Sociology 63993
Exam 1
February 13, 2009

I. True-False. (20 points) Indicate whether the following statements are true or false. If false, briefly explain why.

1. A researcher has written her own computer program to compute regression estimates. She gets F = 17, R^2 = .25, Adjusted R^2 = .27. As far as we can tell, her program is working correctly.

2. Cook's distance is used to test for serial correlation.

3. One of the rare times when pairwise deletion of missing data is desirable is when skip patterns have caused data for some cases to be missing.

4. Random measurement error results in biased estimates of means, correlations and covariances.

5. Robust regression routines work best when it is the DVs that have outliers rather than the IVs.

II. Short answer. Discuss all <u>three</u> of the following problems. (15 points each, 45 points total.) In each case, the researcher has used Stata to test for a possible problem, concluded that there is a problem, and then adopted a strategy to address that problem. Explain (a) what problem the researcher was testing for, and why she concluded that there was a problem, (b) the rationale behind the solution she chose, i.e. how does it try to address the problem, and (c) one alternative solution she could have tried, and why. (NOTE: a few sentences on each point will probably suffice – you don't have to repeat everything that was in the lecture notes.)

II-1.

```
. reg   warmlt2 yr89 male white age ed prst

      Source |       SS       df       MS              Number of obs =    2293
-------------+------------------------------           F(  6,  2286) =   22.07
       Model |  14.1569236     6  2.35948727           Prob > F      =  0.0000
    Residual |  244.374258  2286  .106900375           R-squared     =  0.0548
-------------+------------------------------           Adj R-squared =  0.0523
       Total |  258.531182  2292    .1127972           Root MSE      =  .32696

------------------------------------------------------------------------------
     warmlt2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yr89 |  -.0905367    .014188    -6.38   0.000    -.1183594   -.0627139
        male |   .0355746   .0137434     2.59   0.010     .0086236    .0625255
       white |   .0460708   .0209917     2.19   0.028      .004906    .0872357
         age |   .0018563   .0004363     4.25   0.000     .0010006    .0027119
          ed |  -.0131147    .002827    -4.64   0.000    -.0186586   -.0075709
        prst |   .0004411   .0005846     0.75   0.451    -.0007054    .0015875
       _cons |   .1680543   .0413187     4.07   0.000     .0870283    .2490803
------------------------------------------------------------------------------
```

```
. hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of warmlt2

         chi2(1)      =   306.86
         Prob > chi2  =   0.0000

. tab1 warmlt2, nolabel

-> tabulation of warmlt2

      1=SD; |
   0=D,A,SA |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      1,996       87.05       87.05
          1 |        297       12.95      100.00
------------+-----------------------------------
      Total |      2,293      100.00

. reg warmlt2 yr89 male white age ed prst, robust

Linear regression                               Number of obs =    2293
                                                F(  6,  2286) =   21.21
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0548
                                                Root MSE      =  .32696

------------------------------------------------------------------------------
             |               Robust
     warmlt2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yr89 |  -.0905367   .0130228    -6.95   0.000    -.1160745   -.0649988
        male |   .0355746   .0139546     2.55   0.011     .0082096    .0629395
       white |   .0460708   .0183061     2.52   0.012     .0101726    .0819691
         age |   .0018563   .0004533     4.10   0.000     .0009673    .0027452
          ed |  -.0131147   .0031327    -4.19   0.000     -.019258   -.0069715
        prst |   .0004411   .0006136     0.72   0.472    -.0007622    .0016443
       _cons |   .1680543   .0421927     3.98   0.000     .0853144    .2507942
------------------------------------------------------------------------------
```

II-2.

```
. reg y x1 x2 x3 x4

      Source |       SS       df       MS              Number of obs =    2293
-------------+------------------------------           F(  4,  2288) =   24.60
       Model |   81.427377     4  20.3568442           Prob > F      =  0.0000
    Residual |   1893.3236  2288  .827501575           R-squared     =  0.0412
-------------+------------------------------           Adj R-squared =  0.0396
       Total |  1974.75098  2292  .861584198           Root MSE      =  .90967

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0001393    .003306     0.04   0.966    -.0063436    .0066223
          x2 |  -.0043145   .0033019    -1.31   0.191    -.0107895    .0021605
          x3 |  -.0025131   .0032995    -0.76   0.446    -.0089835    .0039573
          x4 |  -.0044104   .0033055    -1.33   0.182    -.0108925    .0020716
       _cons |   3.106225   .0539982    57.52   0.000     3.000334    3.212116
------------------------------------------------------------------------------
```

```
. test x1 = x2 = x3 = x4

 ( 1)  x1 - x2 = 0
 ( 2)  x1 - x3 = 0
 ( 3)  x1 - x4 = 0

       F(  3,  2288) =    0.31
            Prob > F =    0.8152

. gen x1234 = x1 + x2 + x3 + x4

. reg y x1234

      Source |       SS       df       MS              Number of obs =    2293
-------------+------------------------------           F(  1,  2291) =   97.55
       Model |   80.647724     1   80.647724           Prob > F      =  0.0000
    Residual |  1894.10326  2291  .826758296           R-squared     =  0.0408
-------------+------------------------------           Adj R-squared =  0.0404
       Total |  1974.75098  2292  .861584198           Root MSE      =  .90926

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       x1234 |  -.0027758   .0002811    -9.88   0.000     -.003327   -.0022247
       _cons |   3.106433   .0539674    57.56   0.000     3.000602    3.212263
------------------------------------------------------------------------------
```

II-3.

```
. reg price mpg weight length foreign

      Source |       SS       df       MS              Number of obs =     875
-------------+------------------------------           F(  4,   870) =  174.43
       Model |  1.0147e+09     4   253674918           Prob > F      =  0.0000
    Residual |  1.2653e+09   870     1454327           R-squared     =  0.4451
-------------+------------------------------           Adj R-squared =  0.4425
       Total |  2.2800e+09   874  2608654.65           Root MSE      =    1206

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         mpg |  -38.37705   10.34107    -3.71   0.000    -58.67342   -18.08068
      weight |  -.3910697   .2983449    -1.31   0.190    -.9766296    .1944903
      length |   61.42098   7.731232     7.94   0.000     46.24694    76.59503
     foreign |   1893.053   89.09917    21.25   0.000     1718.179    2067.928
       _cons |  -4470.567   943.7682    -4.74   0.000    -6322.895   -2618.238
------------------------------------------------------------------------------

. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       price |      1850    6165.257    2930.291       3291      15906
         mpg |      1850     21.2973    5.747833         12         41
      weight |       875    2312.571     342.109       1760       2930
      length |      1850    187.9324    22.12136        142        233
     foreign |      1850    .2972973    .4571921          0          1

. impute weight mpg length foreign, gen(xweight)
 52.70% (975) observations imputed
```

```
. reg price mpg xweight length foreign

      Source |       SS       df       MS              Number of obs =    1850
-------------+------------------------------           F(  4,  1845) =  240.20
       Model |  5.4367e+09     4  1.3592e+09           Prob > F      =  0.0000
    Residual |  1.0440e+10  1845  5658506.19           R-squared     =  0.3424
-------------+------------------------------           Adj R-squared =  0.3410
       Total |  1.5877e+10  1849  8586606.22           Root MSE      =  2378.8

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         mpg |  -143.8506   17.54604    -8.20   0.000    -178.2628   -109.4384
     xweight |   -.391066   .5884892    -0.66   0.506    -1.545241    .7631088
      length |   68.06994   13.55269     5.02   0.000     41.48971    94.65017
     foreign |   2611.786   156.4679    16.69   0.000     2304.913    2918.658
       _cons |  -3244.343   1307.184    -2.48   0.013     -5808.06   -680.6273
------------------------------------------------------------------------------
```

III. Computation and interpretation. (35 points total)

A graduate student wants to do her dissertation on the determinants of women's socio-economic status (SES). To see whether the idea is worth pursuing, she is analyzing a few key variables that were collected as part of a nationwide study of 488 women. Her measures include the following:

Variable	Description
ses	Socio-Economic Status scale. Ranges from a low of 0 to a high of 100.
nev_mar	Coded 1 if the woman has never been married, 0 otherwise
rural	Coded 1 if the respondent lives in a rural area, 0 otherwise
school	Number of years of schooling respondent has completed
tenure	Number of years respondent has worked in her current job

An analysis of the data yields the following results. [NOTE: You'll need some parts of the following to answer the questions, but other parts are extraneous. You'll have to figure out which is which.]

```
. reg  ses nev_mar rural school tenure

      Source |       SS       df       MS              Number of obs =     488
-------------+------------------------------           F(  4,   483) =   75.44
       Model |  29626.8441     4  7406.71104           Prob > F      =  [  1 ]
    Residual |  47422.5089   483  98.1832482           R-squared     =  [  2 ]
-------------+------------------------------           Adj R-squared =  0.3794
       Total |   77049.353   487  158.212224           Root MSE      =  9.9087

------------------------------------------------------------------------------
         ses |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     nev_mar |  -.1388159   1.001324    -0.14   0.890    -2.106304    1.828673
       rural |  -4.743383   1.025829    [ 3 ]   0.000    -6.759023   -2.727744
      school |   1.943179   .1719365    11.30   0.000     1.605343    2.281015
      tenure |   [   4  ]   .1232743     8.16   0.000     .7639161    1.248356
       _cons |   17.19019   2.273869     7.56   0.000     12.72229    21.65808
------------------------------------------------------------------------------

. pcorr2 ses nev_mar rural school tenure

(obs=488)

Partial and Semipartial correlations of ses with

    Variable |     Partial      SemiP    Partial^2     SemiP^2       Sig.
-------------+------------------------------------------------------------
     nev_mar |     -0.0063     -0.0049      0.0000      0.0000       0.890
       rural |     -0.2059     -0.1651      0.0424      0.0272       0.000
      school |      0.4573      0.4034      0.2091      0.1628       0.000
      tenure |      0.3481      0.2914      0.1212      0.0849       0.000

. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         ses |       488    43.32709    12.57824   2.465307    84.2362
     nev_mar |       488    .2868852    .4527717          0          1
       rural |       488     .272541    .4457236          0          1
      school |       488    12.71107     2.70533          0         18
      tenure |       488    2.752732    3.776793          0      21.75

. test  nev_mar rural school tenure

 ( 1)  nev_mar = 0
 ( 2)  rural = 0
 ( 3)  school = 0
 ( 4)  tenure = 0

       F(  4,   483) =   75.44
            Prob > F =    0.0000

. collin nev_mar rural school tenure

  Collinearity Diagnostics

                     SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
   nev_mar      1.02    1.01    0.9808      0.0192
     rural      1.04    1.02    0.9643      0.0357
    school      1.07    1.04    [  5 ]      0.0682
    tenure      1.08    1.04    0.9301      0.0699
----------------------------------------------------
  Mean VIF      1.05
```

```
. estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(12)     =      6.91
         Prob > chi2  =    0.8637

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |       6.91     12    0.8637
            Skewness |       1.50      4    0.8272
            Kurtosis |       6.72      1    0.0096
---------------------+-----------------------------
               Total |      15.12     17    0.5868
---------------------------------------------------

. test school = tenure

 ( 1)  school - tenure = 0

       F(  1,   483) =   16.28
            Prob > F =    0.0001
```

a) (10 pts) Fill in the missing quantities [1] – [5].

b) (25 points) Answer the following questions about the analysis and the results, explaining how the printout supports your conclusions.

1. Summarize the key results. What percentage of the women have never been married? How many live in rural areas? What types of women have the highest SES scores, and which types of women have the lowest?

2. The researcher was worried that missing data, heteroskedasticity, and/or multicollinearity might be problematic. Based on the results, are they?

3. The researcher had hypothesized that years in current job (tenure) would have a significantly larger effect on ses than would years in school (school). Do the results support her hypothesis?

4. The researcher debated whether or not to include the variable rural in her model. If she had not included it, how would the R^2 have been affected?

5. The researcher's daughter has just graduated from high school. She wants to spend the next four years living on a farm taking a richly deserved vacation from school and work. According to the researcher's model, if her daughter instead spends those years going to college at UCLA in Los Angeles, what will be the expected impact on her socio-economic status?