

```

1 *****;
2 *** Air pollution and mortality ***;
3 *** Data on mortality from all causes for the years ***;
4 *** 1959 to 1961. ***;
5 *****;
6
7 dm'log;clear;output;clear';
8 options nodate nocenter nonumber ps=512 ls=99 nolabel;
9 ODS HTML style=minimal rs=none
10 ! body='C:\Geaghan\Current\EXST3201\Fall2005\SAS\AirPollution01.html' ;
NOTE: Writing HTML Body file: C:\Geaghan\Current\EXST3201\Fall2005\SAS\AirPollution01.html
11
12 Title1 'Chapter 11 : Air pollution and mortality study';
13 filename input1 'C:\Geaghan\Current\EXST3201\Datasets\ASCII\ex1123.csv';
14
15 data AirPollution; length city $ 18; infile input1 missover DSD dlm="," firstobs=2;
16 input CITY $ MORT PRECIP EDUC NONWHITE NOX SO2;
17 label Mort = 'Mortality (all causes, age adjusted per 100,000 pop'
18 Precip = 'Annual precipitation (inches)'
19 Educ = 'Median number of school years completed'
20 Nonwhite = 'Percentage of nonwhite population in 1960'
21 NOx = 'Oxides of Nitrogen'
22 SO2 = 'Sulfur dioxide'
23 City = 'City, State';
24
25 datalines;
NOTE: The infile INPUT1 is:
File Name=C:\Geaghan\Current\EXST3201\Datasets\ASCII\ex1123.csv,
RECFM=V,LRECL=256
NOTE: 60 records were read from the infile INPUT1.
The minimum record length was 36.
The maximum record length was 85.
NOTE: The data set WORK.AIRPOLLUTION has 60 observations and 7 variables.
NOTE: DATA statement used (Total process time):
real time 0.01 seconds
cpu time 0.02 seconds
24 run;
25 proc sort DATA=AirPollution; by city; run;
NOTE: There were 60 observations read from the data set WORK.AIRPOLLUTION.
NOTE: The data set WORK.AIRPOLLUTION has 60 observations and 7 variables.
NOTE: PROCEDURE SORT used (Total process time):
real time 0.01 seconds
cpu time 0.02 seconds
26 PROC PRINT DATA=AirPollution; TITLE2 'Data Listing'; RUN;
NOTE: There were 60 observations read from the data set WORK.AIRPOLLUTION.
NOTE: The PROCEDURE PRINT printed page 1.
NOTE: PROCEDURE PRINT used (Total process time):
real time 0.14 seconds
cpu time 0.04 seconds

```

Chapter 11 : Air pollution and mortality study
Data Listing

Obs	city	MORT	PRECIP	EDUC	NONWHITE	NOX	SO2
1	Akron, OH	921.87	36	11.4000	8.8000	15	59
2	Albany, NY	997.88	35	11.0000	3.5000	10	39
3	Allentown, PA	962.35	44	9.8000	0.8000	6	33
4	Atlanta, GA	982.29	47	11.1000	27.1000	8	24
5	Baltimore, MD	1071.29	43	9.6000	24.4000	38	206
6	Birmingham, AL	1030.38	53	10.2000	38.5000	32	72
7	Boston, MA	934.70	43	12.1000	3.5000	32	62
8	Bridgeport, CT	899.53	45	10.6000	5.3000	4	4
9	Buffalo, NY	1001.90	36	10.5000	8.1000	12	37
10	Canton, OH	912.35	36	10.7000	6.7000	7	20
11	Chattanooga, TN	1017.61	52	9.6000	22.2000	8	27

Obs	city	MORT	PRECIP	EDUC	NONWHITE	NOX	SO2
12	Chicago, IL	1024.89	33	10.9000	16.3000	63	278
13	Cincinnati, OH	970.47	40	10.2000	13.0000	26	146
14	Cleveland, OH	985.95	35	11.1000	14.7000	21	64
15	Columbus, OH	958.84	37	11.9000	13.1000	9	15
16	Dallas, TX	860.10	35	11.8000	14.8000	1	1
17	Dayton, OH	936.23	36	11.4000	12.4000	4	16
18	Denver, CO	871.77	15	12.2000	4.7000	8	28
19	Detroit, MI	959.22	31	10.8000	15.8000	35	124
20	Flint, MI	941.18	30	10.8000	13.1000	4	11
21	Fort Worth, TX	891.71	31	11.4000	11.5000	1	1
22	Grand Rapids, MI	871.34	31	10.9000	5.1000	3	10
23	Greensboro, NC	971.12	42	10.4000	22.7000	3	5
24	Hartford, CT	887.47	43	11.5000	7.2000	3	10
25	Houston, TX	952.53	46	11.4000	21.0000	5	1
26	Indianapolis, IN	968.66	39	11.4000	15.6000	7	33
27	Kansas City, MO	919.73	35	12.0000	12.6000	4	4
28	Lancaster, PA	844.05	43	9.5000	2.9000	7	32
29	Los Angeles, CA	861.83	11	12.1000	7.8000	319	130
30	Louisville, KY	989.27	30	9.9000	13.1000	37	193
31	Memphis, TN	1006.49	50	10.4000	36.7000	18	34
32	Miami, FL	861.44	60	11.5000	13.5000	1	1
33	Milwaukee, WI	929.15	30	11.1000	5.8000	23	125
34	Minneapolis, MN	857.62	25	12.1000	2.0000	11	26
35	Nashville, TN	961.01	45	10.1000	21.0000	14	78
36	New Haven, CT	923.23	46	11.3000	8.8000	3	8
37	New Orleans, LA	1113.06	54	9.7000	31.4000	17	1
38	New York, NY	994.65	42	10.7000	11.3000	26	108
39	Philadelphia, PA	1015.02	42	10.5000	17.5000	32	161
40	Pittsburgh, PA	991.29	36	10.6000	8.1000	59	263
41	Portland, OR	893.99	37	12.0000	3.6000	21	44
42	Providence, RI	938.50	42	10.1000	2.2000	4	18
43	Reading, PA	946.18	41	9.6000	2.7000	11	89
44	Richmond, VA	1025.50	44	11.0000	28.6000	9	48
45	Rochester, NY	874.28	32	11.1000	5.0000	4	18
46	Saint Louis, MO	953.56	34	9.7000	17.2000	15	68
47	San Diego, CA	839.71	10	12.1000	5.9000	66	20
48	San Francisco, CA	911.70	18	12.2000	13.7000	171	86
49	San Jose, CA	790.73	13	12.2000	3.0000	32	3
50	Seattle, WA	899.26	35	12.2000	5.7000	7	20
51	Springfield, MA	904.16	45	11.1000	3.4000	4	20
52	Syracuse, NY	950.67	38	11.4000	3.8000	5	25
53	Toledo, OH	972.46	31	10.7000	9.5000	7	25
54	Utica, NY	912.20	40	10.3000	2.5000	2	11
55	Washington, DC	967.80	41	12.3000	25.9000	28	102
56	Wichita, KS	823.76	28	12.1000	7.5000	2	1
57	Wilmington, DE	1003.50	45	11.3000	12.1000	11	42
58	Worcester, MA	895.70	45	11.1000	1.0000	3	8
59	York, PA	911.82	42	9.0000	4.8000	8	49
60	Youngstown, OH	954.44	38	10.7000	11.7000	13	39

```

27
28     PROC REG DATA=AirPollution; Title2 'Fit of Air Pollution with REG';
29     MODEL MORT = PRECIP EDUC NONWHITE NOX SO2 / VIF;
30     RUN;
31

```

NOTE: The PROCEDURE REG printed page 2.

NOTE: PROCEDURE REG used (Total process time):

```

    real time      0.10 seconds
    cpu time       0.04 seconds

```

Chapter 11 : Air pollution and mortality study
Fit of Air Pollution with REG

The REG Procedure

Model: MODEL1

Dependent Variable: MORT

```

Number of Observations Read      60
Number of Observations Used      60

```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	152888	30578	21.90	<.0001
Error	54	75385	1396.01827		
Corrected Total	59	228273			

```

Root MSE      37.36333      R-Square      0.6698
Dependent Mean 940.35677      Adj R-Sq      0.6392
Coeff Var     3.97331

```

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Variance Inflation
Intercept	1	1000.10208	92.39811	10.82	<.0001	0
PRECIP	1	1.37920	0.70000	1.97	0.0539	2.06457
EDUC	1	-15.07895	7.07060	-2.13	0.0375	1.50971
NONWHITE	1	3.16024	0.62874	5.03	<.0001	1.32969
NOX	1	-0.10756	0.13589	-0.79	0.4321	1.67530
SO2	1	0.35541	0.09140	3.89	0.0003	1.41859

```

32     PROC REG DATA=AirPollution; Title2 'Fit of Air Pollution with REG';
33     MODEL MORT = PRECIP EDUC NONWHITE SO2;
34     output out=next1 r=resid p=yhat lclm=lclm uclm=uclm lcl=lcli ucl=ucli
35     student=student rstudent=rstudent cookd=cookd h=leverage dffits=dffits;
36     RUN;
37

```

NOTE: The data set WORK.NEXT1 has 60 observations and 18 variables.

NOTE: The PROCEDURE REG printed page 3.

NOTE: PROCEDURE REG used (Total process time):

```

    real time      0.10 seconds
    cpu time       0.04 seconds

```

Chapter 11 : Air pollution and mortality study
Fit of Air Pollution with REG

The REG Procedure

Model: MODEL1

Dependent Variable: MORT

Number of Observations Read 60
 Number of Observations Used 60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	152013	38003	27.41	<.0001
Error	55	76260	1386.54072		
Corrected Total	59	228273			

Root MSE	37.23628	R-Square	0.6659
Dependent Mean	940.35677	Adj R-Sq	0.6416
Coeff Var	3.95980		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	999.31617	92.07861	10.85	<.0001
PRECIP	1	1.61112	0.63358	2.54	0.0138
EDUC	1	-15.77337	6.99211	-2.26	0.0281
NONWHITE	1	3.06093	0.61400	4.99	<.0001
SO2	1	0.32718	0.08387	3.90	0.0003

```
38      proc print data=next1; var city resid yhat student rstudent cookd leverage dffits;
39      run;
```

NOTE: There were 60 observations read from the data set WORK.NEXT1.

NOTE: The PROCEDURE PRINT printed page 4.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.16 seconds
cpu time	0.04 seconds

Chapter 11 : Air pollution and mortality study
 Fit of Air Pollution with REG

Obs	city	resid	yhat	student	rstudent	cookd	leverage	dffits
1	Akron, OH	-1.8702	923.74	-0.05086	-0.05039	0.00001	0.02477	-0.00803
2	Albany, NY	92.2032	905.67	2.51726	2.65170	0.04242	0.03239	0.48513
3	Allentown, PA	33.4816	928.87	0.94701	0.94611	0.01960	0.09849	0.31272
4	Atlanta, GA	-8.4670	990.76	-0.23713	-0.23509	0.00098	0.08050	-0.06956
5	Baltimore, MD	12.0326	1059.26	0.34990	0.34709	0.00422	0.14708	0.14413
6	Birmingham, AL	-34.8401	1065.22	-1.02756	-1.02810	0.04353	0.17090	-0.46676
7	Boston, MA	25.9647	908.74	0.74803	0.74499	0.01688	0.13104	0.28930
8	Bridgeport, CT	-22.6217	922.15	-0.62458	-0.62108	0.00444	0.05389	-0.14824
9	Buffalo, NY	73.3065	928.60	2.00124	2.05936	0.02671	0.03227	0.37607
10	Canton, OH	-3.2465	915.59	-0.08864	-0.08784	0.00005	0.03260	-0.01613
11	Chattanooga, TN	9.1562	1008.46	0.25757	0.25537	0.00129	0.08862	0.07963
12	Chicago, IL	3.4817	1021.40	0.10737	0.10640	0.00073	0.24169	0.06007
13	Cincinnati, OH	-19.9664	990.43	-0.55311	-0.54959	0.00392	0.06019	-0.13908
14	Cleveland, OH	39.3936	946.56	1.06934	1.07077	0.00496	0.02122	0.15766
15	Columbus, OH	42.6084	916.23	1.17057	1.17460	0.01274	0.04442	0.25325
16	Dallas, TX	-55.1076	915.21	-1.51545	-1.53398	0.02231	0.04631	-0.33804
17	Dayton, OH	15.5434	920.69	0.42299	0.41981	0.00096	0.02613	0.06877
18	Denver, CO	17.1706	854.60	0.48902	0.48561	0.00596	0.11082	0.17144
19	Detroit, MI	-8.6209	967.84	-0.23739	-0.23534	0.00058	0.04882	-0.05332
20	Flint, MI	20.1864	920.99	0.55930	0.55577	0.00403	0.06050	0.14103
21	Fort Worth, TX	-13.2644	904.97	-0.36381	-0.36092	0.00114	0.04127	-0.07488
22	Grand Rapids, MI	-24.8758	896.21	-0.68357	-0.68022	0.00439	0.04487	-0.14744
23	Greensboro, NC	-2.9373	974.06	-0.08190	-0.08116	0.00010	0.07241	-0.02268
24	Hartford, CT	-25.0452	912.51	-0.69075	-0.68743	0.00522	0.05185	-0.16075
25	Houston, TX	-5.6891	958.22	-0.15793	-0.15652	0.00034	0.06409	-0.04096
26	Indianapolis, IN	27.7839	940.88	0.75681	0.75383	0.00330	0.02796	0.12786
27	Kansas City, MO	13.4275	906.30	0.36982	0.36690	0.00142	0.04924	0.08350
28	Lancaster, PA	-94.0410	938.09	-2.67375	-2.84035	0.17276	0.10780	-0.98731

29	Los Angeles, CA	-30.7567	892.59	-0.89885	-0.89726	0.02977	0.15555	-0.38510
30	Louisville, KY	-5.4728	994.74	-0.15747	-0.15607	0.00073	0.12887	-0.06003
31	Memphis, TN	-32.7995	1039.29	-0.96088	-0.96019	0.03508	0.15964	-0.41850
32	Miami, FL	-94.8005	956.24	-2.81722	-3.01770	0.35633	0.18333	-1.42977
33	Milwaukee, WI	-2.0666	931.22	-0.05702	-0.05650	0.00004	0.05255	-0.01331
34	Minneapolis, MN	-5.7431	863.37	-0.15925	-0.15783	0.00034	0.06198	-0.04057
35	Nashville, TN	-41.2964	1002.31	-1.13568	-1.13873	0.01254	0.04636	-0.25108
36	New Haven, CT	-1.5084	924.74	-0.04166	-0.04128	0.00002	0.05436	-0.00990
37	New Orleans, LA	83.3006	1029.76	2.43113	2.54983	0.21397	0.15327	1.08484
38	New York, NY	26.5155	968.13	0.72660	0.72344	0.00435	0.03954	0.14679
39	Philadelphia, PA	7.4175	1007.61	0.20691	0.20510	0.00068	0.07315	0.05762
40	Pittsburgh, PA	-9.6714	1000.96	-0.29641	-0.29394	0.00531	0.23218	-0.16164
41	Portland, OR	-1.0717	895.06	-0.02985	-0.02958	0.00001	0.07064	-0.00816
42	Providence, RI	18.2043	920.30	0.50795	0.50450	0.00410	0.07366	0.14226
43	Reading, PA	-5.1466	951.33	-0.14544	-0.14414	0.00045	0.09685	-0.04720
44	Richmond, VA	25.5562	999.95	0.71597	0.71276	0.00905	0.08110	0.21174
45	Rochester, NY	-22.7007	896.98	-0.62020	-0.61670	0.00269	0.03377	-0.11529
46	Saint Louis, MO	-22.4290	975.99	-0.63473	-0.63125	0.00890	0.09945	-0.20977
47	San Diego, CA	-9.4638	849.17	-0.27838	-0.27604	0.00310	0.16649	-0.12337
48	San Francisco, CA	5.7473	905.95	0.16343	0.16198	0.00065	0.10811	0.05639
49	San Jose, CA	-47.2571	837.99	-1.36711	-1.37824	0.05996	0.13822	-0.55198
50	Seattle, WA	12.0026	887.26	0.33326	0.33055	0.00153	0.06449	0.08679
51	Springfield, MA	-9.5281	913.68	-0.26467	-0.26242	0.00098	0.06528	-0.06935
52	Syracuse, NY	50.1384	900.53	1.37476	1.38623	0.01604	0.04070	0.28552
53	Toledo, OH	54.7197	917.74	1.50326	1.52110	0.02099	0.04437	0.32778
54	Utica, NY	-0.3447	912.55	-0.00957	-0.00948	0.00000	0.06324	-0.00246
55	Washington, DC	-16.2074	984.01	-0.47334	-0.46998	0.00818	0.15444	-0.20086
56	Wichita, KS	-53.0900	876.85	-1.46715	-1.48306	0.02536	0.05563	-0.35994
57	Wilmington, DE	59.1455	944.36	1.62322	1.64836	0.02337	0.04246	0.34710
58	Worcester, MA	-6.7148	902.41	-0.18807	-0.18642	0.00062	0.08068	-0.05522
59	York, PA	-43.9304	955.75	-1.27716	-1.28469	0.05608	0.14669	-0.53265
60	Youngstown, OH	14.1052	940.34	0.38281	0.37982	0.00062	0.02084	0.05541

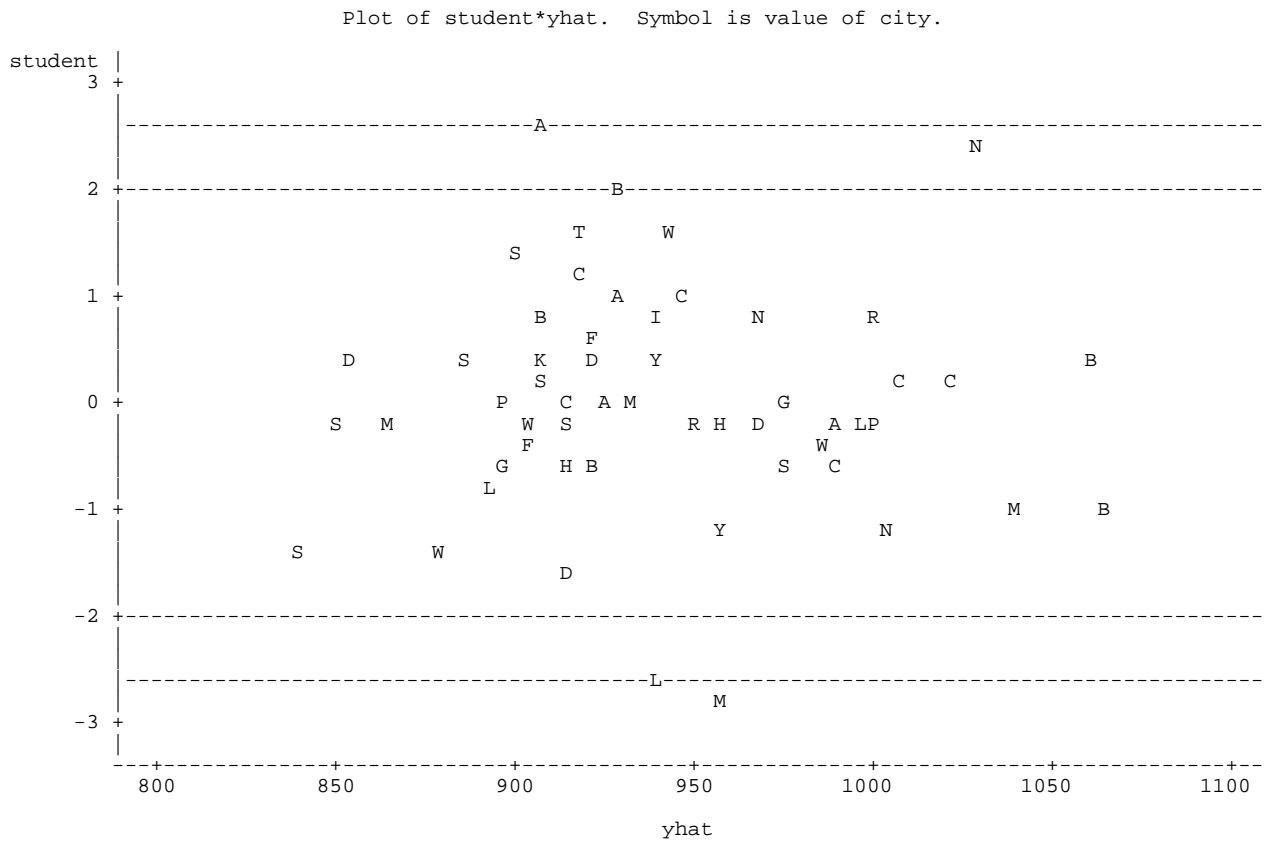
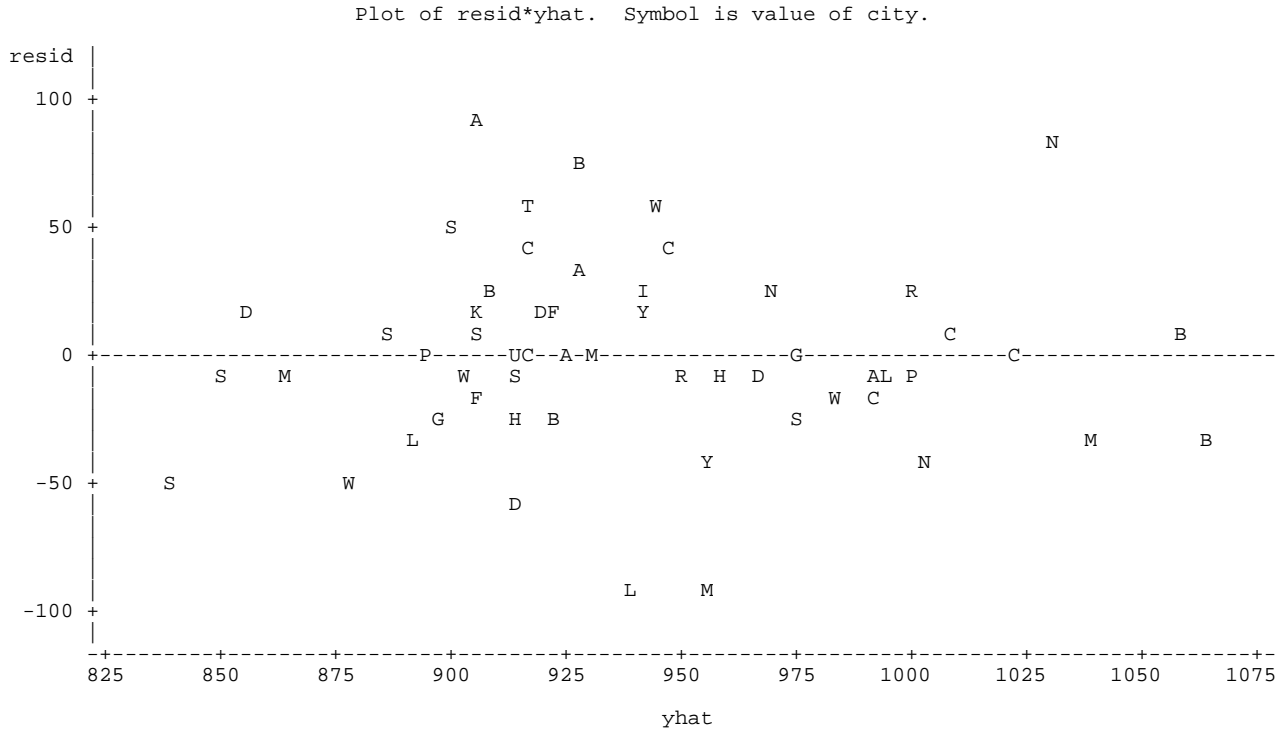
```

40
41     options ps=60 ls=111;
42     proc plot data=next1; TITLE2 'Various plot with group variable';
43         plot resid * yhat = city / vref=0;
44     options ps=44 ls=99;
NOTE: There were 60 observations read from the data set WORK.NEXT1.
NOTE: The PROCEDURE PLOT printed page 5.
NOTE: PROCEDURE PLOT used (Total process time):
      real time          0.05 seconds
      cpu time           0.01 seconds

45     proc plot data=next1; TITLE2 'Various plot with group variable';
46         plot student * yhat = city / vref= -2.6, -2, 2, 2.6;
47         plot rstudent * yhat = city / vref=-2.6, -2, 2, 2.6;
48         plot leverage * yhat = city / vref=0.18; *** 2p/n = 2*5/60 = 0.18 ***;
49         plot cookd * yhat = city / vref=1;
50         plot dffits * yhat = city / vref=1;
51     RUN;
52     !     OPTIONS PS=256 ls=132;
NOTE: There were 60 observations read from the data set WORK.NEXT1.
NOTE: The PROCEDURE PLOT printed pages 6-10.
NOTE: PROCEDURE PLOT used (Total process time):
      real time          0.10 seconds
      cpu time           0.01 seconds

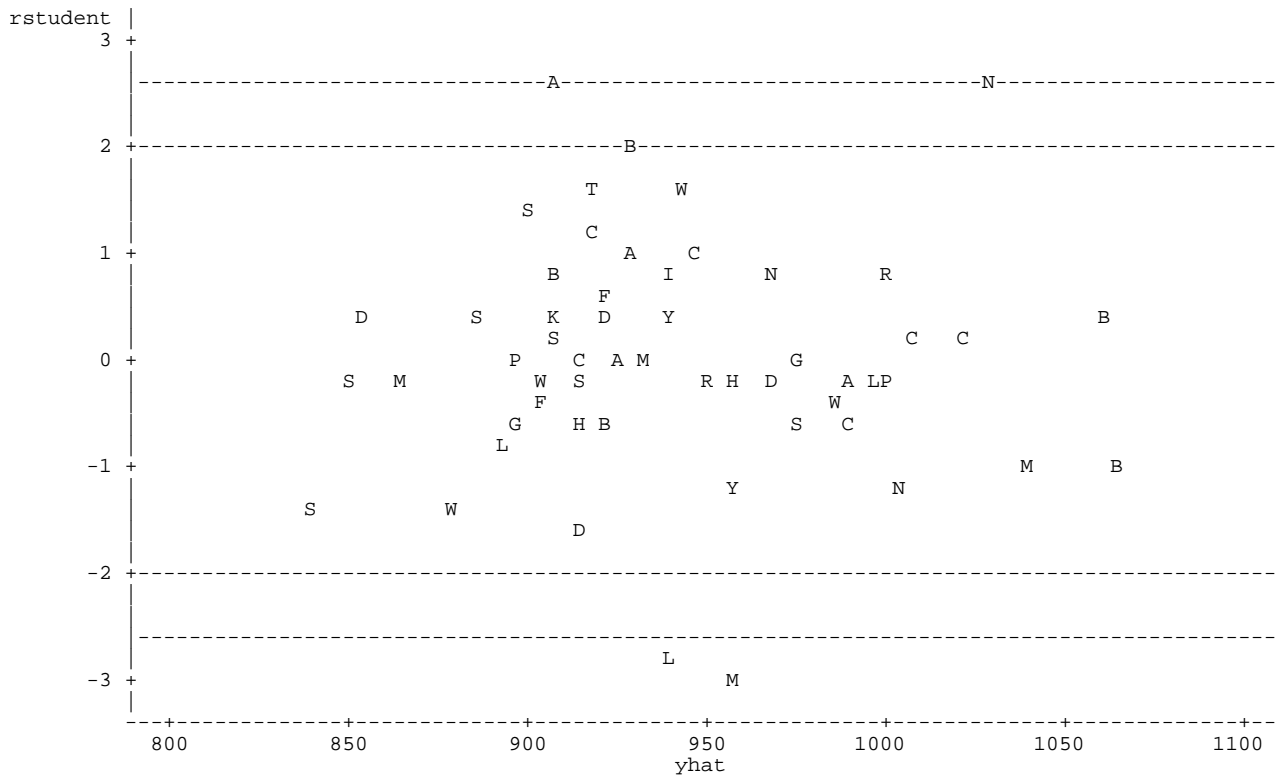
```

Chapter 11 : Air pollution and mortality study
 Various plot with group variable

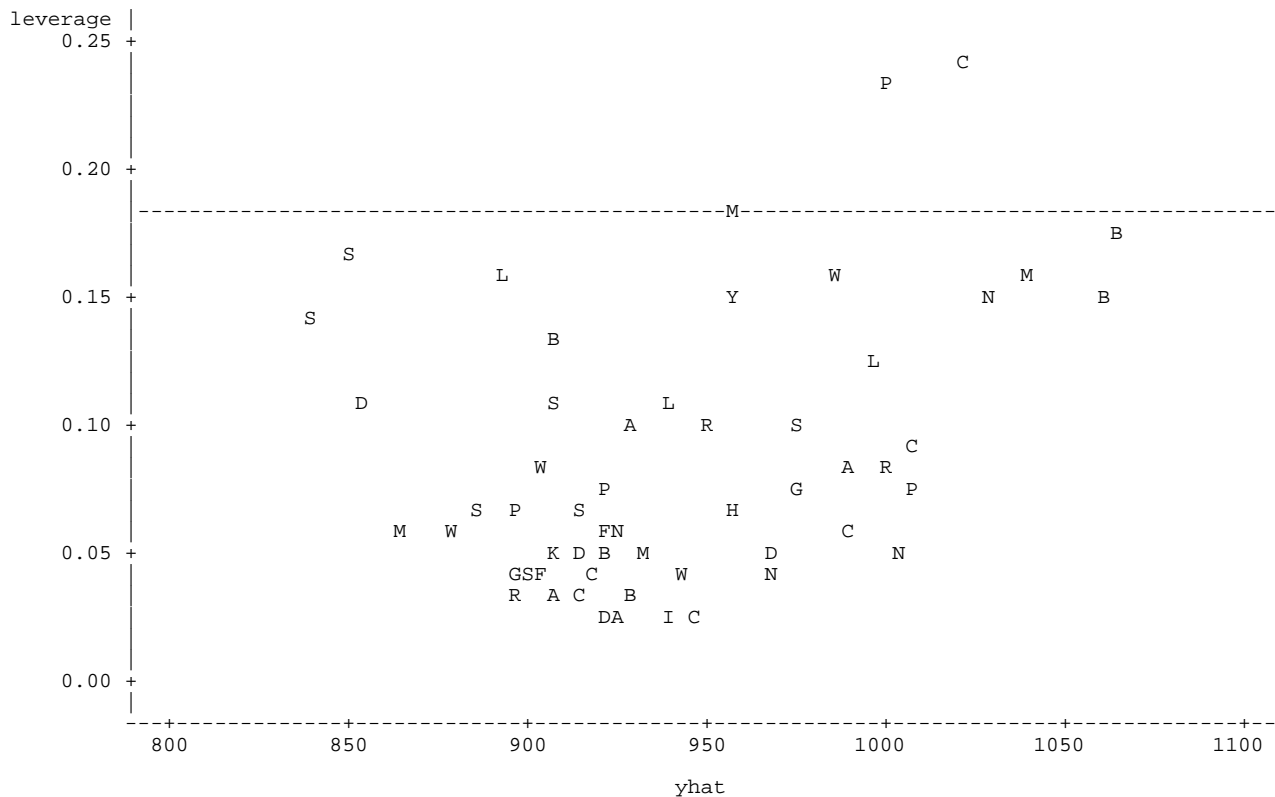


Chapter 11 : Air pollution and mortality study
 Various plot with group variable

Plot of $rstudent \cdot \hat{y}$. Symbol is value of city.

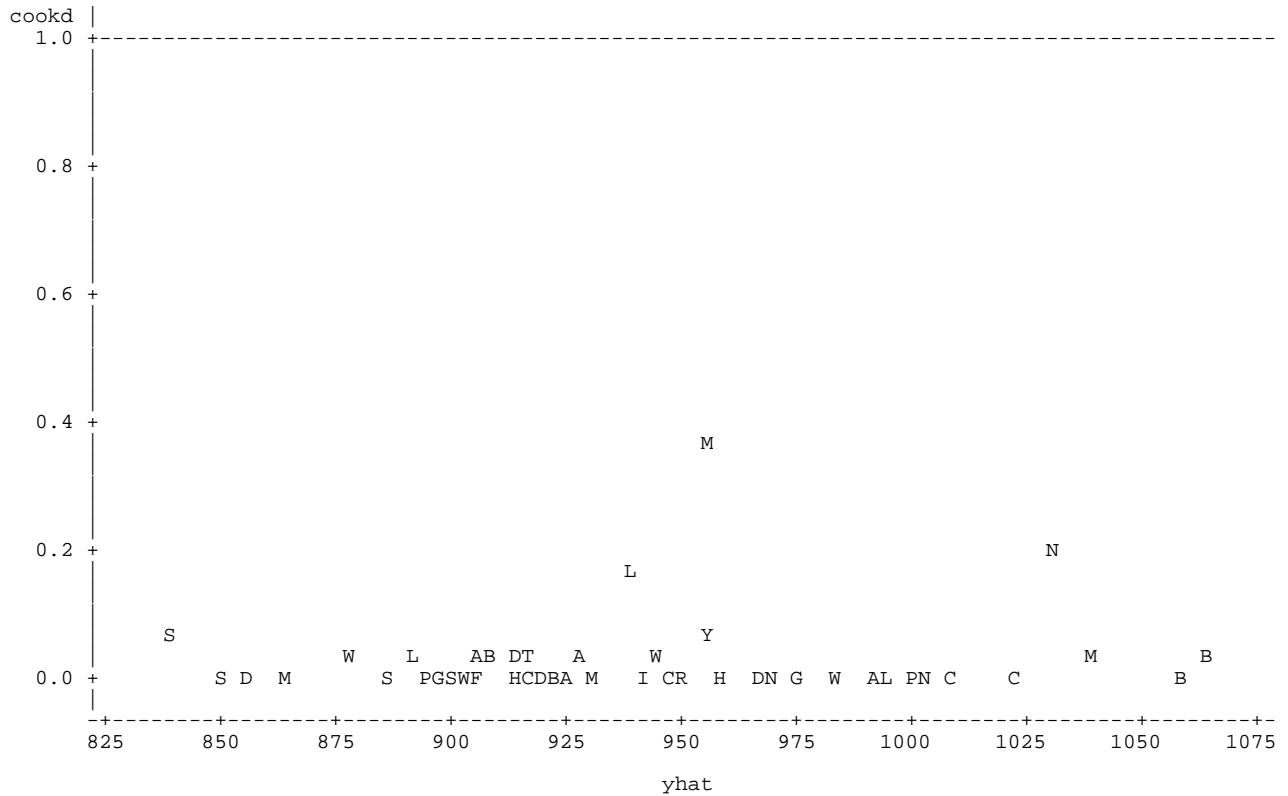


Plot of $leverage \cdot \hat{y}$. Symbol is value of city.



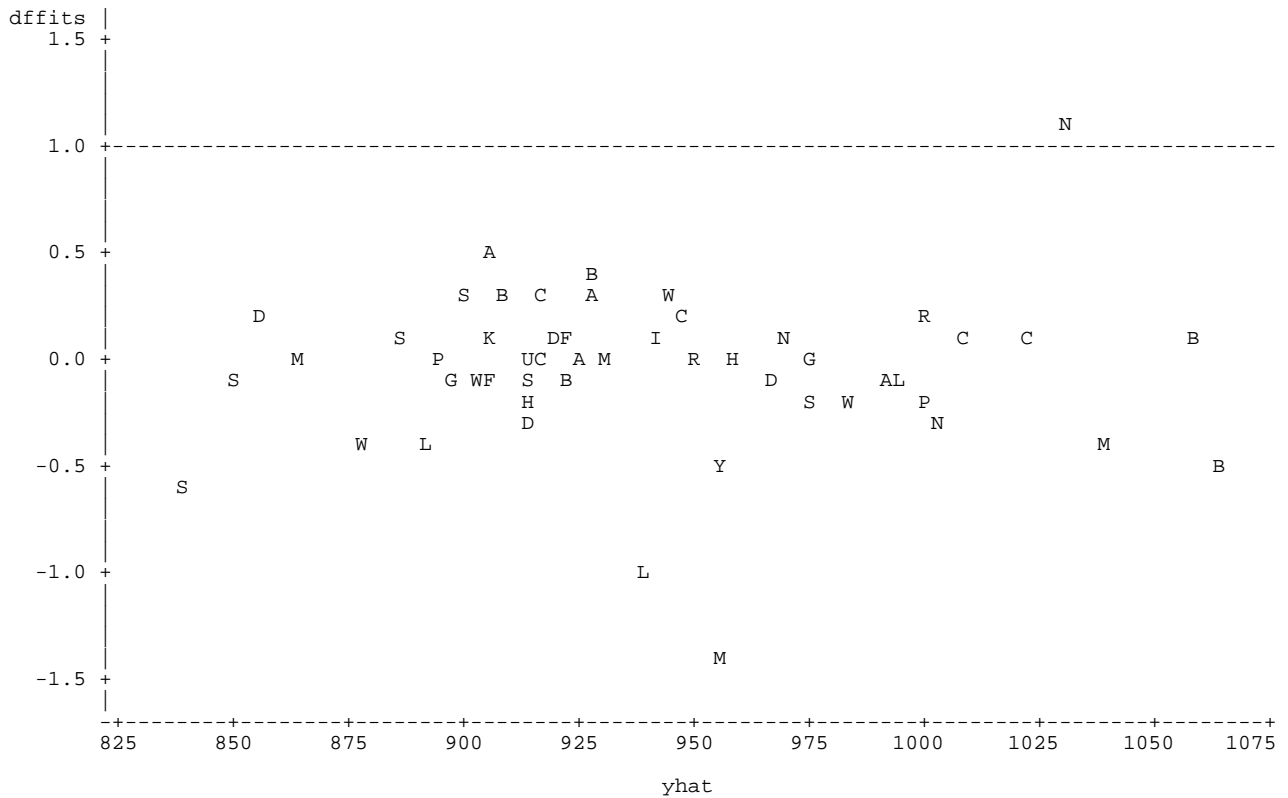
Chapter 11 : Air pollution and mortality study
 Various plot with group variable

Plot of cookd*yhat. Symbol is value of city.



NOTE: 15 obs hidden.

Plot of dffits*yhat. Symbol is value of city.



NOTE: 8 obs hidden.


```

52          PROC UNIVARIATE DATA=NEXT1 NORMAL PLOT; VAR resid; RUN;
NOTE: The PROCEDURE UNIVARIATE printed page 11.
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time          0.12 seconds
      cpu time           0.05 seconds
53
54          ods html close;
55          quit;
NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time          1.95 seconds
      cpu time           0.67 seconds

```

Chapter 11 : Air pollution and mortality study
Various plot with group variable

The UNIVARIATE Procedure
Variable: resid

Moments

N	60	Sum Weights	60
Mean	0	Sum Observations	0
Std Deviation	35.951884	Variance	1292.53796
Skewness	0.0312791	Kurtosis	1.10303795
Uncorrected SS	76259.7396	Corrected SS	76259.7396
Coeff Variation	.	Std Error Mean	4.64136826

Basic Statistical Measures

Location		Variability	
Mean	0.00000	Std Deviation	35.95188
Median	-2.50195	Variance	1293
Mode	.	Range	187.00371
		Interquartile Range	38.88520

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 0	Pr > t	1.0000
Sign	M -5	Pr >= M	0.2451
Signed Rank	S -21	Pr >= S	0.8787

Tests for Normality

Test	--Statistic--	-----p Value-----	
Shapiro-Wilk	W 0.971756	Pr < W	0.1777
Kolmogorov-Smirnov	D 0.093961	Pr > D	>0.1500
Cramer-von Mises	W-Sq 0.110757	Pr > W-Sq	0.0825
Anderson-Darling	A-Sq 0.640932	Pr > A-Sq	0.0924

Quantiles (Definition 5)

Quantile	Estimate
100% Max	92.20319
99%	92.20319
95%	66.22597
90%	46.37344
75% Q3	17.68746
50% Median	-2.50195
25% Q1	-21.19774
10%	-42.61341
5%	-54.09883
1%	-94.80052
0% Min	-94.80052

Extreme Observations

-----Lowest-----		-----Highest-----	
Value	Obs	Value	Obs
-94.8005	32	54.7197	53
-94.0410	28	59.1455	57
-55.1076	16	73.3065	9
-53.0900	56	83.3006	37
-47.2571	49	92.2032	2

```

Stem Leaf   Boxplot
 9  2
 8  3
 7  3
 6
 5 059
 4  3
 3 39
 2 06678
 1 2234678
 0 3679
-0 998766553322210
-1 6300
-2 553320
-3 531
-4 741
-5 53
-6
-7
-8
-9 54
-----+-----+-----+

```

Multiply Stem.Leaf by 10**+1

```

1  0
1  0
1  |
3  |
1  |
2  |
5  |
7  +-----+
4  |  +  |
15 *-----*
4  |  |  |
6  +-----+
3  |
3  |
2  |
2  0

```

Normal Probability Plot

