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

1. A researcher runs the following commands:

   . reg health female black rural

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 10335</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>442.139589</td>
<td>3</td>
<td>147.379863</td>
<td>F( 3, 10331) = 104.34</td>
</tr>
<tr>
<td>Residual</td>
<td>14592.8818</td>
<td>10331</td>
<td>1.41253333</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>15035.0214</td>
<td>10334</td>
<td>1.4549082</td>
<td>Adj R-squared = 0.0291</td>
</tr>
</tbody>
</table>

   health | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
   -------|-------|-----------|---|-----|---------------------|
   female | -0.082975 | .0234247 | -3.54 | 0.000 | -.128892 -.037058 |
   black  | -0.5846982 | .0387848 | -15.08 | 0.000 | -.660724 -.5086724 |
   rural  | -0.2782157 | .0246852 | -11.27 | 0.000 | -.3266035 -.2298279 |
   _cons  | 3.621027   | .0201806 | 179.43 | 0.000 | 3.581469 3.660585 |

   . pcorr2 health female black rural

   (obs=10335)

   Partial and Semipartial correlations of health with

   | Variable | Partial | SemiP | Partial^2 | SemiP^2 | Sig. |
   --------|---------|-------|-----------|---------|------|
   female  | -0.0348 | -0.0343 | 0.0012 | 0.0012 | 0.000 |
   black   | -0.1467 | -0.1461 | 0.0215 | 0.0214 | 0.000 |
   rural   | -0.1102 | -0.1092 | 0.0121 | 0.0119 | 0.000 |

   If she now does a backwards stepwise regression, the variable female will be dropped from the model.

2. Serial correlation causes OLS estimates to be biased.

3. The null and alternative hypotheses are

   H₀: \( \beta_{female} = 0 \)
   Hₐ: \( \beta_{female} > 0 \)

   In her analysis, the researcher finds that
. reg health female

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 10335</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>15.4391056</td>
<td>1</td>
<td>15.4391056</td>
<td>F( 1, 10333) = 10.62</td>
</tr>
<tr>
<td>Residual</td>
<td>15019.5823</td>
<td>10333</td>
<td>1.45355485</td>
<td>Prob &gt; F = 0.0011</td>
</tr>
<tr>
<td>Total</td>
<td>15035.0214</td>
<td>10334</td>
<td>1.4549082</td>
<td>Adj R-squared = 0.0009</td>
</tr>
</tbody>
</table>

| health | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|--------|-------|-----------|-------|------|------------------|
| female | -0.077398 | 0.0237484 | -3.26 | 0.001 | -.1239495 -.0308465 |
| _cons  | 3.454471  | 0.0172076 | 200.75 | 0.000 | 3.420741 3.488202  |

If the researcher is using the .01 level of significance, she should NOT reject the null.

4. If you regress Y on X, and $R^2 = 0$, this means that there is no relationship between Y and X.

5. The reason OLS is not optimal when multicollinearity is present is that it gives equal weight to all observations when, in fact, observations with larger disturbance variance contain less information than observations with smaller disturbance variance.

II. Short answer. Discuss all three 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.

. regress warm yr89 male white age ed prst

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 1146</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>124.547769</td>
<td>6</td>
<td>20.7579616</td>
<td>F( 6, 1139) = 28.12</td>
</tr>
<tr>
<td>Residual</td>
<td>840.748042</td>
<td>1139</td>
<td>.738145779</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>965.295812</td>
<td>1145</td>
<td>.84305311</td>
<td>Adj R-squared = 0.1244</td>
</tr>
</tbody>
</table>

| warm   | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|--------|-------|-----------|-------|------|------------------|
| yr89   | .1960047 | 0.0533578 | 3.67  | 0.000 | .0913141 .3006953 |
| male   | -.352923  | 0.0510827 | -6.91 | 0.000 | -.4531499 -.2526962 |
| white  | -.2104459 | 0.0806038 | -2.61 | 0.009 | -.3685945 -.0522973 |
| age    | -.0117717 | .0016364 | -7.19 | 0.000 | -.0149823 -.0085611 |
| ed     | .0244172  | 0.0104879 | 2.33  | 0.020 | .0038394 .044995 |
| prst   | .0022522  | .0022071 | 1.02  | 0.308 | -.0020781 .0065826 |
| _cons  | 3.022827  | 0.154743  | 19.53 | 0.000 | 2.719214 3.32644 |

Sociology 63993—Exam 1—Page 2
. sum warm yr89 male white age ed prst, sep(7)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>warm</td>
<td>2293</td>
<td>2.607501</td>
<td>.9282156</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>yr89</td>
<td>2293</td>
<td>.3986044</td>
<td>.4897178</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>male</td>
<td>2293</td>
<td>.4648932</td>
<td>.4988748</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>white</td>
<td>2293</td>
<td>.8765809</td>
<td>.3289894</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>age</td>
<td>1146</td>
<td>44.34555</td>
<td>16.69399</td>
<td>19</td>
<td>89</td>
</tr>
<tr>
<td>ed</td>
<td>2293</td>
<td>12.21805</td>
<td>3.160827</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>prst</td>
<td>2293</td>
<td>39.58526</td>
<td>14.49226</td>
<td>12</td>
<td>82</td>
</tr>
</tbody>
</table>

. mi set mlong

. mi register imputed age
(1147 m=0 obs. now marked as incomplete)

. mi register regular warm yr89 male white ed prst

. mi impute regress age warm yr89 male white ed prst, add(100) rseed(123)

Univariate imputation               Imputations =      100
Linear regression                   added =      100
Imputed: m=1 through m=100           updated =      0

<table>
<thead>
<tr>
<th>Variable</th>
<th>complete</th>
<th>incomplete</th>
<th>imputed</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>1146</td>
<td>1147</td>
<td>1147</td>
<td>2293</td>
</tr>
</tbody>
</table>

(complete + incomplete = total; imputed is the minimum across m of the number of filled in observations.)

. mi estimate: regress warm yr89 male white age ed prst

Multiple-imputation estimates                 Imputations =      100
Linear regression                           Number of obs =       2293
Average RVI = 0.1909                        Complete DF =       2286
DF adjustment: Small sample                  DF:             min     =     276.39
                                                  avg     =  1477.89
                                                  max     =  2126.35
Model F test: Equal FMI                      F(   6, 2031.6) =      47.78
Within VCE type: OLS                         Prob > F =     0.0000

| Variable | Coef.     | Std. Err.  | t      | P>|t| | [95% Conf. Interval] |
|----------|-----------|------------|--------|-----|---------------------|
| yr89     | .2510464  | .0384582   | 6.53   | 0.000| .1756255 .3264674   |
| male     | -.3395849 | .0371148   | -9.15  | 0.000| -.41237 -.2667999  |
| white    | -.1807239 | .0568923   | -3.18  | 0.002| -.2922958 -.0691519 |
| age      | -.0119027 | .016758    | -0.70  | 0.000| -.0152017 -.0086037 |
| ed       | .0266213  | .008149    | 3.27   | 0.001| .0106352 .0426075   |
| prst     | .003211   | .0016106   | 1.99   | 0.046| .0000522 .0063998   |
| _cons    | 2.897334  | .1345408   | 21.53  | 0.000| 2.633084 3.161584   |
II-2.

. reg psyscale female black rural

Source | SS     df      MS               Number of obs = 10335
---------+------------------------------------------------------------------
Model    | 325591.507  3 108530.502           F(  3, 10331) = 19.43
Residual | 57707488.4 10331 5585.85698           Prob > F = 0.0000
---------+------------------------------------------------------------------
Total    | 58033079.9 10334 5615.7422           R-squared = 0.0056
         |                        Adj R-squared = 0.0053
---------+------------------------------------------------------------------

psyscale | Coef.  Std. Err.   t    P>|t|     [95% Conf. Interval]
---------+------------------------------------------------------------------
    female | - 3.223899   1.473059  -2.19 0.029    -6.11138    -3.364178
     black |  -13.703    2.438975  -5.62 0.000    -18.48386    -9.922137
    rural  |  -8.843438   1.552324  -5.70 0.000    -11.88629   -5.800583
     _cons |     61.179    1.269053  48.21 0.000     58.69141    63.66658
---------+------------------------------------------------------------------

. predict rstandard, rstandard
(2 missing values generated)

. extremes rstandard psyscale female black rural

+----------------------------------------+
<table>
<thead>
<tr>
<th>obs:      rstandard  psyscale  female  black  rural</th>
</tr>
</thead>
<tbody>
<tr>
<td>4840    -0.9971863  -13.33856   0     0     0</td>
</tr>
<tr>
<td>4932    -0.94367    -12.56431   1     0     0</td>
</tr>
<tr>
<td>939.    -0.9420472  -9.218133   0     0     0</td>
</tr>
<tr>
<td>5794    -0.9130703  -10.27764   1     0     0</td>
</tr>
<tr>
<td>1724    -0.9067409  -6.579773   0     0     0</td>
</tr>
</tbody>
</table>
+----------------------------------------+

. replace psyscale = psyscale/100 in 12
(1 real change made)

. reg psyscale female black rural

Source | SS    df       MS               Number of obs = 10335
---------+------------------------------------------------------------------
Model    | 489028.274  3 163009.425           F(  3, 10331) = 84.63
Residual | 19899499.5 10331  1926.19296           Prob > F = 0.0000
---------+------------------------------------------------------------------
Total    | 20388527.7 10334  1972.95604           R-squared = 0.0240
         |                        Adj R-squared = 0.0237
---------+------------------------------------------------------------------

psyscale | Coef.  Std. Err.   t    P>|t|     [95% Conf. Interval]
---------+------------------------------------------------------------------
    female |  -4.310959    .865018  -4.98 0.000    -6.006562    -2.615357
     black |  -19.25608   1.432229 -13.44 0.000    -22.06352   -16.44863
    rural  |  -8.594151   .9115645  -9.43 0.000    -10.38099   -6.807308
     _cons |    61.64872    .7452205   82.73 0.000     60.18794    63.1095
---------+------------------------------------------------------------------
II-3.

```
. reg y x1 x2 x3 x4 x5 x6 x7 x8 x9

Source |       SS   df       MS              Number of obs =    3975
---------+------------------------------ F(  9,  3965) =  1.11
Model |  56489.8865     9  6276.65406           Prob > F      =  0.3481
Residual |  22323032.1  3965  5630.02071             R-squared =  0.0003
---------+------------------------------ Adj R-squared =  0.0025
Total |  22379522  3974  5631.48515           Root MSE      =  75.033

F(  9,  3965) =    1.11
Model |  56489.8865     9  6276.65406           Prob > F      =  0.3481
Residual |  22323032.1  3965  5630.02071             R-squared =  0.0003
---------+------------------------------ Adj R-squared =  0.0025
Total |  22379522  3974  5631.48515           Root MSE      =  75.033

y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------
  x1 |   1.441756   2.724341     0.53   0.597   -3.899484    6.782996
  x2 |    2.83684   2.691752     1.05   0.292   -2.440507    8.114187
  x3 |    2.851331   3.009868     0.95   0.344   -3.049704    8.752366
  x4 |   0.5989067   2.5459     0.24   0.814   -4.392488    5.590302
  x5 |   1.117681   2.795055     0.40   0.689   -4.362199    6.59756
  x6 |  -0.5185488   2.903095    -0.18   0.858   -6.210253    5.173156
  x7 |   1.200175   2.720099     0.44   0.659   -4.132749    6.533099
  x8 |   4.261204   3.082074    1.38   0.167   -1.781395   10.3038
  x9 |  -0.6321259   2.850720    -0.22   0.825   -6.22114    4.956888
_cons |  -3.166028   2.651549    -1.19   0.233   -8.364556    2.0325
```

```
Test scale = mean(unstandardized items)

Item  |  Obs  |  Sign | item-test  | item-rest  | interitem  | covariance  | alpha
      |      |      | correlation| correlation|           |            |      
---------+--------+--------+------------+------------+-----------+------------+------
x1     | 3975   |  +    | 0.4752     | 0.3094     |           | .0591653   | 0.7351
x2     | 3975   |  +    | 0.5814     | 0.4238     | .0543645  | 0.7167     |
x3     | 3975   |  +    | 0.5262     | 0.3809     | .0575461  | 0.7237     |
x4     | 3975   |  +    | 0.5010     | 0.3259     | .0577938  | 0.7336     |
x5     | 3975   |  +    | 0.6197     | 0.4706     | .0527024  | 0.7084     |
x6     | 3975   |  +    | 0.6540     | 0.5156     | .0513415  | 0.7006     |
x7     | 3975   |  +    | 0.5304     | 0.3664     | .0566869  | 0.7263     |
x8     | 3975   |  +    | 0.6009     | 0.4629     | .0543247  | 0.7107     |
x9     | 3975   |  +    | 0.6440     | 0.4998     | .0515661  | 0.7031     
---------+--------+--------+------------+------------+-----------+------------+------
Test scale |       |      |            |            |           | .0550546   | 0.7412
```

```
. reg y xscale

Source |       SS   df       MS              Number of obs =    3975
---------+------------------------------ F(  1,  3973) =  7.98
Model |  44837.8037     1  44837.8037           Prob > F      =  0.0048
Residual |  22334684.2  3973  5621.61696             R-squared =  0.0018
---------+------------------------------ Adj R-squared =  0.0020
Total |  22379522  3974  5631.48515           Root MSE      =  74.977

F(  1,  3973) =  7.98
Model |  44837.8037     1  44837.8037           Prob > F      =  0.0048
Residual |  22334684.2  3973  5621.61696             R-squared =  0.0018
---------+------------------------------ Adj R-squared =  0.0020
Total |  22379522  3974  5631.48515           Root MSE      =  74.977

y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------
  xscale |  12.32454   4.363945     2.82   0.005    3.768761   20.88033
  _cons |  -2.539008   2.398131    -1.06   0.290   -7.24069    2.162674
```

Sociology 63993—Exam 1—Page 5
III. Computation and interpretation. (35 points total) Despite recent setbacks, the Obama administration is determined to pass health care reform this year. In order to do this, it thinks it first needs to assess where the public support and opposition to health care reform lies. It has therefore commissioned a survey of 10,337 Americans and measured the following:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>hcare</td>
<td>Support for health care reform. Ranges from a low of 0 to a high of 200</td>
</tr>
<tr>
<td>health</td>
<td>Overall health of the respondent. Ranges from 0 (very poor health) to 100 (very good health).</td>
</tr>
<tr>
<td>age</td>
<td>Age of the respondent, in years</td>
</tr>
<tr>
<td>gop</td>
<td>Coded 1 if the respondent is a Republican, 0 otherwise</td>
</tr>
<tr>
<td>black</td>
<td>Coded 1 if the respondent is black, 0 otherwise</td>
</tr>
</tbody>
</table>

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    |
-------------------------------+-----------+---------+-----------+-------+--------|
hcare             |   10337   | 71.90088| 15.35515  | 30.84 | 175.88 |
health            |   10337   | 57.34875| 9.660012  | 25    | 89.5   |
age               |   10337   | 47.5637 | 17.21678  | 20    | 74     |
gop               |   10337   | .525104 | .4993935  | 0     | 1      |
black             |   10337   | .1050595| .3066449  | 0     | 1      |
```

```
. collin health age gop black

Collinearity Diagnostics

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF</th>
<th>VIF</th>
<th>Tolerance</th>
<th>R-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>health</td>
<td>2.12</td>
<td>1.46</td>
<td>[1]</td>
<td>0.5283</td>
</tr>
<tr>
<td>age</td>
<td>1.09</td>
<td>1.04</td>
<td>0.9209</td>
<td>0.0791</td>
</tr>
<tr>
<td>gop</td>
<td>2.03</td>
<td>1.42</td>
<td>0.4926</td>
<td>0.5074</td>
</tr>
<tr>
<td>black</td>
<td>1.00</td>
<td>1.00</td>
<td>0.9988</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

Mean VIF 1.56
```

[Part of output deleted]
. reg hcare health age gop black, beta

Source |       SS     df       MS              Number of obs =  10337
-------------+------------------------------+-----------------------------
Model |   620082.606     4  155020.652           Prob > F      =  0.0000
Residual |  1816944.64 10332  175.856044 R-squared     =   [ 2 ]
        +-------------------------------+-----------------------------
Total |  2437027.25 10336               Adj R-squared     = ------
        +-------------------------------+-----------------------------

Root MSE      =  13.261

------------------------------------------------------------------------------
hcare |      Coef.   Std. Err.      t    P>|t|                  Beta
-------------+------------------------------------------------------
  health  |  -0.7485279    0.01966   -38.07   0.000      -0.4709032
  age     |    0.1237255    0.00000        0.000     .1387257
  gop     |  -1.540187    0.3721392   -4.14   0.000      -0.500913
  black   |   3.679295    0.4256284     8.64   0.000       0.0734762
  _cons   |    0.9741076   112.27000   0.0000  
------------------------------------------------------------------------------

. test age

( 1)  age = 0

  F(  1, 10332) =  245.60
  Prob > F =  0.0000

. test black = -gop

( 1)  gop + black = 0

  F(  1, 10332) =   14.51
  Prob > F =  0.0001

. estat hettest

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

  chi2(1)   =  1.65
  Prob > chi2 =  0.1986

. estat imtest, white

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

  chi2(12)   = 127.51
  Prob > chi2 =  0.0000

Cameron & Trivedi's decomposition of IM-test

+--------------------------------+----------------+-----------------+------------------+
|                                | chi2 | df  | p    |
|--------------------------------+-----+-----+------|
| Heteroskedasticity            | 127.51 | 12  | 0.0000|
| Skewness                      | 264.14 | 4   | 0.0000|
| Kurtosis                      | 41.56  | 1   | 0.0000|
| --------------------------------+-----+-----+------|
| Total                          | 431.22 | 17  | 0.0000|
+--------------------------------+----------------+-----------------+-------------------+
. pcorr2 hcare health age gop black

(obs=10337)

Partial and Semipartial correlations of hcare with

<table>
<thead>
<tr>
<th>Variable</th>
<th>Partial</th>
<th>SemiP</th>
<th>Partial^2</th>
<th>SemiP^2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>health</td>
<td>-0.3508</td>
<td>-0.3234</td>
<td>0.1230</td>
<td>0.1046</td>
<td>0.000</td>
</tr>
<tr>
<td>age</td>
<td>0.1524</td>
<td>0.1331</td>
<td>0.0232</td>
<td>0.0177</td>
<td>0.000</td>
</tr>
<tr>
<td>gop</td>
<td>-0.0407</td>
<td>-0.0352</td>
<td>0.0017</td>
<td>0.0012</td>
<td>0.000</td>
</tr>
<tr>
<td>black</td>
<td>0.0847</td>
<td>0.0734</td>
<td>0.0072</td>
<td>0.0054</td>
<td>0.000</td>
</tr>
</tbody>
</table>

. alpha health age gop black

Test scale = mean(unstandardized items)
Reversed item: black

Average interitem covariance: 6.326097
Number of items in the scale: 4
Scale reliability coefficient: 0.2172

. alpha health age gop black, s

Test scale = mean(standardized items)
Reversed item: black

Average interitem correlation: 0.1570
Number of items in the scale: 4
Scale reliability coefficient: 0.4269

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) (5 pts) Summarize the key findings. What groups are most supportive of health care reform and which groups are least supportive?

c) (20 pts) Before she began her analyses, the researcher had several expectations. Indicate whether you think the evidence tends to support or not support her ideas. Be sure to cite evidence from the printout to justify your conclusions.

1. Despite Obama’s recent setbacks, a majority of Americans still do not consider themselves to be Republicans.

2. Heteroskedasticity are not present in these data.

3. Black support for health care reform is stronger than GOP opposition to it.

4. black is the least important variable in determining support for health care reform.

5. It would be a bad idea to try to create a single scale from her independent variables.