Sociology 63993
Exam 1 Answer Key [DRAFT]
February 12, 2010

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

      Source |       SS       df       MS              Number of obs =   10335
-------------+------------------------------           F(  3, 10331) =  104.34
       Model |  442.139589     3  147.379863            Prob > F      =  0.0000
    Residual |  14592.8818 10331  1.41253333           R-squared     =  0.0294
-------------+------------------------------           Adj R-squared =  0.0291
       Total |  15035.0214 10334   1.4549082           Root MSE      =  1.1885

------------------------------------------------------------------------------
      health |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -.082975   .0234247    -3.54   0.000     -.128892    -.037058
       black |  -.5846982   .0387848   -15.08   0.000     -.660724   -.5086724
       rural |  -.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.

False (unless she is using an incredibly small P value). All variables are highly significant so none would get dropped in backwards stepwise regression. To confirm,

```
. sw, pr(.001): reg health female black rural
                  begin with full model
p < 0.0010        for all terms in model

      Source |       SS       df       MS              Number of obs =   10335
-------------+------------------------------           F(  3, 10331) =  104.34
       Model |  442.139589     3  147.379863            Prob > F      =  0.0000
    Residual |  14592.8818 10331  1.41253333           R-squared     =  0.0294
-------------+------------------------------           Adj R-squared =  0.0291
       Total |  15035.0214 10334   1.4549082           Root MSE      =  1.1885

------------------------------------------------------------------------------
      health |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -.082975   .0234247    -3.54   0.000     -.128892    -.037058
       black |  -.5846982   .0387848   -15.08   0.000     -.660724   -.5086724
       rural |  -.2782157   .0246852   -11.27   0.000    -.3266035   -.2298279
       _cons |   3.621027   .0201806   179.43   0.000     3.581469    3.660585
------------------------------------------------------------------------------
```

2. Serial correlation causes OLS estimates to be biased.

False. Estimates are unbiased and consistent but less efficient.

3. The null and alternative hypotheses are

H_0: $\beta_{female} = 0$
H_A: $\beta_{female} > 0$

In her analysis, the researcher finds that

```
. reg health female

      Source |       SS       df       MS              Number of obs =   10335
-------------+------------------------------           F(  1, 10333) =   10.62
       Model |  15.4391056     1  15.4391056           Prob > F      =  0.0011
    Residual |  15019.5823 10333  1.45355485           R-squared     =  0.0010
-------------+------------------------------           Adj R-squared =  0.0009
       Total |  15035.0214 10334   1.4549082           Root MSE      =  1.2056

------------------------------------------------------------------------------
      health |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   -.077398   .0237484    -3.26   0.001    -.1239495   -.0308465
       _cons |   3.454471   .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.

True. In fact, it doesn't matter what significance level you are using. The alternative hypothesis said that the effect of female would be positive when in fact the estimated effect is negative. You only reject the null if the alternative is better, and in this case the alternative is worse.

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

False. It just means there is no *linear* relationship. Something like a curvilinear relationship is still possible.

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.

False. Substitute "heteroskedasticity" for "multicollinearity."

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.

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

      Source |       SS       df       MS              Number of obs =    1146
-------------+------------------------------           F(  6,  1139) =   28.12
       Model |  124.547769     6  20.7579616           Prob > F      =  0.0000
    Residual |  840.748042  1139  .738145779           R-squared     =  0.1290
-------------+------------------------------           Adj R-squared =  0.1244
       Total |  965.295812  1145   .84305311           Root MSE      =  .85915

------------------------------------------------------------------------------
        warm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        yr89 |   .1960047   .0533578     3.67   0.000     .0913141    .3006953
        male |   -.352923   .0510827    -6.91   0.000    -.4531499   -.2526962
       white |  -.2104459   .0806038    -2.61   0.009    -.3685945   -.0522973
         age |  -.0117717   .0016364    -7.19   0.000    -.0149823   -.0085611
          ed |   .0244172   .0104879     2.33   0.020     .0038394     .044995
        prst |   .0022522   .0022071     1.02   0.308    -.0020781    .0065826
       _cons |   3.022827    .154743    19.53   0.000     2.719214     3.32644
------------------------------------------------------------------------------

. sum warm yr89 male white age ed prst, sep(7)

    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 |      2293    .8765809    .3289894          0          1
         age |      1146    44.34555    16.69399         19         89
          ed |      2293    12.21805    3.160827          0         20
        prst |      2293    39.58526    14.49226         12         82

. 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

               |              Observations per m
               |-----------------------------------------------
      Variable |   complete   incomplete   imputed |     total
---------------+-----------------------------------+----------
           age |       1146         1147      1147 |      2293
------------------------------------------------------------------
(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

------------------------------------------------------------------------------
        warm |      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   .0016758    -7.10   0.000    -.0152017   -.0086037
          ed |   .0266213    .008149     3.27   0.001     .0106352    .0426075
        prst |    .003211   .0016106     1.99   0.046     .0000522    .0063698
       _cons |   2.897334   .1345408    21.53   0.000     2.633084    3.161584
------------------------------------------------------------------------------
```

The researcher notes that one variable (and one variable only), `age`, is missing data for half the cases. She therefore decides to use multiple imputation to plug in estimates for the missing values. She could have used single imputation, but such an approach does not take into account the fact that the imputed values are estimated rather than observed. With multiple imputation you come up with (in this case) 100 different estimates for each missing value, which takes into account the fact that no single estimate is perfect. She could have also stuck with the original listwise deletion. This loses her half the data though, and notice how much more significant the T values are in the multiple imputation. She could have dropped age but that might have led to problems of specification error. Cohen and Cohen's missing data dummy approach makes sense if the missing value really is non-existent, but obviously everyone has a true value for age. Substituting the mean for age is another and probably inferior alternative.

II-2.

```
. reg psyscale female black rural

      Source |       SS       df       MS              Number of obs =   10335
-------------+------------------------------           F(  3, 10331) =   19.43
       Model |  325591.507     3  108530.502           Prob > F      =  0.0000
    Residual |  57707488.4 10331  5585.85698           R-squared     =  0.0056
-------------+------------------------------           Adj R-squared =  0.0053
       Total |  58033079.9 10334   5615.7422           Root MSE      =  74.739

------------------------------------------------------------------------------
    psyscale |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -3.223899   1.473059    -2.19   0.029     -6.11138   -.3364178
       black |    -13.703   2.438975    -5.62   0.000    -18.48386   -8.922137
       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

  +-----------------------------------------------------------+
  |  obs:   rstandard    psyscale   female   black   rural |
  |-----------------------------------------------------------|
  | 4840.   -.9971863   -13.33856        0       0       0 |
  | 4932.     -.94367   -12.56431        1       0       0 |
  |  939.   -.9420472   -9.218133        0       0       0 |
  | 5794.   -.9130703   -10.27764        1       0       0 |
  | 1724.   -.9067409   -6.579773        0       0       0 |
  +-----------------------------------------------------------+

  +-----------------------------------------------------------+
  | 4573.   1.233326   130.7467        0       1       1 |
  |  378.   1.251217   137.7192        1       1       0 |
  | 4092.   1.280469   143.1267        0       1       0 |
  | 4566.   1.289685   134.9561        0       1       1 |
  |   12.   82.27164   6190.022        1       1       0 |
  +-----------------------------------------------------------+

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

. reg psyscale female black rural

      Source |       SS       df       MS              Number of obs =   10335
-------------+------------------------------           F(  3, 10331) =   84.63
       Model |  489028.274     3  163009.425           Prob > F      =  0.0000
    Residual |  19899499.5 10331  1926.19296           R-squared     =  0.0240
-------------+------------------------------           Adj R-squared =  0.0237
       Total |  20388527.7 10334  1972.95604           Root MSE      =  43.888

------------------------------------------------------------------------------
    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
------------------------------------------------------------------------------
```

The `predict` command (which computes the standardized residuals) and the `extremes` command show that case 12 is an extreme outlier, i.e. 82.7 is an enormous standardized residual. Further, we see that the observed value of `psyscale` for case 12 is 6190, which is far larger than any of the other listed values for `psyscale`. The researcher decides to address the problem by dividing `psyscale` by 100 for case 12 only. Presumably she thinks the value for case 12 was entered incorrectly. Hopefully she has checked this assumption first! If you are certain the problem is a coding error, then this is the optimal solution. If you are not so sure, you might just drop case 12; or you might use robust regression techniques like `rreg` and `qreg` that are less sensitive to outliers. (Incidentally, notice how much the Anova table changes when you change case 12; in particular, the Mean Square Total – i.e. the variance of psyscale – drops from 5615.74 down to 1972.96. Also note how the coefficients and t-values change. Even though it is only one case out of 10,000, the outlier has a huge effect.)

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     =  0.0025
-------------+------------------------------           Adj R-squared =  0.0003
       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 |   .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 |  -.5185488   2.903098    -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 |  -.6321259    2.85072    -0.22   0.825     -6.22114    4.956888
       _cons |  -3.166028   2.651549    -1.19   0.233    -8.364556      2.0325
------------------------------------------------------------------------------

. alpha x1 x2 x3 x4 x5 x6 x7 x8 x9, i gen(xscale)

Test scale = mean(unstandardized items)

                                                      average
                            item-test     item-rest       interitem
Item          | Obs  Sign   correlation   correlation     covariance      alpha
--------------+-----------------------------------------------------------------
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     =  0.0020
-------------+------------------------------           Adj R-squared =  0.0018
       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
------------------------------------------------------------------------------
```

The researcher is probably disappointed that none of her variables have statistically significant effects. She might suspect that the problem is unreliability, i.e. each item

suffers from random measurement error. She might also think that all items measure the same concept, in which case she is basically entering the same variable 9 different times into her regression when she really ought to only do so once. She therefore decides to see if the 9 items can legitimately be combined into a scale. Doing so produces a fairly large Cronbach's alpha. When she enters the scale into the regression, she gets a statistically significant result.

She could have also done a Wald test or an incremental F test to see if the items can legitimately be summed together, e.g.

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

 ( 1)  x1 - x2 = 0
 ( 2)  x1 - x3 = 0
 ( 3)  x1 - x4 = 0
 ( 4)  x1 - x5 = 0
 ( 5)  x1 - x6 = 0
 ( 6)  x1 - x7 = 0
 ( 7)  x1 - x8 = 0
 ( 8)  x1 - x9 = 0

       F(  8,  3965) =    0.26
            Prob > F =    0.9788
```

Based on these results, creating an additive scale out of the items seems reasonable.

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:

Variable	Description
hcare	Support for health care reform. Ranges from a low of 0 to a high of 200
health	Overall health of the respondent. Ranges from 0 (very poor health) to 100 (very good health).
age	Age of the respondent, in years
gop	Coded 1 if the respondent is a Republican, 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
-------------+--------------------------------------------------------
       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

                       SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
    health      2.12    1.46    [ 1 ]       0.5283
       age      1.09    1.04    0.9209      0.0791
       gop      2.03    1.42    0.4926      0.5074
     black      1.00    1.00    0.9988      0.0012
----------------------------------------------------
  Mean VIF      1.56

[Part of output deleted]

. reg  hcare health age gop black, beta

      Source |       SS       df       MS              Number of obs =   10337
-------------+------------------------------           F(  4, 10332) =  881.52
       Model |  620082.606     4  155020.652           Prob > F      =  0.0000
    Residual |  1816944.64 10332  175.856044           R-squared     =  [ 2 ]
-------------+------------------------------           Adj R-squared =  ------
       Total |  2437027.25 10336    [ 3 ]              Root MSE      =  13.261

------------------------------------------------------------------------------
       hcare |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
      health |  -.7485279     .01966   -38.07   0.000                -.4709032
         age |   .1237255   --------    [ 4 ]   0.000                 .1387257
         gop |  -1.540187   .3721392    -4.14   0.000                -.0500913
       black |   3.679295   .4256284     8.64   0.000                 .0734762
       _cons |   [ 5 ]      .9741076   112.27   0.000                        .
------------------------------------------------------------------------------

. 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

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |     127.51     12    0.0000
            Skewness |     262.14      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

    Variable |     Partial      SemiP    Partial^2     SemiP^2       Sig.
-------------+------------------------------------------------------------
      health |     -0.3508     -0.3234      0.1230      0.1046       0.000
         age |      0.1524      0.1331      0.0232      0.0177       0.000
         gop |     -0.0407     -0.0352      0.0017      0.0012       0.000
       black |      0.0847      0.0734      0.0072      0.0054       0.000

. 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.)

First off, here are the uncensored results:

```
. collin health age gop black

  Collinearity Diagnostics

                     SQRT                    R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
    health      2.12    1.46    0.4717      0.5283
       age      1.09    1.04    0.9209      0.0791
       gop      2.03    1.42    0.4926      0.5074
     black      1.00    1.00    0.9988      0.0012
----------------------------------------------------
  Mean VIF      1.56

. reg  hcare health age gop black, beta

      Source |       SS       df       MS              Number of obs =   10337
-------------+------------------------------           F(  4, 10332) =  881.52
       Model |  620082.606     4  155020.652           Prob > F      =  0.0000
    Residual |  1816944.64 10332  175.856044           R-squared     =  0.2544
-------------+------------------------------           Adj R-squared =  0.2542
       Total |  2437027.25 10336    235.7805           Root MSE      =  13.261

------------------------------------------------------------------------------
       hcare |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
      health |  -.7485279     .01966   -38.07   0.000                -.4709032
         age |   .1237255   .0078948    15.67   0.000                 .1387257
         gop |  -1.540187   .3721392    -4.14   0.000                -.0500913
       black |   3.679295   .4256284     8.64   0.000                 .0734762
       _cons |   109.3654   .9741076   112.27   0.000                        .
------------------------------------------------------------------------------
```

To confirm that Stata got it right:

[1] = Tol_{health} = $1/VIF_{health}$ = 1/2.12 = .4717

[2] = R^2 = SSR/SST = 620082.606 / 2437027.25 = .2544. Those who prefer to do things the painfully hard way can compute

$$R^2 = \frac{F*K}{(N-K-1)+(F*K)} = \frac{881.52*4}{(10337-4-1)+(881.52*4)} = \frac{3526.08}{13858.08} = .2544$$

[3] = MSTotal = SSTotal / DFTotal = 2437027.25 / 10336 = 235.78. Or, if you prefer, keep in mind that MSTotal is the same as the variance of `hcare`, so you can just square the value of the standard deviation reported by the summarize command, e.g. Var(hcare) = SD(hcare)^2 = 15.35515^2 = 235.78.

[4] = T_{age}. This is slightly tricky because the standard error has been deleted from the printout. However, note that the `test age` command is also a test of whether or not the effect of age is zero; because the effect of age is positive, you can take the positive square root of the F value reported by the `test` command to get the corresponding T value. So, T_{age} = Sqrt(245.6) = 15.67.

However, somebody who feels that life should be more challenging than that and who happens to remember the formulas for semipartial correlations (see the Review of Multiple Regression handout) could note that

$$sr_k = \frac{T_k * \sqrt{1-R^2_{YH}}}{\sqrt{N-K-1}} => T_k = sr_k * \frac{\sqrt{N-K-1}}{\sqrt{1-R^2_{YH}}}$$

Ergo,

$$=> T_{age} = sr_{age} * \frac{\sqrt{N-K-1}}{\sqrt{1-R^2_{YH}}} = .1331 * \frac{\sqrt{10337-4-1}}{\sqrt{1-.2544}}$$

$$= .1331 * \frac{\sqrt{10332}}{\sqrt{.7456}} = .1331 * 117.717 = 15.67$$

Another alternative for those who refuse to take the easy way out: using info from the `regress`, `collin` and `pcorr2` commands you can compute

$$s_{b_k} = \sqrt{\frac{1-R^2_{YH}}{(1-R^2_{X_k G_k}) * (N-K-1)}} * \frac{s_y}{s_{X_k}} = \sqrt{\frac{1-.2544}{(1-.0791) * (10337-4-1)}} * \frac{15.355}{17.217}$$

$$= \sqrt{\frac{.7456}{(.9209) * (10332)}} * \frac{15.355}{17.217} = 0.0078949$$

$$=> \mathrm{T}_{age} = b_{age} / s_{b\,age} = .1237255/0.0078949 = 15.67$$

[5] = $b_{constant} = t_{constant} * se_{constant} = 112.27 * .9741076 = 109.36$.

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

All variables have statistically significant effects. By looking at the signs of the coefficients, we see that the better someone's health is, the less likely he or she is to support health care reform. Republicans are also less likely to support reform. Older people, and blacks, are more likely to support reform.

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.

Not true, at least in this sample. The mean for `gop` is .525 (see the output from the `summarize` command), which means that a little over half of all respondents identify themselves as Republicans.

2. Heteroskedasticity are not present in these data.

The initial `estat hettest` command looked good (chi-square = 1.65 with 1 d.f., p = .1986); it indicated there were no linear forms of heteroskedasticity, e.g. as one of the Xs goes up the residual variance does not go up. However, the `estat imtest` command does a broader test, and the highly significant test statistic (127.51 with 12 d.f., p = 0.0000) indicates that some form of heteroskedasticity is present, e.g. maybe as one of the Xs goes off to extreme values in either direction, the residual variances get bigger (sort of like an hourglass shape). The researcher may wish to investigate this further and decide what if anything should be done about it.

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

The hypothesized effects of `black` and `gop` are in opposite directions, i.e. the researcher expected the effect of `black` to be positive while the effect of `gop` was negative. Looking at the coefficients, the signs of the effects (`black` = 3.68 and `gop` = -1.54) are as predicted. Also as expected, the effect of `black` is larger in magnitude (stronger) than the effect of `gop`. We still need a formal test of whether these differences are statistically significant. The **`test black = -gop`** command does this. The highly significant F value (14.51) indicates that the effects differ in magnitude (i.e. are not equally strong), and by looking at the coefficients we see that it is the `black` effect that is larger.

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

Several criteria suggest that `gop` is the least important variable (but still highly significant). Both variables are measured as dichotomies and the coefficient for `black` (3.68) is larger than the coefficient for `gop` (-1.54). The T value for `black` (8.64) is larger than the t value for `gop` (-4.14). The standardized coefficient for `black` is also larger (.0735 versus -.0501). Finally, the squared semipartial for `black` is .0054 compared to only .0012 for `gop`, meaning that `black` makes a larger unique contribution to R^2 than `gop` does.

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

It looks like she got that one right. In her first `alpha` command, she tested whether a simple additive scale would work, and the Cronbach's alpha was very low (.2172). Of course, an additive scale probably doesn't make much sense given the different metrics the variables are measured in. The 2nd `alpha` command therefore standardized the items first (i.e. did z-score transformations) and then added them together. This worked a little better, but the Cronbach's alpha was still very low, only .4269.

Appendix: Stata Commands for Exam 1. Here are the commands I used to generate the Stata output on the exam. Alas, I haven't really conducted any new nationwide studies, but I have manipulated and sometimes disguised other data sets I have sitting around.

```
version 11

* Problem I-1
webuse nhanes2f, clear
reg health female black rural
pcorr2 health female black rural
sw, pr(.001): reg health female black rural

* Problem I-2
webuse nhanes2f, clear
reg health female

* Problem II-1
use "http://www.indiana.edu/~jslsoc/stata/spex_data/ordwarm2.dta", clear
* Randomly switch 50% of the data on age to missing
* swor command is available from SSC
set seed 345
swor 1147, gen(mdvar) keep
replace age = . if mdvar
* Run analyses
regress warm yr89 male white age ed prst
sum warm yr89 male white age ed prst, sep(7)
mi set mlong
mi register imputed age
mi register regular warm yr89 male white ed prst
mi impute regress age warm yr89 male white ed prst, add(100) rseed(123)
mi estimate: regress warm yr89 male white age ed prst

* Problem II-2
webuse nhanes2f, clear
* Disguise data, create new DV
set seed 123
gen psyscale = health ^ 3 + rnormal() * 5
* Create extreme outlier
replace psyscale = psyscale * 100  in 12
* Run analyses
reg psyscale female black rural
predict rstandard, rstandard
extremes rstandard psyscale female black rural
replace psyscale = psyscale/100 in 12
reg psyscale female black rural
* Additional checks
predict rstandard2, rstandard
extremes rstandard2 psyscale female black rural

* Problem II-3
use "D:\SOC63993\Statafiles\anomia.dta", clear
* Cleverly disguise the data
renpfix anomia x
set seed 456
* Create a new variable that has the desired relationship with the Xs
gen y = x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + rnormal() * 75
* Run analyses
reg y  x1 x2 x3 x4 x5 x6 x7 x8 x9
alpha x1 x2 x3 x4 x5 x6 x7 x8 x9, i gen(xscale)
reg y xscale
* Alternative approach -- Wald test
quietly reg y  x1 x2 x3 x4 x5 x6 x7 x8 x9
test x1 = x2 = x3 = x4 = x5 = x6 = x7 = x8 = x9
```

```
* Problem III
webuse nhanes2f, clear
keep weight height age female black
* Cleverly disguise the data
ren weight hcare
gen health = 225 - height
gen gop = female
drop  height female
order  hcare health age gop black
* Run analyses
sum
collin health age gop black
reg  hcare health age gop black, beta
test age
test black = -gop
estat hettest
estat imtest, white
pcorr2 hcare health age gop black
alpha health age gop black
alpha health age gop black, s
```