QxQ Models: Nonlinear Model Using Regression

Each year any of the city's firefighters who would like can take the Captain's Exam. The same procedure is used each year. Four weeks before the date of the Captain's Exam everybody who plans to take the Exam attends a session during which they take a pretest and receive their score. Copies of "study questions" (5 questions in a bundle) are available to be checked out from the public libraries (which is how we monitored practice) during the four weeks leading up to the Captain's exam. After several analyses revealed strong relationships between practicing and passing the Exam, we decided to explore what predicts practicing for the Exam. We chose to examine how pretest performance and prior experience (how many times they'd taken the Captain's Exam previously) were related to the number of items studied.

Data Preparation

To maximize the interpretability of the regression weights, we mean-centered both of the predictors, prior experience and initial performance. We computed quadratic terms for each predictor, as the square of the mean-centered values. We also calculated a linear interaction term and all three of the possible quadratic interaction terms.

Descriptive Statistics

	N	Mean	Std. Deviation
priorexp	30	6.3667	3.18924
initperf	30	70.4972	10.28455
Valid N (listwise)	30		

```
* compute the centered quant predictors.
compute   priorexp_c =  priorexp   -   6.3667.         ← mean center prior experience
compute   initperf_c  =  initperf    -   70.4972.      ← mean center initial performance (Pretest)

* compute quadratic terms.
compute   pexpc_sq  =  priorexp_c ** 2.                ← square mean-centered prior experience
compute   initc_sq     =  initperf_c ** 2.             ← square mean-centered initial performance

* compute interactions.
compute   priorexpc_initperfc_int    =  priorexp_c  *   initperf_c.   ← linear priorexp / linear initial performance
compute   pexpcsq_initperfc_int     =  pexpc_sq    *   initperf_c.    ← quad priorexp / linear initial performance
compute   priorexpc_initcsq_int      =  priorexp_c   *   initc_sq.    ← linear priorexp / quad initial performance
compute   pexpcsq_initcsq_int       =  pexpc_sq     *   initc_sq.     ← quad priorexp / quad initial performance

exe.

*hierarchical regression model – entering linear terms and
  then adding quadratic terms.
regression
  /statistics   coeff  r  anova  change            ← request weights, fit, model tests & R²Δ
  /dependent      numpract5                         ← specify criterion variable
  /method = enter       priorexp_c   initperf_c     ← specify variables to enter into 1st step
                        priorexpc_initperfc_int
  /method = enter       pexpc_sq    initc_sq        ← specify variables to add as 2nd step
                        pexpcsq_initperfc_int
                        priorexpc_initcsq_int
                        pexpcsq_initcsq_int.
```

Analysis Results

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.676[a]	.456	.394	3.29469	.456	7.275	3	26	.001
2	.827[b]	.683	.563	2.79764	.227	3.012	5	21	.033

a. Predictors: (Constant), priorexpc_initperfc_int, initperf_c, priorexp_c

b. Predictors: (Constant), priorexpc_initperfc_int, initperf_c, priorexp_c, initc_sq, pexpc_sq, priorexpc_initcsq_int, pexpcsq_initperfc_int, pexpcsq_initcsq_int

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	236.911	3	78.970	7.275	.001[b]
	Residual	282.229	26	10.855		
	Total	519.139	29			
2	Regression	354.777	8	44.347	5.666	.001[c]
	Residual	164.363	21	7.827		
	Total	519.139	29			

a. Dependent Variable: numpract5

b. Predictors: (Constant), priorexpc_initperfc_int, initperf_c, priorexp_c

c. Predictors: (Constant), priorexpc_initperfc_int, initperf_c, priorexp_c, initc_sq, pexpc_sq, priorexpc_initcsq_int, pexpcsq_initperfc_int, pexpcsq_initcsq_int

The linear model accounts for a significant 46% of the variance.

Adding in the quadratic main effect and interaction terms increased the variance accounted for by about 23%.

Both the quadratic model and the increase in R^2 are statistically significant.

Entering Quadratic & Interaction into Your Models

Several other hierarchical modeling approaches could be applied here. It is important for you to know which modeling approach is "expected" in your research area (not that you have to comply if you have a reasonable alternative, but you don't want your audience's discontent to be a surprise in the middle of a presentation or review process!).

Remember that the linear, quadratic, linear interaction and quadratic interaction terms are likely to be collinear! So, which regression weights (effects) are significant in a model can strongly depend upon which other effects are included in the model. In particular, quadratic terms (products of linear terms) and interaction terms (other products of linear terms and quadratic terms) can be highly collinear.

So, these different hierarchical entry orders can produce interestingly different results in the intermediate steps. Some adopt the strategy that if, at a particular step in the hierarchical sequence, the additional terms don't increase the variance significantly, then further models are not examined! So, order can really matter & many steps with fewer terms entered on each step is more likely to lead you to stop before finding all the "contributors". Please, remember that once the full model is in place it will be the same no matter what were the intermediate steps!! So, it is good advice to look at the

```
*linear effects then quadratic effects.
  /method = enter    priorexp_c   initperf_c
                     priorexpc_initperfc_int
  /method = enter    pexpc_sq    initc_sq
                     pexpcsq_initperfc_int
                     priorexpc_initcsq_int
                     pexpcsq_initcsq_int.
```

```
*linear main – quadratic main – linear interaction – quadratic int.
/method = enter    priorexp_c   initperf_c
/method = enter    pexpc_sq    initc_sq
/method = enter    priorexpc_initperfc_int
/method = enter    pexpcsq_initperfc_int
                   priorexpc_initcsq_int
                   pexpcsq_initcsq.
```

```
*main effects then interactions.
 /method = enter   priorexp_c   initperf_c
                   pexpc_sq    initc_sq
  /method = enter  priorexpc_initperfc_int
                   pexpcsq_initperfc_int
                   priorexpc_initcsq_int
                   pexpcsq_initcsq.
```

```
*linear main – linear interaction - quadratic main — quadratic int.
/method = enter    priorexp_c   initperf_c
/method = enter    priorexpc_initperfc_int
/method = enter    pexpc_sq    initc_sq
/method = enter    pexpcsq_initperfc_int
                   priorexpc_initcsq_int
                   pexpcsq_initcsq.
```

Regression weights

Since the quadratic model significantly increased the fit of the model, we'll concentrate on it.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	12.781	.638		20.029	.000
	priorexp_c	-.367	.219	-.277	-1.678	.105
	initperf_c	-.066	.064	-.161	-1.040	.308
	priorexpc_initperfc_int	.056	.021	.428	2.623	.014
2	(Constant)	10.956	.913		11.998	.000
	priorexp_c	-.904	.257	-.682	-3.517	.002
	initperf_c	-.073	.077	-.178	-.957	.349
	priorexpc_initperfc_int	-.026	.033	-.201	-.811	.426
	pexpc_sq	.188	.069	.543	2.736	.012
	initc_sq	.003	.006	.081	.445	.661
	pexpcsq_initperfc_int	.017	.011	.712	1.565	.132
	priorexpc_initcsq_int	.002	.003	.314	.863	.398
	pexpcsq_initcsq_int	.001	.001	1.023	2.242	.036

a. Dependent Variable: numpract5

Remember → with interaction terms included in the model, the regression weights are all "simple efffects" → telling the relationship between that variable and the criterion when controlling all other variables in the model at "0"!!

Prior experience - linear — There is a significant negative relationship between prior experience and number of practices for those who had an average (mean=0) pretest score.

Prior experience- quad — There is a positive quadratic (u-shaped) component to the relationship as well

Initial performance - linear — There is no linear relationship between initial performance and number of practices for those who had the average (mean=0) prior experience

Initial performance - quad — There is a no quadratic (u-shaped) component to the relationship either

Linear interaction
- The linear relationship between prior experience and number of practices does not have a linear difference across different amounts of initial performance
- The linear relationship between initial performance and number of practices does not have a linear difference across different amounts of prior experience

Quadratic prior experience – linear initial performance interaction
- The quadratic relationship between prior experience and number of practices does not have a linear difference across different amounts of initial performance
- The linear relationship between initial performance and number of practices does not have a quadratic difference across different amounts of prior experience

Linear prior experience – quadratic initial performance interaction
- The linear relationship between prior experience and number of practices does not have a quadratic difference across different amounts of initial performance
- The quadratic relationship between initial performance and number of practices does not have a linear difference across different amounts of prior experience

Quadratic prior experience – quadratic initial performance interaction
- The quadratic relationship between prior experience and number of practices has a u-shaped quadratic difference across different amounts of initial performance
- The quadratic relationship between initial performance and number of practices has a u-shaped quadratic difference across different amounts of prior experience

QxQ Using GLM

It is possible to get this model via GLM. All the terms (mean-centered, quadratic & interaction) are entered as "Covariates" – there is no "BY" statement, because there are no categorical variables in the model.

The regression weights will be exactly the same as from the regression model, and the F-tests will all be equivalent to the regression weight t-tests ($t^2 = F$).

The advantages to using the regression model are that you get the Beta weights for more direct comparison of the unique contribution of the various terms and you can test hierarchical models.

```
UNIANOVA numpract5 WITH priorexp_c initperf_c pexpc_sq initc_sq priorexpc_initperfc_int
                        pexpcsq_initperfc_int priorexpc_initcsq_int pexpcsq_initcsq_int
  /METHOD=SSTYPE(3)
  /PRINT=PARAMETER
  /DESIGN=priorexp_c initperf_c pexpc_sq initc_sq priorexpc_initperfc_int
          pexpcsq_initperfc_int priorexpc_initcsq_int pexpcsq_initcsq_int.
```

Tests of Between-Subjects Effects

Dependent Variable: numpract5

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	354.777[a]	8	44.347	5.666	.001
Intercept	1126.747	1	1126.747	143.960	.000
priorexp_c	96.810	1	96.810	12.369	.002
initperf_c	7.174	1	7.174	.917	.349
pexpc_sq	58.583	1	58.583	7.485	.012
initc_sq	1.548	1	1.548	.198	.661
priorexpc_initperfc_int	5.151	1	5.151	.658	.426
pexpcsq_initperfc_int	19.180	1	19.180	2.451	.132
priorexpc_initcsq_int	5.832	1	5.832	.745	.398
pexpcsq_initcsq_int	39.348	1	39.348	5.027	.036
Error	164.363	21	7.827		
Total	5857.807	30			
Corrected Total	519.139	29			

a. R Squared = .683 (Adjusted R Squared = .563)

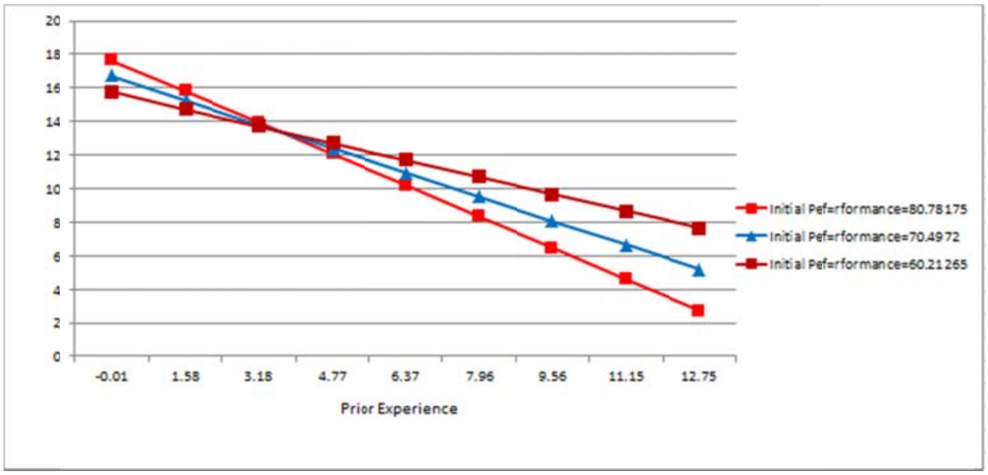
Parameter Estimates

Dependent Variable: numpract5

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	10.956	.913	11.998	.000	9.057	12.855
priorexp_c	-.904	.257	-3.517	.002	-1.439	-.370
initperf_c	-.073	.077	-.957	.349	-.233	.086
pexpc_sq	.188	.069	2.736	.012	.045	.332
initc_sq	.003	.006	.445	.661	-.010	.015
priorexpc_initperfc_int	-.026	.033	-.811	.426	-.094	.041
pexpcsq_initperfc_int	.017	.011	1.565	.132	-.006	.039
priorexpc_initcsq_int	.002	.003	.863	.398	-.003	.008
pexpcsq_initcsq_int	.001	.001	2.242	.036	9.676E-005	.003

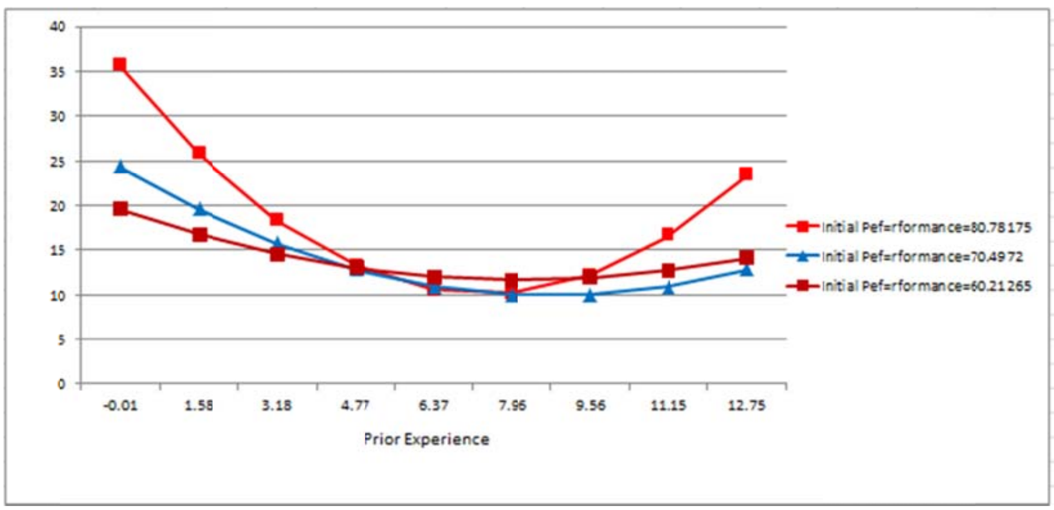
Plotting the Linear and Nonlinear Models

Remember that the linear model fit the data quite well – R^2 = 46%



However, the depiction of the model including the quadratic terms is strikingly different!

Notice, in particular, that the linear model suggests that the relationship between initial performance and number of practices is reversed from low to high prior experience! In the quadratic model, this relationship is different for those with "middle" and "low" initial performance, but those with high initial performance practice more at both ends of the prior experience continuum. This shift is the expression of the quadratic interaction – the shape of the practice-initial performance relationship changes at different values of prior experience.



A huge question was why the overall quadratic relationship between prior experience and # practices??? Further research revealed it was basically a "motivational thing". There was a stipend available for taking the pretest and the Captain's Test – so lots of folks took it each year.

- Those with low prior experience were "fast-trackers" who practiced more and show a strong relationship between how well thet pretested and how many times they studied.
- Those with intermediate prior experience would often admit they were spending the testing time to get the stipend or support friends, so they didn't study much.
- Those with high prior experience were reaching that "up or out" stage of their career, and only those with better pretest scores (though notice, generally lower than the "fast-trackers") invested in much practice.

We would never have explored this had we only the linear model depiction of these data. Changes in the stipend and practice procedures were implemented, resulting in a 15% increase in folks passing the test, and an increase of 22% in the performance ratings of the selected Captains!