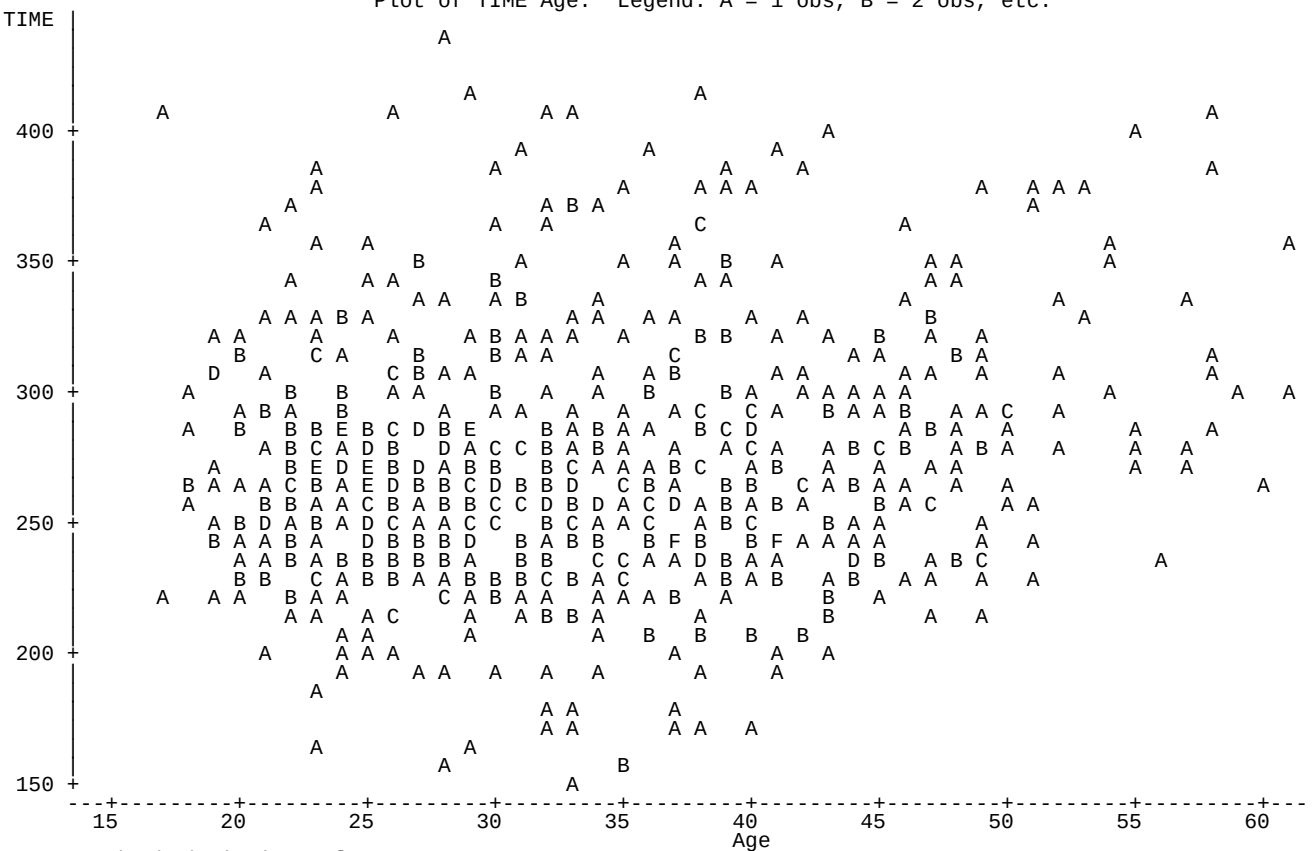
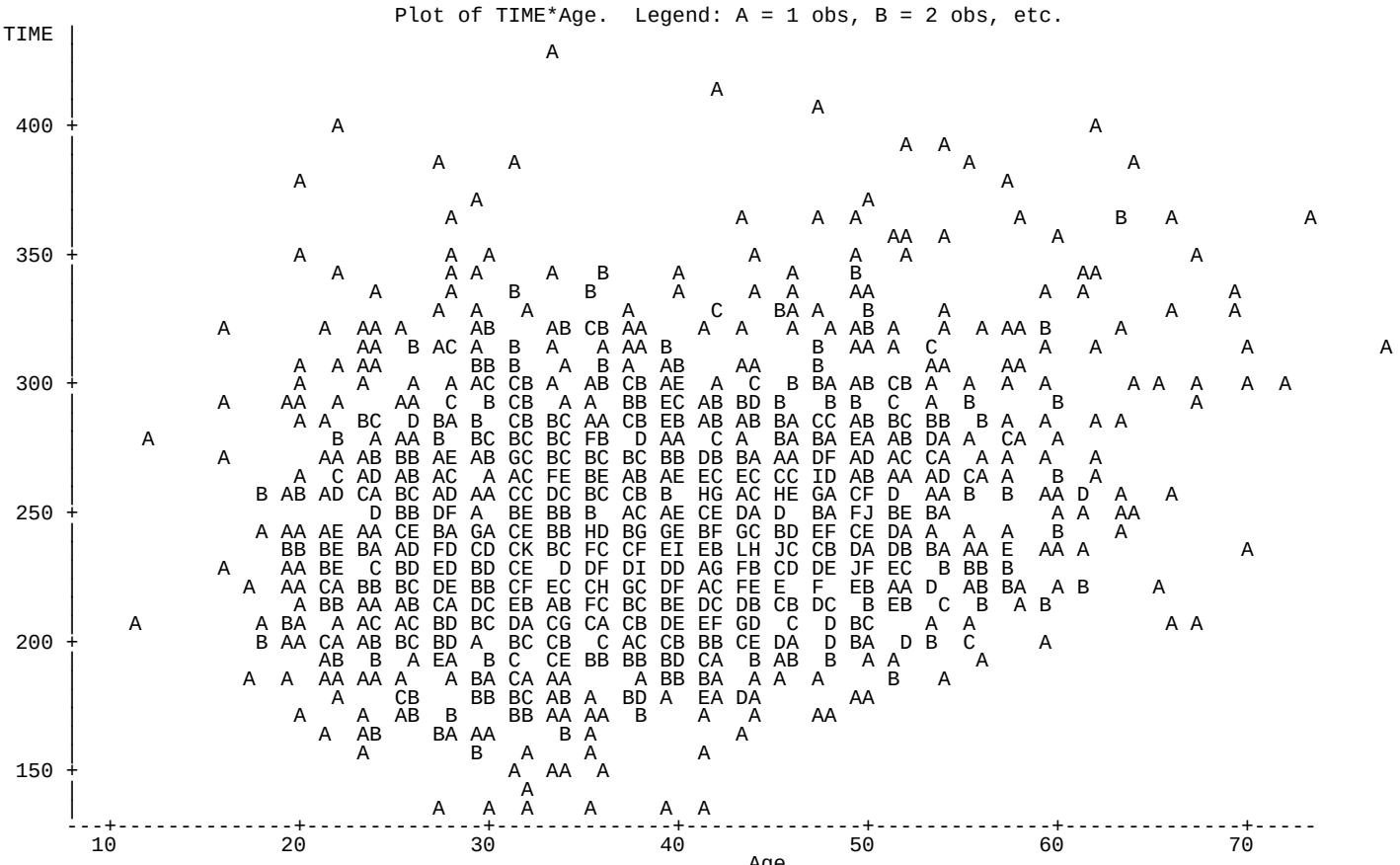
```
dm'log;clear;output;clear';
OPTIONS PS=512 LS=101 NOCENTER NODATE NONUMBER FORMCHAR="|----|+|---+=|-/\<>*";
ODS HTML style=minimal body='Appendix09 LOF & Rsquare Suppliment.html';
ODS listing;

TITLE1 'EXST7015: Marathon Footrace from Pennsylvania 2002';
filename input 'Appendix09 MReg-10K Polynomial.DAT';

*******************************************;
*** Finish times in a race                ***;
*** Data taken from various sites on the  ***;
*** internet reporting race results       ***;
*******************************************;

data PA2002; length hometown $ 23 gender $ 3; infile input missover;
  input Marathon $ Age gender $ TIME HomeTown $ 35-57;
  if age eq 99 then age = .;
  *(apparently 99 represents missing for 5 PA race participants);
*---+----1----+----2----+----3----+----4----+----5----+----6;
cards; run;
;

data PA2002; set PA2002; if marathon = 'PA052002';
proc sort data=PA2002; by gender age; run;

TITLE2 'Scatter plot of raw data';
proc plot data=PA2002; by gender; plot time*age;
options ls=132 ps=65; run; options ps=256 ls=80;

proc glm data=PA2002; BY gender;
   TITLE2 'Quadratic model fitted to raw data - separate by gender';
   model time= age age*age;
run;

proc means data=PA2002 noprint; by gender age; var time;
   output out=MeansByGender n=n mean=timemean std=std; run;

TITLE2 'Scatter plot of means';
proc plot data=MeansByGender; by gender; plot timemean*age;
options ls=132 ps=65; run; options ps=256 ls=80;

proc glm data=MeansByGender; BY gender;
   TITLE2 'Quadratic model fitted on UNweighted means - separate by gender';
   model timemean = age age*age;
run;

proc glm data=MeansByGender; BY gender; weight n;
   TITLE2 'Quadratic model fitted on weighted means - separate by gender';
   model timemean = age age*age;
run;

proc print data=MeansByGender; var gender age n timemean std; run;

proc glm data=PA2002; BY gender; class age;
   TITLE2 'Analysis of Variance of AGE - separate by gender';
   model time = age;
run;

data PA2002; set PA2002; age_again = age;
proc glm data=PA2002; BY gender; class age_again;
   TITLE2 'Analysis of Variance of AGE - separate by gender';
   model time = age age*age age_again;
run;

ODS HTML close;
run; quit;
```

EXST7015: Marathon Footrace from Pennsylvania 2002
Scatter plot of raw data

gender=F

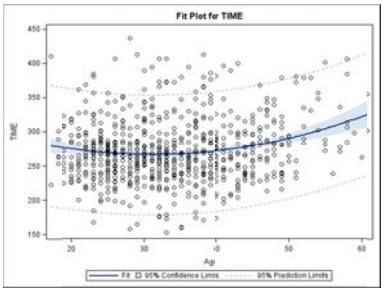


NOTE: 2 obs had missing values.



NOTE: 3 obs had missing values.

EXST7015: Marathon Footrace from Pennsylvania 2002
Quadratic model fitted to raw data - separate by gender



The GLM Procedure

gender=F
Number of Observations Read 753
Number of Observations Used 751

Dependent Variable: TIME

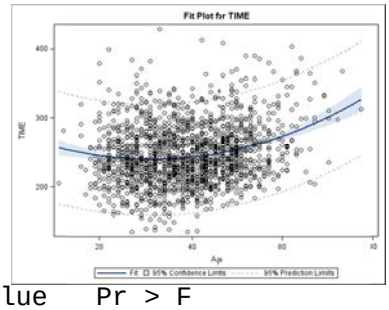
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	61257.907	30628.954	15.29	<.0001
Error	748	1498623.844	2003.508		
Corrected Total	750	1559881.751			

R-Square	Coeff Var	Root MSE	TIME Mean
0.039271	16.47384	44.76056	271.7068

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Age	1	26569.82118	26569.82118	13.26	0.0003
Age*Age	1	34688.08613	34688.08613	17.31	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Age	1	25598.41747	25598.41747	12.78	0.0004
Age*Age	1	34688.08613	34688.08613	17.31	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	331.9717727	20.64142442	16.08	<.0001
Age	-4.2594366	1.19162916	-3.57	0.0004
Age*Age	0.0682107	0.01639299	4.16	<.0001



gender=M
Number of Observations Read 1808
Number of Observations Used 1805

Dependent Variable: TIME

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	173791.926	86895.963	50.94	<.0001
Error	1802	3073916.990	1705.836		
Corrected Total	1804	3247708.916			

R-Square	Coeff Var	Root MSE	TIME Mean
0.053512	16.75563	41.30177	246.4949

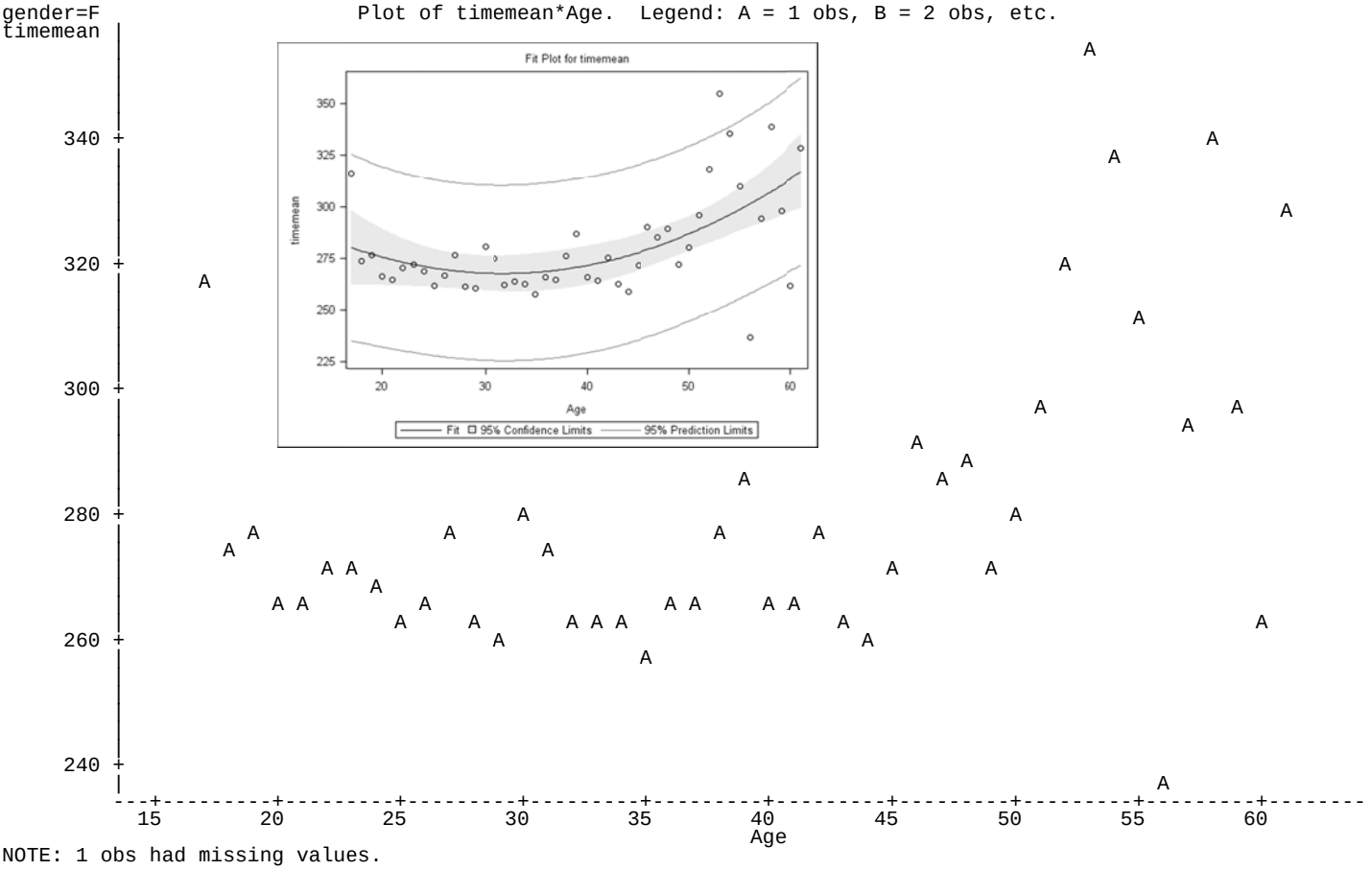
Source	DF	Type I SS	Mean Square	F Value	Pr > F
Age	1	110167.2110	110167.2110	64.58	<.0001
Age*Age	1	63624.7152	63624.7152	37.30	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Age	1	37836.81687	37836.81687	22.18	<.0001
Age*Age	1	63624.71523	63624.71523	37.30	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	280.7072514	10.94798326	25.64	<.0001
Age	-2.6374056	0.56000035	-4.71	<.0001
Age*Age	0.0420317	0.00688229	6.11	<.0001

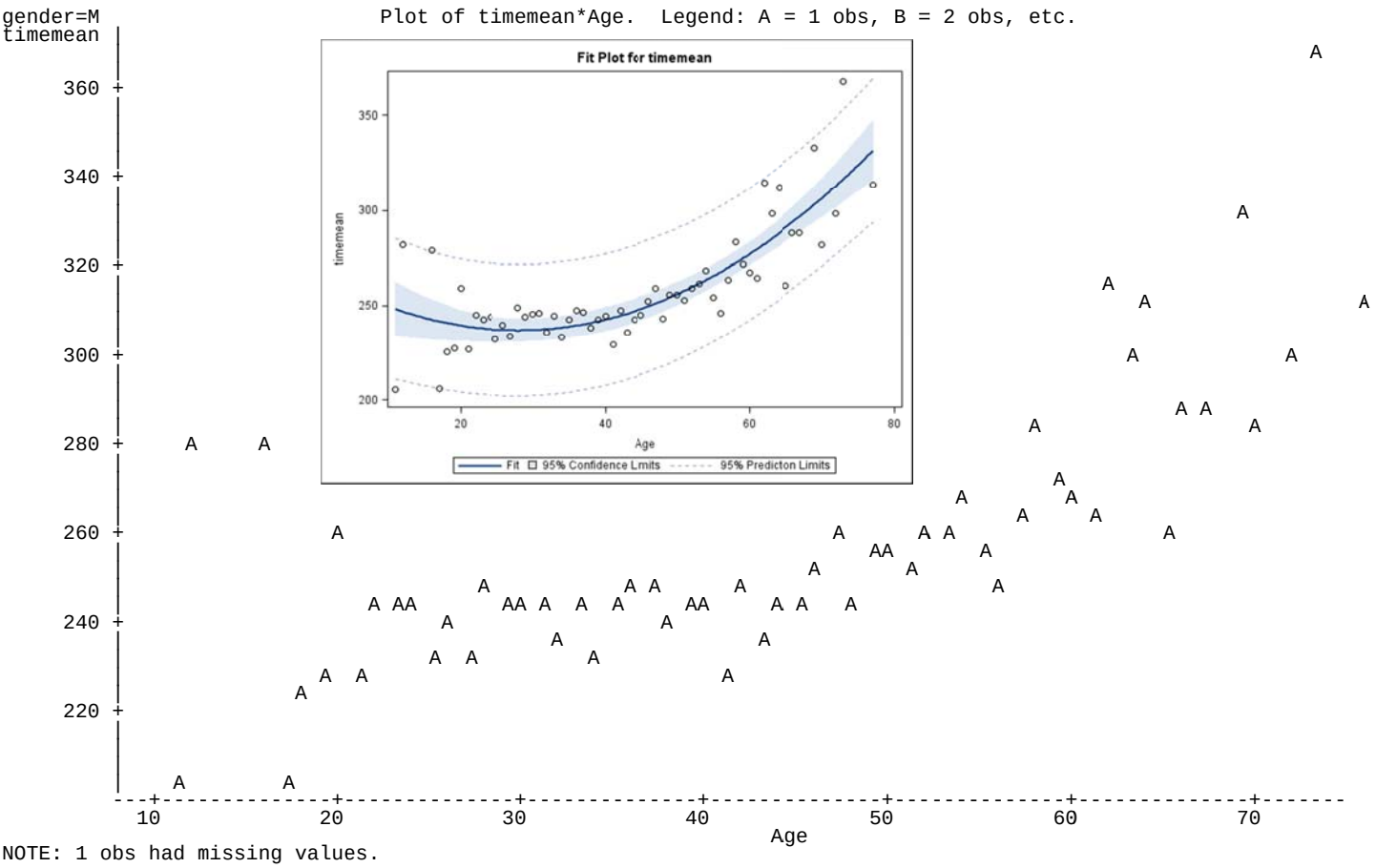
EXST7015: Marathon Footrace from Pennsylvania 2002
Scatter plot of means

gender=F



NOTE: 1 obs had missing values.

gender=M



NOTE: 1 obs had missing values.

EXST7015: Marathon Footrace from Pennsylvania 2002
Quadratic model fitted on UNweighted means - separate by gender

The GLM Procedure

gender=F

Number of Observations Read	46
Number of Observations Used	45

Dependent Variable: timemean

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	8548.78315	4274.39158	10.08	0.0003
Error	42	17818.72839	424.25544		
Corrected Total	44	26367.51154			

R-Square	Coeff Var	Root MSE	timemean Mean
0.324217	7.337506	20.59746	280.7147

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Age	1	5210.561378	5210.561378	12.28	0.0011
Age*Age	1	3338.221774	3338.221774	7.87	0.0076

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Age	1	2164.392158	2164.392158	5.10	0.0292
Age*Age	1	3338.221774	3338.221774	7.87	0.0076

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	325.6661219	29.20891418	11.15	<.0001
Age	-3.6279436	1.60622455	-2.26	0.0292
Age*Age	0.0571346	0.02036832	2.81	0.0076

gender=M

Number of Observations Read	60
Number of Observations Used	59

Dependent Variable: timemean

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	34412.54203	17206.27102	58.99	<.0001
Error	56	16335.32833	291.70229		
Corrected Total	58	50747.87036			

R-Square	Coeff Var	Root MSE	timemean Mean
0.678108	6.620654	17.07929	257.9699

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Age	1	26435.95373	26435.95373	90.63	<.0001
Age*Age	1	7976.58830	7976.58830	27.34	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Age	1	3203.331856	3203.331856	10.98	0.0016
Age*Age	1	7976.588297	7976.588297	27.34	<.0001

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	267.6384295	13.19520859	20.28	<.0001
Age	-2.2006212	0.66407060	-3.31	0.0016
Age*Age	0.0393577	0.00752646	5.23	<.0001

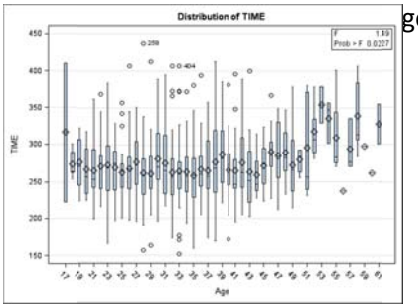
```
EXST7015: Marathon Footrace from Pennsylvania 2002
Quadratic model fitted on weighted means - separate by gender

The GLM Procedure

gender=F
Number of Observations Read          46
Number of Observations Used          45

Dependent Variable: timemean
Weight: n
                                   Sum of
Source                     DF       Squares    Mean Square   F Value   Pr > F
Model                       2     61257.9073     30628.9537     17.97   <.0001
Error                      42     71588.0980      1704.4785
Corrected Total            44    132846.0053

R-Square     Coeff Var      Root MSE    timemean Mean
0.461120      15.19481      41.28533         271.7068

Source                     DF      Type I SS    Mean Square   F Value   Pr > F
Age                         1    26569.82118    26569.82118     15.59   0.0003
Age*Age                     1    34688.08613    34688.08613     20.35   <.0001

Source                     DF    Type III SS    Mean Square   F Value   Pr > F
Age                         1    25598.41747    25598.41747     15.02   0.0004
Age*Age                     1    34688.08613    34688.08613     20.35   <.0001

                                    Standard
Parameter          Estimate           Error    t Value    Pr > |t|
Intercept      331.9717727     19.03881485      17.44      <.0001
Age             -4.2594366      1.09911053      -3.88      0.0004
Age*Age          0.0682107      0.01512023       4.51      <.0001

gender=M
Number of Observations Read          60
Number of Observations Used          59

Dependent Variable: timemean
Weight: n
                                   Sum of
Source                     DF       Squares    Mean Square   F Value   Pr > F
Model                       2    173791.9262     86895.9631     44.70   <.0001
Error                      56    108870.6920      1944.1195
Corrected Total            58    282662.6182

R-Square     Coeff Var      Root MSE    timemean Mean
0.614839      17.88766      44.09217         246.4949

Source                     DF      Type I SS    Mean Square   F Value   Pr > F
Age                         1    110167.2110    110167.2110     56.67   <.0001
Age*Age                     1     63624.7152     63624.7152     32.73   <.0001

Source                     DF    Type III SS    Mean Square   F Value   Pr > F
Age                         1    37836.81687    37836.81687     19.46   <.0001
Age*Age                     1    63624.71523    63624.71523     32.73   <.0001

                                    Standard
Parameter          Estimate           Error    t Value    Pr > |t|
Intercept      280.7072514     11.68764276      24.02      <.0001
Age             -2.6374056      0.59783468      -4.41      <.0001
Age*Age          0.0420317      0.00734727       5.72      <.0001
```

EXST7015: Marathon Footrace from Pennsylvania 2002
Quadratic model fitted on weighted means - separate by gender

Obs	gender	Age	n	timemean	std
1	F	.	2	275.075	63.958
2	F	17	2	316.255	132.547
3	F	18	5	274.174	19.616
4	F	19	11	277.092	34.523
5	F	20	14	266.557	34.947
6	F	21	18	265.066	39.517
7	F	22	25	270.766	37.979
8	F	23	32	272.543	49.630
9	F	24	27	269.064	36.196
10	F	25	37	262.329	31.977
11	F	26	33	267.132	42.219
12	F	27	26	276.793	38.262
13	F	28	26	261.597	51.488
14	F	29	30	260.943	42.763
15	F	30	32	281.060	44.176
16	F	31	23	275.303	47.161
17	F	32	32	262.506	52.472
18	F	33	29	264.058	56.881
19	F	34	25	263.160	41.738
20	F	35	21	258.202	52.349
21	F	36	21	266.474	42.702
22	F	37	29	265.042	45.724
23	F	38	30	276.415	60.488
24	F	39	24	286.910	47.466
25	F	40	27	266.216	40.108
26	F	41	21	264.632	47.833
27	F	42	11	275.758	53.701
28	F	43	17	263.006	49.536
29	F	44	15	259.344	28.021
30	F	45	17	271.878	30.926
31	F	46	12	290.278	35.826
32	F	47	15	285.339	43.975
33	F	48	11	289.601	37.319
34	F	49	14	272.380	46.268
35	F	50	7	280.397	15.286
36	F	51	5	295.962	73.942
37	F	52	5	318.588	39.516
38	F	53	2	354.740	34.450
39	F	54	3	335.793	29.165
40	F	55	4	310.038	61.171
41	F	56	1	237.050	.
42	F	57	3	294.500	36.135
43	F	58	5	339.088	53.048
44	F	59	1	298.050	.
45	F	60	1	261.950	.
46	F	61	2	328.135	37.922
47	M	.	3	255.943	49.579
48	M	11	1	205.470	.
49	M	12	1	281.800	.
50	M	16	4	278.938	38.000
51	M	17	2	205.950	25.456
52	M	18	6	225.348	28.741
53	M	19	11	227.424	31.152
54	M	20	18	258.831	53.618
55	M	21	20	227.018	36.350
56	M	22	38	245.126	41.808
57	M	23	23	242.531	47.606
58	M	24	37	243.947	42.772
59	M	25	27	232.040	38.193
60	M	26	47	239.536	36.343
61	M	27	46	233.891	43.783
62	M	28	58	248.744	45.173
63	M	29	43	243.912	50.860
64	M	30	41	245.474	48.675
65	M	31	65	245.981	47.391
66	M	32	66	235.893	37.748
67	M	33	46	244.307	49.855
68	M	34	58	233.462	41.370
69	M	35	59	242.669	42.640
70	M	36	55	247.132	42.024
71	M	37	45	246.236	39.719
72	M	38	64	238.338	36.309
73	M	39	56	242.350	39.900
74	M	40	69	244.520	37.099
75	M	41	56	229.503	38.029
76	M	42	61	247.215	40.187
77	M	43	69	235.805	34.416
78	M	44	54	242.267	40.908
79	M	45	53	244.811	30.586
80	M	46	36	252.149	39.202
81	M	47	60	258.881	39.681
82	M	48	47	243.054	33.631
83	M	49	54	255.475	41.401
84	M	50	56	255.607	38.493
85	M	51	42	252.615	37.906
86	M	52	32	258.827	45.817
87	M	53	29	261.462	35.076
88	M	54	22	268.478	50.121
89	M	55	19	254.231	45.257
90	M	56	13	246.088	37.578
91	M	57	22	263.264	38.488
92	M	58	6	283.555	60.619
93	M	59	12	271.836	47.213
94	M	60	12	267.285	34.823
95	M	61	10	264.194	42.416
96	M	62	6	314.518	52.725
97	M	63	7	298.454	53.140
98	M	64	3	312.410	67.903
99	M	65	2	260.525	55.713
100	M	66	4	288.125	71.192
101	M	67	4	288.083	61.504
102	M	69	2	332.925	1.549
103	M	70	3	282.143	40.941
104	M	72	1	298.400	.
105	M	73	1	367.500	.
106	M	77	1	313.670	.

Comparison of fits to Raw data and fits to means (weighted and unweighted)

	Females			Males		
Statistic	**Raw data**	**Means**	**Weighted means**	**Raw data**	**Means**	**Weighted means**
b_0	331.9717727	325.6661219	331.9717727	280.7072514	267.6384295	280.7072514
b_1	-4.2594366	-3.6279436	-4.2594366	-2.6374056	-2.2006212	-2.6374056
b_2	0.0682107	0.0571346	0.0682107	0.0420317	0.0393577	0.0420317
R^2	0.039271	0.324217	0.46112	0.053512	0.678108	0.614839
MSE	2003.508	424.25544	1704.4785	1705.836	291.70229	1944.1195



EXST7015: Marathon Footrace from Pennsylvania 2002
Analysis of Variance of AGE - separate by gender

The GLM Procedure

gender=F

Class Level Information

Class	Levels	Values
Age	45	17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61

Number of Observations Read 753
Number of Observations Used 751

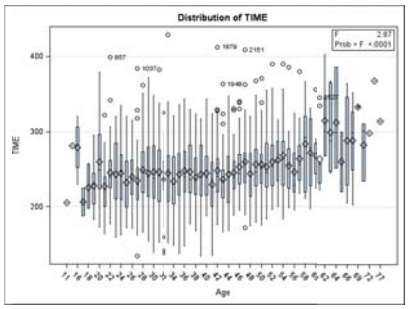
Dependent Variable: TIME

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	44	132846.005	3019.227	1.49	0.0227
Error	706	1427035.746	2021.297		
Corrected Total	750	1559881.751			

R-Square	Coeff Var	Root MSE	TIME Mean
0.085164	16.54682	44.95884	271.7068

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Age	44	132846.0053	3019.2274	1.49	0.0227

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Age	44	132846.0053	3019.2274	1.49	0.0227



gender=M

Class Level Information

Class	Levels	Values
Age	59	11 12 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 69 70 72 73 77

Number of Observations Read 1808
Number of Observations Used 1805

Dependent Variable: TIME

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	58	282662.618	4873.493	2.87	<.0001
Error	1746	2965046.298	1698.194		
Corrected Total	1804	3247708.916			

R-Square	Coeff Var	Root MSE	TIME Mean
0.087034	16.71805	41.20915	246.4949

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Age	58	282662.6182	4873.4934	2.87	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Age	58	282662.6182	4873.4934	2.87	<.0001

Test of Lack of Fit – a measure of adequacy of the model

Regression on means for Females

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	61257.9073	30628.9537	17.97	<.0001
Error	42	71588.098	1704.4785		
Corrected Total	44	132846.0053			**R^2 = 0.461120**

Regression on raw data for Females

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	61257.907	30628.954	15.29	<.0001
Error	748	1498623.844	2003.508		
Corrected Total	750	1559881.751			**R^2 = 0.039271**

Analysis of variance on AGE using raw data for Females

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	44	132846.005	3019.227	1.49	0.0227
Error	706	1427035.746	2021.297		
Corrected Total	750	1559881.751			**R^2 = 0.085164**

Test of lack of fit for Females

Source	d.f.	SSE	MSE	F	P>F
Reduced model	748	1498623.844			
Full model	706	1427035.746			
Difference	42	71588.098	1704.478524	0.843259772	0.7493
Full model	706	1427035.746	2021.297091		

Regression on means for Males

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	173791.9262	86895.9631	44.7	<.0001
Error	56	108870.692	1944.1195		
Corrected Total	58	282662.6182			**R^2 = 0.614839**

Regression on raw data for Males

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	173791.926	86895.963	50.94	<.0001
Error	1802	3073916.99	1705.836		
Corrected Total	1804	3247708.916			**R^2 = 0.053512**

Analysis of variance on AGE using raw data for Males

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	58	282662.618	4873.493	2.87	<.0001
Error	1746	2965046.298	1698.194		
Corrected Total	1804	3247708.916			**R^2 = 0.087034**

Test of lack of ft for Males

Source	d.f.	SSE	MSE	F	P>F
Reduced model	1802	3073916.99			
Full model	1746	2965046.298			
Difference	56	108870.692	1944.1195	1.144816069	0.2180
Full model	1746	2965046.298	1698.193756		

Model to test LOF for Quadratic model fitted to raw data - separate by gender

It is possible to test LOF in a single model by including the two sets of the independent variable and putting one in the class statement. The regression is fitted first with the quantitative X variable (not in the class statement) and the categorical version (the one in the class statement) is fitted last to account for any remaining variation which is LOF.

```
The GLM Procedure

gender=F

Class Level Information
Class        Levels  Values
age_again        45  17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
                     40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61
Number of Observations Read         753
Number of Observations Used         751
```

The usual Deviations from regression are = $\Sigma(Y_{ij} - \hat{Y}_{i.})^2$

```
Dependent Variable: TIME                Sum of
Source                     DF          Squares    Mean Square   F Value   Pr > F
Model                      44     132846.009       3019.227      1.49   0.0227
Error                     706    1427035.742       2021.297
```
PE = $\Sigma(Y_{ij} - \bar{Y}_{i.})^2$

```
Corrected Total           750    1559881.751
```
CSS = $\Sigma(Y_{ij} - \bar{Y}_{..})^2$

```
R-Square     Coeff Var      Root MSE     TIME Mean
0.085164      16.54682      44.95884      271.7068

Source           DF      Type I SS    Mean Square   F Value   Pr > F
Age               1    26569.82118    26569.82118     13.14   0.0003
Age*Age           1    34688.08613    34688.08613     17.16   <.0001
age_again        42    71588.10138     1704.47860      0.84   0.7493
```
LOF = $\Sigma(\bar{Y}_{i.} - \hat{Y}_{i.})^2$

```
Source           DF    Type III SS    Mean Square   F Value   Pr > F
Age               0        0.00000              .         .      .
Age*Age           0        0.00000              .         .      .
age_again        42    71588.10133     1704.47860      0.84   0.7493

gender=M
Class Level Information
Class        Levels  Values
age_again        59  11 12 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
                     34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53
                     54 55 56 57 58 59 60 61 62 63 64 65 66 67 69 70 72 73 77
Number of Observations Read        1808
Number of Observations Used        1805

Dependent Variable: TIME                Sum of
Source                     DF          Squares    Mean Square   F Value   Pr > F
Model                      58     282662.620       4873.493      2.87   <.0001
Error                    1746    2965046.296       1698.194
Corrected Total          1804    3247708.916

R-Square     Coeff Var      Root MSE     TIME Mean
0.087034      16.71805      41.20915      246.4949

Source           DF      Type I SS    Mean Square   F Value   Pr > F
Age               1    110167.2110    110167.2110     64.87   <.0001
Age*Age           1     63624.7152     63624.7152     37.47   <.0001
age_again        56    108870.6938      1944.1195      1.14   0.2180

Source          DF    Type III SS    Mean Square   F Value   Pr > F
Age              0         0.0000              .         .      .
Age*Age          0         0.0000              .         .      .
age_again                 56    108870.6938      1944.1195      1.14   0.2180
```

Compare the tests of "age_again" to the tests of Lack of Fit on the preceding page.