Seasonality Regression

Overview

Introduction

In a previous document, we observed that we seemed to see a decrease in PSHR reports published in QST over time. We also observed what appeared to be some sort of seasonality dependence.

An analysis of that seasonality was undertaken … this is the result of that analysis.

In this section

Following is a list of topics in this section:

Reasons to Explore Seasonality

Introduction

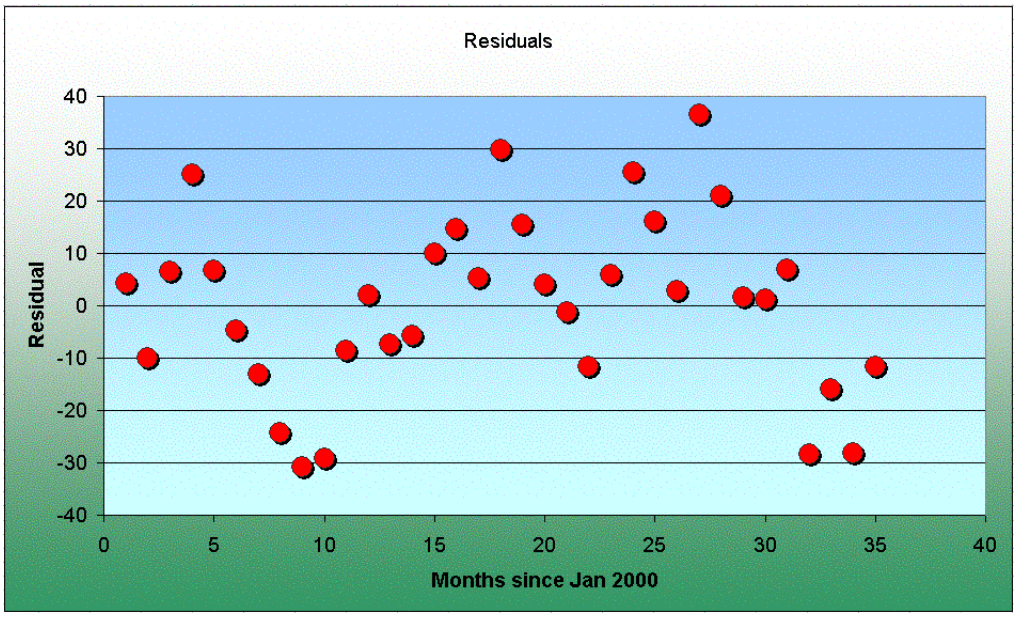
In an earlier analysis of PSHR data, we discovered a decrease in PSHRs published in QST of about 1.6 per month over the years 2000 through 2002. The R value for the regression was about 0.7, which is not bad, but only explains half of the variation. A Seasonal effect would be expected for this sort of report, as stations have different availability at different times of the year, as weather patterns are seasonal, and as public service events tend to favor the better weather.

Continued on next page

Reasons to Explore Seasonality, Continued

Residuals

If we plot the residuals, that is, the difference between the actual and predicted number of reports, we see an effect that has a cyclic appearance.



Preparing data for the regression

Introduction

In the initial analysis, we used a linear equation. The solution of the best parameters for a linear equation is relatively straightforward. Indeed, even a linear equation with a number of independent variables requires no special tools; the necessary analysis can be performed by Excel.

However, we don't expect the seasonal effect to be linear. To perform the analysis without resorting to specialized tools requires that we transform the data.

Selecting the Form of the Equation

Before we can transform the data, we need to decide how we will transform the data. In the initial analysis, we assumed that the equation was of the form:

$$Y = A + B * X$$

Where Y is the number of PSHRs reported in QST, X is the month of the issue, counted since January, 2000. A and B are constants which are determined to best fit the line to the data.

We might choose any number of periodic functions, but a sine wave is simple, and as an amateur, appealing. The seasonal effect will have an amplitude and a phase shift. The period will be constrained to be one year. So the resulting expected equation is:

$$Y = A + B * X + C * SIN(X + D)$$

Since we can't transform the phase shift to result in a linear regression, the value for D was determined, by trial and error, to be 0.44 (the phase angle is in radians). This results in an equation of the form:

$$Y = A + B * X + C * SIN(X + 0.44)$$

Which looks exactly like:

$$Y = A + B * X_1 + C * X_2$$

which is linear in 2 independent variables. X_1 is simply our old X, the number of months since January 2000. X_2 is now sin(X+0.44).

Regression Results

Introduction

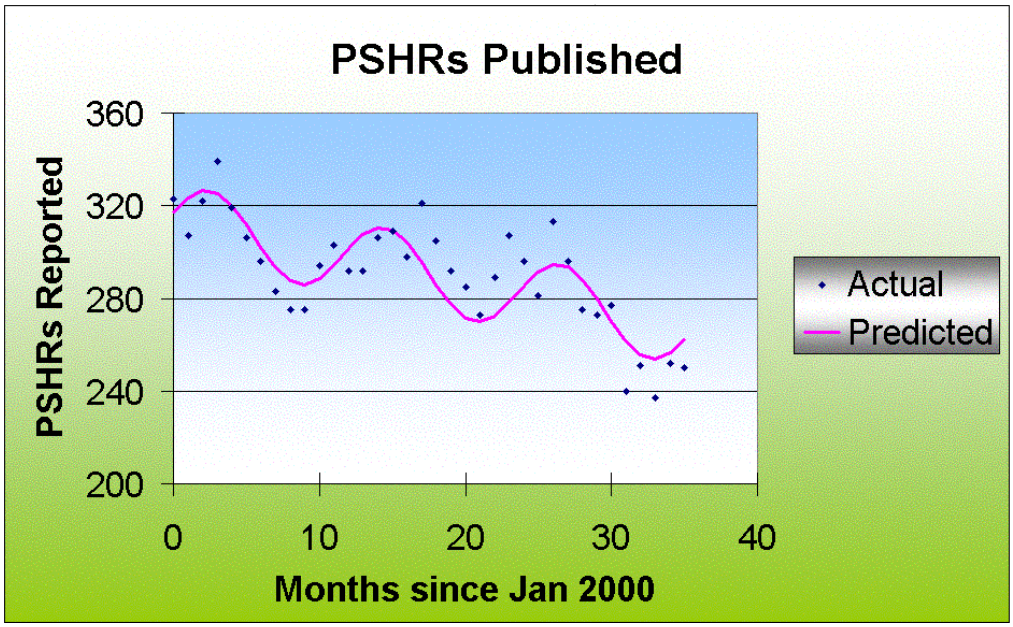
Once we have decided on the form of the equation, it is a relatively simple matter to calculate the second X variable, and then perform a multiple regression, with the dependent variable (Y) being the number of PSHRs reported, X_1 being the month, and X_2 being the sine of the month+0.44.

Resulting Equation

Once the regression has been performed, the resulting equation is:

```
Y = 313.6 – 1.33 * X + 16.2 * sin(X+0.44)
```

Like the original equation, it still shows a negative slope, but the seasonal variation accounts for a variation of about 16 reports either side of the decreasing line.



The result shows the data fitting the prediction quite well. Indeed, given the nature of this data, we would be surprised to have a much closer fit.

Regression Metrics

The R value for this regression is 0.85, which indicates that about three quarters of the variation in the data is accounted for by the equation. Given the relatively small sample and the large number of factors that may affect these reports, a better fit would be somewhat suspicious.

As we look at the curve, and especially at the residuals graph earlier, a sine function might not be the best way to represent the seasonal variation, but the function does a pretty good job, and it is simple, and for that reason, appealing.