Polynomial Regression

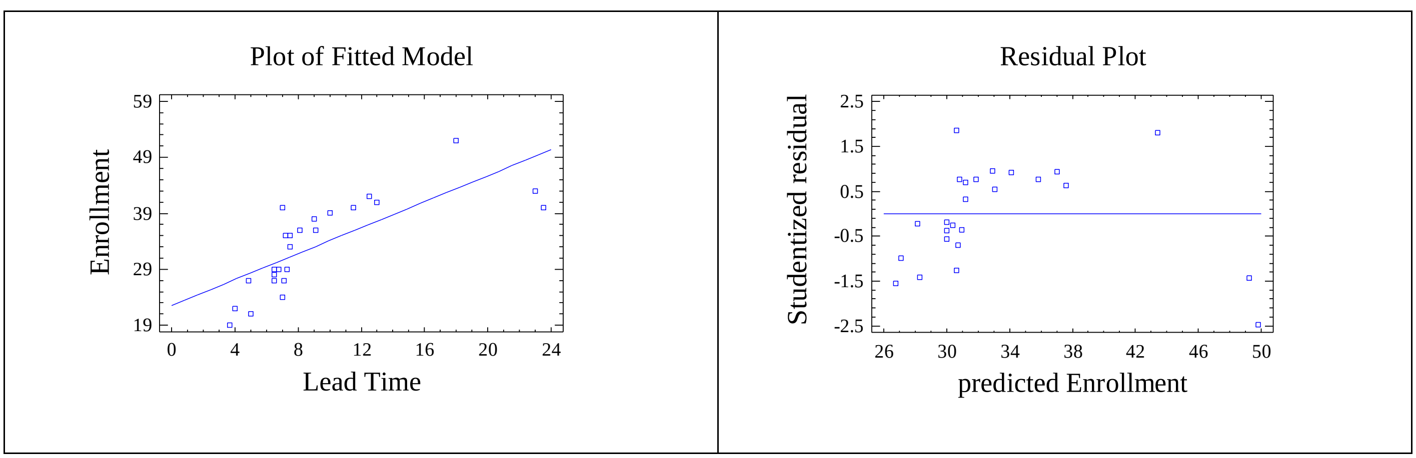
I. Quadratic Models

Example: An organization that conducts management seminar programs wishes to examine the relationship between seminar enrollments and the lead time of seminar announcements (the number of weeks before the seminar that the first promotional material is mailed). The results for 25 seminars is contained in the file SEMINAR. The results of a simple regression of enrolment vs. lead time appears below.

```
Dependent variable: Enrollment
Independent variable: Lead Time
-----------------------------------------------------------------------------
                                    Standard          T
Parameter          Estimate           Error        Statistic          P-Value
-----------------------------------------------------------------------------
Intercept           22.4911          2.25486        9.97451            0.0000
Slope               1.16159         0.212688        5.46148            0.0000
-----------------------------------------------------------------------------


                           Analysis of Variance
-----------------------------------------------------------------------------
Source                   Sum of Squares     Df  Mean Square     F-Ratio
-----------------------------------------------------------------------------
Model                           881.399      1      881.399       29.83
Residual                        679.641     23      29.5496
-----------------------------------------------------------------------------
Total (Corr.)                   1561.04     24

Correlation Coefficient = 0.751414
R-squared = 56.4623 percent
R-squared (adjusted for d.f.) = 54.5694 percent
Standard Error of Est. = 5.43595
```



From this analysis, we can see that, although the *P*-value for the model is small, a curve might fit the data better than a line. This is due largely to the observations with lead times of 23 weeks and 23.5 weeks (the points at the far right of the graphs). These high leverage observations may seem detrimental to the model, and we might be tempted to remove them in order to "improve" the straight line fit. In fact, however, these points provide important information. They suggest that beginning the mailings *too* early may be counterproductive, i.e., enrollment may actually start to decrease for extremely long lead times.

A closer look at the graphs suggests that the **Quadratic** model $\boldsymbol{Y} = \beta_0 + \beta_1 X + \beta_2 X^2 + \boldsymbol{\varepsilon}$ would be a better fit. The quadratic model is a special case of a polynomial model, where a polynomial is used to model the relationship between $\boldsymbol{X}$ and $\boldsymbol{Y}$. The easiest way to fit a polynomial model in Statgraphics is to follow: *Relate* > *One Factor* > *Polynomial Regression*. Statgraphics will automatically fit a quadratic model to the variables. The Statgraphics' output for the seminar data are presented below.

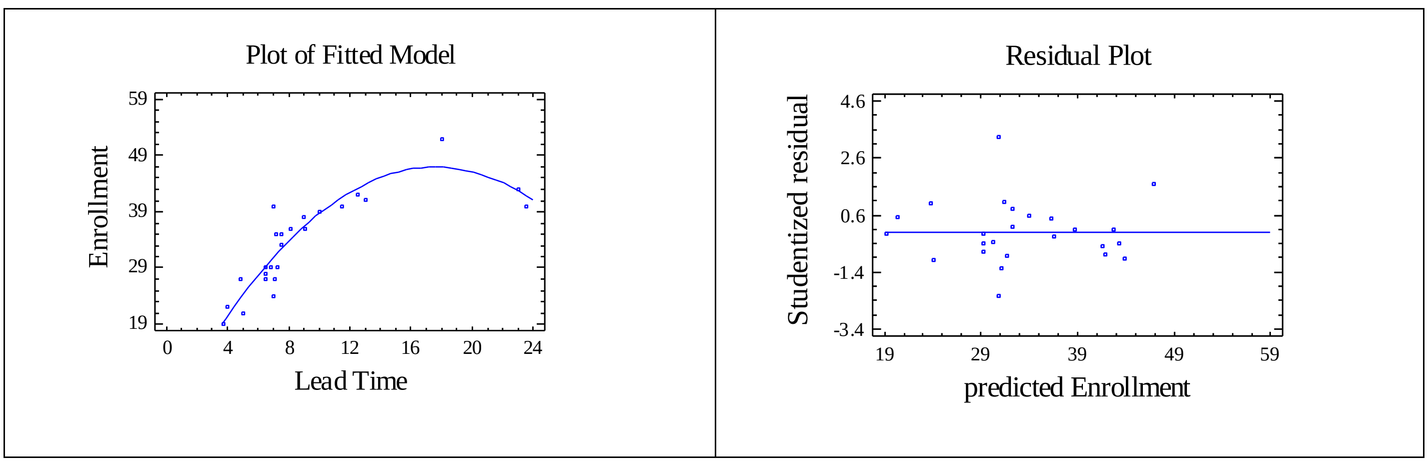
Dependent variable: Enrollment

Parameter	Estimate	Standard Error	T Statistic
CONSTANT	2.40963	3.61875	0.665875
Lead Time	5.0669	0.661291	7.66213
Lead Time^2	-0.144053	0.0238903	-6.02977

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio
Model	1304.83	2	652.414	56.02
Residual	256.213	22	11.646	
Total (Corr.)	1561.04	24		

R-squared = 83.587 percent
R-squared (adjusted for d.f.) = 82.095 percent
Standard Error of Est. = 3.41263



The parabola is clearly a better fit than the line computed in simple regression, and the residual plot is more random. (You should verify that the studentized residuals are plausibly normal.) The *P*-value for the Lead Time^2 term in the model is 0.0000, indicating that the quadratic term is significant. Below are listed some of the features of polynomial regression.

- The rest of the output retains the same interpretation as in other regression models, with the (fairly obvious) exception of slope interpretation. The usual interpretation of β_1 and β_2 as marginal slopes isn't appropriate since one can hardly vary X while holding X^2 constant, and vice versa.
- Sometimes it is necessary to use a polynomial of degree greater than two to fit data. This can be accomplished using the right mouse button to access *Analysis Options* and changing "Order." In practice, polynomials of order greater than 3 (a Cubic model) are rarely used.
- Forecasts can be obtained as in simple regression by using the *Forecasts* option under *Tabular Options*.