Complex Regression Models with Interactions

We decided to continue our study of the relationships among amount and difficulty of exam practice with exam performance in the first graduate research methods/data analysis course by including the program Psychology graduate students were in (1=experimental 2=developmental and 3=clinical programs), their future employment intentions (1=quantitative, 2=research), the number of stats courses they had taken before the current one, and a measure of academic performance motivation.

The univaraite stats for our quantitative predictors is shown at the right.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
prac	143	1.00	10.00	5.8182	2.23807
pristats	143	.00	5.00	2.3986	1.04234
motv	143	24.00	81.00	51.0629	12.10530
Valid N (listwise)	143				

Based on literature reviews and pilot studies, we chose to explore certain nonlinear and interaction effects in the model. The variable preparations for the regression analysis are shown below

```
*mean-centering quant variables.
compute prac_mcen = prac - 5.8182.
compute pristat_mcen = pristats - 2.3986.
compute motv_mcen = motv - 51.629.

*computing quadratic terms for quant variables.
compute prac_mcquad = (prac - 5.8182) ** 2.
compute pristat_mcquad = (pristats - 2.3986) ** 2.
compute motv_mcquad = (motv - 51.0629) ** 2.

*dummy code for job program.
if (prog_1exp_2dev_3clin = 1) prog_1exp_0dev_0clin = 1.
if (prog_1exp_2dev_3clin = 2) prog_1exp_0dev_0clin = 0.
if (prog_1exp_2dev_3clin = 3) prog_1exp_0dev_0clin = 0.

if (prog_1exp_2dev_3clin = 1) prog_0exp_1dev_0clin = 0.
if (prog_1exp_2dev_3clin = 2) prog_0exp_1dev_0clin = 1.
if (prog_1exp_2dev_3clin = 3) prog_0exp_1dev_0clin = 0.

*dummy code for job interest.
if (jobint1qnt_2rsh = 1) jobint1qnt0rsh = 1.
if (jobint1qnt_2rsh = 2) jobint1qnt0rsh = 0.

*dummy code for practice difficulty.
if (prac1e2s = 1) prac1e0s=1.
if (prac1e2s = 2) prac1e0s=0.

*code for job interest X practice difficulty interaction.
compute jobint_practdif_int = jobint1qnt0rsh * prac1e0s.

*practice X motivation interactions.
compute prac_motv_linlinint = prac_mcen * motv_mcen.
compute prac_motv_quadlinint = prac_mcquad * motv_mcen.
compute prac_motv_linquadint = prac_mcen * motv_mcquad.
compute prac_motv_quadquadint = prac_mcquad * motv_mcquad.

*practice difficulty X #practices interations.
compute  practdif_linprac_int = prac1e0s * prac_mcen.
compute   practdif_quadprac_int = prac1e0s * prac_mcquad.
exe.
```

Clinical is comparison group

1st code compares experimental to clinical
2nd code compares developmental to clinical

Research is the comparison group

Interaction between dummy coded binary variables

The “full set” of interactions between two quantitative variables

Linear and quadratic interactions between a binary and a quantitative variable

```
REGRESSION
  /DEPENDENT testperfc
  /METHOD=ENTER
    prac_mcen prac_mcquad                                    ←
    motv_mcen motv_mcquad                                    ← centered quantitative variables & quadratic terms
    pristat_mcen pristat_mcquad                              ←
      prog_1exp_0dev_0clin prog_0exp_1dev_0clin              ←
      jobint1qnt0rsh                                         ← dummy-coded 3-group and binary variables
      prac1e0s                                               ←
    jobint_practdif_int                                      ← interaction of two quantitative variables
      prac_motv_linlinint prac_motv_quadlinint               ←
      prac_motv_linquadint prac_motv_quadquadint             ← linear & quadratic interactions of 2 quantitative variables
    practdif_linprac_int practdif_quadprac_int.              ← linear and quadratic interactions of binary and
                                                               quantitative variable
```

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.867[a]	.752	.719	10.08879

a. Predictors: (Constant), practdif_quadprac_int, pristat_mcquad, prac_motv_linlinint, prog_0exp_1dev_0clin, jobint1qnt0rsh, pristat_mcen, practdif_linprac_int, motv_mcen, motv_mcquad, prac_motv_linquadint, prac1e0s, prog_1exp_0dev_0clin, prac_mcquad, prac_motv_quadlinint, jobint_practdif_int, prac_mcen, prac_motv_quadquadint

The model accounts for nearly 75% of the variance of exam performance, which is statistically significant.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	38624.896	17	2272.053	22.322	.000[b]
	Residual	12722.970	125	101.784		
	Total	51347.866	142			

a. Dependent Variable: testperfc

b. Predictors: (Constant), practdif_quadprac_int, pristat_mcquad, prac_motv_linlinint, prog_0exp_1dev_0clin, jobint1qnt0rsh, pristat_mcen, practdif_linprac_int, motv_mcen, motv_mcquad, prac_motv_linquadint, prac1e0s, prog_1exp_0dev_0clin, prac_mcquad, prac_motv_quadlinint, jobint_practdif_int, prac_mcen, prac_motv_quadquadint

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	86.292	2.860		30.168	.000
	prac_mcen	2.994	.731	.352	4.093	.000
	prac_mcquad	-.311	.280	-.095	-1.109	.269
	motv_mcen	.713	.105	.454	6.774	.000
	motv_mcquad	-.039	.007	-.369	-5.630	.000
	pristat_mcen	-1.434	.869	-.079	-1.650	.101
	pristat_mcquad	-.588	.653	-.042	-.900	.370
	prog_1exp_0dev_0clin	-2.000	2.625	-.047	-.762	.448
	prog_0exp_1dev_0clin	-5.750	2.135	-.149	-2.693	.008
	jobint1qnt0rsh	10.745	2.385	.278	4.505	.000
	prac1e0s	-20.481	3.122	-.535	-6.560	.000
	jobint_practdif_int	10.808	3.566	.243	3.031	.003
	prac_motv_linlinint	.017	.034	.027	.510	.611
	prac_motv_quadlinint	.003	.013	.018	.241	.810
	prac_motv_linquadint	-.005	.002	-.176	-2.448	.016
	prac_motv_quadquadint	.001	.001	.093	1.081	.282
	practdif_linprac_int	-4.859	.832	-.388	-5.840	.000
	practdif_quadprac_int	-.159	.331	-.040	-.481	.631

a. Dependent Variable: testperfc

Interpreting the multiple regression weights

As we tour these interpretations, remember because of the coding and centering we used, the "comparison group" is clinical students having a research interest using the similar difficulty practices, and who had the average amount of practice, the average amount of motivation and the average number of prior stats courses. Also, you have to be careful about which effects can be generalized to other groups, depending on whether or not they are involved in an interaction.

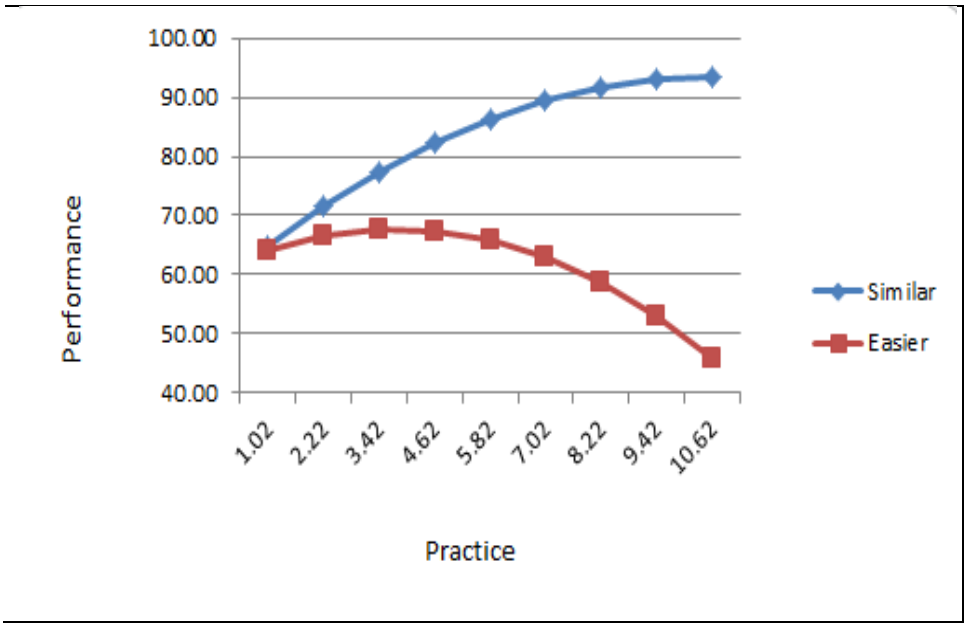
prac_mcen	Each practice is expected to increase performance by 2.994%, for clinical students with a research interest using the similar difficulty practices, and who had average motivation & number of prior stats courses.
prac_mcquad	There is no quadratic component to the relationship between practice and performance, for clinical students with a research interest using the similar difficulty practices, and who had average motivation & number of prior stats courses.
motv_mcen	Each 1-unit increase in motivation score is expected to increase performance by .713%, for clinical students with a research interest using the similar difficulty practices, and who had the average amount of practice and the average number of prior stats courses.
motv_mcquad	There is an inverted-U-shaped quadratic component to the relationship between motivation and test performance, for clinical students with a research interest using the similar difficulty practices, and who had the average amount of practice and the average number of prior stats courses.
pristat_mcen	There is no relationship between number of prior stats courses taken and test performance, for clinical students with a research interest using the similar difficulty practices, and who had the average amount of practice and the average motivation.
pristat_mcquad	There is no quadratic component to the relationship between motivation and performance, for clinical students with a research interest using the similar difficulty practices, and who had the average amount of practice and the average motivation.
prog_1exp_0dev_0clin	There is no performance difference between clinical and experimental students with a research interest using the similar difficulty practices, and who had the average amount of practice, the average number of prior stats courses, and the average motivation.
prog_0exp_1dev_0clin	Developmental students performed 5.75% poorer than clinical students with a research interest using the similar difficulty practices, and who had the average amount of practice, the average number of prior stats courses, and the average motivation.
jobint1qnt0rsh	Those interested in a quant job have an expected score 10.745 higher than those interested in a research job, for clinical students using the similar difficulty practices, and who had the average amount of practice, the average amount of motivation and the average number of prior stats courses.
prac1e0s	Those who used the easier practice scores 20.481 point lower than those who used the same difficulty practice, among clinical students having a research interest and who had the average amount of practice, the average amount of motivation and the average number of prior stats courses.
jobint_practdif_int	Those interested in a quant job scores 10.808 higher than those interested in a research job, among clinical students who had the average amount of practice, the average amount of motivation and the average number of prior stats courses.
prac_motv_linlinint	There is no linear interaction of practice and motivation for clinical students having a research interest using the similar difficulty practices, and who had the average amount of practice, the average amount of motivation and the average number of prior stats courses.
prac_motv_quadlinint	There is no quadratic practice by linear motivation interaction for clinical students having a research interest using the similar difficulty practices, and the average number of prior stats courses.
prac_motv_linquadint	There is a linear practice by quadratic motivation interaction for clinical students having a research interest using the similar difficulty practices, and the average number of prior stats courses. (I wouldn't work too hard to articulate the pattern of this comples interaction. I'd show them the plot – see below!)
prac_motv_quadquadint	There is no quadratic practice by quadratic motivation interaction for clinical students having a research interest using the similar difficulty practices, and the average number of prior stats courses.
practdif_linprac_int	The slope of the performance-practice regression line is 4.859 less for those in the easier practice condition, for clinical students having a research, the average amount of motivation and the average number of prior stats courses.
practdif_quadprac_int.	The shape of the performance practice regression line is the sme for those in the easier and same difficulty practice conditions, for clinical students having a research, the average amount of motivation and the average number of prior stats courses.

With models this complex, plotting specific nonlinear and interaction effects can greatly enhance the interpretation of the regression weights. Here are some additional details that further elaborate and describe the model!

The interaction of #Practices & Practice difficulty was of particular interest in this analysis. Using the "2xQ nonlinear" tab of the plotting computator, we obtained the following.

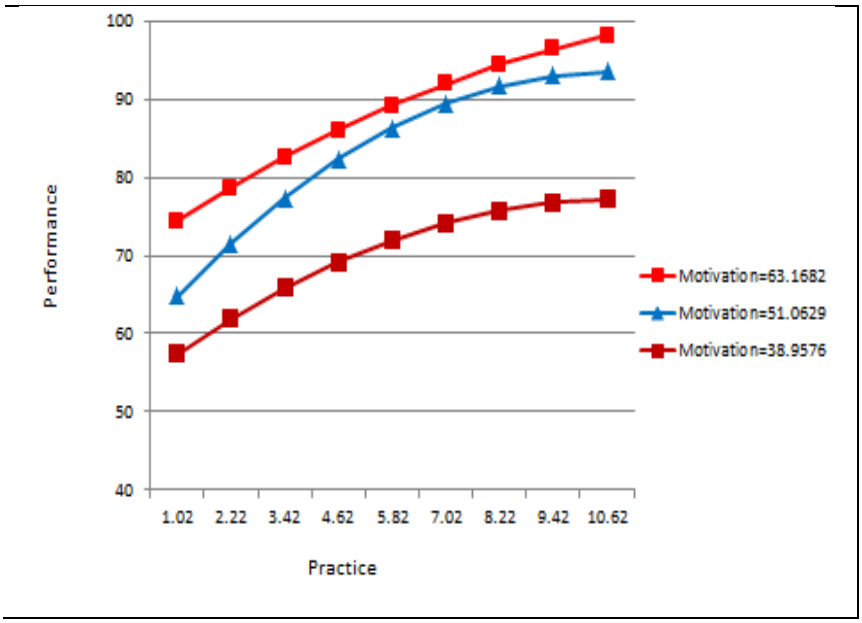
Performance was similar after 1 trial, but diverged sharply from there! The performance difference between the groups increased with each additional practice. Practice led to continual improvement for the Similar group, with performance asymptote apparent at around 9 practices. Practice led to an initial small performance increase, but after 4 practices performance decreased with each additional practice.

height z=0	constant	86.292
slope z=0	b(x)	2.994
curve z=0	b(x^2)	-0.311
height dif z=1	b(z)	-20.481
slope dif z=1	b(xz)	-4.859
curve dif z=1	b(x^2z)	-0.159
	x(mean)	5.8182
	x(std)	2.3986



The complex interaction between #Practices and Motivation is also easier to see when plotted. Using the "QxQ nonlinear" tab we obtained the following. Additional practice continued to lead to improved performance for all motivational levels, but while the relationship between motivation and performance was nearly linear for low amounts of practice, at higher amounts of practice, those with average levels of motivation performed similar to those who were highly motivated.

height x1-=mean x2=mean	constant	86.292
y-x1 slope for x2=mean	b(x1)	2.994
y-x1 curve for x2=mean	b($x1^2$)	-0.311
y-x1 linear height dif for dif x2	b(x2)	0.713
y-x1 nonlinear height dif for dif x2	b($x2^2$)	-0.039
y-x1 linear slope dif for dif x2	b(x1x2)	0.017
y-x1 linear curve dif for dif x2 values	b($x1^2$x2)	0.003
y-x1 nonlinear slope dif for dif x2 values	b(x1$x2^2$)	-0.005
y-x1 nonlinear curve dif for dif x2 values	b($x1^2x2^2$)	0.001
	x1(mean)	5.8182
	x1(std)	2.3986
	x2(mean)	51.0629
	x2(std)	12.1053



SPSS GLM Analysis

We obtained the same model, and a bit more info about it, using GLM! The important difference between running this model in multiple regression and in GLM is that we used dummy-coded categorical variables in multiple regression, but we will use the original categorical variables in the GLM and SPSS will do the coding for us. We will, however, still do the mean centering and compute the quadratic terms. We also have to construct the interaction terms within the Design subcommand!

```
UNIANOVA testperf                                          ← list the DV
            BY prac1e2s                                    ← list the categorical variables – SPSS will code
                 jobint1qnt_2rsh                              these with the highest valued group as the
                 prog_1exp_2dev_3clin                         comparison group
           WITH prac_mcen  pristat_mcen  motv_mcen          ← list the mean-centered quant variables and the
                  prac_mcquad pristat_mcquad motv_mcquad       quad terms

 /METHOD=SSTYPE(3)                                         ← asks for unique effects model (same as mreg)
 /PRINT = PARAMETER                                        ← gets the regression weights

 /PLOT=PROFILE(prac1e2s*jobint1qnt_2rsh)                   ← plot of practice difficulty X job interest
                                                             interaction

 /EMMEANS TABLES ( jobint1qnt_2rsh by prac1e2s)
            COMPARE (prac1e2s)                             ← gets the simple effect pairwise comparisons to
                                                             describe the difficulty X job interest interaction

 /EMMEANS TABLES ( prog_1exp_2dev_3clin)
            COMPARE (prog_1exp_2dev_3clin)                 ← gets the corrected/expected means and
                                                             comparisons among the program groups
 /DESIGN=                                                  ← specifies the model – notice that the
  prac_mcen prac_mcquad                                      interactions are "built from" the main effect terms
  motv_mcen motv_mcquad
  pristat_mcen pristat_mcquad
  prog_1exp_2dev_3clin
  jobint1qnt_2rsh
  prac1e2s
  jobint1qnt_2rsh*prac1e2s
  motv_mcen*prac_mcen  motv_mcen*prac_mcquad
  motv_mcquad*prac_mcen  motv_mcquad*prac_mcquad
  prac1e2s*prac_mcen prac1e2s*prac_mcquad.
```

Tests of Between-Subjects Effects

Dependent Variable: testperf

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	38624.896[a]	17	2272.053	22.322	.000
Intercept	225531.743	1	225531.743	2215.793	.000
prac_mcen	113.794	1	113.794	1.118	.292
prac_mcquad	382.139	1	382.139	3.754	.055
motv_mcen	4671.171	1	4671.171	45.893	.000
motv_mcquad	3226.132	1	3226.132	31.696	.000
pristat_mcen	277.154	1	277.154	2.723	.101
pristat_mcquad	82.395	1	82.395	.810	.370
prog_1exp_2dev_3clin	787.453	2	393.727	3.868	.023
jobint1qnt_2rsh	8097.357	1	8097.357	79.555	.000
prac1e2s	3774.517	1	3774.517	37.084	.000
prac1e2s * jobint1qnt_2rsh	934.942	1	934.942	9.186	.003
prac_mcen * motv_mcen	26.518	1	26.518	.261	.611
motv_mcen * prac_mcquad	5.921	1	5.921	.058	.810
prac_mcen * motv_mcquad	609.748	1	609.748	5.991	.016
prac_mcquad * motv_mcquad	119.044	1	119.044	1.170	.282
prac1e2s * prac_mcen	3470.827	1	3470.827	34.100	.000
prac1e2s * prac_mcquad	23.576	1	23.576	.232	.631
Error	12722.970	125	101.784		
Total	847217.828	143			
Corrected Total	51347.866	142			

a. R Squared = .752 (Adjusted R Squared = .719)

The F-tests in the ANOVA table parallel the t-tests of the regression weights, except for the career interest variable, which is expressed as a 3-group comparison in the F-tests and dummy code-pairwise comparisons in the t-tests.

The regression weights are the same values and interpretations as were obtained from the multiple regression model earlier.

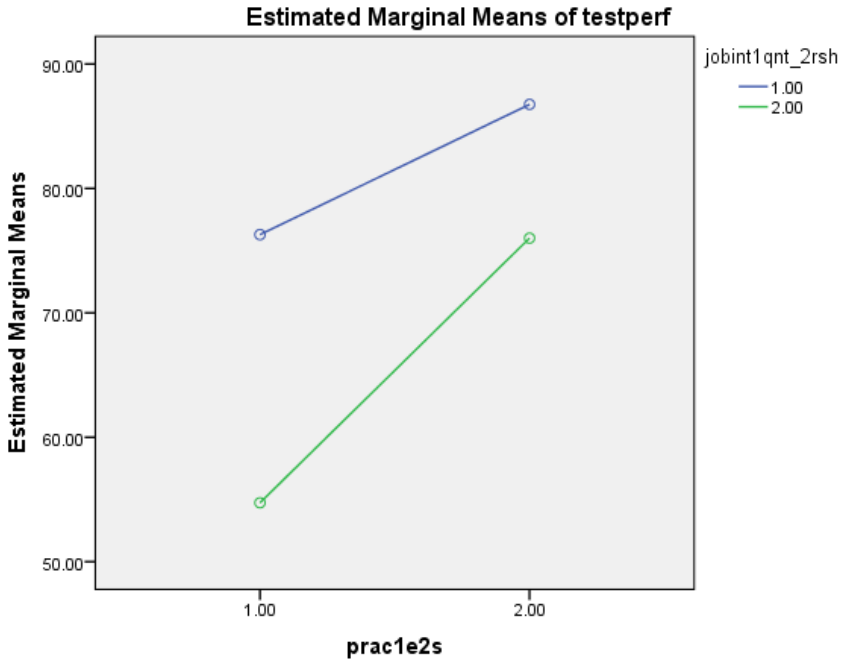
Parameter Estimates

Dependent Variable: testperf

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	86.292	2.860	30.168	.000	80.630	91.953
prac_mcen	2.994	.731	4.093	.000	1.546	4.441
prac_mcquad	-.311	.280	-1.109	.269	-.866	.244
motv_mcen	.713	.105	6.774	.000	.505	.922
motv_mcquad	-.039	.007	-5.630	.000	-.053	-.025
pristat_mcen	-1.434	.869	-1.650	.101	-3.154	.286
pristat_mcquad	-.588	.653	-.900	.370	-1.881	.705
[prog_1exp_2dev_3clin=1.00]	-2.000	2.625	-.762	.448	-7.196	3.196
[prog_1exp_2dev_3clin=2.00]	-5.750	2.135	-2.693	.008	-9.976	-1.524
[prog_1exp_2dev_3clin=3.00]	0[a]	.	.	.	.	.
[jobint1qnt_2rsh=1.00]	10.745	2.385	4.505	.000	6.025	15.466
[jobint1qnt_2rsh=2.00]	0[a]	.	.	.	.	.
[prac1e2s=1.00]	-20.481	3.122	-6.560	.000	-26.660	-14.302
[prac1e2s=2.00]	0[a]	.	.	.	.	.
[prac1e2s=1.00] * [jobint1qnt_2rsh=1.00]	10.808	3.566	3.031	.003	3.750	17.866
[prac1e2s=1.00] * [jobint1qnt_2rsh=2.00]	0[a]	.	.	.	.	.
[prac1e2s=2.00] * [jobint1qnt_2rsh=1.00]	0[a]	.	.	.	.	.
[prac1e2s=2.00] * [jobint1qnt_2rsh=2.00]	0[a]	.	.	.	.	.
prac_mcen * motv_mcen	.017	.034	.510	.611	-.050	.084
motv_mcen * prac_mcquad	.003	.013	.241	.810	-.023	.029
prac_mcen * motv_mcquad	-.005	.002	-2.448	.016	-.010	-.001
prac_mcquad * motv_mcquad	.001	.001	1.081	.282	-.001	.002
[prac1e2s=1.00] * prac_mcen	-4.859	.832	-5.840	.000	-6.505	-3.212
[prac1e2s=2.00] * prac_mcen	0[a]	.	.	.	.	.
[prac1e2s=1.00] * prac_mcquad	-.159	.331	-.481	.631	-.814	.496
[prac1e2s=2.00] * prac_mcquad	0[a]	.	.	.	.	.

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

One advantage of using GLM is that it give more complete information about the categorical variables than does he multiple regression, especially for interaction patterns. Plus, GLM will allow you to get plots of the cell means representing the interactions of categorical variables.



Covariates appearing in the model are evaluated at the following values: prac_mcen = .0000, pristat_mcen = .0000, motv_mcen = .0000, prac_mcquad = 4.9739, pristat_mcquad = 1.0789, motv_mcquad = 145.5135

1. jobint1qnt_2rsh * prac1e2s

Estimates

Dependent Variable: testperf

jobint1qnt_2rsh	prac1e2s	Mean	Std. Error
1.00	1.00	76.286[a]	1.924
	2.00	86.751[a]	1.523
2.00	1.00	54.733[a]	2.052
	2.00	76.006[a]	1.918

a. Covariates appearing in the model are evaluated at the following values: prac_mcen = .0000, pristat_mcen = .0000, motv_mcen = .0000, prac_mcquad = 4.9739, pristat_mcquad = 1.0789, motv_mcquad = 145.5135.

Pairwise Comparisons

Dependent Variable: testperf

jobint1qnt_2rsh	(I) prac1e2s	(J) prac1e2s	Mean Difference (I-J)	Std. Error	Sig.[b]
1.00	1.00	2.00	-10.465*	2.498	.000
	2.00	1.00	10.465*	2.498	.000
2.00	1.00	2.00	-21.273*	2.844	.000
	2.00	1.00	21.273*	2.844	.000

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 plot and the pairwise comparisons both show that people consistently performed better when using the similar difficulty practices than the easier practices, and this smaller for those with a quantitative interest than those with a research interest.

Notice that those with a research interest who used the similar difficulty practices performed similarly to those with a quantitative interest who used the easier practices!

Pairwise comparisons also provide useful information about hos those in different programs differed.

There were no interactions with program, so these are “descriptive” results!

2. prog_1exp_2dev_3clin

Estimates

Dependent Variable: testperf

prog_1exp_2dev_3clin	Mean	Std. Error
1.00	74.027[a]	1.960
2.00	70.278[a]	1.399
3.00	76.028[a]	1.595

a. Covariates appearing in the model are evaluated at the following values: prac_mcen = .0000, pristat_mcen = .0000, motv_mcen = .0000, prac_mcquad = 4.9739, pristat_mcquad = 1.0789, motv_mcquad = 145.5135.

Pairwise Comparisons

Dependent Variable: testperf

(I) prog_1exp_2dev_3clin	(J) prog_1exp_2dev_3clin	Mean Difference (I-J)	Std. Error	Sig.[b]
1.00	2.00	3.750	2.404	.121
	3.00	-2.000	2.625	.448
2.00	1.00	-3.750	2.404	.121
	3.00	-5.750*	2.135	.008
3.00	1.00	2.000	2.625	.448
	2.00	5.750*	2.135	.008

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).