Sociology 63993
Exam 1
February 17, 2012

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

1. Cohen and Cohen's dummy variable adjustment method is useful when variables like gender or age have missing values.

2. R^2 is biased downwards.

3. The more "tolerant" a variable is (i.e. the less highly correlated it is with the other IVs), the smaller its unique contribution to R^2 will be.

4. When you have more than one independent variable, random measurement error can cause coefficients to be biased either upward or downward.

5. A Durbin-Watson statistic of 4 or greater indicates that the case is an extreme outlier.

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.

```
. use "http://www.nd.edu/~rwilliam/stats3/statafiles/rwm11.dta", clear
(German Health Care Panel Data, Riphahn Wambach Million (2003), Greene (2007))

. reg newhsat female age handdum educ married working if year==1984

      Source |       SS       df       MS              Number of obs =    3874
-------------+------------------------------           F(  6,  3867) =  141.19
       Model |   4483.9589     6  747.326483           Prob > F      =  0.0000
    Residual |  20468.7796  3867  5.29319359           R-squared     =  0.1797
-------------+------------------------------           Adj R-squared =  0.1784
       Total |  24952.7385  3873  6.44274168           Root MSE      =  2.3007

------------------------------------------------------------------------------
     newhsat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -.186618   .0870046    -2.14   0.032    -.3571973   -.0160387
         age |  -.0388289   .0035637   -10.90   0.000    -.0458159    -.031842
     handdum |  -2.341489   .1236374   -18.94   0.000     -2.58389   -2.099088
        educ |   .1089876   .0173546     6.28   0.000     .0749626    .1430125
     married |   .2048268   .0918166     2.23   0.026     .0248133    .3848402
     working |   .2955985   .0912629     3.24   0.001     .1166704    .4745265
       _cons |   7.409179   .2928951    25.30   0.000     6.834935    7.983422
------------------------------------------------------------------------------
```

```
. estat hettest

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

         chi2(1)      =    55.33
         Prob > chi2  =   0.0000
```

```
. reg newhsat female age handdum educ married working if year==1984, robust

Linear regression                                      Number of obs =    3874
                                                       F(  6,  3867) =  129.51
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1797
                                                       Root MSE      =  2.3007

------------------------------------------------------------------------------
             |               Robust
     newhsat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -.186618   .0903374    -2.07   0.039    -.3637316   -.0095044
         age |  -.0388289   .0036116   -10.75   0.000    -.0459097   -.0317482
     handdum |  -2.341489   .1363509   -17.17   0.000    -2.608816   -2.074163
        educ |   .1089876   .0161526     6.75   0.000     .0773191    .1406561
     married |   .2048268   .0946607     2.16   0.031      .019237    .3904165
     working |   .2955985   .0974066     3.03   0.002     .1046252    .4865717
       _cons |   7.409179   .2968794    24.96   0.000     6.827124    7.991234
------------------------------------------------------------------------------
```

II-2.

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

      Source |       SS       df       MS              Number of obs =    1290
-------------+------------------------------           F(  6,  1283) =   27.43
       Model |  124.537637     6  20.7562729           Prob > F      =  0.0000
    Residual |  970.982518  1283  .756806327           R-squared     =  0.1137
-------------+------------------------------           Adj R-squared =  0.1095
       Total |  1095.52016  1289  .849899267           Root MSE      =  .86995

------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yr89 |   .2238435   .0502656     4.45   0.000     .1252318    .3224553
        male |  -.2846409    .048767    -5.84   0.000    -.3803127   -.1889691
       white |  -.2322106    .074535    -3.12   0.002    -.3784345   -.0859867
         age |  -.0086944    .001544    -5.63   0.000    -.0117234   -.0056654
          ed |   .0399421   .0098042     4.07   0.000     .0207081    .0591761
        prst |   .0019401   .0020726     0.94   0.349     -.002126    .0060063
       _cons |   2.688483   .1437892    18.70   0.000     2.406395    2.970571
------------------------------------------------------------------------------
```

```
. sum warm yr89 male white age ed prst

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        warm |      2293    2.607501    .9282156          1          4
        yr89 |      2293    .3986044    .4897178          0          1
        male |      2293    .4648932    .4988748          0          1
       white |      1712    .8785047    .3267975          0          1
         age |      2293    44.93546    16.77903         18         89
-------------+--------------------------------------------------------
          ed |      1709     12.1849    3.179042          0         20
        prst |      2293    39.58526    14.49226         12         82
```

```
. mi set mlong
. mi register imputed white ed
(1003 m=0 obs. now marked as incomplete)
```

```
. mi register regular warm yr89 male age prst
. mi impute chained (logit) white (regress) ed = warm yr89 male age prst, add(50) rseed(1234)

Conditional models:
             white: logit white ed warm yr89 male age prst
                ed: regress ed i.white warm yr89 male age prst

Performing chained iterations ...

Multivariate imputation                     Imputations =       50
Chained equations                                added =       50
Imputed: m=1 through m=50                        updated =        0

Initialization: monotone                     Iterations =      500
                                                burn-in =       10

             white: logistic regression
                ed: linear regression

------------------------------------------------------------------
                  |               Observations per m
                  |----------------------------------------------
         Variable |   Complete   Incomplete   Imputed |     Total
------------------+-----------------------------------+----------
             white |       1712          581       581 |      2293
                ed |       1709          584       584 |      2293
------------------------------------------------------------------
(complete + incomplete = total; imputed is the minimum across m
 of the number of filled-in observations.)

. mi estimate, dots: reg warm yr89 male white age ed prst

Imputations (50):
  .........10.........20.........30.........40.........50 done

Multiple-imputation estimates                   Imputations     =         50
Linear regression                               Number of obs   =       2293
                                                Average RVI     =     0.1214
                                                Largest FMI     =     0.3154
                                                Complete DF     =       2286
DF adjustment:   Small sample                   DF:     min     =     380.43
                                                        avg     =    1356.48
                                                        max     =    2270.39
Model F test:       Equal FMI                   F(   6, 2035.1) =      46.46
Within VCE type:          OLS                   Prob > F        =     0.0000

------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yr89 |   .2607513   .0380632     6.85   0.000     .1861084    .3353942
        male |  -.3335333   .0366975    -9.09   0.000    -.4054975   -.2615691
       white |  -.1599008    .068205    -2.34   0.020    -.2940068   -.0257948
         age |  -.0098668   .0011996    -8.22   0.000    -.0122195   -.0075141
          ed |   .0340434   .0087675     3.88   0.000     .0168149    .0512719
        prst |   .0023128   .0016503     1.40   0.161    -.0009248    .0055503
       _cons |   2.735879   .1198425    22.83   0.000     2.500675    2.971083
------------------------------------------------------------------------------
```

II-3.

```
. reg docvis educ ses female

      Source |       SS       df       MS              Number of obs =   27326
-------------+------------------------------           F(  3, 27322) =  138.65
       Model |  13264.8729     3   4421.6243           Prob > F      =  0.0000
    Residual |   871315.75 27322  31.8906284           R-squared     =  0.0150
-------------+------------------------------           Adj R-squared =  0.0149
       Total |  884580.623 27325  32.3725754           Root MSE      =  5.6472

------------------------------------------------------------------------------
      docvis |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   -.366295   .2382335    -1.54   0.124    -.8332447    .1006548
         ses |   .1992199   .2378011     0.84   0.402    -.2668824    .6653222
      female |   1.023064   .0695639    14.71   0.000     .8867151    1.159413
       _cons |   4.535222   .1913001    23.71   0.000     4.160265     4.91018
------------------------------------------------------------------------------

. corr docvis educ ses female
(obs=27326)

             |   docvis     educ      ses   female
-------------+------------------------------------
      docvis |   1.0000
        educ |  -0.0847   1.0000
         ses |  -0.0843   0.9981   1.0000
      female |   0.1023  -0.1831  -0.1832   1.0000


. reg docvis educ female

      Source |       SS       df       MS              Number of obs =   27326
-------------+------------------------------           F(  2, 27323) =  207.63
       Model |  13242.4908     2  6621.24541           Prob > F      =  0.0000
    Residual |  871338.132 27323  31.8902804           R-squared     =  0.0150
-------------+------------------------------           Adj R-squared =  0.0149
       Total |  884580.623 27325  32.3725754           Root MSE      =  5.6471

------------------------------------------------------------------------------
      docvis |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.1671061   .0149471   -11.18   0.000    -.1964031   -.1378091
      female |   1.022659   .0695618    14.70   0.000      .886314    1.159003
       _cons |   4.585648   .1815832    25.25   0.000     4.229735     4.94156
------------------------------------------------------------------------------
```

III. Computation and interpretation. (35 points total) President Obama's plan to provide free birth control to most women has proven to be far more controversial than he expected. The President has therefore commissioned a study of 7,500 Americans to see where the public stands. The variables are

Variable	Description
bcontrol	Support for Obama's birth control policy. Ranges from a low of 0 (strongly oppose the policy) to a high of 100 (strongly favor)
catholic	Coded 1 if the respondent is Catholic, 0 otherwise
female	Coded 1 if the respondent is female, 0 otherwise

health	Overall health of the respondent. Ranges from 0 (very poor health) to 100 (very good health).
liberal	How liberal is the respondent? Ranges from 0 (very conservative) to 100 (very liberal).

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
-------------+--------------------------------------------------------
    bcontrol |      7500    43.13119    9.173105     18.504     95.118
    catholic |      7500    .5262667    .4993429          0          1
      female |      7500    .1141333    .3179943          0          1
      health |      7500    57.41967    9.648723         25         87
     liberal |      7500    62.04155    22.21342         26       96.2

. reg bcontrol catholic female health liberal

      Source |       SS       df       MS              Number of obs =    7500
-------------+------------------------------           F(  4,  7495) =    [1]
       Model |  153728.687     4  38432.1718           Prob > F      =  0.0000
    Residual |  477281.098  7495  63.6799331           R-squared     =    [2]
-------------+------------------------------           Adj R-squared =
       Total |  631009.786  7499  84.1458575           Root MSE      =    7.98

------------------------------------------------------------------------------
    bcontrol |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    catholic |  -.8740108   .2610363    -3.35   0.001    -1.385715   -.3623065
      female |   2.189993    .289954     [3]    0.000     1.621602    2.758385
      health |  -.4399353   .0138063   -31.86   0.000    -.4669995   -.4128712
     liberal |   .0540596   .0043219    12.51   0.000     .0455874    .0625318
       _cons |    65.2482   .6861085    95.10   0.000     63.90324    66.59317
------------------------------------------------------------------------------

. collin catholic female health liberal
(obs=7500)

  Collinearity Diagnostics

                       SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
  catholic      2.00    1.41     [4]        0.5002
    female      1.00    1.00    0.9989      0.0011
    health      2.09    1.45    0.4785      0.5215
   liberal      1.09    1.04    0.9213      0.0787
----------------------------------------------------
  Mean VIF      1.54

                        Cond
       Eigenval         Index
---------------------------------
    1     3.6921          1.0000
    2     0.8578          2.0746
    3     0.3751          3.1373
    4     0.0671          7.4161
    5     0.0078         21.6930
---------------------------------
 Condition Number        21.6930
 Eigenvalues & Cond Index computed from scaled raw sscp (w/ intercept)
 Det(correlation matrix)    0.4780
```

```
. test liberal

 ( 1)  liberal = 0

       F(  1,  7495) =    [5]
            Prob > F =    0.0000

. test female = -catholic

 ( 1)  catholic + female = 0

       F(  1,  7495) =   11.55
            Prob > F =    0.0007

. test catholic female health liberal

 ( 1)  catholic = 0
 ( 2)  female = 0
 ( 3)  health = 0
 ( 4)  liberal = 0

       F(  4,  7495) =  603.52
            Prob > F =    0.0000

. alpha catholic female health liberal

Test scale = mean(unstandardized items)

Average interitem covariance:     7.936035
Number of items in the scale:            4
Scale reliability coefficient:      0.1862

. alpha catholic female health liberal, s

Test scale = mean(standardized items)

Average interitem correlation:      0.1506
Number of items in the scale:            4
Scale reliability coefficient:      0.4149

. predict rstandard, rstandard
. extremes rstandard rstandard bcontrol catholic female health liberal

  +-------------------------------------------------------------------------------------+
  |  obs:   rstandard   rstandard   bcontrol   catholic   female     health   liberal |
  |-------------------------------------------------------------------------------------|
  | 4166.   -2.774446   -2.774446      28.44          0        1     49.703      92.3 |
  | 5909.   -2.773224   -2.773224     27.966          0        0         45      85.8 |
  |  710.   -2.753186   -2.753186     18.504          0        0     67.203      88.4 |
  | 4022.   -2.747619   -2.747619      29.94          0        1   46.30099      88.4 |
  | 2839.   -2.601393   -2.601393       24.9          0        0   55.40199      88.4 |
  +-------------------------------------------------------------------------------------+

  +-------------------------------------------------------------------------------+
  | 4213.   5.357062   5.357062     78.924          1        0     68.703      37.7 |
  |  742.   5.520198   5.520198     83.346          1        1     66.703        39 |
  | 2097.   5.758528   5.758528     89.742          1        0   57.30099      85.8 |
  | 1592.   5.899137   5.899137     95.118          0        1   48.80099        39 |
  | 6511.   5.932644   5.932644     88.926          1        1   63.90199      58.5 |
  +-------------------------------------------------------------------------------+

. pcorr bcontrol catholic female health liberal
(obs=7500)

Partial and semipartial correlations of bcontrol with

             Partial   Semipartial      Partial   Semipartial   Significance
   Variable |    Corr.         Corr.      Corr.^2       Corr.^2          Value
------------+------------------------------------------------------------------
   catholic |  -0.0386       -0.0336       0.0015        0.0011         0.0008
     female |   0.0869        0.0759       0.0076        0.0058         0.0000
     health |  -0.3454       -0.3201       0.1193        0.1025         0.0000
    liberal |   0.1430        0.1257       0.0204        0.0158         0.0000
```

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 the President's policy and which are least supportive?

2. The researchers were worried that outliers might be problematic. Based on the results, do you see any reasons to be concerned?

3. The researchers were concerned that the items may suffer from random measurement error. Would you encourage them to create a scale out of the items in order to deal with the problem?

4. How would the R^2 value change if the variable liberal were dropped from the model? Do you think that would be a good idea?

5. The President's advisors believe that Female support for health care reform is stronger than Catholic opposition to it. Do you think they are right?

c) (1 point extra credit) As soon as the President started reading the results, he became concerned that something might be seriously wrong with the data. Why?