A Discussion of an Alternative Method for Modeling Cyclical Phenomena

Russel Brown
Summa Health Care

Isadore Newman
University of Akron

The purpose of this paper is to provide an answer to the question of the relative effectiveness of the cosine function versus a polynomial function in the description and stability of prediction of a specific set of longitudinal data. If the data conforms to a known function (such as the cosine function), can we test for that function more effectively (that is to say by the stability of the weights upon cross-validation) than using a polynomial function for developing prediction equations?

The purpose of this paper presentation is to describe a methodological approach to facilitate the ease of the description and prediction of cyclical events. The proposed method will be compared with traditional methods for fitting curvilinear relationships, and the potential benefits of the proposed method will be outlined. Finally, an example will be provided to demonstrate the method and its advantages over traditional curve fitting methods for the description and prediction of cyclical events.

Introduction

There are many phenomena with cyclical patterns that are of interest to statisticians, psychologists, and epidemiologists. Seasonal Affective Disorder is a clear example of a cyclical phenomena that is of growing interest in psychological research for both children and adults (Glod, & Baisden, 1999; Rohan & Sigmon, 2000). In turn, Partonen, Piiroinen, Loennqvist, and Jouko (2000) have demonstrated that additional psychological symptoms may have a seasonal pattern. Accidental death has also been shown to be tied to an annual pattern (Coren, 1996). On the other hand, there are ready examples of phenomena that cycle on a monthly basis. For example, suicide rates have been shown to cycle on a monthly basis with suicides being more frequent during the first and second weeks of the month (Phillips & Ryan, 2000). Clearly, there are numerous additional phenomena that cycle on a monthly or daily basis.

In the regression literature, authors have traditionally suggested that these phenomena be predicted by curve fitting or log-linear techniques (Cohen & Cohen, 1983; McNeil, Newman, & Kelly, 1996; Pedhazur, Pedhazur-Schmelkin, 1991). A model of a cyclical phenomena (Cosine) using a polynomial equation will take the general form:

$$\text{Model } Y = a_0 U + a_1 X + a_2 X^2 + \Sigma(a_{1+2i} X^{1+2i} + a_{2+2i} X^{2+2i}) + \text{Error}$$
$$\text{for } i \text{ cycles}$$

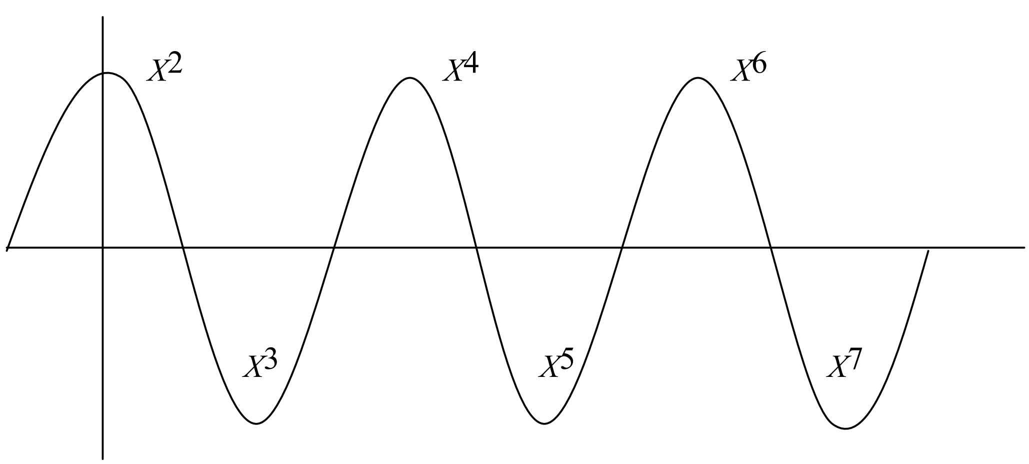

Figure 1. General polynomial representation of a cyclical function.

As can be seen in Figure 1, the first cycle requires a model with a 4th degree polynomial, and each additional cycle requires an addition of two variables to model. For example, a series of 4 cycles would require a model extending to a 10th degree polynomial. If one uses the trigonometric function (cosine), this relationship can be modeled with two predictor variables:

$$\text{Model } Y = a_0 U + a_1 \text{COS}(X) + a_2 X + \text{Error}$$

In this presentation, we will outline a method for using trigonometric functions (cosine) to predict cyclical phenomena; although this technique is not new, few examples have been provided in the literature. A method will be discussed for utilizing the sine or cosine function to predict a cyclical phenomena. In order to facilitate this discussion, the method will be anchored to an example: modeling temperature based on the time of year. Using this example, we will outline the general steps necessary to model a cyclical phenomena. In turn, the example will be used to demonstrate some of the advantages of the use of trigonometric functions in modeling cyclical phenomena.

Method

The use of a trigonometric function to model a cyclical phenomena requires the estimation of two variables: (1) the period, and (2) the amplitude of the function. The period is the time taken to make one complete cycle (a complete oscillation). For the Sine and Cosine functions, the period is measured in radians and is equal to 2π. The amplitude, on the other hand, is the displacement (the distance from the crest to the trough) of the cycle.

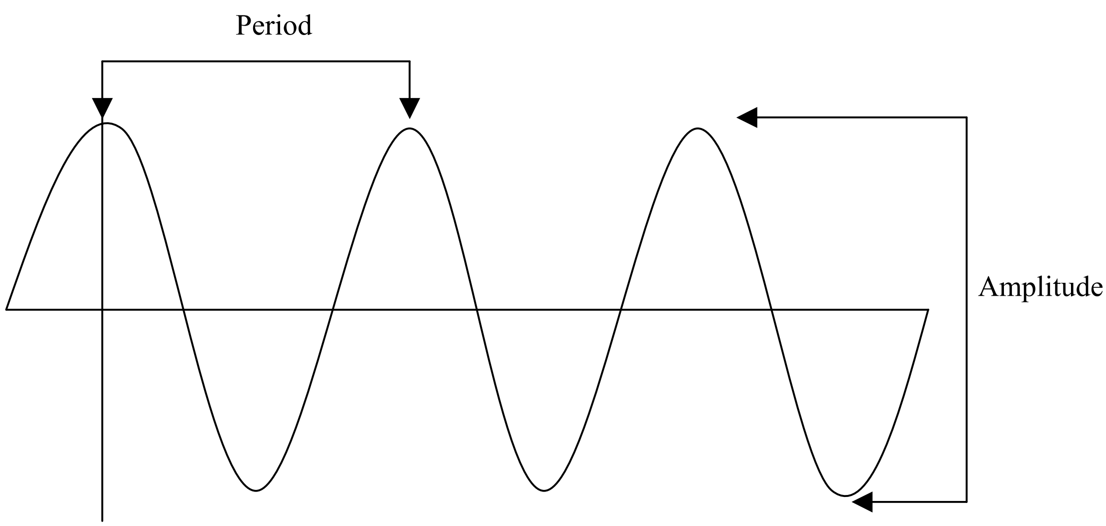

Figure 2. Parameters of a wave function.

One could estimate the period of a cyclical function through either theoretical or empirical means. The choice would be driven by the researcher's question and by whether theory is suggestive of a fixed cycle. Some psychological phenomena (e.g. seasonal affective disorder) are clearly linked to an annual cycle (period = 1 year). In other cases, one could estimate the period empirically. This can be achieved by plotting the dependent measure across time. For this presentation, we plotted temperature readings across time. As would be expected, there was a clear cyclical pattern of warmer temperatures in the summer and cooler temperatures in the winter.

At this point, we had a clear cycle with peaks occurring in July and troughs occurring in January. The period for this cycle is 12 months; therefore, all time measurements were divided by 12 and multiplied by 2π in order to convert the time measurement to a radian scale.

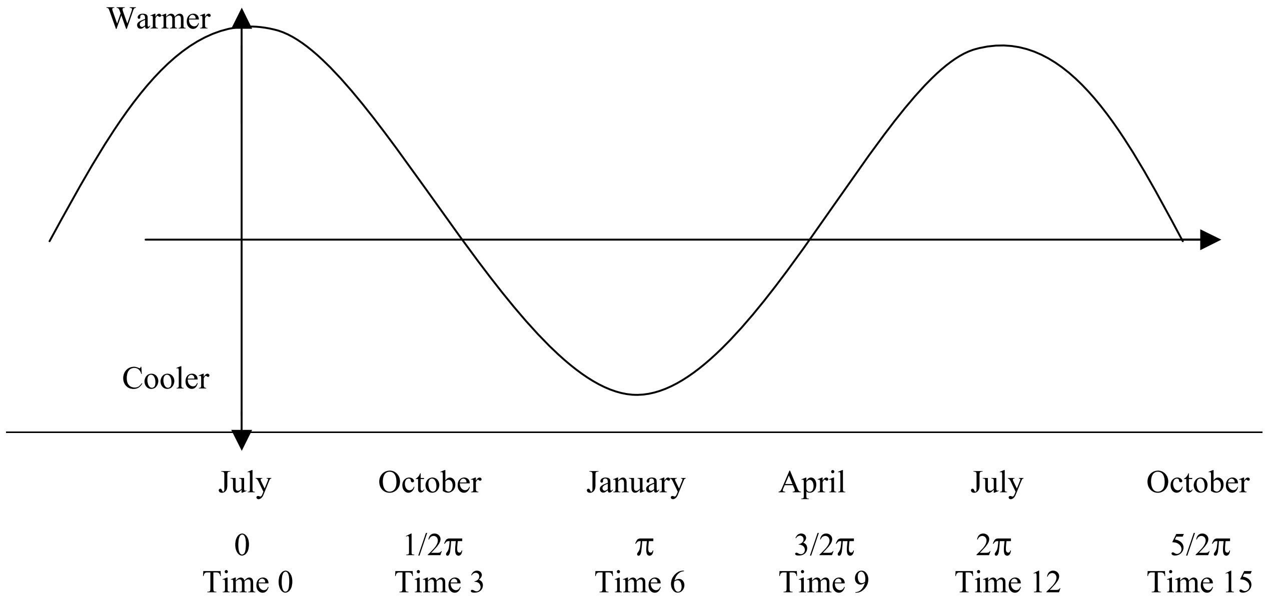

Figure 3. Plotting temperature across time using a radian representation of time.

Once the period was determined and the time measurements were converted to a radian scale, we were able to create a regression model to fit the temperature data:

$$\text{Model } Y = a_0 U + a_1 \text{COS}(X) + a_2 X + \text{Error}$$

where Y = Temperature and X = Measurement Time in Radians

The amplitude of the wave is, then, represented by the weight (a_1) of the COS(X) variable. The least squares regression solution for this model calculates the amplitude (a_1) in such a manner that the error sums of squares is minimized. This regression model can then be used with either the whole data set (i.e. ten year's data) or with smaller subsets (e.g. one year's data) within the whole data set.

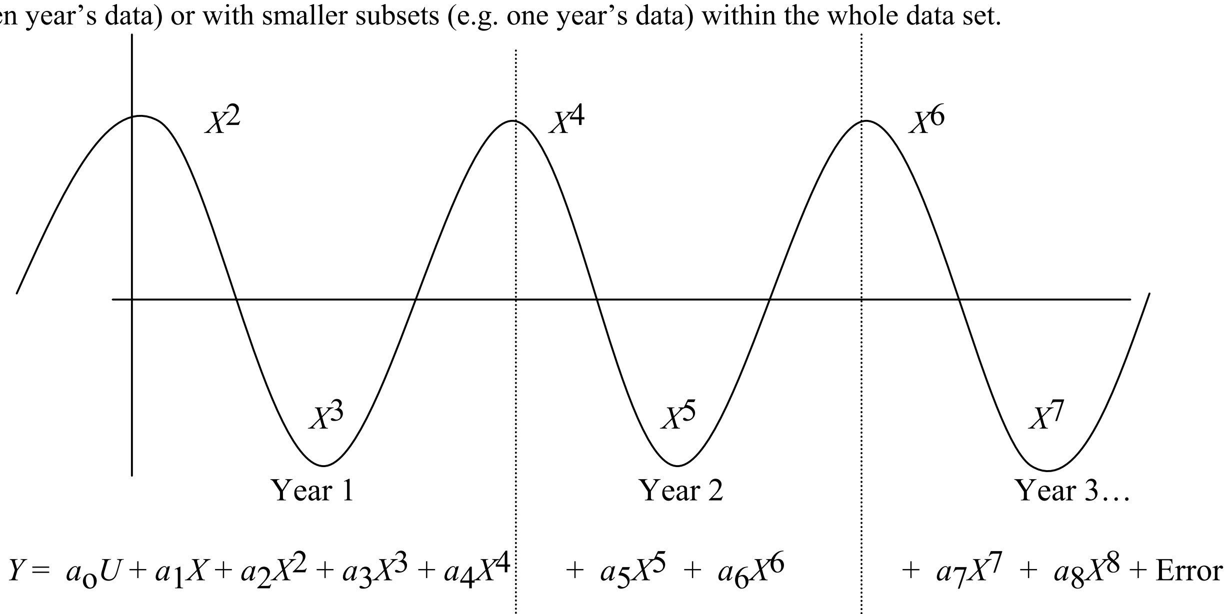

Figure 4. Polynomial representation of a cyclical function.

In order to model the same phenomena using a polynomial, it is necessary to take into account the number of cycles to be modeled when developing a regression equation to fit the data. Each inflection point requires the additional of a polynomial in order to be modeled. Therefore, the model needed to fit one year's data would be as follows:

$$\text{Model} \quad Y = a_0 U + a_1 X + a_2 X^2 + a_3 X^3 + a_4 X^4 + \text{Error}$$

where Y = Temperature and X = Measurement Time in Radians

Each additional year would require the addition of two polynomials in order to model the inflection points in the curve. Ten year's data would require a model extending to the 22nd power of the time measurement!

Results

In order to demonstrate the differences between these modeling methods, both were applied to a longitudinal temperature data set. The data for this analysis were collected from a web site maintained by Utah Climate Center, Utah State University (http://climate.usu.edu/free/default2.htm). This web site contains weather data for hundreds of sites nationally. The temperature data for this study were taken, specifically, from a site in Colorado (Akron – site # 5011403) and consist of daily maximum temperature readings for a ten year span beginning January 1, 1980.

The data were first modeled using a polynomial function and subsequently modeled using the cosine function. The analysis began with a sample of data from a single year; and, then, proceeded to a sample from five year's data. Each sample, ten percent of the cases, was drawn randomly from the time range sampled. Cross-validations were based on random, matched (ten percent), samples. The results of these analyses are in Table 1.

In the first year, the polynomial function accounted for slightly more variance than did the cosine function. As can be seen in the column of shrunken R-squares, both functions were stable when cross-validated. At five years, the cosine function accounted for more variance in the temperature data and was considerably more stable when cross-validated (12.2% versus 95.9% shrinkage).

The problem of shrinkage with the polynomial functions is further exacerbated when one attempts to model additional cycles (years) of data. To demonstrate this point, we attempted to model ten years of daily maximum temperatures using both polynomial and cosine functions. Again, a ten percent sample of the cases, was drawn randomly from the time range (ten years) sampled. Cross-validations were based on random, matched (ten percent), samples.

To demonstrate the explanatory power of the polynomials, the data was first modeled with only a second degree polynomial. Each subsequent model added an additional polynomial to the model until the final model included the second to eighth degree polynomials. As can be seen from Table 2, the addition of higher order polynomials incrementally adds to the variance accounted for by the model, but these models are not stable when cross-validated. In turn, the polynomial models did not account for nearly as much variance as did the cosine function, which was also stable when cross-validated.

Table 1. Modeling One and Five Years Data: Polynomials versus Cosine Function

Model	*R*-square	Adjusted *R*-square	Shrunken *R*-square	Percent of Shrinkage
Polynomial 1 year	.7622	.7334	.7367	3.8%
Polynomial 5 years	.3044	.2651	.0123	95.9%
Cosine 1 year	.7311	.7157	.6951	4.9%
Cosine 5 years	.6551	.6514	.5751	12.2%

Table 2. Modeling Ten Years Data: Polynomials versus Cosine Function

Model	Largest Polynomial in the Model	*R*-square	Shrunken *R*-square
Polynomial	X^2	.0018	.0016
Polynomial	X^3	.0042	.0007
Polynomial	X^4	.0072	.0003
Polynomial	X^5	.0182	.0005
Polynomial	X^6	.0498	.0007
Polynomial	X^7	.0508	.0010
Polynomial	X^8	.0883	.0007
Cosine	NA	.6718	.6537

Note: A weight of zero is applied to higher order polynomials (>8th power) when they are modeled.

Discussion

The purpose of this paper was to provide an answer to the question of the relative effectiveness of the cosine function versus a polynomial function in the description and stability of prediction of a specific set of longitudinal data. With the present data set, the cosine function clearly provided a more stable prediction of the maximum daily temperatures in Akron, Colorado between the years of 1980 and 1990. The benefit of using a cosine function to predict the temperature scores was particularly evident when more than one year of data was modeled. When only one year's data was modeled, there was a slight advantage to using a polynomial model to predict temperature. The polynomial models appear to capitalize on unique variance in the sample. On the other hand, the cosine function appears to be relatively stable across samples and time. Therefore, if the data conforms to a known function, using the function to model the data, especially when numerous cycles are to be modeled, would give one a more stable prediction and hence greater confidence when generalizing beyond the specific sample used in a given study. Currently, we are beginning to work to determine the point at which there is a clear advantage for using one method rather than the other.

There are times, however, when the two modeling methods could be used in a complementary fashion. If one is unsure as to whether a known function exists within a data set, one could model a small interval (2 or 3 cycles) of the data using a polynomial. If the polynomial model is suggestive of a know function (e.g. cosine), one could then model the following intervals using the known function. In this manner, the two modeling methods could be used in conjunction with one and other.

References

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlational analysis for the behavioral sciences*. (2nd ed.). Hillside, NJ; Erlbaum.

Coren, S. (1996). Accidental death and the shift to daylight savings time. *Perceptual & Motor Skills*, *83*(3, Pt 1), 921-922.

Glod, C., & Baisden, N. (1999). Seasonal affective disorder in children and adolescents. *Journal of the American Psychiatric Nurses Association*, *5*(1), 29-33.

McNeil, K., Newman, I., & Kelly, F. (1996). *Testing research hypotheses with the general linear model*. Carbondale, IL: Southern Illinois Press.

Partonen, T., Piiroinen, P., & Loennqvist, J. (2000). Season-dependent symptoms in consultation-liaison patients. *International Journal of Psychiatry in Clinical Practice*, *4*(2),151-154.

Pedhazur, E., & Pedhazur-Schmelkin. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Phillips, D., & Ryan, N. (2000). An abrupt shift in U.S. suicide levels around the month boundary. In R.W. Maris (Ed.), *Review of Suicidology, 2000*. (pp. 34-44). New Yorkz; Guilford Press.

Rohan, K., & Sigmon, S. (2000). Seasonal mood patterns in a northeastern college sample. *Journal of Affective Disorders*, *59*(2), 85-96.

Send correspondence to: Russell Brown, 200 Strecker Drive, Tallmadge, OH 44278
Email: rcbrown@uakron.edu.