Sociology 63993
Exam 1
February 18, 2011

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

1. A data set contains a few extreme outliers. It is usually best to use Stata's `rreg` (Robust Regression) routine to deal with the problem.

2. The independent variables in an analysis include X1, X2, and X1X2 (i.e. X1 * X2). X1 has missing data (and hence X1X2 does too). If multiple imputation is being used, you should first compute X1X2, and then impute the missing values for X1 and X1X2.

3. Cronbach's Alpha is used to test for serial correlation.

4. The less true variability there is in a population, the higher the reliability of measures will tend to be.

5. The most extreme outliers on Y (i.e. the cases where Y is furthest from the mean) will always have the most influence on the regression line.

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.

```
. sum income white male age fathered

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      income |       812    16.96983    8.464258         .5         25
       white |       812     .864532    .3424337          0          1
        male |       812    .4864532    .5001245          0          1
         age |       812    38.53695    11.92651         18         81
    fathered |       695    11.44173    3.838113          0         20
```

. fre fathered

fathered -- HIGHEST YEAR SCHOOL COMPLETED, FATHER

		Freq.	Percent	Valid	Cum.
Valid	0	5	0.62	0.72	0.72
	2	4	0.49	0.58	1.29
	3	10	1.23	1.44	2.73
	4	12	1.48	1.73	4.46
	5	10	1.23	1.44	5.90
	6	38	4.68	5.47	11.37
	7	17	2.09	2.45	13.81
	8	84	10.34	12.09	25.90
	9	28	3.45	4.03	29.93
	10	30	3.69	4.32	34.24
	11	21	2.59	3.02	37.27
	12	224	27.59	32.23	69.50
	13	20	2.46	2.88	72.37
	14	64	7.88	9.21	81.58
	15	9	1.11	1.29	82.88
	16	71	8.74	10.22	93.09
	17	7	0.86	1.01	94.10
	18	15	1.85	2.16	96.26
	19	10	1.23	1.44	97.70
	20	16	1.97	2.30	100.00
	Total	695	85.59	100.00	
Missing	.a R is from Fatherless Family	117	14.41		
Total		812	100.00		

. gen one = 1
. gen mdfathered = missing(fathered)
. impute fathered one, gen(fathered2)
14.41% (117) observations imputed
. fre fathered2 mdfathered

fathered2 -- imputed fathered

		Freq.	Percent	Valid	Cum.
Valid	0	5	0.62	0.62	0.62
	2	4	0.49	0.49	1.11
	3	10	1.23	1.23	2.34
	4	12	1.48	1.48	3.82
	5	10	1.23	1.23	5.05
	6	38	4.68	4.68	9.73
	7	17	2.09	2.09	11.82
	8	84	10.34	10.34	22.17
	9	28	3.45	3.45	25.62
	10	30	3.69	3.69	29.31
	11	21	2.59	2.59	31.90
	11.44173	117	14.41	14.41	46.31
	12	224	27.59	27.59	73.89
	13	20	2.46	2.46	76.35
	14	64	7.88	7.88	84.24
	15	9	1.11	1.11	85.34
	16	71	8.74	8.74	94.09
	17	7	0.86	0.86	94.95
	18	15	1.85	1.85	96.80
	19	10	1.23	1.23	98.03
	20	16	1.97	1.97	100.00
	Total	812	100.00	100.00	

```
mdfathered
-----------------------------------------------------------
              |      Freq.    Percent      Valid       Cum.
--------------+--------------------------------------------
Valid   0     |        695      85.59      85.59      85.59
        1     |        117      14.41      14.41     100.00
        Total |        812     100.00     100.00
-----------------------------------------------------------
```

. reg income white male age fathered2 mdfathered

```
      Source |       SS       df       MS              Number of obs =     812
-------------+------------------------------           F(  5,   806) =   30.26
       Model |  9184.30275     5  1836.86055           Prob > F      =  0.0000
    Residual |   48918.708   806  60.6931861           R-squared     =  0.1581
-------------+------------------------------           Adj R-squared =  0.1528
       Total |  58103.0108   811  71.6436631           Root MSE      =  7.7906

------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       white |   .1521136   .8260281     0.18   0.854    -1.469306    1.773534
        male |   5.267875   .5502797     9.57   0.000     4.187725    6.348026
         age |   .1752915   .0240181     7.30   0.000     .1281461    .2224368
   fathered2 |   .2555826   .0811945     3.15   0.002     .0962049    .4149603
  mdfathered |  -1.122087    .797704    -1.41   0.160    -2.687909    .4437358
       _cons |   4.757922     1.6178     2.94   0.003     1.582324     7.93352
------------------------------------------------------------------------------
```

II-2.

. reg warm ed age prst

```
      Source |       SS       df       MS              Number of obs =    4586
-------------+------------------------------           F(  3,  4582) =  103.01
       Model |  249.541491     3  83.1804971           Prob > F      =  0.0000
    Residual |  3699.96047  4582  .807499012           R-squared     =  0.0632
-------------+------------------------------           Adj R-squared =  0.0626
       Total |  3949.50196  4585  .861396284           Root MSE      =  .89861

------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .0374512   .0054324     6.89   0.000     .0268012    .0481013
         age |  -.0094214   .0008435   -11.17   0.000    -.0110751   -.0077677
        prst |   .0018836   .0011332     1.66   0.097     -.000338    .0041052
       _cons |   2.498711   .0748558    33.38   0.000     2.351958    2.645465
------------------------------------------------------------------------------
```

. estat hettest

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

         chi2(1)      =     7.00
         Prob > chi2  =   0.0081
```

```
. reg  warm  ed age prst male

      Source |       SS       df       MS              Number of obs =    4586
-------------+------------------------------           F(  4,  4581) =  125.23
       Model |  389.311386     4  97.3278466           Prob > F      =  0.0000
    Residual |  3560.19058  4581    .7771645           R-squared     =  0.0986
-------------+------------------------------           Adj R-squared =  0.0978
       Total |  3949.50196  4585  .861396284           Root MSE      =  .88157

------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .0368867   .0053295     6.92   0.000     .0264383    .0473351
         age |  -.0099226   .0008284   -11.98   0.000    -.0115466   -.0082986
        prst |   .0025542   .0011128     2.30   0.022     .0003726    .0047359
        male |  -.3508326   .0261607   -13.41   0.000    -.4021202    -.299545
       _cons |   2.664683   .0744719    35.78   0.000     2.518682    2.810683
------------------------------------------------------------------------------

. estat hettest

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

         chi2(1)      =     0.03
         Prob > chi2  =   0.8613
```

II-3.

```
. reg price w1 w2 w3

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  3,    70) =   10.48
       Model |   196801072     3  65600357.4           Prob > F      =  0.0000
    Residual |   438264324    70  6260918.91           R-squared     =  0.3099
-------------+------------------------------           Adj R-squared =  0.2803
       Total |   635065396    73  8699525.97           Root MSE      =  2502.2

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          w1 |   1.998095   1.423422     1.40   0.165    -.8408306     4.83702
          w2 |   .9836392   .9768691     1.01   0.317    -.9646648    2.931943
          w3 |  -.9777821   .9785287    -1.00   0.321    -2.929396    .9738319
       _cons |   114.4055   1177.767     0.10   0.923    -2234.576    2463.387
------------------------------------------------------------------------------

. corr price w1 w2 w3
(obs=74)

             |    price       w1       w2       w3
-------------+------------------------------------
       price |   1.0000
          w1 |   0.5386   1.0000
          w2 |   0.5389   0.9347   1.0000
          w3 |   0.4644   0.9299   0.8695   1.0000
```

```
. sw, pe(.05): reg price w1 w2 w3
                      begin with empty model
p = 0.0000 <  0.0500  adding  w2

      Source |       SS       df       MS              Number of obs =      74
-------------+------------------------------           F(  1,    72) =   29.46
       Model |   184420235     1   184420235           Prob > F      =  0.0000
    Residual |   450645161    72  6258960.58           R-squared     =  0.2904
-------------+------------------------------           Adj R-squared =  0.2805
       Total |   635065396    73  8699525.97           Root MSE      =  2501.8

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          w2 |   1.884568   .3471831     5.43   0.000     1.192471    2.576664
       _cons |   474.8814   1087.899     0.44   0.664    -1693.806    2643.569
------------------------------------------------------------------------------
```

III. Computation and interpretation. (35 points total) The Indiana State legislature is considering a measure that would make gay marriage unconstitutional. The Indianapolis Chamber of Commerce opposes the measure because it worries that the resolution will cast the state as intolerant and put off talented workers who might otherwise relocate to Indianapolis. The Chamber has therefore commissioned a study of 10,000 Hoosiers to see where residents of the state stand on the issue. The variables are

Variable	Description
gaymarriage	Support for gay marriage. Ranges from a low of -200 (strongly oppose gay marriage) to a high of 200 (strongly favor)
educ	Years of education
age	Age of the respondent, in years
evangel	Coded 1 if the respondent is an evangelical Christian, 0 otherwise
black	Coded 1 if the respondent is black, 0 otherwise

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.]

```
. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         age |     10337     47.5637    17.21678         20         74
       black |     10337    .1050595    .3066449          0          1
     evangel |     10337    .2907033    .4541088          0          1
        educ |     10337    14.26352    5.043619          5         20
 gaymarriage |     10337    23.12387    50.68773  -188.7194   186.1061
```

```
. reg gaymarriage evangel black educ age, beta

      Source |       SS       df       MS              Number of obs =   10337
-------------+------------------------------           F(  4, 10332) = 3349.61
       Model |  14993619.8     4  3748404.95           Prob > F      =  0.0000
    Residual |  11562101.6 10332  1119.05746           R-squared     =    [1]
-------------+------------------------------           Adj R-squared =
       Total |  26555721.4 10336      [2]              Root MSE      =  33.452

------------------------------------------------------------------------------
 gaymarriage |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
     evangel |  -42.53951   .7288237     [3]    0.000                -.3811094
       black |  -34.44778   1.078767   -31.93   0.000                -.2083983
        educ |   6.174029   .0652522    94.62   0.000                 .6143391
         age |  -.2635312   .0191403   -13.77   0.000                 -.089512
       _cons |     [4]       1.38087   -26.37   0.000                        .
------------------------------------------------------------------------------

. estat hettest

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

         chi2(1)      =    49.70
         Prob > chi2  =   0.0000

. pcorr gaymarriage evangel black educ age
(obs=10337)

Partial and semipartial correlations of gaymarriage with

               Partial   Semipartial      Partial   Semipartial   Significance
   Variable |    Corr.         Corr.      Corr.^2       Corr.^2          Value
------------+------------------------------------------------------------------
    evangel |  -0.4980       -0.3789       0.2480        0.1436         0.0000
      black |  -0.2997       -0.2073       0.0898        0.0430         0.0000
       educ |   0.6813        0.6142       0.4642        0.3773         0.0000
        age |  -0.1342       -0.0894       0.0180        0.0080         0.0000

. predict rstandard, rstandard

. sum rstandard

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   rstandard |     10337   -8.04e-07    1.000047  -3.671386   3.441897

. test evangel black educ age

 ( 1)  evangel = 0
 ( 2)  black = 0
 ( 3)  educ = 0
 ( 4)  age = 0

       F(  4, 10332) =    [5]
            Prob > F =    0.0000

. test evangel = black

 ( 1)  evangel - black = 0

       F(  1, 10332) =   42.49
            Prob > F =    0.0000
```

```
. reg gaymarriage evangel black educ age, beta robust

Linear regression                                      Number of obs =   10337
                                                       F(  4, 10332) = 3387.31
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.5646
                                                       Root MSE      =  33.452

------------------------------------------------------------------------------
             |               Robust
 gaymarriage |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
     evangel |  -42.53951    .723011   -58.84   0.000                -.3811094
       black |  -34.44778   1.087479   -31.68   0.000                -.2083983
        educ |   6.174029   .0642269    96.13   0.000                 .6143391
         age |  -.2635312   .0191713   -13.75   0.000                 -.089512
       _cons |  -36.41955   1.385137   -26.29   0.000                        .
------------------------------------------------------------------------------
```

a) (10 pts) Fill in the missing quantities [1] – [5]. (A few other values have also been blanked out, but you don't need to fill them in.)

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

1. Summarize the key findings. What groups or types of individuals are most supportive of gay marriage and which are least supportive?

2. There was a problem with the study that almost caused the variable age not to be measured. How would R^2 have declined if age was not included in the model?

3. Why did the researchers run the regression a second time? What, if anything, was different about the two regressions? Do the differences have any major effects on the conclusions?

4. Before she began the study, the researcher expected education to be the least important determinant of support for gay marriage. Indicate whether you think the results support or do not support her belief.

5. The statistician preparing the report is very annoyed with her assistant who did the computer runs. She specifically told him that she wanted an incremental F test of the hypothesis that neither evangel nor black affected support for gay marriage, NOT just separate t tests of each coefficient; but she says the output does not contain the information she needs. Explain why you either agree or disagree with her; if you disagree, give her the information she wants.