisons, requested as linear combinations of the fixed slopes, revealed that math was significantly higher for motivation = 4 than motivation = 2 ($p = .012$), and marginally higher for motivation = 4 than motivation = 1

(p = .050). Finally, adding linear and quadratic slopes for writing to the model resulted in a significant model, $F(7, 592)$ = 71.11, MSE = 48.72, p < .001, R^2 = . 457, as well as a significant increase in the model R^2 relative to program type and motivation alone, $F(2, 592)$ = 135.13, p < .001, R^2 change = .248.

Simultaneous Results

Results for the full model are given in Table 1, including effect sizes (Cohen's *d* standardized mean differences and partial correlations), which can be interpreted as follows. The intercept (45.27) represents the expected math score for a student in an academic program with motivation = 1 and writing = 26. Program type contributed .043 to the full model R^2, which was significant, $F(2, 592)$ = 23.16, p < .001. Relative to students in academic programs, students in general or vocational programs had significantly lower math by 3.54 or 4.81, respectively. Additionally, students in vocational programs had nonsignificantly lower math by 1.27 than students in general programs. Motivation contributed .001 to the full model R^2, which was not significant, $F(3, 592)$ = 0.21, p < .001. Thus, it appears that the marginal contribution of motivation at step 2 was eliminated after adding writing at step 3. More specifically, students with motivation = 2 had nonsignificantly lower math by 0.55 than students with motivation = 1, students with motivation = 3 had nonsignificantly higher math by 0.62 than students with motivation = 2, and students with motivation = 4 had nonsignificantly lower math by 0.11 than students with motivation = 3. No other pairwise differences in motivation were significant. Finally, writing contributed .248 to the model R^2, which was significant, $F(2, 592)$ = 135.13, p < .001. The instantaneous linear slope at writing = 26 was nonsignificantly positive, but it became significantly more positive (i.e., accelerated) by twice the quadratic slope of 0.01 per additional unit of writing.

Table 1

Results for Final Model Predicting Math from Program, Motivation, and Writing

Fixed Effects	Est	SE	*p* <	Cohen's *d*	Partial *r*
Intercept	45.266	1.919	.001		
Academic vs General	**−3.539**	0.733	.001	−0.397	−0.195
Academic vs Vocational	**−4.805**	0.763	.001	−0.518	−0.251
General vs Vocational*	−1.267	0.827	.126	−0.126	−0.063
Motivation 1 vs 2	−0.546	1.068	.610	−0.042	−0.021
Motivation 2 vs 3	0.622	0.839	.459	0.061	0.030
Motivation 3 vs 4	−0.113	0.705	.872	−0.013	−0.007
Motivation 1 vs 3*	0.076	1.019	.941	0.006	0.003
Motivation 1 vs 4*	−0.038	0.991	.970	−0.003	−0.002
Motivation 2 vs 4*	0.508	0.802	.527	0.052	0.026
Linear Writing	0.015	0.145	.916	0.009	0.004
Quadratic Writing	0.011	0.003	.001	0.295	0.146

Note: Est = estimate, SE = standard error. Bold values indicate significant fixed effects at alpha = .05, and * indicates slopes computed as a linear combination of fixed slopes from the model. Effect sizes were computed from the slope t test-statistics as follows: Cohen's $d = \frac{2t}{\sqrt{DF_{den}}}$; partial $r = \frac{t}{\sqrt{t^2+DF_{den}}}$.