NOTE U: POLYNOMIAL REGRESSION

EXCEL 2000

CENTERING INPUT DATA:

In most cases, centering the data about the mean (an approximate mean is OK) will improve accuracy of the result. If the problem requires a polynomial beyond a cubic, centering will be the only way it can be done. The standard coefficient error values will not be the values obtained from fitting a polynomial to the original data. A similar process to convert centered standard coefficient error values back to direct standard coefficient error values has not yet been worked out.

1. The centered data set:
    a. Build a worksheet with the Y, X data. Set column A as the Y data and columns B to {m+1}[1], as the X data. Reserve row 1 for labels that identify each of the powers of the X variable.
    b. In column A, row {n+3} enter =AVERAGE(A2:A{n+1}) . Formula copy across to column B.
    c. Copy the cells in row 1 and paste then in row {n+4}
    d. In column A row {n+5} put in the formula =A2-A${n+3}) and formula copy across to column B.
    e. Formula copy the selected cells down to row {2n+4}.
    f. The centered X, Y starting data will be columns A and B from rows {n+4} to {2n+5).
    g. Generate the polynomial terms.
        i. For quadratic: Starting with cell C{n+5} enter =B{n+5} * B{n+5}
        ii. For cubic: Starting with cell D{n+5} enter =B{n+5} * C{n+5}
        iii. For quadratic: Starting with cell E{n+5} enter =B{n+5} * D{n+5}
        iv. Continue this scheme up to the highest power desired.
    h. The range for the Y data will be A{n+4}:A{2n+5}.
    i. The range for the X variables will be B{n+4}:{m+1}{2n+5}, where m is the highest power of the fitting polynomial.
2. Go into Data Analysis, select Regression. Set Y to the range listed in g. above and set X to the range listed in h. above. Set the labels box.

---

[1] The {….} is a notation whose value has to be converted either to a row number or an alphabetic character corresponding to a column designation. n is the number of "points" or observations in the data set, and m is the number of variables in the data set.

3. Set the Regression output to begin to the left of the centered data.

CONVERTING FROM CENTERED COEFFICIENTS TO DIRECT COFFICIENTS:

1. For fitting polynomials to centered X and Y values, the calculations to convert coefficient values and the intercept to the original coordinate system are more complex. It is a translation of the polynomial from the centered X and Y values to the original coordinates. Algebra can be used to show how this is done. Each term involves a sign, a binomial coefficient, the mean value of the X to a power and the regression coefficient.
2. The following instructions are for any polynomial up to a power of 10.
   a. Build up a table on a worksheet in Excel, from column A to column W and from row 1 to row 12. Columns A to M contain the basic information, and columns N to W contain the cells of the cross products.
   b. Put in the following data into the cells.
      i. Cell N1 is the average Y from the data worksheet Cell A{n+3}.
      ii. Cell N3 is the average X from the data worksheet Cell B{n+3}
      iii. Cells A1:A11 are the Data Analysis Regression output sheet cells corresponding to the column of fitted coefficient values. Put in a zero for any coefficients beyond the ones fitted.
      iv. The center of the table represents the standard binomial coefficients. Column B is a sign control on the term.
      v. In cells N4:N11, put in the indicated formula. These are the powers of the average X.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Intercept	+1	1	0	0	0	0	0	0	0	0	0	0	average y
2	Coef A	-1	1	-1	0	0	0	0	0	0	0	0	0	1
3	Coef B	+1	1	-2	1	0	0	0	0	0	0	0	0	average x
4	Coef C	-1	1	-3	3	-1	0	0	0	0	0	0	0	M3*M3
5	Coef D	+1	1	-4	6	-4	1	0	0	0	0	0	0	M4*M3
6	Coef E	-1	1	-5	10	-10	5	-1	0	0	0	0	0	M5*M3
7	Coef F	+1	1	-6	15	-20	15	-6	1	0	0	0	0	M6*M3
8	Coef G	-1	1	-7	21	-35	35	-21	7	-1	0	0	0	M7*M3
9	Coef H	+1	1	-8	28	-56	70	-56	28	-8	1	0	0	M8*M3
10	Coef I	-1	1	-9	36	-84	126	-126	84	-36	9	-1	0	M9*M3
11	Coef J	+1	1	-10	45	-120	210	-252	210	-120	45	-10	1	M10*M3
12														M11*M3

vi. Put in the following formulas on the diagonals. Then formula copy down to row 11 in each column. It will not properly formula copy across a row.

vii. In cell O12, enter =SUM(O1:O11) and formula copy across to cell Y12. These cells will contain the translated direct coefficient values for the polynomial

	O	P	Q	R	S	T	U	V	W	X	Y
1	A1+N1										
2	A2*B2* C1*N3	A2*B2* D2*N2									
3	A3*B3* C2*N4	A3*B3* D3*N3	A3*B3* E3*N2								
4			A4*B4* E4*N3	A4*B4* F4*N2							
5				A5*B5* F5*N3	A5*B5* G5*N2						
6					A6*B6* G6*N3	A6*B6* H6*N2					
7						A7*B7* H7*N3	A7*B7* I7*N2				
8							A8*B8* I8*N3	A8*B8* J8*N2			
9								A9*B9* J9*N3	A9*B9* K9*N2		
10									A10*B10 * K10*N3	A10*B10* L10*N2	
11										A11*B11* L11*N3	A11*B11* M11*N2
12	SUM (O1: O11)	SUM (P2:P11)	SUM(Q3 :Q11)	SUM (R4:R11)	SUM (S5:S11)	SUM (T6:T11)	SUM (U7:U11)	SUM (V8:V11)	SUM(W9 : W11)	SUM (X10:X11)	SUM (Y11:Y11)

c. Transfer the values in cells O12 to Y12 to the regression output sheet. The standard errors of the coefficients will remain the same.

FILIP POLYNOMIAL FIT

Centering the 82 observations and running Data Analysis – Regression on the data (as a multivariate regression) gave a set of parameter values. The coefficients were translated back to the original coordinate system using the method described above. Other parameters values came from the centered data run. The standard deviations of coefficient values in the original coordinate system cannot be recovered from the centered data standard deviations of coefficient values.

The comparison is given in LRE values.

Parameter	Estimate	Standard Deviation of Estimate
B0	10.28	0
B1	10.28	0
B2	10.27	0
B3	10.27	0
B4	10.26	0
B5	10.25	0
B6	10.24	0
B7	10.23	0
B8	9.25	0
B9	10.40	0
B10	10.20	0

Residual Standard Deviation	14.51
R-Squared	15.95

Analysis of Variance Table

Source	df	SS	MS	F ratio
Regression	16	14.83	14.80	14.33
Residual	16	14.27	14.18	

EXCEL 2003:

The new regression algorithm is robust enough to do polynomial regression directly. No special methods are needed.