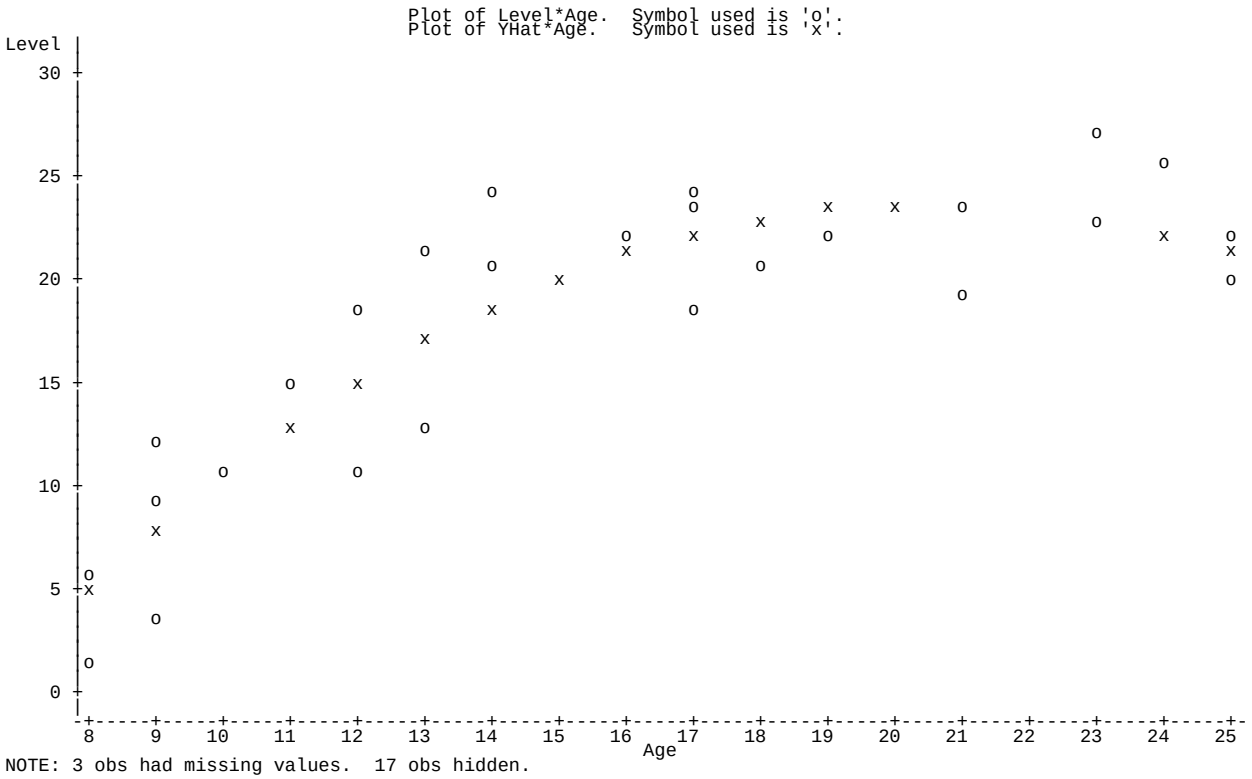
The SAS program I used to obtain the analyses for my answers is given below.

```
dm'log;clear;output;clear';
options nodate nocenter nonumber ps=512 ls=99 nolabel;
ODS HTML style=minimal rs=none
body='C:\Geaghan\Current\EXST7034\Fall2005\SAS\SteroidLevel01.html' ;

Title1 'Steroid use level - KNNL, 2005, Chapter 8, Problem 6';

data SteroidLevel; infile cards missover;
   input Level Age;
     label Level = 'Steroid level'
           Age = 'Subject age';
   Age2 = age*age;
   AgeAgain = Age;
   XMean =  15.7777778; /* mean calculated before extra X values added */
   dAge = age - XMean;
   dAge2 = dAge*dAge;
datalines; run;
 27.1   23.0
. . .
 20.6   18.0
   .    10
   .    15
   .    20
;

PROC PRINT DATA=SteroidLevel; TITLE2 'Data Listing'; RUN;

Title2 'Fit of quadratic model (X-Xmean) - PROC REG';
PROC REG DATA=SteroidLevel;
   MODEL Level = dAge dAge2 / clb alpha=0.01;
   output out=next0 r=resid p=YHat lclm=lclm uclm=uclm lcl=lcli ucl=ucli stdp=stdp
stdi=stdi;
RUN;
proc print data=next0; run; options ps=65 ls=132;

options ps=60 ls=111;
proc plot data=next0;  TITLE2 'Plot of the raw data with type variable';
   plot Level * Age = 'o' YHat * Age = 'x' / overlay;
RUN; OPTIONS PS=256;

Title2 'Fit of quadratic model - PROC REG';
PROC REG DATA=SteroidLevel;
   MODEL Level = Age Age2 / alpha=0.00333333 clb;
   output out=next1 r=resid p=YHat lclm=lclm uclm=uclm lcl=lcli ucl=ucli stdp=stdp
stdi=stdi;
RUN;
proc print data=next1; Title3 'Joint interval estimates of females age 10, 15 & 20';
   Var Age Level YHat Resid lclm uclm stdp;
run;

Title2 'Fit of quadratic model on original variables - PROC REG';
PROC REG DATA=SteroidLevel;
   MODEL Level = Age Age2 / clb alpha=0.01;
   output out=next2 r=resid p=YHat lclm=lclm uclm=uclm lcl=lcli ucl=ucli stdp=stdp
stdi=stdi;
RUN;
proc print data=next2; Title3 'Interval estimate of individual females age 15';
   Var Age Level YHat Resid lcli ucli stdi;
run;

options ps=65 ls=132;
proc plot data=next2;
   TITLE2 'Residual plot on X'; plot resid * Age / vref = 0;
   TITLE2 'Residual plot on YHat'; plot resid * YHat / vref = 0;
RUN; OPTIONS PS=256;
PROC UNIVARIATE DATA=NEXT1 NORMAL PLOT; VAR resid; RUN;

Title2 'Fit of quadratic model - PROC GLM';
PROC GLM DATA=SteroidLevel;
   MODEL Level = Age Age*Age;
RUN;

Title2 'Fit of quadratic model - PROC GLM';
PROC GLM DATA=SteroidLevel; class AgeAgain;
   MODEL Level = Age Age*Age AgeAgain;
RUN;
ODS HTML close;
quit;
```

Question 8.6a from KNNL: The quadratic term would appear to be appropriate from the raw data plot. The predicted values have been superimposed on the observed values. All terms were significant and the R^2 was 0.8143. Some higher order term may account for some additional minor inflections, but there is no obvious curvature remaining from the residual plot.

```
Steroid use level - KNNL, 2005, Chapter 8, Problem 6
Plot of the raw data with type variable
```




```
NOTE: 3 obs had missing values.  17 obs hidden.
```

Question 8.6b from KNNL: The quadratic model is fitted and the results are shown below. There does appear to be a quadratic regression relationship at a P = 0.01 level. The full model can be tested H: $E(\mu_Y) = \beta_0 + \beta_1 X + \beta_2 X^2 + e_i$, (P>F < 0.0001), or the fully adjusted quadratic term H_0: $\beta_2 = 0$ (P>|t| < 0.0001). All lower order terms would be kept if this is significant.

```
Steroid use level - KNNL, 2005, Chapter 8, Problem 6
Fit of quadratic model (X-Xmean) - PROC REG

The REG Procedure
Model: MODEL1
Dependent Variable: Level

Analysis of Variance
                                    Sum of           Mean
Source                      DF       Squares         Square    F Value    Pr > F
Model                        2     1046.26586      523.13293      52.63    <.0001
Error                       24      238.54081        9.93920
Corrected Total             26     1284.80667

Root MSE              3.15265    R-Square     0.8143
Dependent Mean       17.64444    Adj R-Sq     0.7989
Coeff Var            17.86766

Parameter Estimates
                  Parameter       Standard
Variable     DF    Estimate          Error    t Value    Pr > |t|       99% Confidence Limits
Intercept     1    21.09416        0.91415      23.08      <.0001       18.53735       23.65097
dAge          1     1.13736        0.11546       9.85      <.0001        0.81442        1.46029
dAge2         1    -0.11840        0.02347      -5.05      <.0001       -0.18404       -0.05276
```

Question 8.6c from KNNL: The Bonferroni joint confidence intervals of the means for girls ages 10, 15 and 20 were calculated with SAS REG using an ALPHA = 0.01.
The Working-Hotelling W is : $W = 3*F_{(\alpha=0.01,3,24)} = 3*\ 4.718 = 14.154$; $W = 3.7622$
The Bonferroni B is : $B = t_{(\alpha=0.01/6,24)} = 3.2584$; more efficient

The SAS produced predicted values are,

Steroid use level - KNNL, 2005, Chapter 8, Problem 6
Fit of quadratic model - PROC REGc`
Joint interval estimates of females age 10, 15 & 20

Obs	Age	Level	YHat	resid	lclm	uclm	stdp
1	23	27.1	23.1325	3.96746	20.1339	26.1312	1.07213
2	19	22.1	23.5297	-1.42965	21.0673	25.9920	0.88039
3	25	21.9	21.5132	0.38683	17.0163	26.0101	1.60779
4	12	10.7	15.1077	-4.40770	12.9223	17.2931	0.78134
5	8	1.4	5.0855	-3.68550	1.1120	9.0590	1.42067
6	12	18.8	15.1077	3.69230	12.9223	17.2931	0.78134
7	11	14.7	12.9574	1.74264	10.6903	15.2244	0.81056
8	8	5.7	5.0855	0.61450	1.1120	9.0590	1.42067
9	17	18.6	22.3074	-3.70739	19.7277	24.8871	0.92234
10	18	20.4	23.0369	-2.63692	20.4995	25.5744	0.90722
11	9	9.2	7.9463	1.25374	4.7855	11.1071	1.13009
12	21	23.4	23.8047	-0.40470	21.3895	26.2199	0.86351
13	10	10.5	10.5702	-0.07021	7.9863	13.1541	0.92384
14	25	19.7	21.5132	-1.81317	17.0163	26.0101	1.60779
15	9	11.8	7.9463	3.85374	4.7855	11.1071	1.13009
16	17	24.6	22.3074	2.29261	19.7277	24.8871	0.92234
17	9	3.4	7.9463	-4.54626	4.7855	11.1071	1.13009
18	23	22.8	23.1325	-0.33254	20.1339	26.1312	1.07213
19	13	21.1	17.0212	4.07876	14.7663	19.2762	0.80621
20	14	24.0	18.6980	5.30202	16.3170	21.0789	0.85127
21	16	21.8	21.3411	0.45894	18.7734	23.9087	0.91801
22	17	23.5	22.3074	1.19261	19.7277	24.8871	0.92234
23	21	19.4	23.8047	-4.40470	21.3895	26.2199	0.86351
24	24	25.6	22.4413	3.15874	18.8025	26.0801	1.30099
25	13	12.8	17.0212	-4.22124	14.7663	19.2762	0.80621
26	14	20.8	18.6980	2.10202	16.3170	21.0789	0.85127
27	18	20.6	23.0369	-2.43692	20.4995	25.5744	0.90722
28	**10**	**.**	**10.5702**	**.**	**7.9863**	**13.1541**	**0.92384**
29	**15**	**.**	**20.1379**	**.**	**17.6409**	**22.6349**	**0.89276**
30	**20**	**.**	**23.7856**	**.**	**21.3871**	**26.1840**	**0.85753**

The confidence intervals can be obtained using the Bonferroni B and the Standard error produced by SAS.

Question 8.6d from KNNL: The confidence interval of an employee age 15 was calculated with SAS GLM using an ALPHA = 0.01. The results are,

```
Steroid use level - KNNL, 2005, Chapter 8, Problem 6
Fit of quadratic model - PROC REG
Interval estimate of individual females age 15
```

Obs	Age	Level	YHat	resid	lcli	ucli	stdi
1	23	27.1	23.1325	3.96746	13.8188	32.4462	3.32996
2	19	22.1	23.5297	-1.42965	14.3745	32.6848	3.27327
3	25	21.9	21.5132	0.38683	11.6149	31.4114	3.53895
4	12	10.7	15.1077	-4.40770	6.0232	24.1922	3.24803
5	8	1.4	5.0855	-3.68550	-4.5862	14.7572	3.45796
6	12	18.8	15.1077	3.69230	6.0232	24.1922	3.24803
7	11	14.7	12.9574	1.74264	3.8528	22.0619	3.25518
8	8	5.7	5.0855	0.61450	-4.5862	14.7572	3.45796
9	17	18.6	22.3074	-3.70739	13.1200	31.4948	3.28480
10	18	20.4	23.0369	-2.63692	13.8613	32.2125	3.28059
11	9	9.2	7.9463	1.25374	-1.4209	17.3134	3.34908
12	21	23.4	23.8047	-0.40470	14.6621	32.9473	3.26877
13	10	10.5	10.5702	-0.07021	1.3816	19.7588	3.28522
14	25	19.7	21.5132	-1.81317	11.6149	31.4114	3.53895
15	9	11.8	7.9463	3.85374	-1.4209	17.3134	3.34908
16	17	24.6	22.3074	2.29261	13.1200	31.4948	3.28480
17	9	3.4	7.9463	-4.54626	-1.4209	17.3134	3.34908
18	23	22.8	23.1325	-0.33254	13.8188	32.4462	3.32996
19	13	21.1	17.0212	4.07876	7.9197	26.1228	3.25410
20	14	24.0	18.6980	5.30202	9.5644	27.8315	3.26556
21	16	21.8	21.3411	0.45894	12.1571	30.5251	3.28359
22	17	23.5	22.3074	1.19261	13.1200	31.4948	3.28480
23	21	19.4	23.8047	-4.40470	14.6621	32.9473	3.26877
24	24	25.6	22.4413	3.15874	12.9022	31.9803	3.41054
25	13	12.8	17.0212	-4.22124	7.9197	26.1228	3.25410
26	14	20.8	18.6980	2.10202	9.5644	27.8315	3.26556
27	18	20.6	23.0369	-2.43692	13.8613	32.2125	3.28059
28	10	.	10.5702	.	1.3816	19.7588	3.28522
29	**15**	**.**	**20.1379**	**.**	**10.9734**	**29.3024**	**3.27662**
30	20	.	23.7856	.	14.6474	32.9237	3.26720

Question 8.6e –From the SAS TYPE I SS above, the quadratic term is highly significant (P>F=0.0001). It would be retained.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	99% Confidence Limits	
Intercept	1	21.09416	0.91415	23.08	<.0001	18.53735	23.65097
dAge	1	1.13736	0.11546	9.85	<.0001	0.81442	1.46029
dAge2	**1**	**-0.11840**	**0.02347**	**-5.05**	**<.0001**	**-0.18404**	**-0.05276**

Question 8.6f – The regression model, using the deviations, is

$$Y_i = 21.09416 + 1.13736D_i - 0.11840D_i^2$$

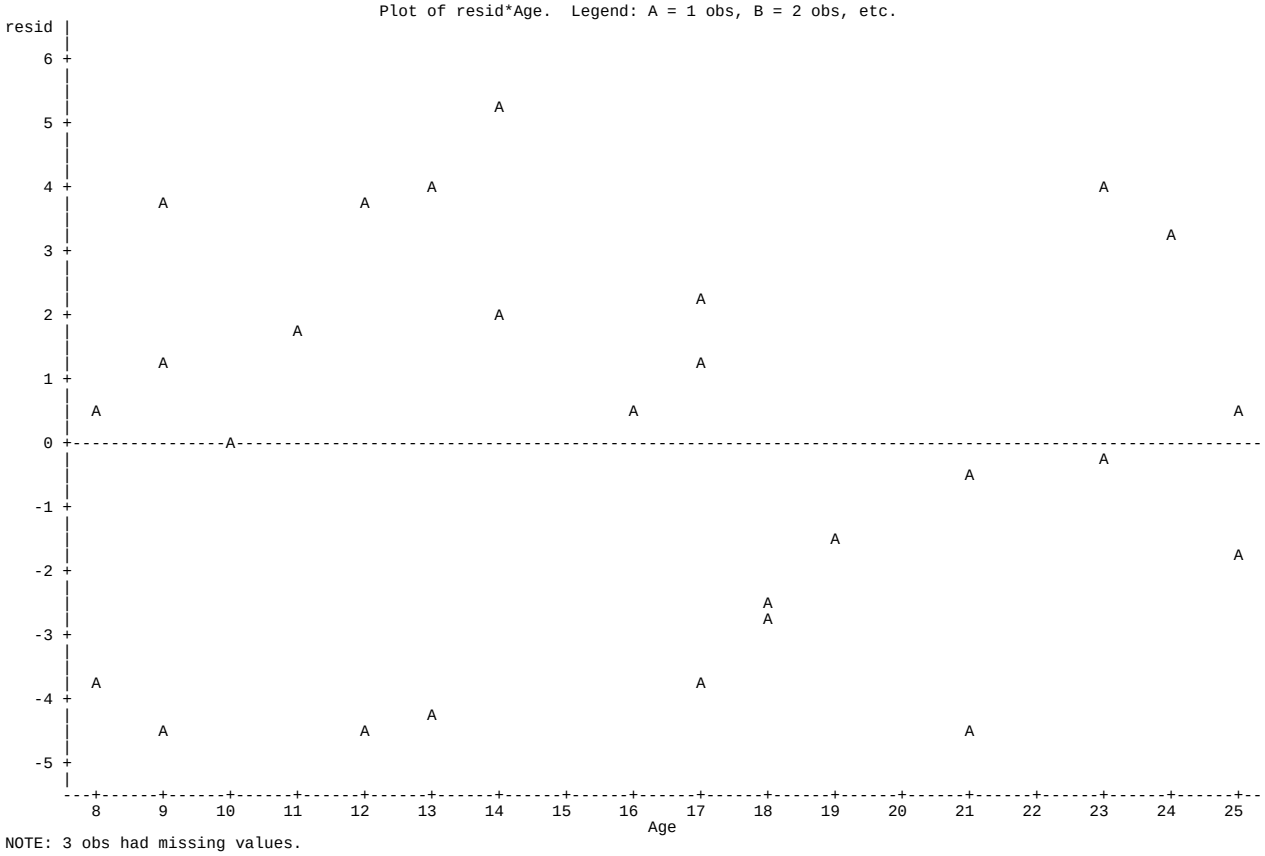
Using the original variables the model is,

$$Y_i = -26.32541254 + 4.87357440D_i - 0.11840125D_i^2$$

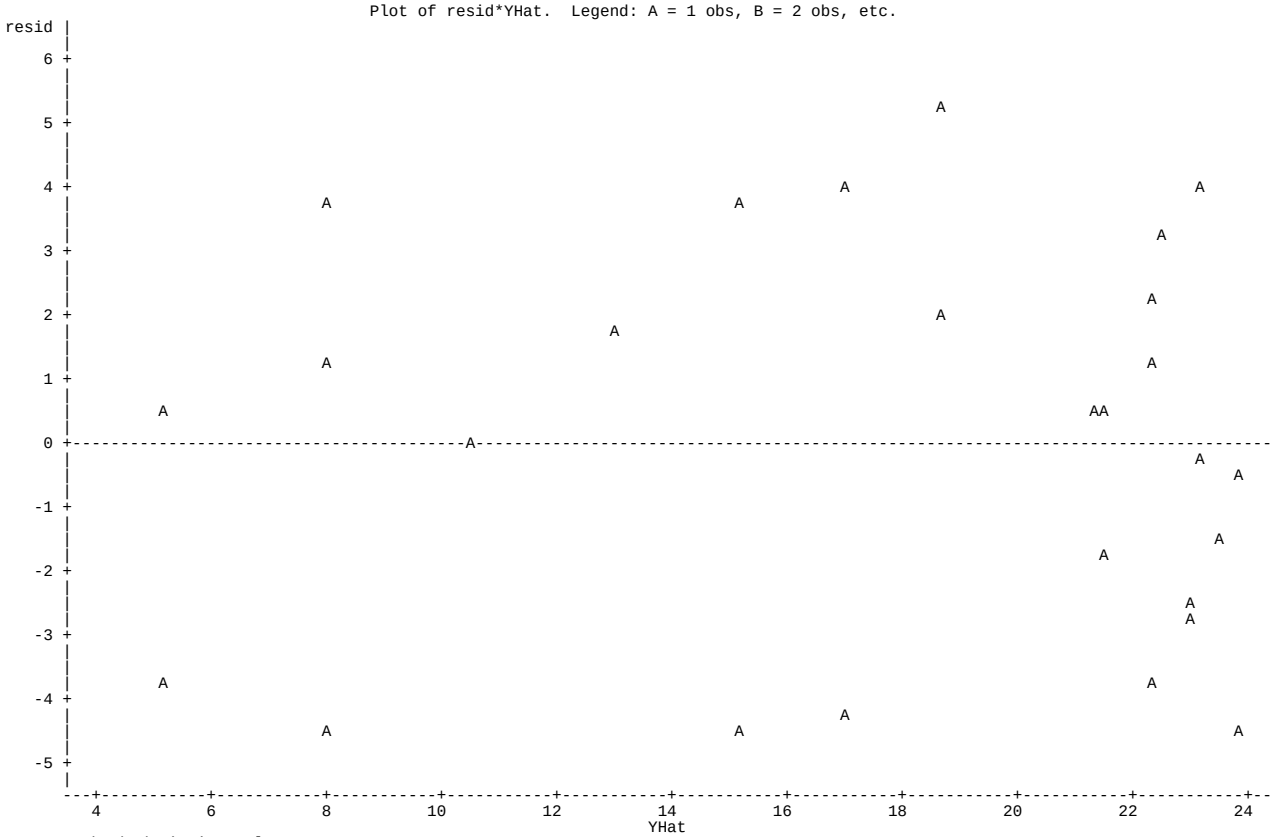
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	**-26.32541254**	**5.88153620**	**-4.48**	**0.0002**
Age	**4.87357440**	**0.77514717**	**6.29**	**<.0001**
Age*Age	**-0.11840125**	**0.02346844**	**-5.05**	**<.0001**

8.7a KNNL) – Residual plot and diagnostics of normality of residuals.

Steroid use level - KNNL, 2005, Chapter 8, Problem 6
Residual plot on YHat



NOTE: 3 obs had missing values.



NOTE: 3 obs had missing values.

Steroid use level - KNNL, 2005, Chapter 8, Problem 6
Residual plot on YHat

The UNIVARIATE Procedure
Variable: resid

Moments

N	27	Sum Weights	27
Mean	0	Sum Observations	0
Std Deviation	3.02896789	Variance	9.1746465
Skewness	-0.0507591	Kurtosis	-1.1620347
Uncorrected SS	238.540809	Corrected SS	238.540809
Coeff Variation	.	Std Error Mean	0.58292514

Basic Statistical Measures

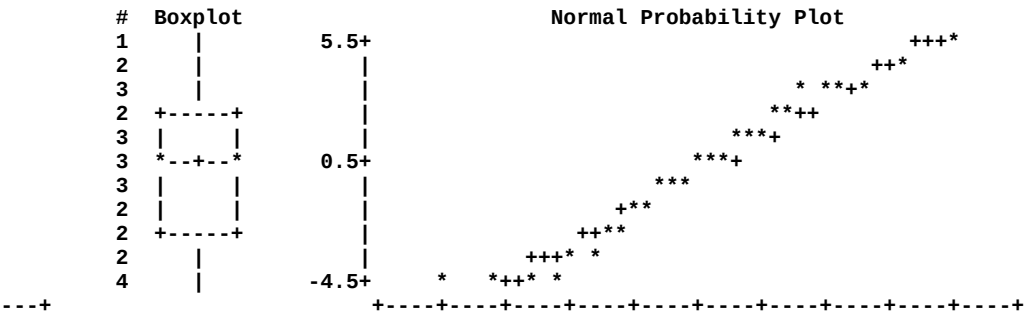
Location		Variability	
Mean	0.000000	Std Deviation	3.02897
Median	0.386831	Variance	9.17465
Mode	.	Range	9.84827
		Interquartile Range	4.92953

Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	0.5	Pr >= |M|	1.0000
Signed Rank	S	-1	Pr >= |S|	0.9814

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	**W**	**0.944619**	**Pr < W**	**0.1584**
Kolmogorov-Smirnov	D	0.110373	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.050658	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.41587	Pr > A-Sq	>0.2500



8.7b KNNL) – Lack of fit was tested by entering a variable equivalent to X in PROC GLM as a CLASSES variable, with the following results. This formal test of Lack of Fit shows that there is no significant departure from the cubic model (P>F = 0.1268).

Lack of Fit
Fit of quadratic model - PROC GLM

The GLM Procedure

Dependent Variable: Level

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	1126.113333	80.436667	6.08	0.0017
Error	12	158.693333	13.224444		
Corrected Total	26	1284.806667			

R-Square	Coeff Var	Root MSE	Level Mean
0.876485	20.61013	3.636543	17.64444

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Age	1	793.2805099	793.2805099	59.99	<.0001
Age*Age	1	252.9853477	252.9853477	19.13	0.0009
AgeAgain	12	79.8474757	6.6539563	0.50	0.8758

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Age	0	0.00000000	.	.	.
Age*Age	0	0.00000000	.	.	.
AgeAgain	12	79.84747573	6.65395631	0.50	0.8758