Curvilinear Regression

Linear	Quadratic	Linear & Quadratic
$\boldsymbol{Y} = \beta_0 + \beta_1 \boldsymbol{X}_1 + \varepsilon$	$\boldsymbol{Y} = \beta_0 + \beta_2 \boldsymbol{X}_1^2 + \varepsilon$	$\boldsymbol{Y} = \beta_0 + \beta_1 \boldsymbol{X}_1 + \beta_2 \boldsymbol{X}_1^2 + \varepsilon$

Nonlinear Relationships

When regression analysis is applied with a continuous variable, the answer being sought is about the nature of the relationship between the independent and dependent variable. Thus, one may determine whether the regression of the dependent variable on the independent variable is linear or curvilinear. Moreover, when the relationship is curvilinear, one may determine its specific form.

Types of Curvilinear (Polynomial) Relationships: One Predictor

Polynomial regression is carried out as an ordinary regression analysis, except that powered terms are included and the analysis is done hierarchically (from the most complex form to the most simple form).

We usually do not try to model beyond a cubic relationship because interpretation is difficult. In fact, Pedhazur (1997) suggests testing for quadratic as highest term.

Interaction terms and higher-order terms that are based on quantitative variables should be centered. The original variables should be untransformed.

$X1$
$X2$
$X1X2 = \left(X1 - \overline{X1}\right)\left(X2 - \overline{X2}\right)$
$X1SQ = \left(X1 - \overline{X1}\right)^2$
$X2SQ = \left(X2 - \overline{X2}\right)^2$

Is there a cubic relationship between Y and X?

Test for a **cubic** relationship while adjusting for a possible quadratic or linear relationship by comparing two models:

Cubic: Attention = Caffeine Caffeine^2 Caffeine^3
Quadratic or Linear: Attention = Caffeine Caffeine^2

Use the cubic model if the test for the cubic form is significant.
Test for a quadratic model if the cubic model is not significant.

Is there a quadratic relationship between Y and X?

Test for a **quadratic** relationship while adjusting for a possible linear relationship by comparing two models:

Quadratic: Attention = Caffeine Caffeine^2
Linear: Attention = Caffeine

Use the quadratic model if the test for the quadratic form is significant.
Test for a linear model if the quadratic model is not significant.

Is there a linear relationship between Y and X?

Test for a **linear** relationship by comparing two models:

Linear: Attention = Caffeine
Intercept Only: Attention = Average of Attention

Use the linear model if the test for the linear form is significant.
Drop the predictor if the linear model is not significant.

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
STRESS	3.0115	.85582	87
EXPERIEN	6.5632	1.36144	87
PRODUCT	41.3678	8.97038	87

Correlations

		STRESS	EXPERIEN	PRODUCT
STRESS	Pearson Correlation	1	-.285**	-.046
	Sig. (2-tailed)	.	.007	.672
	N	87	87	87
EXPERIEN	Pearson Correlation	-.285**	1	.392**
	Sig. (2-tailed)	.007	.	.000
	N	87	87	87
PRODUCT	Pearson Correlation	-.046	.392**	1
	Sig. (2-tailed)	.672	.000	.
	N	87	87	87

**. Correlation is significant at the 0.01 level (2-tailed).

```
COMPUTE sq_stres = (stress-3.0115)**2 .
VARIABLE LABELS sq_stres 'squared stress' .
EXECUTE .
COMPUTE c_stress = (stress-3.0115)**3 .
VARIABLE LABELS c_stress 'Cubic Stress' .
EXECUTE .
```

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	STRESS[a]	.	Enter
2	squared stress[a]	.	Enter
3	Cubic Stress[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: PRODUCT

Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.046[a]	.002	-.010	9.01345
2	.093[b]	.009	-.015	9.03697
3	.164[c]	.027	-.008	9.00720

a. Predictors: (Constant), STRESS

b. Predictors: (Constant), STRESS, squared stress

c. Predictors: (Constant), STRESS, squared stress, Cubic Stress

d. Dependent Variable: PRODUCT

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14.641	1	14.641	.180	.672[a]
	Residual	6905.589	85	81.242		
	Total	6920.230	86			
2	Regression	60.219	2	30.109	.369	.693[b]
	Residual	6860.011	84	81.667		
	Total	6920.230	86			
3	Regression	186.467	3	62.156	.766	.516[c]
	Residual	6733.763	83	81.130		
	Total	6920.230	86			

a. Predictors: (Constant), STRESS

b. Predictors: (Constant), STRESS, squared stress

c. Predictors: (Constant), STRESS, squared stress, Cubic Stress

d. Dependent Variable: PRODUCT

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	42.820	3.554		12.048	.000	35.753	49.886
	STRESS	-.482	1.136	-.046	-.425	.672	-2.740	1.776
2	(Constant)	43.151	3.591		12.017	.000	36.010	50.292
	STRESS	-.809	1.220	-.077	-.663	.509	-3.235	1.617
	squared stress	.902	1.207	.087	.747	.457	-1.499	3.303
3	(Constant)	51.727	7.751		6.674	.000	36.311	67.144
	STRESS	-3.354	2.375	-.320	-1.412	.162	-8.077	1.370
	squared stress	-1.034	1.964	-.100	-.527	.600	-4.940	2.872
	Cubic Stress	1.860	1.491	.380	1.247	.216	-1.106	4.827

a. Dependent Variable: PRODUCT

```
COMPUTE sq_exper = (experien-6.5632)**2 .
EXECUTE .
COMPUTE c_exper = (experien-6.5632)**3 .
VARIABLE LABELS c_exper 'Cubic Experience' .
EXECUTE .
```

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	EXPERIEN[a]	.	Enter
2	SQ_EXPER[a]	.	Enter
3	Cubic Experience[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: PRODUCT

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.392[a]	.154	.144	8.29987
2	.408[b]	.166	.147	8.28687
3	.412[c]	.170	.140	8.32071

a. Predictors: (Constant), EXPERIEN

b. Predictors: (Constant), EXPERIEN, SQ_EXPER

c. Predictors: (Constant), EXPERIEN, SQ_EXPER, Cubic Experience

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1064.759	1	1064.759	15.456	.000[a]
	Residual	5855.471	85	68.888		
	Total	6920.230	86			
2	Regression	1151.765	2	575.882	8.386	.000[b]
	Residual	5768.465	84	68.672		
	Total	6920.230	86			
3	Regression	1173.787	3	391.262	5.651	.001[c]
	Residual	5746.443	83	69.234		
	Total	6920.230	86			

a. Predictors: (Constant), EXPERIEN

b. Predictors: (Constant), EXPERIEN, SQ_EXPER

c. Predictors: (Constant), EXPERIEN, SQ_EXPER, Cubic Experience

d. Dependent Variable: PRODUCT

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	24.405	4.405		5.540	.000	15.646	33.164
	EXPERIEN	2.585	.657	.392	3.931	.000	1.277	3.892
2	(Constant)	24.481	4.399		5.565	.000	15.733	33.229
	EXPERIEN	2.691	.663	.408	4.058	.000	1.372	4.009
	SQ_EXPER	-.421	.374	-.113	-1.126	.264	-1.165	.323
3	(Constant)	28.307	8.095		3.497	.001	12.206	44.407
	EXPERIEN	2.128	1.199	.323	1.775	.080	-.257	4.513
	SQ_EXPER	-.524	.418	-.141	-1.255	.213	-1.356	.307
	Cubic Experience	.118	.210	.109	.564	.574	-.299	.536

a. Dependent Variable: PRODUCT

Curvilinear Regression With Two Predictors

With two or more predictors, we must be concerned with:

- How is X1 related to Y (i.e., linear, quadratic, cubic)?
- How is X2 related to Y (i.e., linear, quadratic, cubic)?
- Do X1 and X2 interact with each other? That is, does the effect of X1 on Y change for various levels of X2?

Interaction is the condition where the relationship of interest is different at different levels (i.e., values) of the extraneous variable(s).

(Draw picture of interaction between Gender & Education for predicting income.)

Is it necessary to include some combination of quadratic or interaction terms in the model?

Test for a **nonlinear** relationship while adjusting for a possible linear relationship by comparing two models:

Nonlinear: Attention = Caffeine Sleep Caffeine^2 Sleep^2 Caffeine*Sleep
Linear: Attention = Caffeine Sleep

Use only linear forms of the predictors if the *F Change* test for nonlinearity is not significant.

Test for the individual importance of the quadratic and interaction terms if the *F Change* for nonlinearity is significant.

The individual importance of the quadratic and interaction terms is determined by the *t* tests for the nonlinear model.

- If all the *t* tests for the quadratic and interaction terms are significant, use the entire nonlinear model.
- If one or more of the *t* tests for the quadratic and interaction terms are not significant, eliminate the term that has the smallest *t* value and re-estimate the model. Continue eliminating the quadratic and interaction terms one-at-a-time and re-estimating the model until the remaining quadratic and/or interaction terms are significant.

```
COMPUTE expstres = (experien-6.5632)*(stress-3.0115) .
EXECUTE .
```

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	EXPERIEN, STRESS[a]	.	Enter
2	SQ_EXPER, squared stress, EXPSTRES, Cubic Experience, Cubic Stress[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: PRODUCT

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	.398[a]	.159	.139	8.32584	.159	7.915	2	84	.001
2	.553[b]	.306	.244	7.79719	.147	3.355	5	79	.008

a. Predictors: (Constant), EXPERIEN, STRESS

b. Predictors: (Constant), EXPERIEN, STRESS, SQ_EXPER, squared stress, EXPSTRES, Cubic Experience, Cubic Stress

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1097.388	2	548.694	7.915	.001[a]
	Residual	5822.842	84	69.320		
	Total	6920.230	86			
2	Regression	2117.330	7	302.476	4.975	.000[b]
	Residual	4802.900	79	60.796		
	Total	6920.230	86			

a. Predictors: (Constant), EXPERIEN, STRESS

b. Predictors: (Constant), EXPERIEN, STRESS, SQ_EXPER, squared stress, EXPSTRES, Cubic Experience, Cubic Stress

c. Dependent Variable: PRODUCT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	21.261	6.367		3.339	.001	8.600	33.921
	STRESS	.751	1.094	.072	.686	.495	-1.426	2.927
	EXPERIEN	2.719	.688	.413	3.952	.000	1.351	4.087
2	(Constant)	38.425	10.209		3.764	.000	18.104	58.746
	STRESS	-1.423	2.101	-.136	-.678	.500	-5.605	2.758
	EXPERIEN	1.454	1.184	.221	1.227	.223	-.904	3.811
	squared stress	-1.844	1.818	-.178	-1.014	.314	-5.463	1.775
	Cubic Stress	1.481	1.333	.303	1.111	.270	-1.172	4.135
	SQ_EXPER	-1.478	.474	-.397	-3.116	.003	-2.422	-.534
	Cubic Experience	.428	.222	.395	1.927	.058	-.014	.870
	EXPSTRES	-3.492	1.087	-.420	-3.213	.002	-5.655	-1.329

a. Dependent Variable: PRODUCT

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	EXPSTRES, EXPERIEN, STRESS, SQ_EXPER, squared stress, Cubic Experience[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: PRODUCT

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.543[a]	.295	.242	7.80863

a. Predictors: (Constant), EXPSTRES, EXPERIEN, STRESS, SQ_EXPER, squared stress, Cubic Experience

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2042.258	6	340.376	5.582	.000[a]
	Residual	4877.972	80	60.975		
	Total	6920.230	86			

a. Predictors: (Constant), EXPSTRES, EXPERIEN, STRESS, SQ_EXPER, squared stress, Cubic Experience

b. Dependent Variable: PRODUCT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	32.958	8.959		3.679	.000	15.129	50.787
	STRESS	.561	1.108	.054	.506	.614	-1.644	2.766
	EXPERIEN	1.255	1.172	.190	1.070	.288	-1.078	3.588
	squared stress	-.322	1.198	-.031	-.269	.789	-2.707	2.062
	SQ_EXPER	-1.469	.475	-.395	-3.09	.003	-2.414	-.524
	Cubic Experience	.473	.219	.436	2.160	.034	.037	.908
	EXPSTRES	-3.628	1.081	-.437	-3.36	.001	-5.780	-1.476

a. Dependent Variable: PRODUCT

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	EXPSTRES, EXPERIEN, STRESS, SQ_EXPER, Cubic Experience[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: PRODUCT

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.543[a]	.294	.251	7.76379

a. Predictors: (Constant), EXPSTRES, EXPERIEN, STRESS, SQ_EXPER, Cubic Experience

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2037.843	5	407.569	6.762	.000[a]
	Residual	4882.386	81	60.276		
	Total	6920.230	86			

a. Predictors: (Constant), EXPSTRES, EXPERIEN, STRESS, SQ_EXPER, Cubic Experience

b. Dependent Variable: PRODUCT

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	32.907	8.905		3.695	.000	15.188	50.626
	STRESS	.475	1.055	.045	.450	.654	-1.624	2.573
	EXPERIEN	1.265	1.165	.192	1.086	.281	-1.053	3.583
	SQ_EXPER	-1.439	.459	-.387	-3.136	.002	-2.352	-.526
	Cubic Experience	.470	.217	.434	2.163	.033	.038	.903
	EXPSTRES	-3.496	.959	-.421	-3.646	.000	-5.404	-1.588

a. Dependent Variable: PRODUCT

Strategies for 3 or More Predictors

With three predictors, we must be concerned with:

- How is *X1* related to *Y* (i.e., linear, quadratic, cubic)?
- How is *X1* related to *Y* (i.e., linear, quadratic, cubic)?
- How is *X1* related to *Y* (i.e., linear, quadratic, cubic)?
- Do *X1* and *X2* interact with each other?
- Do *X1* and *X3* interact with each other?
- Do *X2* and *X3* interact with each other?
- Do *X1, X2,* and *X3* simultaneously interact with each other?

  But, we rarely consider three-way interactions…so exclude from all analyses unless theory dictates otherwise.

Is it necessary to include some combination of quadratic or interaction terms in the model?

Test for a nonlinear relationship while adjusting for a possible linear relationship by comparing two models:

Nonlinear: Attn = Caffeine Sleep Age
Caffeine^2 Sleep^2 Age^2
Caffeine*Sleep Caffeine*Age Sleep*Age

Linear: Attn = Caffeine Sleep Age

Use only linear forms of the predictors if the *F Change* test for nonlinearity is not significant.

Test for the individual importance of the quadratic and interaction terms if the *F Change* for nonlinearity is significant.

The individual importance of the quadratic and interaction terms is determined by the *t* tests for the nonlinear model.

- If all the *t* tests for the quadratic and interaction terms are significant, use the entire nonlinear model.

- If one or more of the *t* tests for the quadratic and interaction terms are not significant, eliminate the term that has the smallest *t* value and re-estimate the model. Continue eliminating the quadratic and interaction terms one-at-a-time and re-estimating the model until the remaining quadratic and/or interaction terms are significant.

An alternative strategy for determining exactly which interaction terms and squared terms are needed would be to conduct a stepwise regression analysis with the linear terms ‘forced’ into the model. **This is only necessary if the *F Change* for the test of nonlinearity is significant.**

Issues in Performing Nonlinear Regression

Interpolation: It is generally acceptable to use the regression equation to predict performance on the dependent variable for values not used in the study, provided they are within the range of those originally used.

Extrapolation: Extrapolation beyond the original range of X values is hazardous and should be avoided. In other words, one should *not* engage in predictions for values of the independent variable that are outside the range used in the study.

- Once you determine the highest-degree polynomial that fits the data, you can calculate the regression equation with the terms that are to be retained.

- Only the test of the regression coefficient associated with the highest-degree polynomial in the equation is meaningful.

- Even when t ratios for the regression coefficients of lower-order polynomials are statistically nonsignificant, the lower-order terms should be retained.

<u>Interpretation of the Regression Coefficients when the model includes higher-order terms</u>

- The regression coefficients do not lend themselves to easy interpretation.

- A single independent variable is represented by two or more terms in the model. Therefore, the usual interpretation of a regression coefficient as the expected change in Y associated with a one unit change in the variable under consideration while holding the other variables constant makes no sense in polynomial regression.

- The relative magnitudes of the regression coefficients cannot be compared because the standard deviations of higher-order terms become increasingly larger, thereby leading to increasingly smaller regression coefficients.

- A linear transformation of X will (e.g., centering), of course, not affect its standard deviation, but the standard deviation of the powered terms will change, as will the correlations among the terms.

Pedhazur (1997, p. 546) recommends <u>not testing</u> beyond the quadratic term because:

- Cubic and higher-order relationships are rare in behavioral research.

- The higher the degree of the polynomial the more it is affected by the reliability of the measure involved and the more difficult it is to interpret.

- When the reliability is not very high, trends may seem to appear when they do not exist, or trends that do exist may be overlooked. Unlike manipulated variables, attribute variables used in the behavioral sciences tend to have only moderate reliabilities.