Polynomial Regression Models

Polynomial Regression is a special case of Multiple Linear Regression. It refers to regression applications in which the independent variables are powers of another variable

$$x_1 = x,\ x_2 = x^2,\ etc$$

A polynomial regression model of order k has the equation

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + \varepsilon$$

where ε is a random variable with mean 0 and variance σ^2.

A prediction equation for this model fitted to data is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \ldots + \hat{\beta}_k x^k$$

Here is a generic ANOVA table for polynomial regression of order k:

Source of Variation	DF	SS	MS
Regression	k	SSR	MSR
Linear	1	SSR1	MSR1
Quadratic	1	SSR2	MSR2
Cubic	1	SSR3	MSR3
…			
Error	n-k-1	SSE	MSE
Total	n-1	SSTot	

An example of a polynomial regression of order two is given by the KWH data. Recall that the plot of residuals versus DRYER showed a curving pattern, suggesting the need for additional variables to account for the curvature. Figure 1 shows a plot of the data.

One possibility is to use polynomial terms in DRYER. Starting with a quadratic, the variables for the multiple regression would now be x_1=DRYER and x_2=DRYER^2. The model equation would be

$$KWH = \beta_0 + \beta_1 DRYER + \beta_2 DRYER^2 + \varepsilon$$

The prediction equation is

$$KWH = 34.39 + 41.23(DRYER) \ -9.33(DRYER)^2$$

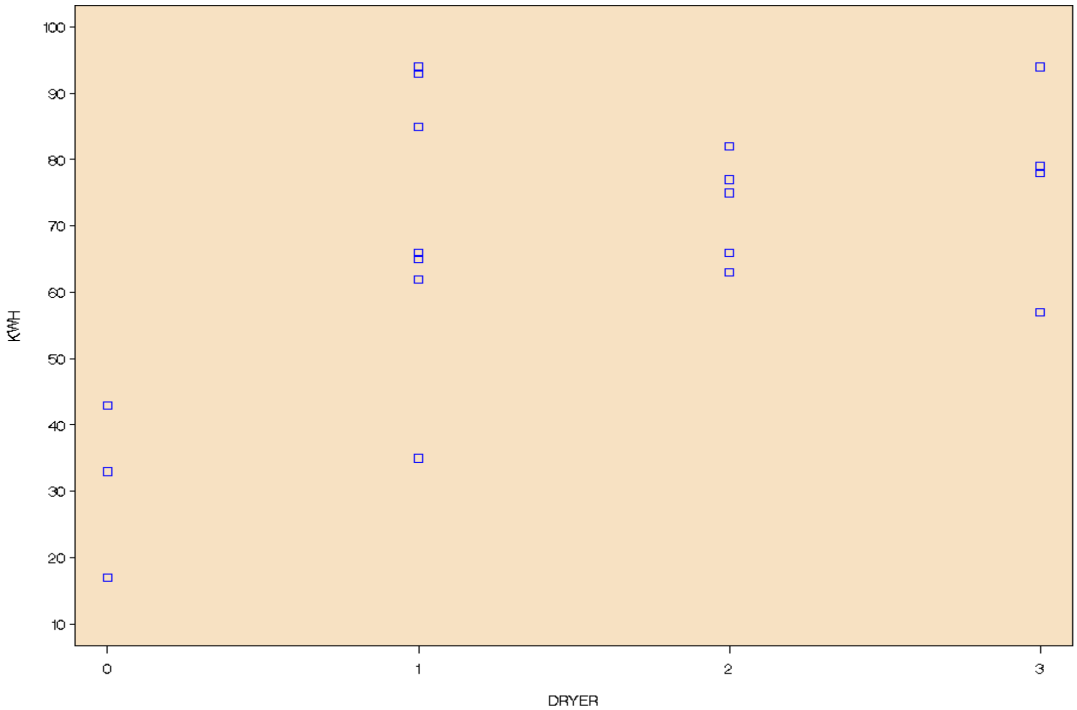


Figure 1. Plot of KWH versus DRYER

Figure 2 shows the curve plotter through the data. You can see the curvature.

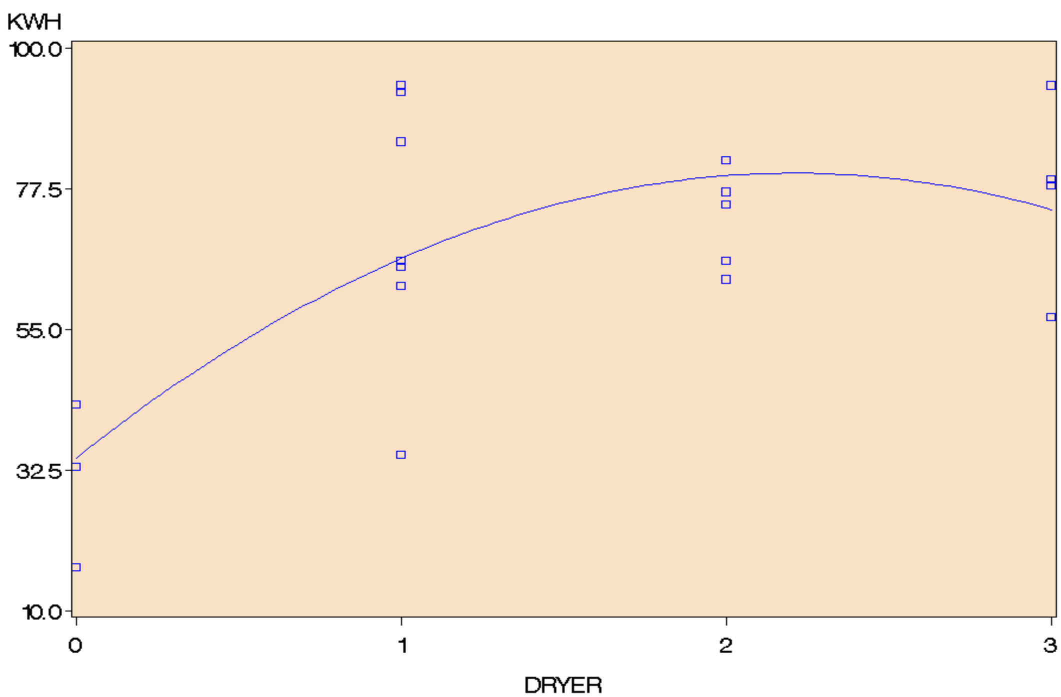


Figure 2. Quadratic curve plotted through the KWH data.

Here is an ANOVA table for the quadratic regression:

ANOVA for Quadratic Regression on DRYER

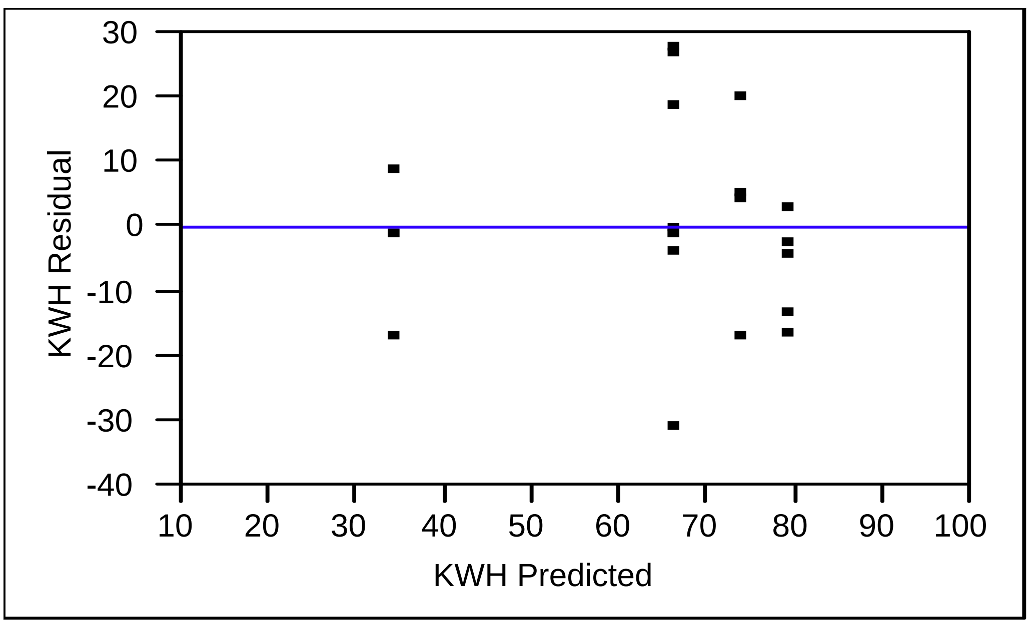
Source of Variation	DF	SS	MS
Regression	2	5149.2	2574.6
Linear	1	3429.7	3429.7
Quadratic	1	1749.5	1749.5
Error	18	4429.4	246.0
Total	20	9578.6	

The sums of squares are sequential; that is, they are

SS(Linear) = 3429.7 and SS(Quadratic|Linear) = 1749.5.

The Quadratic mean square is large relative to MSE (F=1749.5/246.0=1.11) which is "highly significant." In other words, the quadratic model fits significantly better than the linear model.

Figure 3 shows a plot of the residuals from the quadratic model plotted versus DRYER.



You can see that the curved pattern in the residuals has been removed.

Recall the multiple regression model that included both AC and DRYER. Polynomial models can contain powers and products of two or more variables. Here is the polynomial model contained AC and linear and quadratic terms in DRYER.

$$KWH = \beta_0 + \beta_1 AC + \beta_2 DRYER + \beta_3 DRYER^2 + \varepsilon$$

The prediction equation is

$$KWH = 7.07 + 5.12(AC) + 21.30(DRYER) - 2.65(DRYER)^2$$

The coefficients of DRYER and $DRYER^2$ have changed numerically, but the curve in DRYER is not changed very much.

Figure 4 shows a plot of the fitted surface in three dimensions.

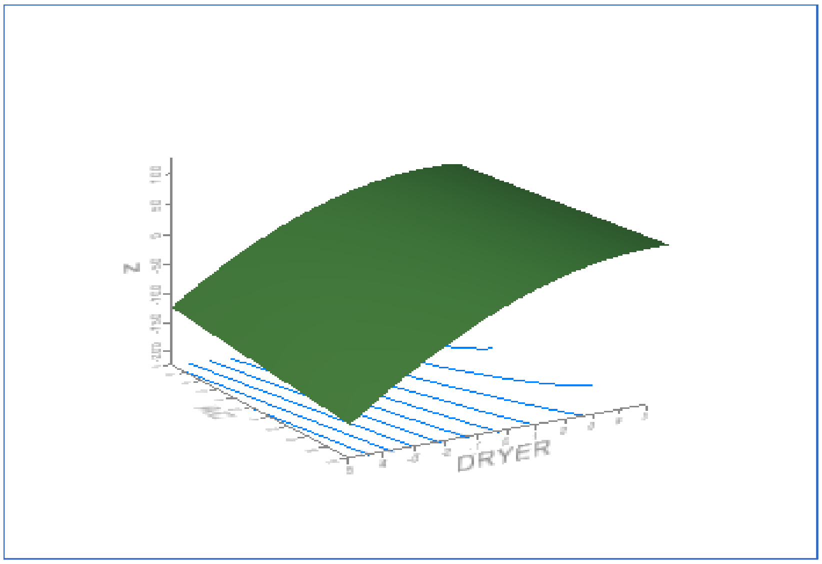


Figure 4. Response surface showing effects of AC and DRYER on KWH.

The next step might be to consider product terms between AC and DRYER, using the model

$$KWH = \beta_0 + \beta_1 AC + \beta_2 DRYER + \beta_3 DRYER^2 + \beta_4 AC \bullet DRYER + \varepsilon$$

The following ANOVA table shows the effects of the AC*DRYER term:

Source of Variation	DF	SS	MS
Regression	4	9423.4	2355.8
AC	1	4266.1	4266.1
DRYER Linear	1	3690.1	3690.1
DRYER Quadratic	1	115.5	115.5
AC*DRYER	1	8.1	8.1
Error	16	155.1	9.7
Total	20	9578.6	

The mean square for AC*DRYER is small relative to MSE, indicating that the product term is not needed in model. Recall that there was no evidence of curvature in residuals plotted versus AC, so there is no need to consider terms in AC^2.

Finally, Figure 5 shows a residual plot for the final model

$$KWH = 7.07 + 5.12(AC) + 21.30(DRYER) - 2.65(DRYER)^2$$

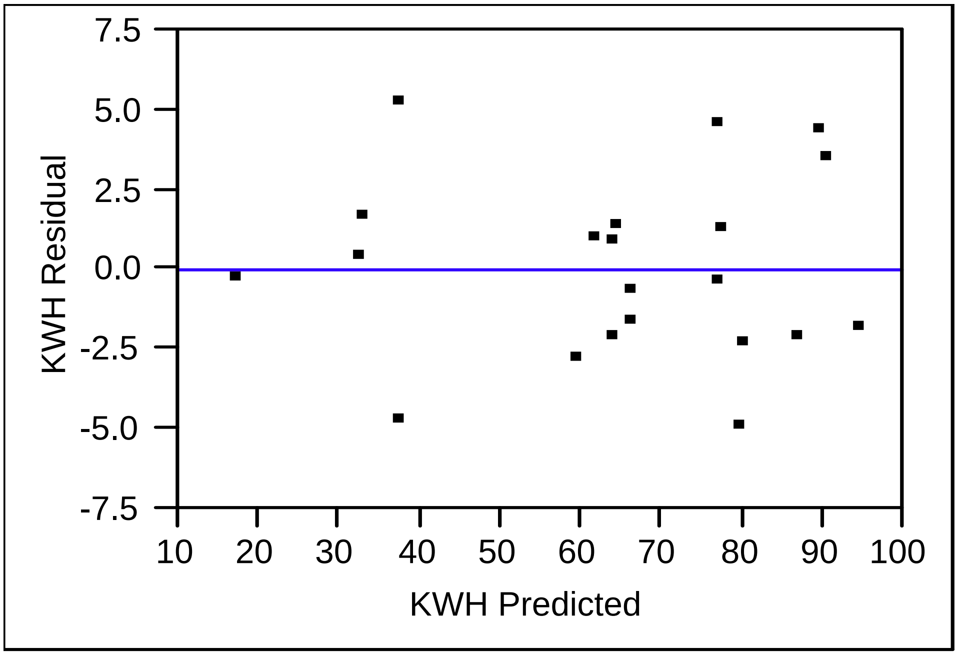


Figure 5. Residual plot for the polynomial model in AC, DRYER and $DRYER^2$.