2xK Using GLM & Regression

The purpose of this study was to examine the relationships of exam Review Attendance and Practice Difficulty with exam performance. Practice Difficulty was a 3-condition variable - practice problems were either about the same difficulty as the exam problems (=3), they were easier than the exam problems (=1), or they were more difficult than the exam problems (=2). Different sections of the course were randomly assigned to receive the three difficulty levels. The schedule showed the class meeting during which the exam review would occur & student's attendance was recorded. The dependent variable was performance on an examination.

SPSS Code

```
unianova  testperf  by  pg1e2h3s    ar1y2n                                          ← lists DV  "by"  IVs
  / method = sstype(3)                                                             ← Type 3 SS (more below)
  / emmeans tables ( pg1e2h3s by ar1y2n ) compare ( ar1y2n )                       ← pairwise simple effects of "ar1y2n" from
                                                                                     the "pg1e2h3s by ar1y2n" interaction
  / emmeans tables ( pg1e2h3s by ar1y2n ) compare ( pg1e2h3s )                     ← pairwise simple effects of "pg1e2h3s" from
                                                                                     the "pg1e2h3s by ar1y2n" interaction
  / emmeans tables ( pg1e2h3s )              compare ( pg1e2h3s )                  ← pairwise comparisons of "pg1e2h3s"
                                                                                     corrected marginal means
  / emmeans tables ( ar1y2n )                compare ( ar1y2n )                    ← pairwise comparisons of "ar1y2n"
                                                                                     corrected marginal means
  / print descriptives parameters                                                  ← get raw/data means and regression weights
  / design =  pg1e2h3s    ar1y2n    pg1e2h3s*ar1y2n.                               ← specify the design (including the interaction
                                                                                     that GLM automatically calculates from the
                                                                                     IVs specified above)
```

The "Descriptive Statistics" are the raw or "uncorrected" means.

The F-tests are based on effects coding (using .5, 0 & -.5 weights) of each main effect and their product terms to represent the interactions.

F-tests for effects that are represented by 2 or more codes (here, the pg main effect and the interaction, each df=2), are the same F you would get from a nested-model $R^2\Delta$ F-test dropping all the codes representing that effect.

Descriptive Statistics

Dependent Variable: testperf

practgrp_1e2h3s	atndrev_1y2n	Mean	Std. Deviation	N
Easier	Yes	44.0000	9.66092	10
	No	61.6667	9.83192	6
	Total	50.6250	12.89380	16
Harder	Yes	81.0000	13.70320	10
	No	41.6667	11.69045	6
	Total	66.2500	23.34524	16
Same	Yes	80.0000	8.94427	6
	No	60.0000	8.16497	10
	Total	67.5000	12.90994	16
Total	Yes	66.5385	21.15510	26
	No	55.4545	12.62170	22
	Total	61.4583	18.44942	48

Tests of Between-Subjects Effects

Dependent Variable: testperf

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11301.250[a]	5	2260.250	20.212	.000
Intercept	169586.806	1	169586.806	1516.532	.000
pg1e2h3s	2210.278	2	1105.139	9.883	.000
ar1y2n	2170.139	1	2170.139	19.406	.000
pg1e2h3s * ar1y2n	6301.944	2	3150.972	28.178	.000
Error	4696.667	42	111.825		
Total	197300.000	48			
Corrected Total	15997.917	47			

a. R Squared = .706 (Adjusted R Squared = .671)

Like all models with an interaction term, the regression weights for the dummy codes describe simple effects of that variable when all other variable = 0, which is for the comparision/reference condition of the other variable

Parameter Estimates

Dependent Variable: testperf

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	60.000	3.344	17.942	.000	53.251	66.749
[pg1e2h3s=1.00]	1.667	5.461	.305	.762	-9.354	12.687
[pg1e2h3s=2.00]	-18.333	5.461	-3.357	.002	-29.354	-7.313
[pg1e2h3s=3.00]	0[a]	.	.	.	.	.
[ar1y2n=1.00]	20.000	5.461	3.662	.001	8.980	31.020
[ar1y2n=2.00]	0[a]	.	.	.	.	.
[pg1e2h3s=1.00] * [ar1y2n=1.00]	-37.667	7.723	-4.877	.000	-53.252	-22.082
[pg1e2h3s=1.00] * [ar1y2n=2.00]	0[a]	.	.	.	.	.
[pg1e2h3s=2.00] * [ar1y2n=1.00]	19.333	7.723	2.503	.016	3.748	34.918
[pg1e2h3s=2.00] * [ar1y2n=2.00]	0[a]	.	.	.	.	.
[pg1e2h3s=3.00] * [ar1y2n=1.00]	0[a]	.	.	.	.	.
[pg1e2h3s=3.00] * [ar1y2n=2.00]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Interpreting the regression weights

constant

- The expected value of the criterion when the value of all predictors = 0
- The expected value of testperf for those in the same condition and did not attend the review
- Those in same condition who did not attend the review scored 60% on the exam

pg1e2h3s=1.00

compares easier & same

- The direction and extent of the expected change in testperf for a 1-unit increase in this predictor, holding the value of the other predictor constant at 0
- The expected difference in testperf between same and easier practices for those who did not attend the review
- The simple effect of same versus easier practices for those who did not attend the review
- Among those who did not attend the review, those with easier practices (mean 61.667%) scored 1.667% better than those with same difficulty practices (mean = 60.00%)

pg1e2h3s=2.00

compares harder & same

- The direction and extent of the expected change in testperf for a 1-unit increase in this predictor, holding the value of the other predictor constant at 0
- The expected difference in testperf between same and harder practices for those who did not attend the review
- The simple effect of same versus harder practices for those who did not attend the review
- Among those who did not attend the review, those with harder practices (mean 41.667%) scored 18.333% poorerr than those with same difficulty practices (mean = 60.00%)

ar1y2n=1.00

- The direction and extent of the expected change in testperf for a 1-unit increase in this predictor, holding the value of the other predictor constant at 0
- The expected difference in testperf for those who did and did not attend the review, among those who had the same difficulty practice
- The simple effect of attending the review for those who had the same difficulty practices
- Among those who had the same difficulty practices, those who did attend (mean = 80%) scored 20% higher on average than those who did not attend (mean = 60&)

pg1e2h3s=1.00
ar1y2n=1.00

- The direction and extent of the difference in the expected effect of one predictor when the other predictor increases by 1 – can be expressed in terms of either variable
- How the simple effect of one variable is expected to change as the value of the other variable increases by one – can be expressed in terms of either variable

SE of practice difficulty (same vs easier)

- SE of same vs easier for those who did not attend review → 60 – 61.667 ➔ 1.667
- SE of same vs easier for those who did attend review → 80 – 44 ➔ -36 dif ➔ -37.667

SE of attending review session

- SE of no vs yes for those with similar difficulty practice → 60 – 80 ➔ 20
- SE of no vs yes for those with easier practice → 61.667 – 44 ➔ -17.667 dif ➔ -37.667

pg1e2h3s=2.00
ar1y2n=1.00

- The direction and extent of the difference in the expected effect of one predictor when the other predictor increases by 1 – can be expressed in terms of either variable
- How the simple effect of one variable is expected to change as the value of the other variable increases by one – can be expressed in terms of either variable

SE of practice difficulty (same vs harder)

- SE of same vs harder for those who did not attend review → 60 – 41.667 ➔ 18.333
- SE of same vs harder for those who did attend review → 80 – 81 ➔ -1 dif ➔ 19.333

SE of attending review session

- SE of no vs yes for those with similar difficulty practice → 60 – 80 ➔ 20
- SE of no vs yes for those with harder practice → 41.667 -81 ➔ -39.333 dif ➔ 19.333

The idea is that we can “recover” the cell means from the regression weights

Same / No Review = the constant ➔ **60.000**

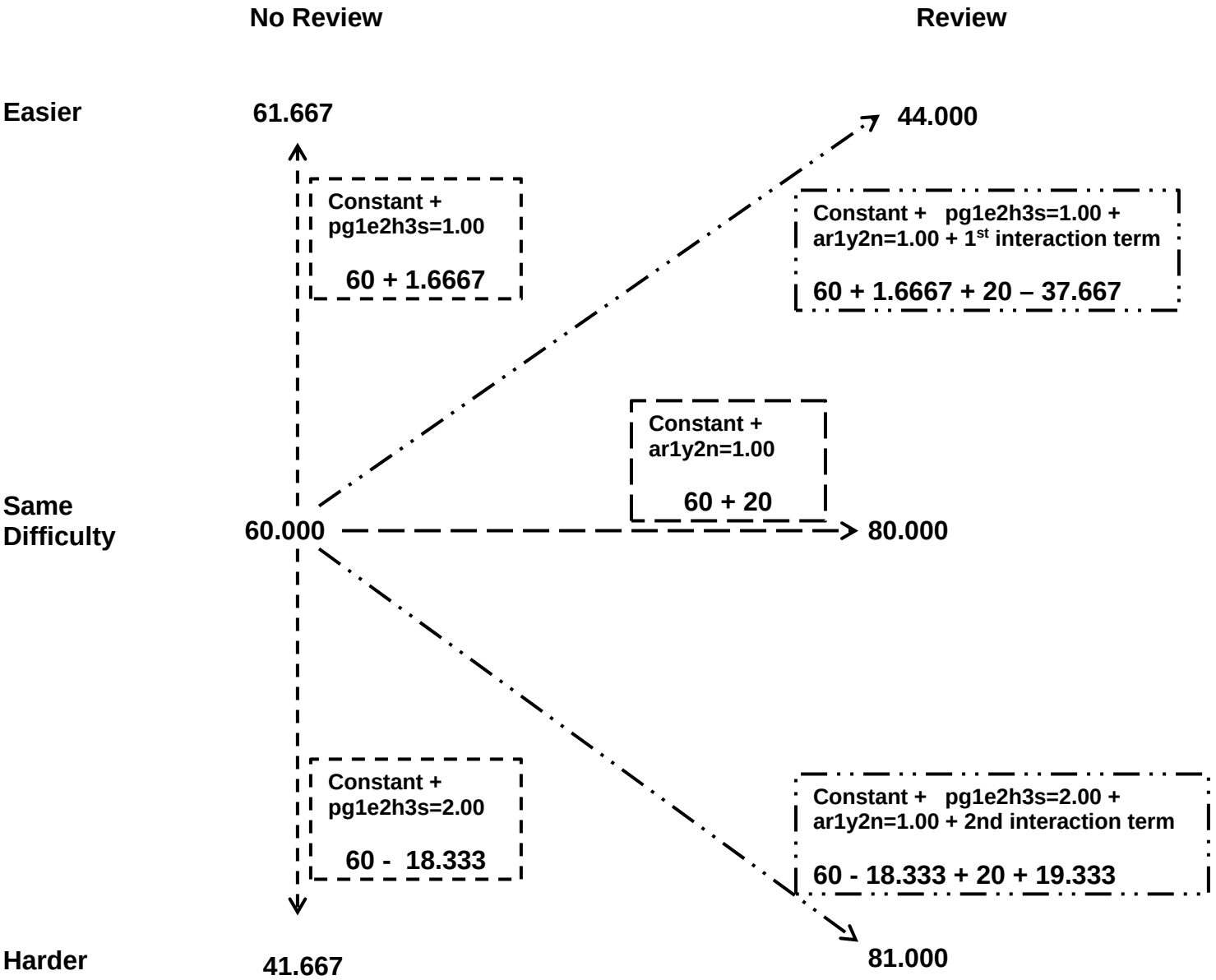
Easier / No Review = constant + SE Same v Easier for No Review (pg1e2h3s=1.00) ➔ 60 + 1.667 = **61.667**

Harder / No Review = constant + SE Same v Harder for No Review (pg1e2h3s=2.00) ➔ 60 + (-18.33) = **41.667**

Same / Yes Review = constant + SE of review for same (ar1y2n=1.00) ➔ 60 + 20 = **80.000**

Easier / Yes Review = Easier/No Review mean ➔ 61.667 + 20 + (-37.667) = **44.000**
+ SE for Review for Same (ar1y2n=1.00)
+ how SE for Easier differs from same (1st interaction term)

Harder / Yes Review = Harder/No Review Mean ➔ 41.667 + 20 + 19.333 = **81.000**
+ SE for Review for same (ar1y2n=1.00)
+ now SE for Harder differs from same (2nd interaction term)



Emmeans results from GLM

In addition to the effect F-tests and the regression weights, GLM can be coaxed into giving us specific pairwise comparisons among any adjacent pair of cell means, and among any set of marginal means. These pairwise comparisons are a nice addition to the regression weights, because they provide significance tests for all comparisons. We would need to perform multiple recordings of the categorical variables to produce all of these comparisons and significance tests via regression weights.

You will usually want both sets of simple effects, as are requested in the GLM code above. One of those sets will be used to describe the pattern of the significant interaction. Each set will be used to determine if the corresponding main effect pattern is descriptive or misleading.

Describing the pairwise simple effects of Review Attendance for each level of Practice Difficulty

/ emmeans tables (pg1e2h3s by ar1y2n) compare (ar1y2n)

Estimates

Dependent Variable: testperf

pg1e2h3s	ar1y2n	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Easier	Yes	44.000	3.344	37.251	50.749
	No	61.667	4.317	52.954	70.379
Harder	Yes	81.000	3.344	74.251	87.749
	No	41.667	4.317	32.954	50.379
Same	Yes	80.000	4.317	71.288	88.712
	No	60.000	3.344	53.251	66.749

The cell means will be the same as given in the "Descriptive Statistics" above.

Univariate Tests

Dependent Variable: testperf

pg1e2h3s		Sum of Squares	df	Mean Square	F	Sig.
Easier	Contrast	1170.417	1	1170.417	10.466	.002
	Error	4696.667	42	111.825		
Harder	Contrast	5801.667	1	5801.667	51.881	.000
	Error	4696.667	42	111.825		
Same	Contrast	1500.000	1	1500.000	13.414	.001
	Error	4696.667	42	111.825		

Each F tests the simple effects of ar1y2n within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

The F-tests tell us that the simple effect of Review Attendance is significant for each level of Practice Difficulty.

With only 2 Review Attendance conditions, the pairwise comparisons are redundant with the F-tests.

Pairwise Comparisons

Dependent Variable: testperf

pg1e2h3s	(I) ar1y2n	(J) ar1y2n	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
Easier	Yes	No	-17.667*	5.461	.002	-28.687	-6.646
	No	Yes	17.667*	5.461	.002	6.646	28.687
Harder	Yes	No	39.333*	5.461	.000	28.313	50.354
	No	Yes	-39.333*	5.461	.000	-50.354	-28.313
Same	Yes	No	20.000*	5.461	.001	8.980	31.020
	No	Yes	-20.000*	5.461	.001	-31.020	-8.980

Based on estimated marginal means

*. The mean difference is significant at the .050 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

The pattern of the interaction is:

Easier Practice
Review < No Review

Same Difficulty Practice
Review > No Review

Harder Practice
Review > No Review

This interactionpattern allows us to anticipate that the main effect of Review Attendance will be **misleading**

Describing the pairwise simple effects of Practice Difficulty for each level of Review Attendance

/ emmeans tables (pg1e2h3s by ar1y2n) compare (pg1e2h3s)

Estimates

Dependent Variable: testperf

pg1e2h3s	ar1y2n	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Easier	Yes	44.000	3.344	37.251	50.749
	No	61.667	4.317	52.954	70.379
Harder	Yes	81.000	3.344	74.251	87.749
	No	41.667	4.317	32.954	50.379
Same	Yes	80.000	4.317	71.288	88.712
	No	60.000	3.344	53.251	66.749

It repeats the same cell means for each emmeans.

Univariate Tests

Dependent Variable: testperf

ar1y2n		Sum of Squares	df	Mean Square	F	Sig.
Yes	Contrast	8258.462	2	4129.231	36.926	.000
	Error	4696.667	42	111.825		
No	Contrast	1578.788	2	789.394	7.059	.002
	Error	4696.667	42	111.825		

Each F tests the simple effects of pg1e2h3s within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

The F-tests tell us that there is a significant simple effect of Practice Difficulty for each condition of Review Attendance.

Pairwise Comparisons

Dependent Variable: testperf

ar1y2n	(I) pg1e2h3s	(J) pg1e2h3s	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
Yes	Easier	Harder	-37.000*	4.729	.000	-46.544	-27.456
		Same	-36.000*	5.461	.000	-47.020	-24.980
	Harder	Easier	37.000*	4.729	.000	27.456	46.544
		Same	1.000	5.461	.856	-10.020	12.020
	Same	Easier	36.000*	5.461	.000	24.980	47.020
		Harder	-1.000	5.461	.856	-12.020	10.020
No	Easier	Harder	20.000*	6.105	.002	7.679	32.321
		Same	1.667	5.461	.762	-9.354	12.687
	Harder	Easier	-20.000*	6.105	.002	-32.321	-7.679
		Same	-18.333*	5.461	.002	-29.354	-7.313
	Same	Easier	-1.667	5.461	.762	-12.687	9.354
		Harder	18.333*	5.461	.002	7.313	29.354

Based on estimated marginal means

*. The mean difference is significant at the .050 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

The pairwise effects describing the interaction are:

	Easier v Same	Easier v Harder	Same v Harder
Did attend the review	44.0 < 80.0	44.0 < 81.0	80.0 = 81.0
Did not attend review	61.7 = 60.0	61.7 > 41.7	60.0 > 41.7

This interaction pattern allows us to anticipate that the main effect pattern of Practice Difficulty will be **misleading**

Describing the Main Effect of Review Attendance

/ emmenas tables (ar1y2n) compare (ar1y2n)

Estimates

Dependent Variable: testperf

ar1y2n	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Yes	68.333	2.134	64.026	72.641
No	54.444	2.320	49.762	59.127

Univariate Tests

Dependent Variable: testperf

	Sum of Squares	df	Mean Square	F	Sig.
Contrast	2170.139	1	2170.139	19.406	.000
Error	4696.667	42	111.825		

The F tests the effect of ar1y2n. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

Pairwise Comparisons

Dependent Variable: testperf

(I) ar1y2n	(J) ar1y2n	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
Yes	No	13.889*	3.153	.000	7.526	20.251
No	Yes	-13.889*	3.153	.000	-20.251	-7.526

Based on estimated marginal means

*. The mean difference is significant at the .050 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

You should notice that the means shown here are not the same as the marginal means from the "Descriptive Statistics" above (which were 66.54 for Yes and 55.45 for No).

Also, the F-test for "ar1y2n" in the ANOVA table above and shown below (which match) are not comparing the data means shown in the "Descriptive Statistics" above.

Because there are unequal sample sizes among the design conditions, the main effects and the interaction are all collinear (nonorthogonal, or correlated). Thus, like all other multiple regressions, the model tests the unique contribution of each effect to the model, controlling for the other effects in the model.

So, in a factorial ANOVA (or regression with two coded categorical variables and their interaction, same thing), the main effects being tested are different than the raw data marginal means, the same as a multiple regression including quantitative variables will test a regression weight that is not the same as the bivariate correlation between a variable and the criterion!

The overall or main effect for Review Attendance is:

Review > No Review

However, we know from the pattern of the interaction, that this is not descriptive for those in the Easier Practice condition.

This main effect must be communicated carefully, because it is potentially misleading.

Describing the Main Effect of Practice Difficulty

/ emmaans tables (pg1e2h3s) compare (pg1e2h3s)

Estimates

Dependent Variable: testperf

pg1e2h3s	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Easier	52.833	2.730	47.323	58.343
Harder	61.333	2.730	55.823	66.843
Same	70.000	2.730	64.490	75.510

As with the other main effect, you should notice that the means shown here are not the same as the marginal means from the "Descriptive Statistics" above (which were 50.6 for Easier, 67.5 for Same and 66.3 for Harder).

The pairwise comparisons show the pattern of the main effect of Practice Difficulty to be:

Easier < Harder < Same

Univariate Tests

Dependent Variable: testperf

	Sum of Squares	df	Mean Square	F	Sig.
Contrast	2210.278	2	1105.139	9.883	.000
Error	4696.667	42	111.825		

The F tests the effect of pg1e2h3s. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

Pairwise Comparisons

Dependent Variable: testperf

(I) pg1e2h3s	(J) pg1e2h3s	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
Easier	Harder	-8.500*	3.861	.033	-16.293	-.707
	Same	-17.167*	3.861	.000	-24.959	-9.374
Harder	Easier	8.500*	3.861	.033	.707	16.293
	Same	-8.667*	3.861	.030	-16.459	-.874
Same	Easier	17.167*	3.861	.000	9.374	24.959
	Harder	8.667*	3.861	.030	.874	16.459

Based on estimated marginal means

*. The mean difference is significant at the .050 level.

b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

However, we know from the pattern of the interaction, that this is not descriptive, either those to attended the review or for those who did not attend the review.

This main effect must be communicated carefully, because it is potentially misleading.

Data Preparation for Regression Analysis

Here's the SPSS syntax code to dummy code the binary grouping variable, to dummy code the 3-category variable and to compute the interaction term.

```
* pract_dc1 compares same=1=>0 with easier = 2 => 1.
if (practgrp = 1) pract_dc1 = 0.
if (practgrp = 2) pract_dc1 = 1.
if (practgrp = 3) pract_dc1 = 0.

*pract_dc2 compare same=1=>0 with harder=3=>1.
if (practgrp = 1) pract_dc2 = 0.
if (practgrp = 2) pract_dc2 = 0.
if (practgrp = 3) pract_dc2 = 1.

* atndrev_dc  no=1=>0   yes=2=>1.
if (atndrev = 1) atndrev_dc = 0.
if (atndrev =2) atndrev_dc = 1.

compute pract_rev_int1 = pract_dc1 * atndrev_dc.
compute pract_rev_int2 = pract_dc2 * atndrev_dc.

exe.
```

IF statements to dummy-code the group variable:

- same is going to be the comparison group, so it is coded "0" for both dummy codes
- dc1 is going to compare easier with same, so easier is coded "1" as the target group & same is coded "0" (harder is also coded "0")
- dc2 is going to compare harder with same, so harder is coded as "1" as the target group & same is codec "0" (easier is also coded "0")

IF statements to dummy-code the binary variable: "yes" is coded "1" as the target group and "no" is coded "0" as the comparison group

The products of each of the dummy codes from the 3-category variable with the dummy coded binary variable are the interaction terms

```
regression
  /statistics coeff r anova
  /dependent testperf
  /method = enter   pract_dc1  pract_dc2 atndrev_dc  pract_rev_int1 pract_rev_int2.
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.840[a]	.706	.671	10.57475

a. Predictors: (Constant), pract_rev_int2, pract_rev_int1, pract_dc2, pract_dc1, atndrev_dc

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11301.250	5	2260.250	20.212	.000[a]
	Residual	4696.667	42	111.825		
	Total	15997.917	47			

a. Predictors: (Constant), pract_rev_int2, pract_rev_int1, pract_dc2, pract_dc1, atndrev_dc
b. Dependent Variable: testperf

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	60.000	3.344		17.942	.000
	pract_dc1	1.667	5.461	.043	.305	.762
	pract_dc2	-18.333	5.461	-.473	-3.357	.002
	atndrev_dc	20.000	5.461	.546	3.662	.001
	pract_rev_int1	-37.667	7.723	-.838	-4.877	.000
	pract_rev_int2	19.333	7.723	.430	2.503	.016

a. Dependent Variable: testperf

The R2, F-test and regression weights are all the same as from the GLM analysis.