# EXST 7015 Fall 2014 Lab 08: Polynomial Regression

### **OBJECTIVES**

Polynomial regression is a statistical modeling technique to fit the curvilinear data that either shows a maximum or a minimum in the curve, or that could show a max or min if you extrapolated the curve beyond your data. The ability to determine a minimum or maximum point based on the experimental data is a useful application of polynomials. The simple polynomial regressions are multiple regression that use power terms of the independent variable (X<sub>i</sub>) with the form of  $Y=\beta_0 + \beta_1 X_i + \beta_2 X_i^2 + ... + \beta_k X_i^k + e_i$ . Notice the subtle difference from multiple linear regression model, here the numbers 2, 3, ...,k represent the powers of the same variable. For data that are shaped like a parabola, you probably won't need more than a quadratic model. If the curve trends up again at one end, you'll need a cubic model. Curves with multiple kinks need even higher-order terms. It is obvious that multi collinearity is unavoidable issue in Polynomial regression because the model terms are related to each other. Use of **sequentially adjusted type I SS** is the solution as presented as the following.

Several hypotheses are tested during polynomial regression which is fitted successively starting with the liner term (a first order polynomial). The first null hypothesis, then, is that a quadratic equation does not fit the data significantly better than a linear equation; the next null hypothesis may be that a cubic equation does not fit the data significantly better than a quadratic equation, and so on. Therefore, the **sequentially adjusted type I SS** should be used when one attempts to test whether a polynomial model is as good as the one with a higher order term. If a particular higher order term is significant, all terms of lower order should be assumed significant and retained in the regression model. There is also a null hypothesis for each equation that says that it does not fit the data significantly better than a horizontal line; in other words, that there is no relationship between the X and Y variables.

It should be noticed that the **fully adjusted regression coefficients** are still used to fit the polynomial regression, which usually leads to no practical explanation of regression coefficients. Further, extrapolation outside the range of the fitted experimental data is untenable, and should not be attempted. This is because the shape of the regression function, as predicted by the model, may not at all accurately represent reality outside the range of the experimental data.

The assumptions for performing polynomial regression are similar to those of ordinary regression. The assumptions of normality and homogeneity can be evaluated by customary diagnostic technique (Shapiro-Wilk, residual plots). Residual and influence statistics still work with these regression models.

## LABORATORY INSTRUCTIONS

#### **Housekeeping Statements**

```
dm 'log; clear; output; clear';
options nodate nocenter pageno = 1 ls=78 ps=53;
title1 'EXST7015 lab 8, Name, Section#';
ods rtf file = 'c:/temp/lab8.rtf';
ods html file = 'c:/temp/lab8.html';
```

#### Data Set:

The data set that we use contains the per capita state and local public expenditures and associated state demographic and economic characteristics for 48 states during the year of 1960. Detailed information can be found at http://lib.stat.cmu.edu/DASL/Datafiles/pubexpendat.html. The variables in the dataset are: **EX**: Per capita state and local public expenditures (\$) **ECAB**: Economic ability index, in which income, retail sales, and the value of output (manufactures, mineral, and agricultural) per capita are equally weighted. **MET**: Percentage of population living in standard metropolitan areas **GROW**: Percent change in population, 1950-1960 **YOUNG**: Percent of population aged 5-19 years **OLD**: Percent of population over 65 years of age **WEST**: Western state (1) or not (0)

In this lab, we will use MET, Percentage of population living in standard metropolitan area, as the independent variable and the expenditure (EX) as the dependent variable. Other variables are dropped from the dataset.

```
data expenditure;
input ex ecab met grow young old west state$;
drop ecab grow young old west;
cards;
256
     85.5 19.7 6.9
                            29.6 11.0
                                       0
                                             ME
275
     94.3 17.7 14.7 26.4 11.2
                                  0
                                       NH
327
     87.0 0.0
                      3.7
                                  28.5
                                       11.2
                                             0
                                                   VΤ
     107.5 85.2 10.2 25.1 11.1
                                  0
297
                                       MA
                            25.3 10.4
256
    94.9 86.2 1.0
                                        0
                                             RI
     121.6 77.6 25.4 25.2
312
                            9.6
                                        0
                                             CT
     111.5 85.5 12.9 24.0
                                       NY
374
                            10.1
                                  0
     117.9 78.9 25.5 24.8 9.2
257
                                        0
                                             NJ
257
     103.1 77.9 7.8
                            25.7 10.0
                                        0
                                             PA
     116.1 68.8 39.9 26.4 8.0
336
                                        0
                                             DE
     93.4 78.2 31.1 27.5
                            7.3
                                        0
                                             MD
269
213
     77.2 50.9 21.9 28.8
                            7.3
                                        0
                                             VA
308
     108.4 73.1 22.2 28.0
                            8.2
                                        0
                                             MI
273
     111.8 69.5
                 21.8
                      26.9
                            9.2
                                        0
                                             OH
     110.8 48.1
                18.3
                      27.5
                            9.6
256
                                        0
                                             IN
287
     120.9 76.9
                15.5
                      25.4
                            9.7
                                        0
                                             IL
290
     104.3 46.3
                14.9
                      27.4
                            10.2 0
                                       WI
     85.1 30.9 -7.4 30.0
                                             WV
217
                            9.3
                                        0
198
   76.8 34.1 0.3
                            29.4 9.6
                                             0
                                                   KΥ
```

195       78.7       24.6       12.4       30.8       6.9       0       NC         183       65.2       32.2       12.9       32.9       6.3       0       SC         222       73.0       46.0       14.4       30.0       7.4       0       GA         223       80.9       65.6       77.2       25.5       11.2       0       FL         217       69.4       45.6       7.0       30.5       8.0       1       AL         329       95.7       51.3       14.4       28.8       10.4       1       MN         224       100.2       33.2       5.3       27.3       11.9       1       IA         231       95.7       51.3       14.4       28.8       10.4       1       MN         232       99.1       57.9       9.8       25.6       11.7       1       MO         369       93.4       10.6       2.9       30.2       9.3       1       ND         302       88.2       12.7       4.6       28.9       10.5       1       SD         210       23.7.4       13.7       26.8       11.0       1       KS<											
183       65.2       32.2       12.9       32.9       6.3       0       SC         222       73.0       46.0       14.4       30.0       7.4       0       GA         283       80.9       65.6       77.2       25.5       11.2       0       FL         217       69.4       45.6       7.0       30.5       8.0       1       AL         231       57.4       8.6       0.5       32.1       8.7       1       MS         232       95.7       51.3       14.4       28.8       10.4       1       MN         232       99.1       57.9       9.8       25.6       11.7       1       MO         369       93.4       10.6       2.9       30.2       9.3       1       ND         302       88.2       12.7       4.6       28.9       10.5       1       SD         210       10.2.3       37.4       13.7       26.8       11.0       1       KS         323       86.0       5.0       21.9       30.3       7.4       1       LA         324       84.9       43.9       6.4       27.4       10.7       1 <td>217</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>8.7</td> <td></td> <td>0</td> <td>TE</td> <td></td>	217						8.7		0	TE	
222       73.0       46.0       14.4       30.0       7.4       0       GA         283       80.9       65.6       77.2       25.5       11.2       0       FL         217       69.4       45.6       7.0       30.5       8.0       1       AL         213       57.4       8.6       0.5       32.1       8.7       1       MS         329       95.7       51.3       14.4       28.8       10.4       1       MN         294       100.2       33.2       5.3       27.3       11.9       1       IA         232       99.1       57.6       6.8       28.9       10.5       SD       30.2       9.3       1       ND         302       88.2       12.7       4.6       28.9       10.5       SD       SD         269       99.1       37.6       6.8       26.6       11.6       1       NB         211       102.2       37.4       13.7       26.8       11.0       1       KS         323       86.0       5.0       21.9       30.3       7.4       1       IA         282       84.9       9.4       31.5 </td <td>195</td> <td></td>	195										
283       80.9       65.6       77.2       25.5       11.2       0       FL         217       69.4       45.6       7.0       30.5       8.0       1       AL         211       57.4       8.6       0.5       32.1       8.7       1       MS         329       95.7       51.3       14.4       28.8       10.4       1       MN         224       100.2       33.2       5.3       27.3       11.9       1       MA         232       99.1       57.9       9.8       25.6       11.7       1       MO         369       93.4       10.6       2.9       30.2       9.3       1       ND         302       88.2       12.7       4.6       28.9       10.5       1       SD         269       99.1       37.6       6.8       28.9       10.1       KS       SZ         323       86.0       5.0       21.9       30.3       7.4       1       LA         198       68.6       19.1       -6.2       29.4       10.7       1       NK         246       98.8       63.4       24.1       28.8       7.8       1 <td>183</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>0</td> <td>SC</td> <td></td> <td></td>	183							0	SC		
217       69.4       45.6       7.0       30.5       8.0       1       AL         231       57.4       8.6       0.5       32.1       8.7       1       MS         232       95.7       51.3       14.4       28.8       10.4       1       MN         294       100.2       33.2       5.3       27.3       11.9       1       MO         294       106.2       33.2       5.3       27.3       11.9       1       MO         232       99.1       57.7       9.8       25.6       11.7       1       MO         302       88.2       12.7       4.6       28.9       10.5       1       SD         269       99.1       37.6       6.8       26.6       11.6       1       NB         21       102.2       37.4       13.7       26.8       10.0       1       AR         233       86.0       5.0       21.9       30.3       7.4       1       LA         198       68.6       19.1       -6.2       29.4       10.7       1       MK         246       98.8       63.4       24.1       28.8       7.8       1 <td>222</td> <td>73.0</td> <td>46.0</td> <td>14.4</td> <td>30.0</td> <td>7.4</td> <td></td> <td>0</td> <td>GA</td> <td></td> <td></td>	222	73.0	46.0	14.4	30.0	7.4		0	GA		
231       57.4       8.6       0.5       32.1       8.7       1       MS         329       95.7       51.3       14.4       28.8       10.4       1       MN         294       100.2       33.2       5.3       27.3       11.9       1       IA         294       100.2       33.2       5.3       27.3       11.9       1       IA         294       100.2       33.2       5.3       27.3       11.9       1       IA         232       99.1       57.9       9.8       25.6       11.7       1       MO         369       93.4       10.6       2.9       30.2       9.3       1       ND         302       88.2       12.7       4.6       28.9       10.5       1       SD         269       99.1       37.6       6.8       26.6       11.6       1       NB         210       102.2       37.4       13.7       26.8       10.9       1       AR         282       84.9       43.9       6.4       27.4       10.7       1       MT         224       98.8       63.4       24.1       28.8       7.8       1<	283	80.9	65.6	77.2	25.5	11.2	0	FL			
329       95.7       51.3       14.4       28.8       10.4       1       MN         294       100.2       33.2       5.3       27.3       11.9       1       IA         232       99.1       57.9       9.8       25.6       11.7       1       MO         369       93.4       10.6       2.9       30.2       9.3       1       ND         362       88.2       12.7       4.6       28.9       10.5       1       SD         269       99.1       37.6       6.8       26.6       11.6       1       NB         291       102.2       37.4       13.7       26.8       11.0       1       KS         323       86.0       5.0       21.9       30.3       7.4       1       LA         102.2       37.4       13.7       26.8       11.0       1       KS         323       86.0       5.0       21.9       30.3       7.4       1       LA         284       94.3       9.6       6.4       27.4       10.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       MY <td>217</td> <td>69.4</td> <td>45.6</td> <td>7.0</td> <td></td> <td>30.5</td> <td>8.0</td> <td></td> <td>1</td> <td>AL</td> <td></td>	217	69.4	45.6	7.0		30.5	8.0		1	AL	
294       100.2       33.2       5.3       27.3       11.9       1       IA         232       99.1       57.9       9.8       25.6       11.7       1       MO         369       93.4       10.6       2.9       30.2       9.3       1       ND         302       88.2       12.7       4.6       28.9       10.5       1       SD         269       99.1       37.6       6.8       26.6       11.6       1       NE         21       102.2       37.4       13.7       26.8       11.0       1       KS         323       86.0       5.0       21.9       30.3       7.4       1       LA         198       68.6       19.1       -6.2       29.4       10.9       1       AR         282       84.9       43.9       6.4       27.4       10.7       1       OK         246       98.8       63.4       24.1       28.8       7.8       1       TX         309       90.2       71.4       74.3       29.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       UT	231	57.4	8.6		0.5		32.1	8.7		1	MS
232       99.1       57.9       9.8       25.6       11.7       1       MO         369       93.4       10.6       2.9       30.2       9.3       1       ND         302       88.2       12.7       4.6       28.9       10.5       1       SD         269       99.1       37.6       6.8       26.6       11.6       1       NB         211       102.2       37.4       13.7       26.8       11.0       1       KS         323       86.0       5.0       21.9       30.3       7.4       1       LA         18       68.6       19.1       -6.2       29.4       10.9       1       AR         282       84.9       43.9       6.4       27.4       10.7       1       OK         246       98.8       63.4       24.1       28.8       7.8       1       NM         309       86.2       27.6       39.4       31.5       5.4       1       NM         309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.0       1 <td>329</td> <td>95.7</td> <td>51.3</td> <td>14.4</td> <td>28.8</td> <td>10.4</td> <td>1</td> <td>MN</td> <td></td> <td></td> <td></td>	329	95.7	51.3	14.4	28.8	10.4	1	MN			
369       93.4       10.6       2.9       30.2       9.3       1       ND         302       88.2       12.7       4.6       28.9       10.5       1       SD         269       99.1       37.6       6.8       26.6       11.6       1       NB         291       102.2       37.4       13.7       26.8       11.0       1       KS         323       86.0       5.0       21.9       30.3       7.4       1       LA         198       68.6       19.1       -6.2       29.4       10.7       1       OK         282       84.9       43.9       6.4       27.4       10.7       1       OK         246       98.8       63.4       24.1       28.8       7.8       1       TX         309       86.2       27.6       39.4       31.5       5.4       1       NM         309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1 <td>294</td> <td>100.2</td> <td>33.2</td> <td>5.3</td> <td></td> <td>27.3</td> <td>11.9</td> <td>1</td> <td>IA</td> <td></td> <td></td>	294	100.2	33.2	5.3		27.3	11.9	1	IA		
302       88.2       12.7       4.6       28.9       10.5       1       SD         269       99.1       37.6       6.8       26.6       11.6       1       NB         291       102.2       37.4       13.7       26.8       11.0       1       KS         323       86.0       5.0       21.9       30.3       7.4       1       LA         198       68.6       19.1       -6.2       29.4       10.9       1       AR         282       84.9       43.9       6.4       27.4       10.7       1       OK         246       98.8       63.4       24.1       28.8       7.8       1       TX         309       86.2       27.6       39.4       31.5       5.4       1       NM         309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       UT         334       98.0       6.8       31.5       28.0       9.0       1 <td>232</td> <td>99.1</td> <td>57.9</td> <td>9.8</td> <td></td> <td>25.6</td> <td>11.7</td> <td>1</td> <td>MO</td> <td></td> <td></td>	232	99.1	57.9	9.8		25.6	11.7	1	MO		
269       99.1       37.6       6.8       26.6       11.6       1       NB         291       102.2       37.4       13.7       26.8       11.0       1       KS         323       86.0       5.0       21.9       30.3       7.4       1       LA         198       68.6       19.1       -6.2       29.4       10.9       1       AR         282       84.9       43.9       6.4       27.4       10.7       1       OK         246       98.8       63.4       24.1       28.8       7.8       1       TX         309       86.2       27.6       39.4       31.5       5.4       1       NM         309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       UT         344       98.0       6.8       31.5       28.0       9.0       1       CO         333       100.4       63.1       19.9       27.5       9.8       1 <td>369</td> <td>93.4</td> <td>10.6</td> <td>2.9</td> <td></td> <td>30.2</td> <td>9.3</td> <td></td> <td>1</td> <td>ND</td> <td></td>	369	93.4	10.6	2.9		30.2	9.3		1	ND	
291       102.2       37.4       13.7       26.8       11.0       1       KS         323       86.0       5.0       21.9       30.3       7.4       1       LA         198       68.6       19.1       -6.2       29.4       10.9       1       AR         282       84.9       43.9       6.4       27.4       10.7       1       OK         246       98.8       63.4       24.1       28.8       7.8       1       TX         309       86.2       27.6       39.4       31.5       5.4       1       NM         309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       ID         454       125.8       0.0       13.7       29.1       7.8       1       WT         334       98.0       6.4       15.7       27.7       10.4       1       OR         421       205.0       7.4.2       77.8       26.6       6.4       1<	302	88.2	12.7	4.6		28.9	10.5	1	SD		
323       86.0       5.0       21.9       30.3       7.4       1       LA         198       68.6       19.1       -6.2       29.4       10.9       1       AR         282       84.9       43.9       6.4       27.4       10.7       1       OK         246       98.8       63.4       24.1       28.8       7.8       1       TX         309       86.2       27.6       39.4       31.5       5.4       1       NM         309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       ID         454       125.8       0.0       13.7       29.1       7.8       1       WY         344       98.0       6.8       31.5       28.0       9.0       1       CO         307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1	269	99.1	37.6	6.8		26.6	11.6	1	NB		
198       68.6       19.1       -6.2       29.4       10.9       1       AR         282       84.9       43.9       6.4       27.4       10.7       1       OK         246       98.8       63.4       24.1       28.8       7.8       1       TX         309       86.2       27.6       39.4       31.5       5.4       1       NM         309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       ID         454       125.8       0.0       13.7       29.1       7.8       1       WY         344       98.0       6.8       31.5       28.0       9.0       1       CO         307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1       WA         421       205.0       74.2       77.8       25.6       6.4       1 <td>291</td> <td>102.2</td> <td>37.4</td> <td>13.7</td> <td>26.8</td> <td>11.0</td> <td>1</td> <td>KS</td> <td></td> <td></td> <td></td>	291	102.2	37.4	13.7	26.8	11.0	1	KS			
282       84.9       43.9       6.4       27.4       10.7       1       OK         246       98.8       63.4       24.1       28.8       7.8       1       TX         309       86.2       27.6       39.4       31.5       5.4       1       NM         309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       ID         454       125.8       0.0       13.7       29.1       7.8       1       WY         344       98.0       6.8       31.5       28.0       9.0       1       CO         307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1       WA         343       98.0       50.4       15.7       27.7       10.4       1       OR         421       205.0       74.2       77.8       25.6       6.4       1 <td>323</td> <td>86.0</td> <td>5.0</td> <td></td> <td>21.9</td> <td>30.3</td> <td>7.4</td> <td></td> <td>1</td> <td>LA</td> <td></td>	323	86.0	5.0		21.9	30.3	7.4		1	LA	
246       98.8       63.4       24.1       28.8       7.8       1       TX         309       86.2       27.6       39.4       31.5       5.4       1       NM         309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       ID         454       125.8       0.0       13.7       29.1       7.8       1       WY         344       98.0       6.8       31.5       28.0       9.0       1       CO         307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1       WA         343       98.0       50.4       15.7       27.7       10.4       1       OR         421       205.0       74.2       77.8       25.6       6.4       1       NV         380       112.6       86.5       48.5       26.2       8.8       1 </td <td>198</td> <td>68.6</td> <td>19.1</td> <td>-6.2</td> <td>29.4</td> <td>10.9</td> <td>1</td> <td>AR</td> <td></td> <td></td> <td></td>	198	68.6	19.1	-6.2	29.4	10.9	1	AR			
309       86.2       27.6       39.4       31.5       5.4       1       NM         309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       ID         454       125.8       0.0       13.7       29.1       7.8       1       WY         344       98.0       6.8       31.5       28.0       9.0       1       CO         307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1       WA         343       98.0       50.4       15.7       27.7       10.4       1       OR         421       205.0       74.2       77.8       25.6       6.4       1       NV         380       112.6       86.5       48.5       26.2       8.8       1       CA         proc       pplot       data=expenditure;       gata at	282	84.9	43.9	6.4		27.4	10.7	1	OK		
309       90.2       71.4       74.3       29.7       6.9       1       AZ         334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       ID         454       125.8       0.0       13.7       29.1       7.8       1       WY         344       98.0       6.8       31.5       28.0       9.0       1       CO         307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1       WA         343       98.0       50.4       15.7       27.7       10.4       1       OR         421       205.0       74.2       77.8       25.6       6.4       1       NV         380       112.6       86.5       48.5       26.2       8.8       1       CA         proc       print       data=expenditure;       point       point       gata=expenditure;       point       gata=expenditure;         plot       ex*met;       title3       Plot o	246	98.8	63.4	24.1	28.8	7.8		1	TX		
334       97.6       22.6       13.4       28.9       9.7       1       MT         284       93.9       0.0       13.3       30.7       8.7       1       ID         454       125.8       0.0       13.7       29.1       7.8       1       WY         344       98.0       6.8       31.5       28.0       9.0       1       CO         307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1       WA         343       98.0       50.4       15.7       27.7       10.4       1       OR         421       205.0       74.2       77.8       25.6       6.4       1       NV         380       112.6       86.5       48.5       26.2       8.8       1       CA         proc       print       data=expenditure;       print;       poot       guidata       piot       guidata       piot         plot       ex*met;       title3       "Plot of raw data";       symbol1       c=red v=dot;       idata       idata       idata       idata       idata <td>309</td> <td>86.2</td> <td>27.6</td> <td>39.4</td> <td>31.5</td> <td>5.4</td> <td></td> <td>1</td> <td>NM</td> <td></td> <td></td>	309	86.2	27.6	39.4	31.5	5.4		1	NM		
284       93.9       0.0       13.3       30.7       8.7       1       ID         454       125.8       0.0       13.7       29.1       7.8       1       WY         344       98.0       6.8       31.5       28.0       9.0       1       CO         307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1       WA         343       98.0       50.4       15.7       27.7       10.4       1       OR         421       205.0       74.2       77.8       25.6       6.4       1       NV         380       112.6       86.5       48.5       26.2       8.8       1       CA         Proc       print       data=expenditure;       print;       proc       gplot       data=expenditure;         plot       ex*met;       title3       "Plot of raw data";       symbol1       c=red v=dot;       symbol1       c=red v=dot;	309	90.2	71.4	74.3	29.7	6.9		1	AZ		
284       93.9       0.0       13.3       30.7       8.7       1       ID         454       125.8       0.0       13.7       29.1       7.8       1       WY         344       98.0       6.8       31.5       28.0       9.0       1       CO         307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1       WA         343       98.0       50.4       15.7       27.7       10.4       1       OR         421       205.0       74.2       77.8       25.6       6.4       1       NV         380       112.6       86.5       48.5       26.2       8.8       1       CA         Proc       print       data=expenditure;       print;       proc       gplot       data=expenditure;         plot       ex*met;       title3       "Plot of raw data";       symbol1       c=red v=dot;       symbol1       c=red v=dot;	334	97.6	22.6	13.4	28.9	9.7		1	MT		
344       98.0       6.8       31.5       28.0       9.0       1       CO         307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1       WA         343       98.0       50.4       15.7       27.7       10.4       1       OR         421       205.0       74.2       77.8       25.6       6.4       1       NV         380       112.6       86.5       48.5       26.2       8.8       1       CA         ;       Proc       print       data=expenditure;       run;         proc       gplot       data=expenditure;       supplot       ex*met;         title3       "Plot of raw data";       symbol1       c=red       v=dot;	284	93.9	0.0		13.3	30.7	8.7		1	ID	
307       92.5       67.5       28.7       31.9       6.7       1       UT         333       100.4       63.1       19.9       27.5       9.8       1       WA         343       98.0       50.4       15.7       27.7       10.4       1       OR         421       205.0       74.2       77.8       25.6       6.4       1       NV         380       112.6       86.5       48.5       26.2       8.8       1       CA         ;       Proc print data=expenditure;       run;       proc gplot data=expenditure;       i       i       i       i         plot ex*met;       title3       "Plot of raw data";       i       i       i       i         symbol1       c=red v=dot;       i       i       i       i       i       i	454	125.8	0.0		13.7	29.1	7.8		1	WY	
<pre>333 100.4 63.1 19.9 27.5 9.8 1 WA 343 98.0 50.4 15.7 27.7 10.4 1 OR 421 205.0 74.2 77.8 25.6 6.4 1 NV 380 112.6 86.5 48.5 26.2 8.8 1 CA ; Proc print data=expenditure; run; proc gplot data=expenditure; plot ex*met; title3 "Plot of raw data"; symbol1 c=red v=dot;</pre>	344	98.0	6.8		31.5	28.0	9.0		1	CO	
<pre>343 98.0 50.4 15.7 27.7 10.4 1 OR 421 205.0 74.2 77.8 25.6 6.4 1 NV 380 112.6 86.5 48.5 26.2 8.8 1 CA ; Proc print data=expenditure; run; proc gplot data=expenditure; plot ex*met; title3 "Plot of raw data"; symbol1 c=red v=dot;</pre>	307	92.5	67.5	28.7	31.9	6.7		1	UT		
<pre>421 205.0 74.2 77.8 25.6 6.4 1 NV 380 112.6 86.5 48.5 26.2 8.8 1 CA ; Proc print data=expenditure; run; proc gplot data=expenditure; plot ex*met; title3 "Plot of raw data"; symbol1 c=red v=dot;</pre>	333	100.4	63.1	19.9	27.5	9.8		1	WA		
<pre>380 112.6 86.5 48.5 26.2 8.8 1 CA ; Proc print data=expenditure; run; proc gplot data=expenditure; plot ex*met; title3 "Plot of raw data"; symbol1 c=red v=dot;</pre>	343	98.0	50.4	15.7	27.7	10.4	1	OR			
; Proc print data=expenditure; run; proc gplot data=expenditure; plot ex*met; title3 "Plot of raw data"; symbol1 c=red v=dot;	421	205.0	74.2	77.8	25.6	6.4		1	NV		
<pre>Proc print data=expenditure; run; proc gplot data=expenditure; plot ex*met; title3 "Plot of raw data"; symbol1 c=red v=dot;</pre>	380	112.6	86.5	48.5	26.2	8.8		1	CA		
<pre>run; proc gplot data=expenditure; plot ex*met; title3 "Plot of raw data"; symbol1 c=red v=dot;</pre>	;										
<pre>proc gplot data=expenditure; plot ex*met; title3 "Plot of raw data"; symbol1 c=red v=dot;</pre>	Proc	print	<b>:</b> dat	<mark>a</mark> =exp	endit	ure;					
<pre>plot ex*met; title3 "Plot of raw data"; symbol1 c=red v=dot;</pre>	run;										
<pre>plot ex*met; title3 "Plot of raw data"; symbol1 c=red v=dot;</pre>	proc	gplot	t dat	a=exp	endit	ure;					
title3 "Plot of raw data"; symboll c=red v=dot;				-							
symboll c=red v=dot;	-			f rou	data						
-						'					
run;	-	OTT C=	=red	v=dot	i						
	run;										

**Proc gplot:** It can produce high-resolution graphic plot, and improve the appearance of the plot by defining plot symbols. **C=red** specifies the color of the symbol. **V=dot** would place a dot for each data point.

Take time to exam the plot, can you spot a maximum or minimum point?

# Fitting a Polynomial Regression Model with a cubic term of MET:

```
proc glm data=expenditure;
title2 "Polynomial regression model with cubic term";
model ex = met met*met met*met / ss1 ss2;
output out=outdatal p=yhat1 r=resid1;
run;
proc plot data=outdata1;
```

```
title3 'residual plot';
plot resid1*yhat1;
run;
proc univariate data=outdata1 normal;
title3 'univariate procedure on residuals';
var resid1;
run;
```

Notice that **PROC GLM** is used for polynomial regression rather than **PROC REG**. While they are essentially the same for the analysis, there is one difference. When using **PROC GLM**, one can enter the higher order terms directly at the model statement, does not need to create new variables in the data step for each of them, which on the contrary has to be done before PROC **REG** works properly.

Carefully exam the F-tests of parameter estimates. Keep in mind that the sequentially adjusted **type I sums of squares** are used for polynomial regression.

# Fitting a Polynomial Regression Model with a Quadratic term of MET:

```
proc glm data=expenditure;
title2 "Polynomial regression model with cubic term";
model ex = met met*met / ss1 ;
output out=outdata2 p=yhat2 r=resid2;
run;
proc plot data=outdata2;
title2 'residual plot';
plot resid2*yhat2;
run;
proc univariate data=outdata2 normal;
title3 'univariate procedure on residuals';
var resid2;
run;
ods rtf close;
run;
quit;
```

## LAB ASSIGNMENT

- 1. Describe the trend in the scatter plot of the raw data: what is the relationship between variables EX and MET?
- Fit a polynomial regression model with cubic term of MET. When you decide whether the cubic term and quadratic term should be included in the model, do you use the Type I SS or Type II SS? Why?
- 3. Is the cubic effect significant? How about quadratic and linear effects?
- 4. Based on your answers to the above questions and the SAS output, which polynomial model do you consider the best? Write down the polynomial model with the estimated coefficient values. Do you keep the linear term in the model? Why?
- 5. Now assume that there is a state where 100 percent of its residents live in standard metropolitan areas. Use the best model to predict the per capita public expenditure of this state. Is there any problem in doing so?