EXST7015 - Statistical Techniques **NAME ____**
October 14, 2010 – Exam 1 — **Geaghan**

READ CAREFULLY. Give an answer in the form of a number or numeric expression where possible. Show calculations for possible partial credit. Tables of the t distribution are provided. Use a critical value of α=0.05 if not specified in a particular question.

1) 2 points each – Answer the following questions as true (T) or false (F) by **circling** the appropriate letter.

T [✓ F] a) The equation $\hat{Y}_i = b_0 + b_1 X_i$ would be a correct equation if the term "+ e_i" was added to it.

[✓ T] F b) For a simple linear regression the value of R^2 can never exceed 1.0.

T [✓ F] c) Prior to doing regression analysis, the raw data values for Y_i should be examined to determine if they are normally distributed in order to meet the usual regression assumptions.

[✓ T] F d) There are no *a priori* expectations or limitations on the distribution of the independent variables in a regression.

[✓ T] F e) Hat values (i.e. h_{ii}) are NOT based on calculations of some change from deleted observations.

T [✓ F] f) Logarithmic transformation of an Y_i variable in regression does not affect homogeneity of variance.

[✓ T] F g) The sum of squares that a variable accounts for, or removes from the remaining error term when it enters the model, is called the extra sum of squares.

[✓ T] F h) Sequential sums of squares (TYPE I) will always sum to a number that is equal to the model sum of squares.

All multiple choice questions have one correct response unless otherwise indicated.
<u>Circle</u> the letter corresponding to the best response.

2) 3 points – For an ordinary multiple regression, with n = 17 observations (Y_1, Y_2, … Y_{17}) and 4 quantitative variables (X_1, X_2, X_3 and X_4) plus an intercept, the size of the X'X matrix (X prime X) will be which of the following?

a) 4 x 17.

b) 4 x 4

c) 17 x 17

[✓ d)] 5 x 5

3) 3 points – Which of the following best describes a partial residual plot of Y and X3 in a 4 variable multiple regression with X1, X2, X3 and X4?

a) plot Y on X3 after adjusting Y for X1, X2, X3 and X4

b) plot Y on X3 after adjusting both Y and X3 for X1, X2, X3 and X4.

c) plot Y on X3 after adjusting Y for X1, X2 and X4

[✓ d)] plot Y on X3 after adjusting both Y and X3 for X1, X2 and X4.

4) 3 points each – For the questions that follow, choose one diagnostic from this list (SAS names used). **Circle** the letter corresponding to your answer.

a) Some diagnostic statistics require that the regression be fitted with an observation deleted in order to compare the result to the regression with the observation included. Which of the statistics below is calculated using a "deleted" observation**?**

a) STUDENT **[✓ b) RSTUDENT]** **c) HATDIAG** **d) VIF**

b) Which of the statistics below is NOT calculated from a "deleted" observation?

a) **COOK's D** **b) DFFITS** **c) DFBETAS** **[✓ d) HATDIAG]**

c) Some statistics are diagnostics for observations and some for variables. Which of the statistics below is a diagnostic for a variable?

a) STUDENT **b) RSTUDENT** **c) HATDIAG** **[✓ d) VIF]**

d) Some normally distributed statistics are "standardized" in order to obtain scores that approximate a Z distribution or t distribution. Which one of the following are "standardized"?

a) R^2 **[✓ b) RSTUDENT]** **c) HATDIAG** **d) VIF**

e) Which of the statistics below could be produced by "standardizing" the Y_i variable and the X_i variables before doing the regression ?

a) RSTUDENT **b) COOK's D** **c) HATDIAG** **[✓ d) Standardized b_i]**

f) Some diagnostic statistics are called "influence diagnostics". Which of the statistics below is **NOT** an influence diagnostic?

a) **COOK's D** **b) DFFITS** **c) DFBETAS** **[✓ d) HATDIAG]**

g) In regression analysis, how would the values of each of the TYPE II sum of squares compare to the values of the TYPE III sum of squares for the same variable?

a) **Type II larger** **b) Type II smaller** **[✓ c) would be the same]** **d) cannot say**

h) In regression analysis, how would the values of each of the TYPE I sum of squares compare to the values of the TYPE II sum of squares for the same variable?

a) **Type II larger** **b) Type II smaller** **c) would be the same** **[✓ d) cannot say]**

i) Which of the diagnostic tools below would NOT be used to examine the observations for potential outliers of Y_i values?

a) STUDENT **b) RSTUDENT** **[✓ c) HATDIAG]** **d) a residual plot**

j) Which of the statistics below would be the better statistic to use as a diagnostic for multicolinearity?

a) STUDENT **b) COOK's D** **c) R^2** **[✓ d) VIF]**

Put your name on the output and RETURN THE OUTPUT WITH YOUR EXAM!

The remaining questions refer to the SAS output pages given separately. These pages contain output for a multiple regression fitted with PROC REG. t–tables are provided with the computer output. Questions requesting a "P value" may not always have a P value associated. If no P value is associated with the answer, state what statistic or graphic was used to answer the question.

A team of investigators are analyzing some crime of crime statistics from 47 states. They make some *a priori* decisions to streamline their multiple regression analysis.

1) they assume that there will be multicolinearity in the analysis if all variables are used.

2) they assume that a stepwise regression will eliminate the multicolinearity issues.

3) They plan to use the reduced set of variables from the stepwise regression to do their analysis.

The computer output provided contains (1) the SAS program, (2) the full model analysis with selected diagnostics, (3) the summary from the stepwise regression and (4) the reduced model analysis with selected diagnostics. **Unless otherwise stated the questions below will be about the reduced model analysis.**

9a) 3 points – The investigators believed that the full model would have multicolinearity? Based the original full model, does the evidence support this believe?

Circle one: [✓ **Yes**] / **No** P value or statistic used = Condition No & VIF .

9b) 3 points – One of the key hypotheses by the investigators was that there would be a relationship between the wealth (i.e. the variable "Wealth") of a state and crime. Based the original full model, does the evidence support this hypothesis?

Circle one: **Yes** / [✓ **No**] P value or statistic used = 0.2533 .

9c) 3 points – The investigators used stepwise regression (forward selection with an option to remove non-significant variables) to reduce the model. Based on the summary of the stepwise regression, what was the first variable entered into the model and what proportion of the total variation did it accounted for?

Variable name Po1 Proportion of variance = 0.4728 .

9d) 3 points – The investigators used stepwise regression (forward selection with an option to remove non-significant variables) to reduce the model. Based on the stepwise regression summary, which variable was entered second in the model and what was its P value at the time it was entered?

Variable name Ineq P value or statistic used = 0.0016 .

The remaining questions should be answered with results of the "Reduced model multiple regression".

9e) 4 points – The investigators used stepwise regression to eliminate the multicolinearity they believed would be present in the original model. Were they correct that multicolinearity, if present originally, was eliminated from the analysis?

Circle one: **Yes** / [✓ **No**] P value or statistic used = Condition Number & VIF.

9f) 4 points – Of all the variables examined, which one, or ones, appear to reduce crime?

Variable names Prob Statistic used = negative slope .

9g) 4 points – Published studies indicate that the number of crimes is reduced as the mean years of schooling (Ed) increases. Does this contention appear to be correct?

Circle one: **Yes** / [✓ **No**] P value or statistic used = positive slope .

9h) 3 points – Is the assumption of "normality" met?

Circle one: [✓ **Yes**] / **No** P value or statistic used = 0.2903 .

9i) 3 points – Is the assumption of "homogeneity" met?

Circle one: [✓ **Yes**] / [✓ **No**] P value or statistic used = residual plot .

9j) 3 points – Which of the variables in the reduced model appears to be the most important to the analysis?

Variable name Po1 Statistic used = t value (& P value) .

9k) 3 points – What is the standard error of the intercept? 816.2794 .

9l) 3 points – The coefficient of the variable "Ineq" suggests that there are an additional 68.31 crimes for each additional percent of income inequality. Place a confidence interval on the effect of each percentage point of income inequality on crime.

$$P(38.90688 \le \text{estimate} \le 97.71363) = 0.95$$

9m) 3 points – The estimated, or predicted, rate of crime for the first state, labeled "a", was 754.96. Place a confidence interval this estimate for this one, particular state.

$$P(309.01 \le \text{estimate} \le 1200.92) = 0.95$$

9n) 3 points – The 99.5% two-tailed t-value for 41 degrees of freedom is 2.97? The investigators have decided to use this value to define an "outlier". Based on this value, do you think there are outliers in this analysis and if so, which observation?

Circle one: [✓ **Yes**] / **No** Observation number = #29 or "C" (Rstudent) .

9o) 3 points – Are there any particularly unusual combinations of values of the X_i variables? If you find an observation that occupies an unusual value in the "X space" give the observation number.

Circle one: [✓ **Yes**] / **No** Observation number = #18 or "r", et al. (hat value).

hii = 2(p/n) = 2(6/47) = 0.2553

Office use only

Do not write in this box

```
dm'log;clear;output;clear';
options ps=512 ls=120 nocenter nodate nonumber nolabel FORMCHAR="|----|+|---+=|-/\<>*";
TITLE1 'Crime rate statistics example';

ODS HTML style=minimal body='C:\Geaghan\Current\EXST7015\Fall2010\SAS\Crime
Rates\CrimeRateStats.HTML' ;

*****************************************************;
*** OzDASL - Australasian Data and Story Library ***;
*** http://www.statsci.org/data/index.html       ***;
**************************************************************************;
*** Criminologists are interested in the effect of punishment regimes on ***;
*** crime rates. This has been studied using aggregate data on 47 states ***;
*** of the USA for 1960. The data set contains the following columns:    ***;
**************************************************************************;

data crime; infile cards missover;
   input M So Ed Po1 Po2 LF MpF Pop NW U1 U2 Wealth Ineq Prob Time Crime ObsID $;
      label M = percentage of males aged 14–24 in total state population
            So = indicator variable for a southern state
            Ed = mean years of schooling of the population aged 25 years or over
            Po1 = per capita expenditure on police protection in 1960
            Po2 = per capita expenditure on police protection in 1959
            LF = labor force participation rate by urban males in the age-group 14-24
            M.F = number of males per 100 females
            Pop = state population in 1960 in hundred thousands
            NW = percentage of nonwhites in the population
            U1 = unemployment rate of urban males 14–24
            U2 = unemployment rate of urban males 35–39
            Wealth = wealth: median value of transferable assets or family income
            Ineq = income inequality: % of families earning below half the median income
            Prob = probability of imprisonment: ratio of commitment number to offenses
            Time = average time in months served by first offenders in state prisons
            Crime = crime rate: number of offenses per 100,000 population in 1960;
datalines; run;

data here

;

proc print data=crime;
    TITLE3 'Raw data list'; run;

proc reg data=crime LINEPRINTER;
   TITLE2 'Original Full model multiple regression';
   model Crime = M So Ed Po1 LF MpF Pop NW U1 Wealth Ineq Prob Time / clb vif collin;
run;

proc reg data=crime LINEPRINTER;
   TITLE2 'Stepwise regression';
   model Crime = M So Ed Po1 LF MpF Pop NW U1 Wealth Ineq Prob Time / selection=stepwise;
run;

proc reg data=crime LINEPRINTER; ID ObsID;
   TITLE2 'Reduced model multiple regression';
   model Crime = M Ed Po1 Ineq Prob / clb vif collin;
   Police:Test PO1 = 100;
   OUTPUT OUT=NEXT1 P=Predicted R=Resid cookd=cooksd dffits=dffits h=hatdiag
      STUDENT=student rstudent=rstudent lclm=lclm uclm=uclm lcl=lcl ucl=ucl;
run;
```

```
proc print data=next1;
   TITLE3 'Listing of observation diagnostics';
   var ObsId crime Predicted Resid student rstudent cooksd dffits lcl ucl lclm uclm; run;
options ps=512 ls=85;

proc univariate data=next1 normal plot; var Resid;
    TITLE3 'Residual analysis'; run;

options ls=120 ps=61;
proc plot data=next1; plot resid*predicted=obsid=ObsID / vref=0;
   TITLE3 'Residual plot'; run;
```

```
Crime rate statistics example
Reduced model multiple regression
Raw data list


Obs     M  So    Ed   Po1   Po2     LF    MpF  Pop    NW     U1   U2  Wealth   Ineq     Prob     Time  Crime  ID
  1  15.1   1   9.1   5.8   5.6  0.510   95.0   33  30.1  0.108  4.1    3940   26.1  0.08460  26.2011    791   a
  2  14.3   0  11.3  10.3   9.5  0.583  101.2   13  10.2  0.096  3.6    5570   19.4  0.02960  25.2999   1635   b
  3  14.2   1   8.9   4.5   4.4  0.533   96.9   18  21.9  0.094  3.3    3180   25.0  0.08340  24.3006    578   c
  4  13.6   0  12.1  14.9  14.1  0.577   99.4  157   8.0  0.102  3.9    6730   16.7  0.01580  29.9012   1969   d
  5  14.1   0  12.1  10.9  10.1  0.591   98.5   18   3.0  0.091  2.0    5780   17.4  0.04140  21.2998   1234   e
  6  12.1   0  11.0  11.8  11.5  0.547   96.4   25   4.4  0.084  2.9    6890   12.6  0.03420  20.9995    682   f
  7  12.7   1  11.1   8.2   7.9  0.519   98.2    4  13.9  0.097  3.8    6200   16.8  0.04210  20.6993    963   g
  8  13.1   1  10.9  11.5  10.9  0.542   96.9   50  17.9  0.079  3.5    4720   20.6  0.04010  24.5988   1555   h
  9  15.7   1   9.0   6.5   6.2  0.553   95.5   39  28.6  0.081  2.8    4210   23.9  0.07170  29.4001    856   i
 10  14.0   0  11.8   7.1   6.8  0.632  102.9    7   1.5  0.100  2.4    5260   17.4  0.04450  19.5994    705   j
 11  12.4   0  10.5  12.1  11.6  0.580   96.6  101  10.6  0.077  3.5    6570   17.0  0.01620  41.6000   1674   k
 12  13.4   0  10.8   7.5   7.1  0.595   97.2   47   5.9  0.083  3.1    5800   17.2  0.03120  34.2984    849   l
 13  12.8   0  11.3   6.7   6.0  0.624   97.2   28   1.0  0.077  2.5    5070   20.6  0.04530  36.2993    511   m
 14  13.5   0  11.7   6.2   6.1  0.595   98.6   22   4.6  0.077  2.7    5290   19.0  0.05320  21.5010    664   n
 15  15.2   1   8.7   5.7   5.3  0.530   98.6   30   7.2  0.092  4.3    4050   26.4  0.06910  22.7008    798   o
 16  14.2   1   8.8   8.1   7.7  0.497   95.6   33  32.1  0.116  4.7    4270   24.7  0.05210  26.0991    946   p
 17  14.3   0  11.0   6.6   6.3  0.537   97.7   10   0.6  0.114  3.5    4870   16.6  0.07630  19.1002    539   q
 18  13.5   1  10.4  12.3  11.5  0.537   97.8   31  17.0  0.089  3.4    6310   16.5  0.11980  18.1996    929   r
 19  13.0   0  11.6  12.8  12.8  0.536   93.4   51   2.4  0.078  3.4    6270   13.5  0.01910  24.9008    750   w
 20  12.5   0  10.8  11.3  10.5  0.567   98.5   78   9.4  0.130  5.8    6260   16.6  0.03480  26.4010   1225   t
 21  12.6   0  10.8   7.4   6.7  0.602   98.4   34   1.2  0.102  3.3    5570   19.5  0.02280  37.5998    742   u
 22  15.7   1   8.9   4.7   4.4  0.512   96.2   22  42.3  0.097  3.4    2880   27.6  0.08950  37.0994    439   v
 23  13.2   0   9.6   8.7   8.3  0.564   95.3   43   9.2  0.083  3.2    5130   22.7  0.03070  25.1989   1216   w
 24  13.1   0  11.6   7.8   7.3  0.574  103.8    7   3.6  0.142  4.2    5400   17.6  0.04160  17.6000    968   x
 25  13.0   0  11.6   6.3   5.7  0.641   98.4   14   2.6  0.070  2.1    4860   19.6  0.06920  21.9003    523   y
 26  13.1   0  12.1  16.0  14.3  0.631  107.1    3   7.7  0.102  4.1    6740   15.2  0.04170  22.1005   1993   z
 27  13.5   0  10.9   6.9   7.1  0.540   96.5    6   0.4  0.080  2.2    5640   13.9  0.03610  28.4999    342   A
 28  15.2   0  11.2   8.2   7.6  0.571  101.8   10   7.9  0.103  2.8    5370   21.5  0.03820  25.8006   1216   B
 29  11.9   0  10.7  16.6  15.7  0.521   93.8  168   8.9  0.092  3.6    6370   15.4  0.02340  36.7009   1043   C
 30  16.6   1   8.9   5.8   5.4  0.521   97.3   46  25.4  0.072  2.6    3960   23.7  0.07530  28.3011    696   D
 31  14.0   0   9.3   5.5   5.4  0.535  104.5    6   2.0  0.135  4.0    4530   20.0  0.04200  21.7998    373   E
 32  12.5   0  10.9   9.0   8.1  0.586   96.4   97   8.2  0.105  4.3    6170   16.3  0.04270  30.9014    754   F
 33  14.7   1  10.4   6.3   6.4  0.560   97.2   23   9.5  0.076  2.4    4620   23.3  0.04950  25.5005   1072   G
 34  12.6   0  11.8   9.7   9.7  0.542   99.0   18   2.1  0.102  3.5    5890   16.6  0.04080  21.6997    923   H
 35  12.3   0  10.2   9.7   8.7  0.526   94.8  113   7.6  0.124  5.0    5720   15.8  0.02070  37.4011    653   I
 36  15.0   0  10.0  10.9   9.8  0.531   96.4    9   2.4  0.087  3.8    5590   15.3  0.00690  44.0004   1272   J
 37  17.7   1   8.7   5.8   5.6  0.638   97.4   24  34.9  0.076  2.8    3820   25.4  0.04520  31.6995    831   K
 38  13.3   0  10.4   5.1   4.7  0.599  102.4    7   4.0  0.099  2.7    4250   22.5  0.05400  16.6999    566   L
 39  14.9   1   8.8   6.1   5.4  0.515   95.3   36  16.5  0.086  3.5    3950   25.1  0.04710  27.3004    826   M
 40  14.5   1  10.4   8.2   7.4  0.560   98.1   96  12.6  0.088  3.1    4880   22.8  0.03880  29.3004   1151   N
 41  14.8   0  12.2   7.2   6.6  0.601   99.8    9   1.9  0.084  2.0    5900   14.4  0.02510  30.0001    880   O
 42  14.1   0  10.9   5.6   5.4  0.523   96.8    4   0.2  0.107  3.7    4890   17.0  0.08890  12.1996    542   P
 43  16.2   1   9.9   7.5   7.0  0.522   99.6   40  20.8  0.073  2.7    4960   22.4  0.05490  31.9989    823   Q
 44  13.6   0  12.1   9.5   9.6  0.574  101.2   29   3.6  0.111  3.7    6220   16.2  0.02810  30.0001   1030   R
 45  13.9   1   8.8   4.6   4.1  0.480   96.8   19   4.9  0.135  5.3    4570   24.9  0.05620  32.5996    455   W
 46  12.6   0  10.4  10.6   9.7  0.599   98.9   40   2.4  0.078  2.5    5930   17.1  0.04660  16.6999    508   T
 47  13.0   0  12.1   9.0   9.1  0.623  104.9    3   2.2  0.113  4.0    5880   16.0  0.05280  16.0997    849   U
```

```
Crime rate statistics example
Original Full model multiple regression

The REG Procedure
Model: MODEL1
Dependent Variable: Crime

Number of Observations Read          47
Number of Observations Used          47

Analysis of Variance
                                    Sum of           Mean
Source                      DF     Squares         Square    F Value    Pr > F
Model                       13     5303054         407927       8.53    <.0001
Error                       33     1577874          47814
Corrected Total             46     6880928

Root MSE            218.66495    R-Square     0.7707
Dependent Mean      905.08511    Adj R-Sq     0.6804
Coeff Var            24.15960

Parameter Estimates
                     Parameter       Standard                              Variance
Variable     DF       Estimate          Error    t Value    Pr > |t|      Inflation       95% Confidence Limits
Intercept     1    -5627.80237     1599.42791      -3.52      0.0013              0    -8881.86292    -2373.74183
M             1       74.94071       42.90448       1.75      0.0900        2.79712      -12.34912      162.23053
So            1       73.95947      151.28993       0.49      0.6282        5.05179     -233.84220      381.76115
Ed            1      137.71236       60.48884       2.28      0.0294        4.40532       14.64689      260.77784
Po1           1      110.42746       24.05473       4.59      <.0001        4.91664       61.48773      159.36718
LF            1     -111.46341     1473.11420      -0.08      0.9401        3.40949    -3108.53678     2885.60996
MpF           1       10.83717       20.93543       0.52      0.6082        3.66139      -31.75628       53.43062
Pop           1       -0.88472        1.34660      -0.66      0.5157        2.52852       -3.62439        1.85495
NW            1        2.04891        6.48129       0.32      0.7539        4.27319      -11.13737       15.23519
U1            1     1286.10195     2751.25126       0.47      0.6432        2.36697    -4311.36083     6883.56473
Wealth        1        0.12429        0.10690       1.16      0.2533       10.23609       -0.09320        0.34178
Ineq          1       77.21985       23.53797       3.28      0.0024        8.48394       29.33150      125.10821
Prob          1    -4307.88573     2288.86666      -1.88      0.0687        2.60559    -8964.61997      348.84850
Time          1        0.51678        6.96375       0.07      0.9413        2.34314      -13.65108       14.68463
```

```
Crime rate statistics example
Original Full model multiple regression

Collinearity Diagnostics
                         Condition   -----------------------------Proportion of Variation----------------------------
Number     Eigenvalue        Index     Intercept             M            So            Ed           Po1            LF

     1       11.97230      1.00000    0.00000274    0.00001966    0.00037927    0.00001659    0.00013480    0.00001009
     2        1.08720      3.31844    0.00000190   4.939959E-7       0.05785    0.00004790       0.00103    0.00001402
     3        0.54833      4.67271    0.00000703    0.00006592       0.00240    0.00006678    0.00059057    0.00003743
     4        0.13003      9.59563   2.191751E-8    0.00005649       0.22952    0.00000118    0.00032869    0.00002344
     5        0.10964     10.44964    0.00000936    0.00016458       0.23919    0.00002832    0.00005488    0.00001729
     6        0.07871     12.33347    0.00001098    0.00040662       0.05325    0.00019317       0.15326    0.00003376
     7        0.03444     18.64498    0.00000397    0.00055329       0.00511    0.00025550    0.00057362    0.00059071
     8        0.02116     23.78679    0.00021365       0.00508       0.00721       0.00402       0.01442       0.00689
     9        0.00986     34.84024    0.00012828    0.00014393       0.06111       0.02882       0.59482       0.00261
    10        0.00360     57.66028    0.00032243       0.59331       0.00903       0.01213       0.01664       0.03073
    11        0.00242     70.31930       0.00166       0.01128       0.13092       0.43510       0.10548       0.01455
    12        0.00147     90.33782       0.00231    0.00031451       0.18404       0.42236       0.03216       0.60102
    13     0.00066791    133.88445       0.34323       0.33530    0.00083729       0.09506       0.00368       0.07623
    14     0.00017379    262.47111       0.65210       0.05330       0.01917       0.00190       0.07682       0.26725

Collinearity Diagnostics
         ------------------------------------------Proportion of Variation-----------------------------------------
Number            MpF           Pop            NW            U1        Wealth          Ineq          Prob          Time

     1     0.00000165    0.00076148    0.00048229    0.00009458    0.00002056    0.00003174    0.00041416    0.00018237
     2     0.00000139       0.00710       0.03064    0.00012781    0.00011688    0.00002960       0.00268    0.00011798
     3     0.00000606       0.27244       0.01215    0.00041612    0.00001536    0.00007109       0.01063    0.00050416
     4    1.648241E-8       0.12955       0.44005    0.00010951    0.00001966    0.00000348       0.07333       0.01282
     5     0.00000532       0.08401       0.09072    0.00008216    0.00017967    0.00024098       0.33274       0.03311
     6     0.00000267       0.09793       0.01901       0.00430       0.00289       0.00661       0.00276       0.04557
     7     0.00000726       0.00394       0.02592       0.28906    0.00026386    0.00060212       0.16283       0.14450
     8     0.00022722       0.09489       0.00911       0.09684       0.00107       0.03636       0.25015       0.50016
     9     0.00005078       0.09443       0.19498    0.00004779       0.06687       0.16076       0.03749       0.03117
    10     0.00010347       0.05648       0.03947    0.00058946       0.04184       0.12505       0.00390       0.03714
    11     0.00001750       0.00565       0.01255       0.03887       0.57409       0.10605    0.00010774       0.08951
    12        0.00194       0.00807       0.05458       0.07918       0.06453       0.28457       0.00191       0.00279
    13        0.06563       0.00273       0.04052       0.17276       0.24495       0.27835       0.10227       0.03265
    14        0.93201       0.14202       0.02983       0.31752       0.00314       0.00126       0.01879       0.06978
```

```
Crime rate statistics example
Stepwise regression

All variables left in the model are significant at the 0.1500 level.

No other variable met the 0.1500 significance level for entry into the model.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Summary of Stepwise Selection
        Variable     Variable     Number     Partial      Model
Step    Entered      Removed      Vars In    R-Square    R-Square     C(p)      F Value    Pr > F
  1     Po1                            1      0.4728      0.4728     32.8690      40.36    <.0001
  2     Ineq                           2      0.1075      0.5803     19.3962      11.27    0.0016
  3     Ed                             3      0.0853      0.6656      9.1185      10.97    0.0019
  4     M                              4      0.0348      0.7004      6.1116       4.88    0.0327
  5     Prob                           5      0.0375      0.7379      2.7144       5.87    0.0199
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
```

```
Crime rate statistics example
Reduced model multiple regression

The REG Procedure
Model: MODEL1
Dependent Variable: Crime

Number of Observations Read          47
Number of Observations Used          47

Analysis of Variance
                                    Sum of           Mean
Source                   DF        Squares         Square    F Value    Pr > F
Model                     5        5077637        1015527      23.09    <.0001
Error                    41        1803290          43983
Corrected Total          46        6880928

Root MSE            209.72050    R-Square     0.7379
Dependent Mean      905.08511    Adj R-Sq     0.7060
Coeff Var            23.17136
```

```
Parameter Estimates
                     Parameter       Standard                              Variance
Variable      DF      Estimate          Error    t Value    Pr > |t|      Inflation       95% Confidence Limits
Intercept      1    -4064.57400      816.27940      -4.98      <.0001              0    -5713.08369    -2416.06431
M              1       79.68867       32.62004       2.44      0.0190        1.75773       13.81117      145.56618
Ed             1      160.15278       43.42207       3.69      0.0007        2.46788       72.46014      247.84543
Po1            1      121.22906       14.06269       8.62      <.0001        1.82675       92.82890      149.62923
Ineq           1       68.31025       14.55943       4.69      <.0001        3.52878       38.90688       97.71363
Prob           1    -3867.27087     1596.55188      -2.42      0.0199        1.37818    -7091.57279     -642.96894

Collinearity Diagnostics

                        Condition   -----------------------------Proportion of Variation-----------------------------
Number   Eigenvalue         Index     Intercept             M            Ed           Po1          Ineq          Prob
    1       5.67805       1.00000    0.00004333    0.00013971    0.00013376       0.00161    0.00033768       0.00358
    2       0.23398       4.92614    0.00004420    0.00000242    0.00080182       0.09644       0.00187       0.30263
    3       0.06349       9.45722    0.00065379       0.00636    0.00012658       0.25901       0.04879       0.68712
    4       0.01934      17.13422       0.00327    0.00077196       0.10399       0.52747       0.18636       0.00113
    5       0.00409      37.26070    0.00062678       0.75305       0.16328       0.00367       0.46334    0.00041977
    6       0.00105      73.68864       0.99537       0.23967       0.73167       0.11179       0.29931       0.00512


Crime rate statistics example
Reduced model multiple regression

The REG Procedure
Model: MODEL1

Test Police Results for Dependent Variable Crime
                                Mean
Source             DF         Square    F Value    Pr > F
Numerator           1         100232       2.28    0.1388
Denominator        41          43983



Crime rate statistics example
Reduced model multiple regression
Listing of observation diagnostics

    Obs
Obs ID  Crime Predicted    Resid  student rstudent hatdiag  cooksd  dffits       lcl   ucl     lclm    uclm
  1  a    791   754.96    36.037  0.18201  0.17985 0.10864 0.00067  0.06279   309.01 1200.92  615.36  894.57
  2  b   1635  1344.11   290.889  1.44403  1.46403 0.07739 0.02915  0.42401   904.49 1783.73 1226.29 1461.93
  3  c    578   423.12   154.882  0.78792  0.78422 0.12148 0.01431  0.29162   -25.41  871.65  275.50  570.74
  4  d   1969  1843.03   125.972  0.67140  0.66684 0.19960 0.01874  0.33301  1379.14 2306.92 1653.80 2032.25
```

```
    Obs
Obs ID  Crime Predicted    Resid  student rstudent hatdiag  cooksd  dffits       lcl   ucl     lclm    uclm
  5  e   1234  1346.78  -112.779 -0.56894 -0.56419 0.10660 0.00644 -0.19488   901.24 1792.32 1208.50 1485.06
  6  f    682   820.29  -138.287 -0.70951 -0.70515 0.13630 0.01324 -0.28012   368.81 1271.77  663.92  976.65
  7  g    963   704.05   258.953  1.26916  1.27896 0.05349 0.01517  0.30403   269.33 1138.76  606.09  802.00
  8  h   1555  1371.26   183.735  0.92422  0.92254 0.10143 0.01607  0.30996   926.76 1815.77 1236.37 1506.16
  9  i    856   771.25    84.755  0.42361  0.41933 0.08986 0.00295  0.13176   329.09 1213.41  644.28  898.21
 10  j    705   818.11  -113.109 -0.56285 -0.55810 0.08180 0.00470 -0.16658   377.59 1258.63  696.97  939.25
 11  k   1674  1170.66   503.338  2.53983  2.73284 0.10705 0.12888  0.94620   725.03 1616.29 1032.09 1309.23
 12  l    849   696.40   152.604  0.75090  0.74684 0.06096 0.00610  0.19028   260.14 1132.65  591.82  800.97
 13  m    511   809.40  -298.398 -1.51029 -1.53507 0.11245 0.04817 -0.54641   362.68 1256.12  667.37  951.43
 14  n    664   728.79   -64.787 -0.32507 -0.32150 0.09692 0.00189 -0.10532   285.20 1172.38  596.93  860.64
 15  o    798   767.19    30.809  0.15447  0.15262 0.09558 0.00042  0.04961   323.87 1210.51  636.25  898.13
 16  p    946   944.09     1.913  0.00964  0.00952 0.10493 0.00000  0.00326   498.88 1389.29  806.89 1081.29
 17  q    539   475.65    63.352  0.32492  0.32135 0.13565 0.00276  0.12731    24.29  927.00  319.65  631.64
 18  r    929   831.73    97.266  0.65489  0.65027 0.49847 0.07105  0.64828   313.27 1350.20  532.71 1130.76
 19  w    750  1229.21  -479.210 -2.39564 -2.55155 0.09024 0.09487 -0.80358   786.97 1671.45 1101.98 1356.44
 20  t   1225  1030.44   194.562  0.95662  0.95560 0.05951 0.00965  0.24037   594.48 1466.40  927.12 1133.75
 21  u    742   810.12   -68.125 -0.34396 -0.34023 0.10811 0.00239 -0.11846   364.28 1255.97  670.86  949.39
 22  v    439   720.91  -281.909 -1.44914 -1.46949 0.13957 0.05677 -0.59184   268.78 1173.04  562.68  879.14
 23  w   1216  1011.39   204.606  1.03504  1.03596 0.11153 0.02241  0.36704   564.86 1457.93  869.95 1152.84
 24  x    968   824.10   143.904  0.70708  0.70270 0.05826 0.00516  0.17478   388.39 1259.80  721.86  926.33
 25  y    523   664.17  -141.172 -0.72233 -0.71805 0.13156 0.01317 -0.27948   213.63 1114.71  510.55  817.80
 26  z   1993  1733.92   259.080  1.39475  1.41153 0.21550 0.08906  0.73981  1266.97 2200.87 1537.30 1930.54
 27  A    342   403.28   -61.277 -0.32101 -0.31747 0.17152 0.00356 -0.14445   -55.15  861.70  227.87  578.68
 28  B   1216  1255.42   -39.420 -0.19999 -0.19764 0.11667 0.00088 -0.07183   807.86 1702.98 1110.75 1400.09
 29  C   1043  1571.24  -528.242 -2.88542 -3.19253 0.23798 0.43335 -1.78410  1099.99 2042.49 1364.63 1777.86
 30  D    696   714.50   -18.502 -0.09593 -0.09476 0.15421 0.00028 -0.04046   259.48 1169.53  548.18  880.82
 31  E    373   411.03   -38.032 -0.19344 -0.19115 0.12112 0.00086 -0.07096   -37.42  859.49  263.63  558.43
 32  F    754   716.59    37.406  0.18397  0.18178 0.06000 0.00036  0.04593   280.53 1152.65  612.84  820.34
 33  G   1072   936.38   135.616  0.66903  0.66446 0.06579 0.00525  0.17632   499.14 1373.63  827.75 1045.02
 34  H    923   981.40   -58.397 -0.28706 -0.28383 0.05909 0.00086 -0.07113   545.52 1417.27  878.44 1084.35
 35  I    653   724.33   -71.326 -0.36664 -0.36273 0.13951 0.00363 -0.14606   272.21 1176.45  566.13  882.52
 36  J   1272  1072.14   199.857  1.12073  1.12434 0.27698 0.08020  0.69590   593.53 1550.76  849.24 1295.05
 37  K    831  1002.66  -171.661 -0.95836 -0.95739 0.27054 0.05677 -0.58305   525.25 1480.07  782.36 1222.96
 38  L    566   607.30   -41.298 -0.20759 -0.20515 0.10016 0.00080 -0.06844   163.05 1051.54  473.26  741.34
 39  M    826   804.07    21.928  0.10973  0.10840 0.09202 0.00020  0.03451   361.48 1246.67  675.59  932.55
 40  N   1151  1158.00    -6.999 -0.03458 -0.03416 0.06865 0.00001 -0.00927   720.16 1595.84 1047.02 1268.97
 41  O    880   828.13    51.869  0.27912  0.27596 0.21485 0.00355  0.14436   361.30 1294.96  631.81 1024.45
 42  P    542   301.04   240.957  1.27344  1.28345 0.18597 0.06175  0.61345  -160.20  762.29  118.40  483.69
 43  Q    823  1038.94  -215.942 -1.09121 -1.09382 0.10962 0.02443 -0.38379   592.79 1485.09  898.72 1179.17
 44  R   1030  1106.67   -76.673 -0.38075 -0.37674 0.07802 0.00204 -0.10959   666.92 1546.42  988.37 1224.97
 45  W    455   493.67   -38.674 -0.19702 -0.19470 0.12397 0.00092 -0.07324    44.65  942.70  344.55  642.80
 46  T    508   878.02  -370.019 -1.82919 -1.88531 0.06965 0.04175 -0.51584   439.98 1316.06  766.24  989.79
 47  U    849   889.05   -40.053 -0.19920 -0.19685 0.08081 0.00058 -0.05837   448.73 1329.37  768.65 1009.46
```

```
Crime rate statistics example
Reduced model multiple regression
Residual analysis

The UNIVARIATE Procedure
Variable:  Resid

Moments
N                          47    Sum Weights                47
Mean                        0    Sum Observations            0
Std Deviation      197.994856    Variance           39201.9629
Skewness           -0.3946102    Kurtosis           0.98133542
Uncorrected SS      1803290.3    Corrected SS        1803290.3
Coeff Variation             .    Std Error Mean      28.880518

Basic Statistical Measures
    Location                    Variability
Mean      0.00000     Std Deviation          197.99486
Median   -6.99881     Variance                   39202
Mode       .          Range                       1032
                      Interquartile Range    220.57621

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        15    Pr >= |S|   0.8759

Tests for Normality
Test                  --Statistic---    -----p Value------
Shapiro-Wilk          W     0.971049    Pr < W      0.2903
Kolmogorov-Smirnov    D     0.115245    Pr > D      0.1179
Cramer-von Mises      W-Sq   0.07433    Pr > W-Sq   0.2434
Anderson-Darling      A-Sq  0.494434    Pr > A-Sq   0.2139

Quantiles (Definition 5)

Quantile        Estimate
100% Max       503.33791
99%            503.33791
95%            259.08028
90%            240.95716
75% Q3         143.90368
50% Median      -6.99881
25% Q1         -76.67253
10%           -281.90902
5%            -370.01855
1%            -528.24222
0% Min        -528.24222

Extreme Observations
------Lowest-----        -----Highest-----
   Value      Obs           Value      Obs
-528.242       29         240.957       42
-479.210       19         258.953        7
-370.019       46         259.080       26
-298.398       13         290.889        2
-281.909       22         503.338       11
```

```
   Stem Leaf  Boxplot                                                     Normal Probability Plot
      5 0                        1     0                525+                                               *
      4                                                    |                                                +++
      4                                                    |                                              ++
      3                                                    |                                           +++
      3                                                    |                                         ++
      2 669                      3     |                   |                                      +*+* *
      2 004                      3     |                   |                                    +**
      1 5589                     4     |                175+                                 ****
      1 0344                     4  +-----+                |                               **
      0 568                      3  |     |                |                            ***
      0 02344                    5  |  +  |                |                         +***
     -0 4444421                  7  *-----*                |                      ****
     -0 877666                   6  +-----+                |                  ****+
     -1 4411                     4     |                   |               ***++
     -1 7                        1     |               -175+              *+++
     -2 2                        1     |                   |             *+
     -2 8                        1     |                   |          +**
     -3 0                        1     |                   |        ++
     -3 7                        1     |                   |     +++ *
     -4                                                    |   ++
     -4 8                        1     0                   |+++    *
     -5 3                        1     0               -525+   *
        ----+----+----+----+                                +----+----+----+----+----+----+----+----+----+----+
    Multiply Stem.Leaf by 10**+2                                -2        -1         0        +1        +2
```

```
Crime rate statistics example
Reduced model multiple regression
Residual plot

                                  Plot of Resid*Predicted.  Symbol is value of ObsID.

Resid |
      |
  600 +
      |
      |
      |                                                            k
      |
      |
      |
  400 +
      |
      |
      |
      |                                                                       b
      |                                g                                                             z
      |        P
  200 +                                                   wt J
      |                                                                        h
      |               c                l      x      G
      |                                                                                                     d
      |                                    i   r
      |                   q                    O
      |                                 F ao M
    0 +-----------------------------------------------p-----------N-----------------------------------------------------
      |               E    W     L      D         U                     B
      |              A                  In    u        H
      |                                                        R
      |                                       j                               e
      |                              y        f
      |                                                  K
 -200 +
      |                                                    Q
      |
      |                                 v     m
      |
      |
      |                                           T
 -400 +
      |
      |
      |                                                                w
      |                                                                                    C
      |
      |
 -600 +
      |
      ---+-----------+-----------+-----------+-----------+-----------+-----------+-----------+-----------+-----------+--
        200         400         600         800        1000        1200        1400        1600        1800        2000
                                                           Predicted
```

t - tables : Probability of a larger absolute value (two tailed test)

d.f.	***0.500***	***0.400***	***0.300***	***0.200***	***0.100***	***0.050***	***0.020***	***0.010***	***0.002***	***0.001***	***d.f.***
1	1.000	1.376	1.963	3.078	6.314	**12.706**	31.821	**63.656**	318.289	636.578	**1**
2	0.816	1.061	1.386	1.886	2.920	**4.303**	6.965	**9.925**	22.328	31.600	**2**
3	0.765	0.978	1.250	1.638	2.353	**3.182**	4.541	**5.841**	10.214	12.924	**3**
4	0.741	0.941	1.190	1.533	2.132	**2.776**	3.747	**4.604**	7.173	8.610	**4**
5	**0.727**	**0.920**	**1.156**	**1.476**	**2.015**	**2.571**	**3.365**	**4.032**	**5.894**	**6.869**	**5**
6	0.718	0.906	1.134	1.440	1.943	**2.447**	3.143	**3.707**	5.208	5.959	**6**
7	0.711	0.896	1.119	1.415	1.895	**2.365**	2.998	**3.499**	4.785	5.408	**7**
8	0.706	0.889	1.108	1.397	1.860	**2.306**	2.896	**3.355**	4.501	5.041	**8**
9	0.703	0.883	1.100	1.383	1.833	**2.262**	2.821	**3.250**	4.297	4.781	**9**
10	**0.700**	**0.879**	**1.093**	**1.372**	**1.812**	**2.228**	**2.764**	**3.169**	**4.144**	**4.587**	**10**
11	0.697	0.876	1.088	1.363	1.796	**2.201**	2.718	**3.106**	4.025	4.437	**11**
12	0.695	0.873	1.083	1.356	1.782	**2.179**	2.681	**3.055**	3.930	4.318	**12**
13	0.694	0.870	1.079	1.350	1.771	**2.160**	2.650	**3.012**	3.852	4.221	**13**
14	0.692	0.868	1.076	1.345	1.761	**2.145**	2.624	**2.977**	3.787	4.140	**14**
15	**0.691**	**0.866**	**1.074**	**1.341**	**1.753**	**2.131**	**2.602**	**2.947**	**3.733**	**4.073**	**15**
16	0.690	0.865	1.071	1.337	1.746	**2.120**	2.583	**2.921**	3.686	4.015	**16**
17	0.689	0.863	1.069	1.333	1.740	**2.110**	2.567	**2.898**	3.646	3.965	**17**
18	0.688	0.862	1.067	1.330	1.734	**2.101**	2.552	**2.878**	3.610	3.922	**18**
19	0.688	0.861	1.066	1.328	1.729	**2.093**	2.539	**2.861**	3.579	3.883	**19**
20	**0.687**	**0.860**	**1.064**	**1.325**	**1.725**	**2.086**	**2.528**	**2.845**	**3.552**	**3.850**	**20**
21	0.686	0.859	1.063	1.323	1.721	**2.080**	2.518	**2.831**	3.527	3.819	**21**
22	0.686	0.858	1.061	1.321	1.717	**2.074**	2.508	**2.819**	3.505	3.792	**22**
23	0.685	0.858	1.060	1.319	1.714	**2.069**	2.500	**2.807**	3.485	3.768	**23**
24	0.685	0.857	1.059	1.318	1.711	**2.064**	2.492	**2.797**	3.467	3.745	**24**
25	**0.684**	**0.856**	**1.058**	**1.316**	**1.708**	**2.060**	**2.485**	**2.787**	**3.450**	**3.725**	**25**
26	0.684	0.856	1.058	1.315	1.706	**2.056**	2.479	**2.779**	3.435	3.707	**26**
27	0.684	0.855	1.057	1.314	1.703	**2.052**	2.473	**2.771**	3.421	3.689	**27**
28	0.683	0.855	1.056	1.313	1.701	**2.048**	2.467	**2.763**	3.408	3.674	**28**
29	0.683	0.854	1.055	1.311	1.699	**2.045**	2.462	**2.756**	3.396	3.660	**29**
30	**0.683**	**0.854**	**1.055**	**1.310**	**1.697**	**2.042**	**2.457**	**2.750**	**3.385**	**3.646**	**30**
32	0.682	0.853	1.054	1.309	1.694	**2.037**	2.449	**2.738**	3.365	3.622	**32**
34	0.682	0.852	1.052	1.307	1.691	**2.032**	2.441	**2.728**	3.348	3.601	**34**
36	0.681	0.852	1.052	1.306	1.688	**2.028**	2.434	**2.719**	3.333	3.582	**36**
38	0.681	0.851	1.051	1.304	1.686	**2.024**	2.429	**2.712**	3.319	3.566	**38**
40	**0.681**	**0.851**	**1.050**	**1.303**	**1.684**	**2.021**	**2.423**	**2.704**	**3.307**	**3.551**	**40**
45	0.680	0.850	1.049	1.301	1.679	**2.014**	2.412	**2.690**	3.281	3.520	**45**
50	0.679	0.849	1.047	1.299	1.676	**2.009**	2.403	**2.678**	3.261	3.496	**50**
75	0.678	0.846	1.044	1.293	1.665	**1.992**	2.377	**2.643**	3.202	3.425	**75**
100	0.677	0.845	1.042	1.290	1.660	**1.984**	2.364	**2.626**	3.174	3.390	**100**
∞	**0.674**	**0.842**	**1.036**	**1.282**	**1.645**	**1.960**	**2.326**	**2.576**	**3.090**	**3.290**	**∞**
d.f.	***0.250***	***0.200***	***0.150***	***0.100***	***0.050***	***0.025***	***0.010***	***0.005***	***0.001***	***0.0005***	***d.f.***

t - tables : Probability of a larger value (one tailed test)