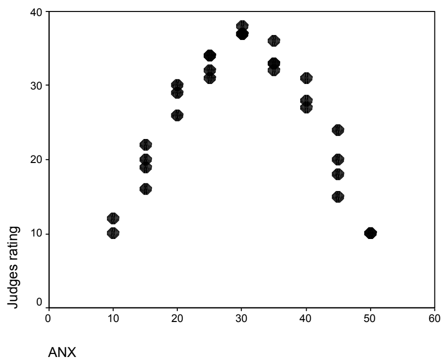
Example of Including Nonlinear Components in Regression

These are real data obtained at a local martial arts tournament. First-time adult competitors were approached during registration and asked to complete an informed consent form, a performance anxiety questionnaire and to tell how many times during the last 24 hours they had practiced the kata they would be performing. The criterion variable was the total of four judges' ratings of their performance.

Looking at Performance Anxiety as a Predictor of Judges Rating



You can see the strong quadratic component to this bivariate relationship.

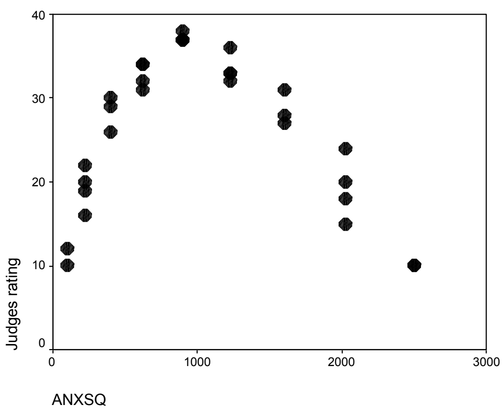
We can try to model this using relationship using a “quadratic term’ which is X^2.

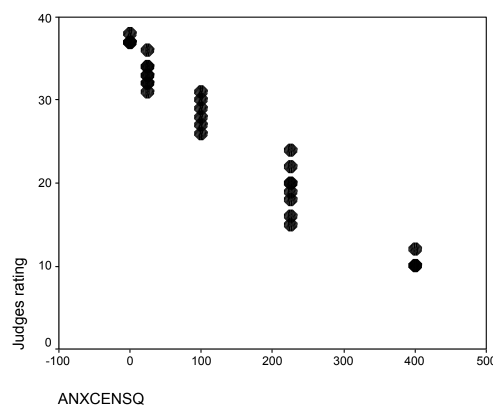
There are two ways to do this: 1) squaring the raw x scores (and 2) squaring the centered x scores (subtracting the mean of x from each x score before squaring)

SPSS Code:

```
compute anxsq = anx ** 2.
compute anx_cen = anx – 30.
compute anxcensq = (anx - 30) ** 2.
```

🡐 squaring gives a "linear + quadratic" term
🡐 mean-centering the original variable
🡐 centering first gives a "pure quadratic" term





Correlations

		time to complete the task -- DV	ANX	ANXSQ	ANXCENSQ
time to complete the task -- DV	Pearson Correlation	1	-.005	-.182	-.970**
	Sig. (2-tailed)	.	.980	.336	.000
	N	42	30	30	30
ANX	Pearson Correlation	-.005	1	.983**	.000
	Sig. (2-tailed)	.980	.	.000	1.000
	N	30	30	30	30
ANXSQ	Pearson Correlation	-.182	.983**	1	.183
	Sig. (2-tailed)	.336	.000	.	.334
	N	30	30	30	30
ANXCENSQ	Pearson Correlation	-.970**	.000	.183	1
	Sig. (2-tailed)	.000	1.000	.334	.
	N	30	30	30	30

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

Since there is no linear component to the bivariate relationship, neither the linear nor the linear+quadratic terms of the predictor are strongly related to performance. But the "pure" quadratic term is.

Notice that the linear and quadratic (anxcensq) terms are uncorrelated!

Notice that the sign of the correlation is “--“ for an inverted quadratic trend (“+” for an U-shaped trend)

Two Versions of the Multiple Regression -- uncentered vs. centered terms

Model Summary

Model	R	R Square
1	.970[a]	.942

a. Predictors: (Constant), ANXSQ, ANX

Coefficients[a]

Model		Unstandardized Coefficients			t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-23.602	2.638		-8.945	.000
	ANX	3.923	.192	5.218	20.486	.000
	ANXSQ	-.065	.003	-5.312	-20.856	.000

a. Dependent Variable: time to complete the task -- DV

Model Summary

Model	R	R Square
1	.970[a]	.942

a. Predictors: (Constant), ANX_CEN, ANXCENSQ

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	35.304	1.221		29.157	.000
	ANX_CEN	-.004	.035	-.005	-.102	.919
	ANXCANSQ	-.065	.333	-.970	-20.856	.000

a. Dependent Variable: time to complete the task -- DV

Notice that while the R^2 for each model are the same the β weights from the two modes are not the same! Why?

- Remember, the multiple regression weights are supposed to reflect the independent contribution of each variable to the model -- after controlling for collinearity among the predictors.
- However, the collinearity between ANX and ANXSQ (the not-centered, linear+quad term) was large enough to "mess up" the mathematics used to compute the β weights for ANX and ANXSQ -- giving a nonsense result.
- The βs from the model using the centered-squared term show the separate contribution of the linear and quadratic terms to the model.

So, there are two good reasons to work with centered terms: 1) they reduce collinearity among the computed predictors and 2) each term is a "pure" version of the orthogonal component it is intended to represent.

Interpreting the regression weights for the centered and the center-and-squared terms

Constant

- expected value of y when value of all predictors = 0
- value of y when x = mean (after mean-centering, mean = 0 & $mean^2 = 0$)
- for this model -- those with anxiety of 30 are expected to have a Judges Rating score of 35.304

Linear Term

- expected change in y for a 1-unit change in x, holding the value of the other variables constant at 0
- the linear component of how y changes as x changes – is non-significant for this model
- for this model -- for each 1 point increase in anxiety, judges rating is expected to decrease .004

Quadratic Term

- a quadratic term is a kind of interaction – x_cen ** 2 = x_cen * x_cen
- it tells about the expected direction and change of the y-x slope changes as the value of x changes
- +b → the slope becomes more positive as x increases; -b → the slope becomes less positive as x increases
- However, there's one more thing… Remember this is "an interaction term! This only tells you the nonlinear contribution of the change in y' for the 1-unit change from x=0 to x=1! The nonlinear contribution for other "shifts" along the x-axis will be different and incremental by this amount!

Using the Plotting Computator

5				
6	***height***	constant	35.304	
7	***slope***	b(x)	-0.004	
8	***curve***	b(x^2)	-0.065	
9				
10				
11		x(mean)	30	
12		x(std)	10	
13				

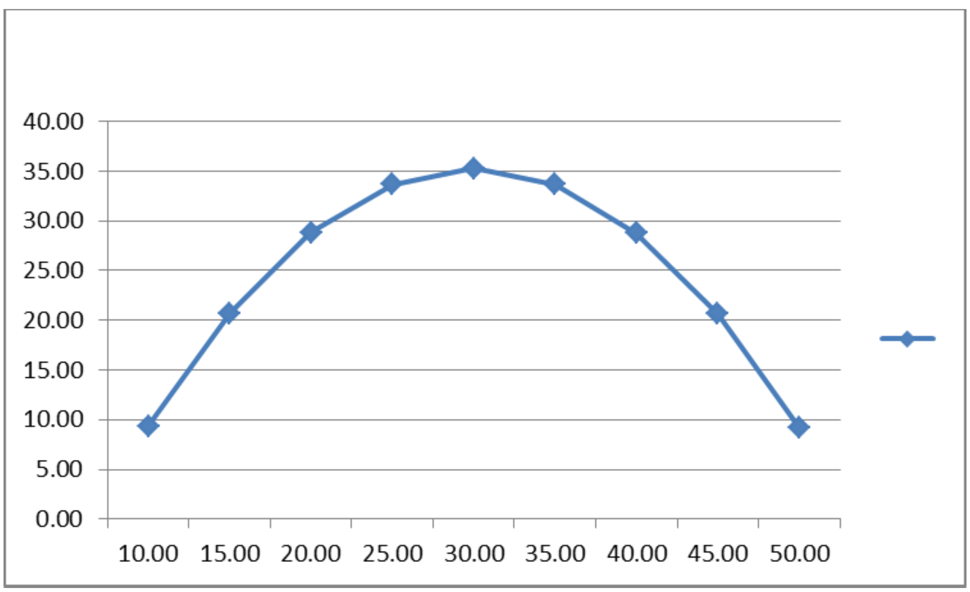
Fill in the predictor (x) mean and standard deviation.

Fill in the constant, linear and nonlinear regression weights.

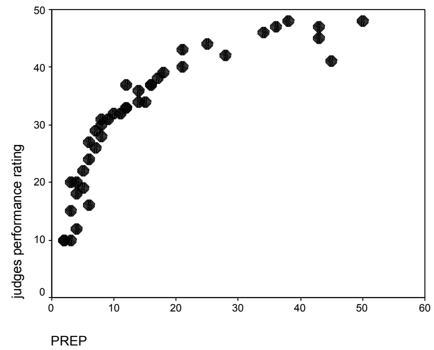
x std range	***-2***	***-1.5***	***-1***	***-0.5***	***0***	***0.5***	***1***	***1.5***	***2***
x-centered	-20.00	-15.00	-10.00	-5.00	0.00	5.00	10.00	15.00	20.00
x-centered^2	400.00	225.00	100.00	25.00	0.00	25.00	100.00	225.00	400.00
x-raw	10.00	15.00	20.00	25.00	30.00	35.00	40.00	45.00	50.00
y' =	9.38	20.74	28.84	33.70	35.30	33.66	28.76	20.62	9.22

The Computator computes the y' for several values of x…

and plots the model.



Working with Preparation (# of times practiced kata in previous 24 hrs) as Predictor of Judges Rating



These data seem to show a combination of a positive linear and an inverted-U-shaped quadratic trend.

```
compute prepsq = prep **2.
compute prep = (prep - 15.5).
```

🡐 computing the "combined" term

🡐 computing the "mean-centered" term

Notice something – this used "computational replacement" which is not recommended & can be very confusing! I suggest you always compute a new variable and label it as the centered variable!

```
compute prpcensq = prep ** 2.
```

🡐 computing the "pure quadratic" term

Correlations

		judges performance rating	PREP	PREPSQ	PRPCENSQ
judges performance rating	Pearson Correlation	1	.845**	.703**	.388*
	Sig. (2-tailed)	.	.000	.000	.011
	N	42	42	42	42
PREP	Pearson Correlation	.845**	1	.967**	.787**
	Sig. (2-tailed)	.000	.	.000	.000
	N	42	42	42	42
PREPSQ	Pearson Correlation	.703**	.967**	1	.918**
	Sig. (2-tailed)	.000	.000	.	.000
	N	42	42	42	42
PRPCENSQ	Pearson Correlation	.388*	.787**	.918**	1
	Sig. (2-tailed)	.011	.000	.000	.
	N	42	42	42	42

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

*. Correlation is significant at the 0.05 level (2-tailed).

This highlights the problem with "computational replacement" → when I went back to look at these data, years later, to make this handout, I had a tough time figuring out the results – sometimes "PREP" was the original term & sometimes the mean-centered term! Make new variables & label them correctly!!

Teaser Alert → Notice that both the raw squared term and the centered & squared term are highly collinear with the original predictor – more about this later!

Here are the two versions of the model: the first using mean-centered terms & the second using the original variables

Model Summary

Model	R	R Square
1	.956[a]	.914

a. Predictors: (Constant), PRPCENSQ, PREP

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Beta	t	Sig.
1	(Constant)	17.969	.841		21.4	.000
	PREP	1.204	.065	.415	18.7	.000
	PRPCENSQ	-.031	.003	-.225	-9.6	.000

a. Dependent Variable: judges performance rating

Once again, the centered version had lower collinearity & "more reasonable" β results -- R^2 was again the same.

Model Summary

Model	R	R Square
1	.956[a]	.914

a. Predictors: (Constant), PRPSQ, PREP

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Beta	t	Sig.
1	(Constant)	10.506	1.288		8.156	.000
	PREP	2.167	.157	2.547	13.83	.000
	PREPSQ	-.031	.003	-1.760	-9.559	.000

a. Dependent Variable: judges performance rating

Curvilinear relationship *OR* skewed predictor?!?

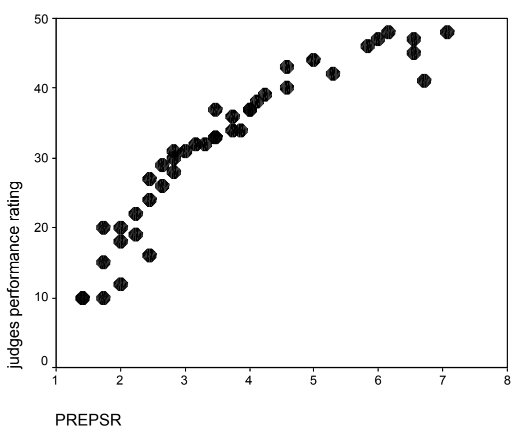
There are three "red flags" that this second example is not what it appears to be (a curvilinear relationship between prep and performance).

1) Re-examine the original scatter plot – notice that there are far fewer cases on the right side of the plot than the left
2) Notice that the mean (15.5) is pretty low for a set of scores with a 0-50 range
3) The mean-centered term is highly collinear with the mean-centered & squared term

All three of these hint at a skewed predictor – the first two suggesting a positive skew.

Looking at the univariate statistics (which we really should have done along with the other elements of data screening sefore we did the fancy-schmancy trend modeling, we found a Skewness = 1.27, which is higher than the common cutoff of .80. What would happen if we "symmetrized" this variable -- say with a square root transformation?

```
compute prepsr = prep ** .5.                ← should decrease the skewing of prep
compute prepsr = (prepsr – 3.586)           ← mean-centering the square-rooted term (replaced!!)
compute prepsrcs = (prepsr – 3.586) ** 2    ← quadratic term for "unskewed" version of prep
```



Correlations

		judges performance rating	PREPSR	PREPSR CS
judges perfomance rating	Pearson Correlation	1	.920	.020
	Sig. (2-tailed)	.	.000	.865
	N	42	42	42
PREPSR	Pearson Correlation	.920	1	.000
	Sig. (2-tailed)	.000	.	1.000
	N	42	42	42
PREPSRCS	Pearson Correlation	.020	.000	1
	Sig. (2-tailed)	.865	1.000	.
	N	42	42	42

This simple transformation resulted in a near-orthogonality between the linear and quadratic terms.

Look at the results from the transformed data → we get a linear scatterplot, a strong linear relationship & no nonlinear relationship,

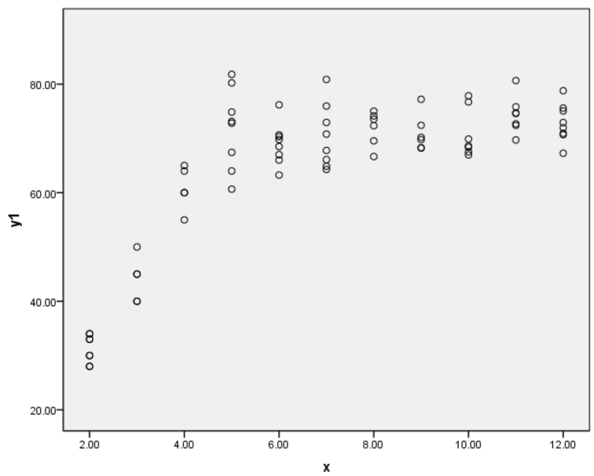
Take-Home Message: It is **very** important to differentiate between "true quadratic components" and "apparent quadratic components" that are produced by a skewed predictor!! Always remember to screen your data and consider the univariate and bivariate data patterns before hurrying onto the multivariate analysis!!!

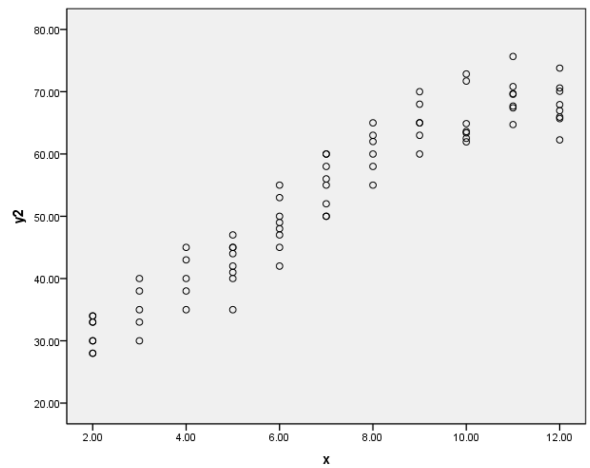
Diminishing Returns Curvilinear Models

Probably the most common linear+quadratic form seem in behavioral research is the “diminishing returns” trend. Initial rapid increases in y with increases in x (a substantial positive linear slope) eventually diminish, so that greater increases in x are need to have substantial increases in y (a lessening positive linear slope & so a negative quadratic trend). Here are two variations of that type of trend – an “early peaking” on the left and a “later peaking” on the right.

Both analyses use the same predictor, which was prepared for the analysis by mean-centering and then squaring the mean-centered version.

compute x_cen = x – 7.1053.
compute x_cen_sq = x_cen ** 2.





Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.903[a]	.815	.810	6.22708

a. Predictors: (Constant), x_cen_sq, x_cen

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12463.126	2	6231.563	160.705	.000[a]
	Residual	2830.682	73	38.776		
	Total	15293.808	75			

a. Predictors: (Constant), x_cen_sq, x_cen
b. Dependent Variable: y1

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	72.958	1.060		68.816	.000
	x_cen	3.194	.225	.716	14.218	.000
	x_cen_sq	-.814	.077	-.530	-10.512	.000

a. Dependent Variable: y1

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.957[a]	.915	.913	4.06551

a. Predictors: (Constant), x_cen_sq, x_cen

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	13024.267	2	6512.133	393.997	.000[a]
	Residual	1206.573	73	16.528		
	Total	14230.839	75			

a. Predictors: (Constant), x_cen_sq, x_cen
b. Dependent Variable: y2

Coefficients[a]

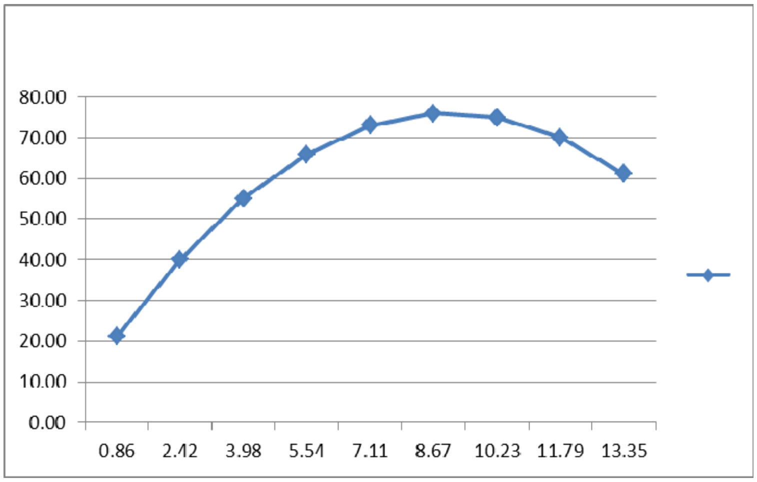
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	55.024	.692		79.494	.000
	x_cen	4.063	.147	.945	27.704	.000
	x_cen_sq	-.189	.051	-.127	-3.736	.000

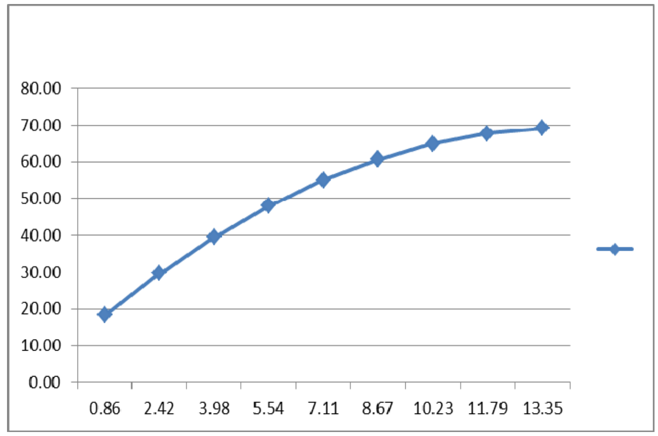
a. Dependent Variable: y2

Something to notice:

Often, there are “individual differences in asymptotic attainment” a term sometimes used by biometricians to mean that different individual’s trends flatten out at different values of x. Econometricians sometimes use the term “individuated point of inflection” which means the same thing.

Since we are exploring this y-x relationship using a between subjects design (rather than using a repeated measures or longitudinal design/analysis) this usually is expressed as there being greater variability in scores around the inflection point. This is easily seen in the y1-x relationship – look at the greater variability in y1 scores for x values of 5-7, Here are the plots of these two models…





Comparison of the two diminishing return models shows some common differences:

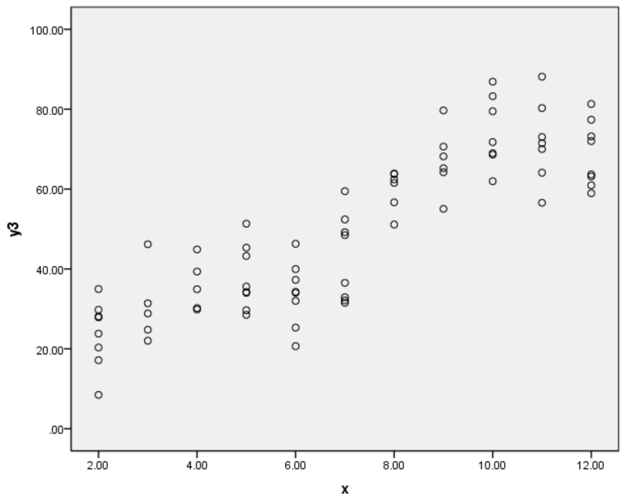
- A relatively stronger quadratic term for the early-peaking model (β = -.53) than the late-peaking model (β = -.127)
- A relatively stronger linear term for the late-peaking model(β = -.945) than the late-peaking model (β = -.716)

Remember to interpret the quadratic term as an interaction!!!

By how much and in what direction does the criterion value change as x increases by 1? It depends! The extent and direction of change in y with a 1-unit change in x depends on the starting value of x!

“Learning Curves”

Perhaps the quintessential nonlinear model in behavioral sciences is the combination of a positive linear trend and an initially-decreasing cubic trend into a “learning curve”. Some developmental and learning psychologists have been so bold as to assert that nearly any time you get linear or linear+quadratic trend, you would have gotten a “learning curve” if you had included lower x values, higher x values, or both! In other words, a linear trend is just the middle of a learning curve, an “accelerating returns” model is just the earlier part and a “diminishing returns” model is just the later part of the learning curve. However, the limited number of important cubic trends in the literature would seem to belie this characterization.



Here is the syntax to
compute x_cen = x – 7.1053.
compute x_cen_sq = x_cen ** 2.
compute x_cen_cube = x_cen ** 3.

Correlations

		x_cen	x_cen_sq	x_cen_cube
x_cen	Pearson Correlation	1	-.028	.092**
	Sig. (2-tailed)		.807	.328
	N	76	76	76
x_cen_sq	Pearson Correlation	-.028	1	-.084
	Sig. (2-tailed)	.807		.471
	N	76	76	76
x_cen_cube	Pearson Correlation	.925**	-.084	1
	Sig. (2-tailed)	.000	.471	
	N	76	76	76

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

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.885[a]	.784	.775	9.42395

a. Predictors: (Constant), x_cen_cube, x_cen_sq, x_cen

Yep, a cubic term is like a 3-way interaction x*x*x !

Just as you can interpret a 3-way interaction as “how the 2-way interaction differs across values of the 3rd variable,” you can interpret a cubic term as how the quadratic term changes across values of x.

“A cubic trend tells how the way that the slope of the y-x relationship that changes across values of x, changes across values of x!” David Owen Shaw

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	23206.025	3	7735.342	87.099	.000[a]
	Residual	6394.377	72	88.811		
	Total	29600.402	75			

a. Predictors: (Constant), x_cen_cube, x_cen_sq, x_cen
b. Dependent Variable: y3

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	49.962	1.611		31.021	.000
	x_cen	8.320	.900	.641	9.246	.000
	x_cen_sq	-.049	.119	-.023	-.410	.683
	x_cen_cube	-.160	.045	-.517	-3.553	.001

a. Dependent Variable: y3

For this model:

- It is easy to see the positive linear trend in the data!
- The negative cubic trend tells us that the quadratic trend is increasingly negative with increasing values of x. Said differently, the quadratic trend at smaller values of x is positive (y-x linear slope gets more positive), while the quadratic trend at larger values of x is negative (y-x linear slope gets more less positive).
- There is a no quadratic trend, because “the positive quadratic trend at low values of x” and “the negative quadratic trend at higher values of x” average out to “no quadratic trend”