POLYNOMIAL REGRESSION (Chapter 9)

We have discussed curvilinear regression from transformations and polynomials

1) Transformations generally more interpretable, often more easily interpreted in terms of a possible functional relationship. (extrapolation, interpretation of parameters)

However, there are cases where the functional relationship is a polynomial. Parabolic shapes exist for a number of relationships, particularly in engineering.

2) Polynomials very flexible, and useful where a model must be developed empirically. They fit a wide range of curvature.

use when the functional relationship is overly complicated, but a repeatable pattern of the dependent variable is likely.

When applied in this context, we generally fit a quadratic, cubic, maybe a quartic, and then see if we can reduce the model by a few terms.

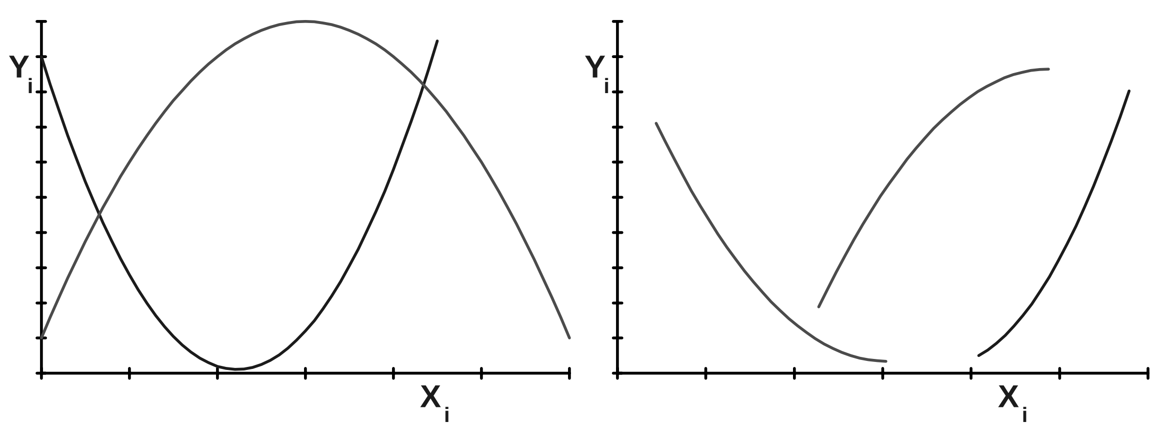
In this case, the polynomial may provide a good approximation of the relationship

Basically, with polynomials we can

1) Determine if there is a curvilinear relationship between the Y_i and X_i.
2) Determine if the curvature is Quadratic? Cubic? Quartic? ...
3) Obtain the curvilinear predictive equation for Y_i on X_i.

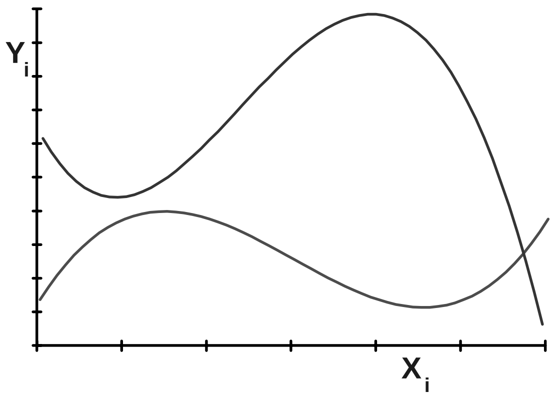
Simplest Polynomial : One independent variable - Second order

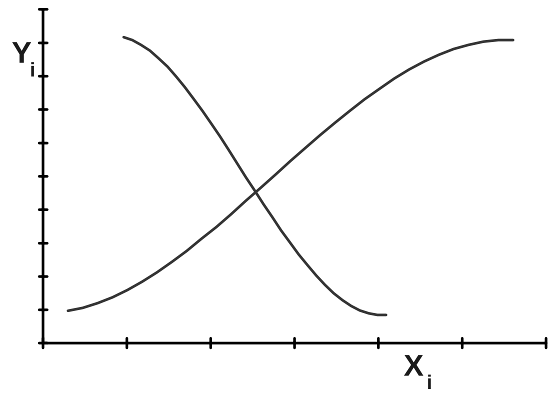
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$



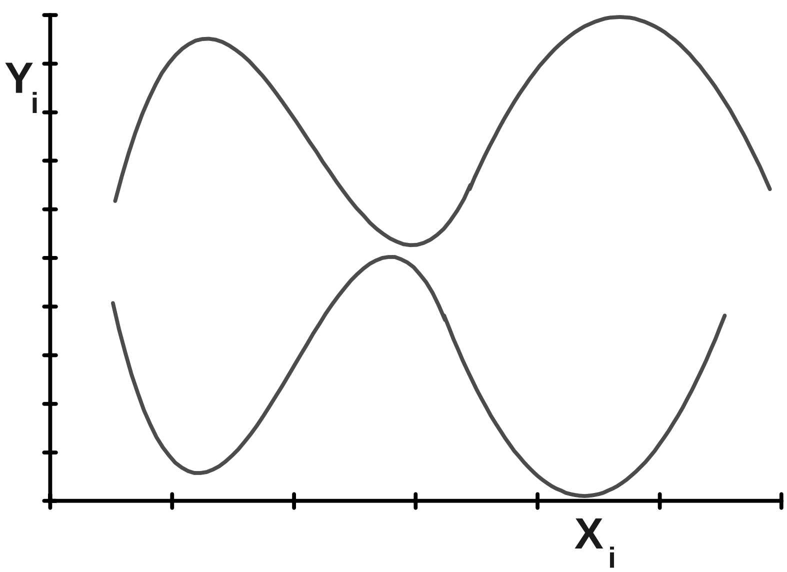
Next larger Polynomial : One independent variable - Third order

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \epsilon_i$$





$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \beta_4 X_i^4 + \epsilon_i$$



NOTES on POLYNOMIAL REGRESSION

1) Polynomial regressions are fitted successively starting with the linear term (a first order polynomial). These are tested in order, so Sequential SS are appropriate.

2) When the highest order term is determined, then all lower order terms are also included.

If for instance we fit a fifth order polynomial, and only the CUBIC term is significant, then we would OMIT THE HIGHER ORDER NON-SIGNIFICANT TERMS, BUT RETAIN THOSE TERMS OF SMALLER ORDER THAN THE CUBIC.

This does not mean that $Y_i = b_0 + b_1X_i + e_i$ is not a useful model, only that this is not a "polynomial".

3) If there are s different values of X_i, then s-1 polynomial terms (plus the intercept) will pass through every point (or the mean of every point if there are more than one observation per X_i value.

It is often recommended that not more than 1/3 of the total number of points (different X_i values) be tied up in polynomial terms.

eg. If we are fitting a polynomial to the 12 months of the year, don't use more than 4 polynomial terms (quartic).

4) All of the assumptions for regression apply to polynomials.

5) Polynomials are WORTHLESS outside the range of observed data, do not try to extend predictions beyond this range. Extrapolation is useless unless the functional relationship is actually a parabola.

INFLECTIONS for Polynomial Regression lines

Linear – straight line, no curve or inflections

Quadratic – one parabolic curve, no inflections

Cubic – two parabolic rates of curvature with the possibility of an

inflection point.

Each additional term allows for another change in the rate of curvature and allows for an additional inflection.

APPLICATIONS of Polynomial Regression lines

1) If enough polynomial terms are used, these curves will fit about anything. However, there is usually no good theoretical reason for using polynomial curves.

eg. Suppose we have a model where we expect an exponential type growth curve to result. We could fit this with a quadratic or cubic or quartic polynomial, but the exponential curve would fit with two advantages.

a) Good interpretation of the regression coefficient (proportional growth)

b) Uses fewer d.f. in a simpler model.

2) Polynomials are useful for testing for the presence of curvature, and the nature of that curvature (inflections or no), and can be used to fit trends with complex curvature where no particular theoretical function is known to be applicable.

This is also a useful “covariable" in designs

3) The successive terms in polynomials are highly correlated. This is not a problem when Sequential SS are used.

4) Recall Lack of Fit. Each individual X_{ji} value has a mean, and the Pure error results from fitting this mean and calculating deviations from this mean.

If a high order polynomial is used such that the order is one less than the number of different X_j values, then Pure error is obtained.

eg. Two different X_j values can be fitted to a line
Three different X_j values can be fitted to a line + quadratic, etc.

Therefore, fitting too high an order polynomial is no more meaningful as a “regression" than fitting Pure error (ie all X_j are categories). Conceptually, we GAIN understanding and interpretability when we fit, say 12 levels, with only 2 or 3 degrees of freedom (hopefully with no Lack of Fit) as opposed to 11.

Therefore, a meaningful fit should be provided by a relatively low order poly (generally no more than $\frac{1}{3}$ of the possible dg, certainly no more than $\frac{1}{2}$), and hopeful such that there is no Lack of Fit. Otherwise, you may as well go to ANOVA with CLASSES X;.

When fitting polynomials, particularly high order polynomials, there are a number of problems.

1) Variables can get very large. Imagine a model done over time (within a year),

We may regress on the julian data (from 1 to 365 days)

If we fit only a 4 order polynomial, by mid-Dec we are up to 350^4=15,006,250

The corresponding regression coefficient will be very small. SAS advises when the regressions coefficients get too small to insure a minimum of precision.

A simple solution is to rescale the regression, use months 1.0 to 12.9, or even year = days*$\frac{1}{365}$ = a decimal year.

2) There is a high correlation between the different variables which are powers of X.

This causes problems with regressions coefficients.

The ability to get a good predictive equation is not impaired, (unless the regression coefficients get too small or large

also, the actual values of the coefficients themselves are unstable, a common consequence of multicolinearity

Fortunately, the solution which helps with multicolinearity will also help rounding error and large or small errors.

a) One possibility is orthogonal polynomials. These are transformations of the powers of X and its powers to extract pure linear, quadratic, cubic, etc.

These are uncorrelated, and especially useful for design problems.
eg. Any three X_i values which are equally spaced can be represented as

Linear	-1	0	1
Quadratic	-1	2	-1

Any four equally spaced X_i values can be replaced by

Linear	-3	-1	1	3
Quadratic	-1	1	1	-1
Cubic	-1	3	-3	1

Any four equally spaced X_i values can be replaced by

Linear	-2	-1	0	1	2
Quadratic	2	-1	-2	-1	2
Cubic	-1	2	0	-2	1
Quartic	1	-4	6	-4	1

These are available in tables for equally spaced values. They can be obtained in SAS with the PROC IML “ORPOL" statement.

These are used by simply entering the coefficients in place of the X or X^2 or X^3

These are commonly used to subdivide TREATMENT SS in a design into components

Orthogonal polynomials
Advantages

1) These are orthogonal, with all of the desirable properties of orthogonal independent variables. The order of entry and adjustment in the model is not important.

2) The have few problems with rounding error or very large (small) regression coefficients

Disadvantages

1) They are abstract. If you want a predicted value for an X which was not observed, you will have difficulty. (eg. if X's are 10, 20 and 30 and you want a predicted value for X=27.
interpolation is not linear and must ge done on each orpol level

2) The values must be obtained and entered by hand in most cases.

b) Another possibility is stabilizing the regression coefficients by just centering them on zero. This is easily done by subtracting the mean of X from each X.

$$x_i = X_i - \bar{X}$$

The higher order terms are then calculated on the x_i

Advantages

1) These are not orthogonal, but will reduce multicolinearity substantially (kind of like a poor mans Orthogonal Polynomial)

2) The have few problems with rounding error or very large (small) regression coefficients

3) They are somewhat abstract, but getting values in between the observed is no problem. These values are easily transformed and detransformed.

Disadvantages

1) This will reduce multicolinearity, but not as much as orthogonal multipliers. This reduction is strongest for lower orders, and deteriorates for higher orders.

2) The order of entry must te considered to some degree.

Detransforming deviation scores

Starting with highest order (where b′ is the original)

$b'_{11} = b_{11}$, this reg coeff has same units in both cases

$b'_1 = b_1 - 2b_{11}\bar{X}$

$b'_0 = b_0 - b_1\bar{X} + b_{11}\bar{X}^2$ for the intercept (we have seen this calculation before)

Applying this to our example,

Reduced model with Di values

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	705.473810	3.20799304	219.911	0.0001
D1	1	54.892857	1.05006221	52.276	0.0001
D2	1	-4.248810	0.60625370	-7.008	0.0001

$b'_{11} = b_{11} = -4.2488$

$b'_1 = b_1 - 2b_{11}\bar{X} = 54.8928 - 2*(-4.2488)*\frac{42}{14} = 80.3856$

$$b'_0 = b_0 - b_1\bar{X} - b_{11}\bar{X}^2$$
$$= 705.4738 - 54.8928*\frac{42}{14} + (-4.2488)\left[\frac{42}{14}\right]^2 = 502.5560$$

These values are the same as those obtained by fitting the polynomial directly on X and X^2

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	502.555952	4.85002959	103.619	0.0001
X1	1	80.385714	3.78605314	21.232	0.0001
X2	1	-4.248810	0.60625370	-7.008	0.0001

These detransformations are possible, and may prevent some rounding error. However, unless the rounding error is extreme, it is easier to fit the X values directly.

Summary of Polynomials

1) A model consisting of successive power terms. Each model will include the highest order term plus all lower order terms (significant or not).

The models are order dependent and Type I SS should be used for interpretation.

2) Polynomial models are an effective and flexible curve fitting technique. They are not usually based on theoretical derivations, and do not usually address an underlying functional model.

These models particularly should not be extended outside the range of observed data!!!!!!!!!!

3) As with many models exhibiting multicolinearity, the regression coefficients and their variance is not likely to be stable.

Use of a transform, $x_i=(X_i - \bar{X})$, may help stabilize
Even if this transformation is not used, it may be necessary to rescale the X_i values to avoid precision and rounding error problems.

4) Orthogonal polynomials are completely uncorrelated and will provide essentially perfect stability and either type SS can be used (they are the same).

5) If enough polynomial terms are added (one for each X_i) then a perfect fit is achieved to each $\bar{X}$ (just as if an ANOVA of X_i was done).

If fewer terms are used, then the difference between the model with the maximum number of power terms and the number of power terms actually employed is LOF.

Aptness of the model and confidence intervals :

1) In terms of judging the adequacy of the regression line, all of the usual diagnostics of the residuals apply.

NOTE that, unlike other multiple regressions, residual plots can be done directly on X, instead of $\hat{Y}$.

2) To determine the highest order polynomial, most of our multiple regression techniques apply.

If fitted on orthogonal polynomials, the highest significant order is determined. All lower orders are retained, all higher orders deleted.

If ordinary least squares are used (using either D or X directly), the variables can be examined in order: X_1; $X_1|X_2$; $X_3|X_1$, X_2; etc. which is equivalent to the TYPE I SS.

Additional use in judging adequacy of the model can be made from evaluating the Lack of Fit, but if there are k different X values, and k polynomial terms are fitted, then there is no Lack of Fit remaining.

(Note: LoF is essentially ΣSSR(X_k^k|all lower order terms)

Therefore, if LoF is significant, it can also be interpreted as "a higher order polynomial is needed".

3) Regression coefficients can be examined, but these are essentially equivalent to the the TYPE III SS, and are not useful except for D (not for X directly). However, the regression coefficients are those which provide the best fit, and except for possible multi-colinearity problems can be employed and used as any other regression coefficients.

There is an additional problem in estimating the standard errors with high correlations. D would behave better here (or orthogonal polynomials).

Since the regression coefficients are closely related as coefficients on different powers of the same X, a joint confidence interval is a natural option.

A Bonferroni simultaneous interval is a logical option

$$\hat{Y}_h \pm B^* s_{\hat{Y}_h}$$

where $B = t_{1-\frac{\alpha}{2g};n-2\,df}$

and g is the number of cases

4) The predictive ability of polynomials does not suffer from the multicolinearity aspects of the variables. Predictions and their confidence intervals (including simultaneous intervals) are the same as for other multiple regressions.

My philosophy on fitting polynomials

1) Any better model? (Airplane handout)

2) Fit 1 or 2 more levels than I feel necessary, no more than $\frac{1}{2}$ of df.

3) If a single variable, X, is to be fitted alone, use X, X^2, X^3, etc.

Otherwise, consider Orthogonal Polynomials; also if part of an ANOVA

4) Do all Interpretation with Type I SS, if alone. Consider consequences otherwise.

5) Do not use regression coefficients for interpretation or testing, only to fit predicted values.

I do look at the signs to tell if inflections are present.

Two independent variable - Second order

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \beta_3 X_{2i} + \beta_4 X_{2i}^2 + \beta_5 X_{1i} X_{2i} + \epsilon_i$$

it may be clearer to change the subscripting to

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_{11} X_{1i}^2 + \beta_{22} X_{2i}^2 + \beta_{12} X_{1i} X_{2i} + \epsilon_i$$

Two independent variable - higher orders would proceed as follows

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_{11} X_{1i}^2 + \beta_{22} X_{2i}^2 + \beta_{12} X_{1i} X_{2i}$$

$$\beta_{111} X_{1i}^3 + \beta_{222} X_{2i}^3 + \beta_{112} X_{1i}^2 X_{2i} + \beta_{122} X_{1i} X_{2i}^2 + \epsilon_i$$

note that this contains all 3-way levels

X_1*X_1*X_1, X_2*X_2*X_2, X_1*X_1*X_2, X_1*X_2*X_2,

the next higher order would have all 4 way levels, but we rarely go this far

Three independent variable - Second order

$$Y_i=\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i} + \beta_3 X_{3i}+\beta_{11} X_{1i}^2 + \beta_{22} X_{2i}^2 + \beta_{33} X_{1i}^2 + \beta_{12} X_{1i} X_{2i}$$

$$+ \beta_{13} X_{1i} X_{3i} + \beta_2 X_{2i} X_{3i} + \epsilon_i$$

The use of Polynomials suggests that the variables should be order dependent (if not orthogonal)

However, the use of 2 or more different X variables suggests that the variables should be adjusted for each other.

Who wins?

Hopefully the two X variables are somewhat independent, so which comes first is not to important

Or the objective is prediction, so hypothesis testing is not so critical

However, the sequential approach for polynomials should be utilized as much as possible. In analyzing a second order response surface, you text uses the following progression (although on the D transformation);

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \beta_3 X_{2i} + \beta_4 X_{2i}^2 + \beta_5 X_{1i} X_{2i} + \epsilon_i$$

1) $SSR(X_1)$

2) $SSR(X_2|X_1)$

3) $SSR(X_1^2|X_1,X_2)$

4) $SSR(X_2^2|X_1,X_2,X_1^2)$

5) $SSR(X_1X_2|X_1,X_2,X_1^2,X_2^2)$

This retains the sequential nature of the squares and interactions,
But it is somewhat arbitrary as to whether X_1 comes before or after X_2

All aspects of the analysis proceed pretty much as any multiple regression,

1) Residual analysis is feasible,

2) Hypothesis testing is sequential for variable retention

eg. X_1 is retained if either a higher order term for X_1 OR AN INTERACTION TERM CONTAINING X_1 IS USED IN THE MODEL.

see handout (refer to figure 9.8 in text; a flat surface)

Response Surface Methodology

These are usually two (or more) dimensional representations of a plane, where a polynomial approach is used to fit a complex shape

See text for discussion

1) Factorial designs

2) Optimal response

1 dimension : quadratic - first derivative

the optimum can be estimated even if outside the range of the observed data, this gives an area of X variables for future research

2 dimensions - same idea