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

1. The independent variables in a model include X1, X2, and X1*X2. X1 and X2 both have missing values. If multiple imputation is used for X1 and X2, then passive imputation should be used to impute values for X1*X2.

2. A researcher runs the following analysis:

```
. alpha v1 v2 v3, i
```

Test scale = mean(unstandardized items)

<table>
<thead>
<tr>
<th>Item</th>
<th>Obs</th>
<th>Sign</th>
<th>item-test correlation</th>
<th>item-rest correlation</th>
<th>average interitem covariance</th>
<th>alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>v1</td>
<td>2500</td>
<td>+</td>
<td>0.7296</td>
<td>0.3393</td>
<td>.0357863</td>
<td>0.2613</td>
</tr>
<tr>
<td>v2</td>
<td>2500</td>
<td>+</td>
<td>0.6693</td>
<td>0.2537</td>
<td>.0634239</td>
<td>0.4150</td>
</tr>
<tr>
<td>v3</td>
<td>2500</td>
<td>+</td>
<td>0.6820</td>
<td>0.2610</td>
<td>.060012</td>
<td>0.4036</td>
</tr>
</tbody>
</table>

Test scale | .0530741 | 0.4610

Based on these results, v1 should be dropped from the scale.

3. In a bivariate regression, if a case is an extreme outlier on Y, then the closer its value on X is to the mean of X, the more impact the case will have on the slope coefficient.

4. While random measurement error in the independent variables is problematic, random measurement error in the dependent variable has no adverse consequences.

5. Marital satisfaction is a key independent variable in the analysis. However, some subjects are not married. The Cohen and Cohen dummy variable adjustment technique is one way of dealing with this problem.

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.)
. logit diabetes wgt female black age, nolog

Logistic regression                               Number of obs   =       8316
LR chi2(4)      =     347.55
Prob > chi2     =     0.0000
Pseudo R2       =     0.1263

------------------------------------------------------------------------------
diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
wgt |   .0242748   .0035619     6.82   0.000     .0172936    .0312561
female |   .4823663   .1217499     3.96   0.000     .2437409    .7209917
black |   .8426339    .151018     5.58   0.000     .5466441    1.138624
age |   .0623913   .0042484    14.69   0.000     .0540645    .0707181
  _cons |  -8.586301     .40814   -21.04   0.000    -9.386241   -7.786361
------------------------------------------------------------------------------

. sum diabetes wgt female black age

Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
diabetes |     10335    .0482825     .214373          0          1
wgt |      8316    71.83235    15.51516      30.84     175.88
female |     10335    .5250121    .4993982          0          1
black |     10335    .1050798    .3066711          0          1
age |     10335    47.56584    17.21752         20         74

. mi set mlong
. mi register imputed wgt
(2019 m=0 obs. now marked as incomplete)
. mi impute pmm wgt diabetes female black age, add(10) knn(5) rseed(2232)

Univariate imputation                       Imputations =       10
Predictive mean matching                          added =       10
Imputed: m=1 through m=10                       updated =        0
Nearest neighbors =        5

|               Observations per m
|----------------------------------------------
|----------------------------------------------
|----------------------------------------------
|----------------------------------------------
|----------------------------------------------
|----------------------------------------------
Variable |   Complete   Incomplete   Imputed |     Total
-------------------+-----------------------------------+----------
wgt |       8316         2019      2019 |     10335

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

. mi estimate, dots: logit diabetes wgt female black age

Imputations (10):
...........10 done

Multiple-imputation estimates                  Imputations =       10
Logistic regression  Number of obs =     10335
Average RVI =  0.0374
Largest FMI =  0.1465
DF adjustment: Large sample
   DF:      min =    442.32
           avg =    369971.66
           max =    1543449.26
Model F test:       Equal FMI                     F(   4,15282.4) =      78.10
Within VCE type:          OIM                     Prob > F        =     0.0000

| diabetes | Coef. | Std. Err. |      t | P>|t|   | [95% Conf. Interval] |
|----------|-------|-----------|-------|------|-------------------|
| wgt      | 0.0241408 | 0.0031895 | 7.57  | 0.000 | 0.0178724 - 0.0304092 |
| female   | 0.3806918 | 0.0993647 | 3.83  | 0.000 | 0.1859312 - 0.5754525 |
| black    | 0.6186384 | 0.1288599 | 4.80  | 0.000 | 0.3660765 - 0.8712002 |
| age      | 0.0615972 | 0.0038579 | 15.97 | 0.000 | 0.0540359 - 0.0691585 |
| _cons    | -8.426041 | 0.3708681 | -22.72| 0.000 | -9.15351 - -7.698571  |

II-2.

. reg hscore weight

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 6000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>789854.23</td>
<td>1</td>
<td>789854.23</td>
<td>F( 1, 5998) = 5546.54</td>
</tr>
<tr>
<td>Residual</td>
<td>854144.85</td>
<td>5998</td>
<td>142.404943</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>1643999.08</td>
<td>5999</td>
<td>274.045521</td>
<td>R-squared = 0.4804</td>
</tr>
</tbody>
</table>

| hscore | Coef. | Std. Err. |      t | P>|t|  | [95% Conf. Interval] |
|--------|-------|-----------|-------|------|-------------------|
| weight | 0.7383213 | 0.0099137 | 74.48 | 0.000 | 0.718887 - 0.7577557 |
| _cons  | 0.197513 | 0.7323928 | 0.27  | 0.787 | -1.23824 - 1.633266 |

. dfbeta
    _dfbeta_1: dfbeta(weight)

. extremes _dfbeta_1 hscore weight

+---------------------------------------+---------------------------------------+---------------------------------------+---------------------------------------+
| obs:   _dfbeta_1     hscore   weight | obs:   _dfbeta_1     hscore   weight | obs:   _dfbeta_1     hscore   weight | obs:   _dfbeta_1     hscore   weight |
|--------|-----------|--------|--------|-----------|--------|--------|-----------|--------|-----------|--------|--------|-----------|--------|--------|-----------|--------|--------|-----------|--------|--------|
| 4652.  | -1.002746 | 382    | 34.93  | 382    | 34.93  | 382    | 34.93  | 382    | 34.93  | 382    | 34.93  | 382    | 34.93  | 382    | 34.93  | 382    | 34.93  | 382    | 34.93  | 382    | 34.93  |
| 4906.  | -1.199883 | 51.05366 | 118.84 | 51.05366 | 118.84 | 51.05366 | 118.84 | 51.05366 | 118.84 | 51.05366 | 118.84 | 51.05366 | 118.84 | 51.05366 | 118.84 | 51.05366 | 118.84 | 51.05366 | 118.84 |
| 4616.  | -.0789586 | 59.08268 | 115.33 | 59.08268 | 115.33 | 59.08268 | 115.33 | 59.08268 | 115.33 | 59.08268 | 115.33 | 59.08268 | 115.33 | 59.08268 | 115.33 | 59.08268 | 115.33 | 59.08268 | 115.33 |
| 5076.  | -.0768352 | 73.24126 | 126.44 | 73.24126 | 126.44 | 73.24126 | 126.44 | 73.24126 | 126.44 | 73.24126 | 126.44 | 73.24126 | 126.44 | 73.24126 | 126.44 | 73.24126 | 126.44 | 73.24126 | 126.44 |
| 5164.  | .0792918 | 112.779 | 120.77 | 112.779 | 120.77 | 112.779 | 120.77 | 112.779 | 120.77 | 112.779 | 120.77 | 112.779 | 120.77 | 112.779 | 120.77 | 112.779 | 120.77 | 112.779 | 120.77 |
| 4139.  | .1081116 | 121.6261 | 123.72 | 121.6261 | 123.72 | 121.6261 | 123.72 | 121.6261 | 123.72 | 121.6261 | 123.72 | 121.6261 | 123.72 | 121.6261 | 123.72 | 121.6261 | 123.72 | 121.6261 | 123.72 |
| 1732.  | .1235975 | 131.2471 | 144.24 | 131.2471 | 144.24 | 131.2471 | 144.24 | 131.2471 | 144.24 | 131.2471 | 144.24 | 131.2471 | 144.24 | 131.2471 | 144.24 | 131.2471 | 144.24 | 131.2471 | 144.24 |
| 5611.  | .1358087 | 140.1575 | 159.48 | 140.1575 | 159.48 | 140.1575 | 159.48 | 140.1575 | 159.48 | 140.1575 | 159.48 | 140.1575 | 159.48 | 140.1575 | 159.48 | 140.1575 | 159.48 | 140.1575 | 159.48 |

. drop _dfbeta_1
. drop in 4652
(1 observation deleted)
. reg hscore weight

Source |       SS       df       MS              Number of obs =    5999
-------------+------------------------------           F(  1,  5997) = 6669.63
Model |  808826.687     1  808826.687           Prob > F      =  0.0000
Residual |  727256.701  5997  121.270085           R-squared     =  0.5266
-------------+------------------------------           Adj R-squared =  0.5265
Total |  1536083.39  5998  256.099264           Root MSE      =  11.012

------------------------------------------------------------------------------
hscore |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
weight |   .7474949   .0091529    81.67   0.000      .729552    .7654378
    _cons |  -.5244449    .676231    -0.78   0.438    -1.850101     .801211
------------------------------------------------------------------------------

. dfbeta
      _dfbeta_1: dfbeta(weight)

. extremes _dfbeta_1 hscore weight

II/3.

. reg y x

Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  1,   198) =  824.88
Model |  499764.343     1  499764.343           Prob > F      =  0.0000
Residual |  119960.988   198  605.863576           R-squared     =  0.8064
-------------+------------------------------           Adj R-squared =  0.8055
Total |  619725.331   199  3114.19765           Root MSE      =  24.614

------------------------------------------------------------------------------
y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
x |   5.027104    .175034    28.72   0.000     4.681934    5.372274
    _cons |   .7003925    1.74319     0.40   0.688    -2.737208    4.137993
------------------------------------------------------------------------------
. estat hettest

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

\[ \text{chi2(1)} = 3.36 \]
\[ \text{Prob > chi2} = 0.0670 \]

. estat imtest

Cameron & Trivedi's decomposition of IM-test

<table>
<thead>
<tr>
<th>Source</th>
<th>chi2</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heteroskedasticity</td>
<td>149.71</td>
<td>2</td>
<td>0.0000</td>
</tr>
<tr>
<td>Skewness</td>
<td>18.73</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.19</td>
<td>1</td>
<td>0.1385</td>
</tr>
<tr>
<td>Total</td>
<td>170.64</td>
<td>4</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

. scatter y x

. scatter y x, by(female)
III. Computation and interpretation. (35 points total) The Republican Party is dismayed that it has lost the popular vote in 5 of the last 6 presidential elections. The party leadership strongly suspects that Vice-President Joe Biden will be the Democratic Party nominee in 2016. It has therefore commissioned political consultant Dick Morris to assess Biden’s strengths and weaknesses. Morris has conducted a random survey of more than 5,000 registered voters. The variables he has collected data on are

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>biden</td>
<td>How much the respondent likes Biden. Scores potentially range from a low of 0 to a high of 200</td>
</tr>
<tr>
<td>m47</td>
<td>Is the respondent a member of the 47%, i.e. the proportion of the population that does not pay federal income taxes (although most pay other types of taxes)? Coded 1 if yes, 0 otherwise</td>
</tr>
</tbody>
</table>
obamacare
How much does the respondent support Obama’s health care program? Scores range from a low of 0 to a high of 10.

tea party
Does the respondent consider himself or herself a member of the Tea Party? 1 = Tea Party, 0 = not Tea Party

black
Respondent’s race (1 = black, 0 = not black)

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
-----------------+-----------------------------------
   biden | 5032  72.01731  15.41968    30.84     158.53
obamacare | 5032  4.684237  1.383346  .5998579   9.042428
      m47 | 5032  .4789348  .4996057      0        1
     teaparty | 5032  .1405008  .3475404      0        1
     black | 5032  .1065183  .3085305      0        1
```

```
. alpha m47 teaparty black
Test scale = mean(unstandardized items)
Average interitem covariance:  .0011512
Number of items in the scale:  3
Scale reliability coefficient:  0.0219
```

```
. collin obamacare m47 teaparty black
(obs=5032)
Collinearity Diagnostics

<table>
<thead>
<tr>
<th>Variable</th>
<th>VIF</th>
<th>SQRT VIF</th>
<th>Tolerance</th>
<th>R-Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>obamacare</td>
<td>1.97</td>
<td>1.40 [1]</td>
<td>0.4927</td>
<td></td>
</tr>
<tr>
<td>m47</td>
<td>1.95</td>
<td>1.40</td>
<td>0.5118</td>
<td>0.4882</td>
</tr>
<tr>
<td>teaparty</td>
<td>1.02</td>
<td>1.01</td>
<td>0.9837</td>
<td>0.0163</td>
</tr>
<tr>
<td>black</td>
<td>1.00</td>
<td>1.00</td>
<td>0.9986</td>
<td>0.0014</td>
</tr>
</tbody>
</table>

Mean VIF   1.49
```
. reg biden m47 obamacare teaparty black, l(99)

Source | SS       | df | MS              | Number of obs = 5032
--------+----------+----+----------------+-----------------------------
Model    | 314848.283 | 4  | 78712.0707     | F(4, 5027) = [2]             |
Residual | 881355.261 | 5027| 175.324301    | Prob > F = 0.0000            |
Total    | 1196203.54  | 5031|                  | Adj R-squared = 0.2626       |
        |           |    |                | Root MSE = 13.241            |

------------------------------------------------------------------------------
 biden | Coef.  Std. Err.      t    P>|t|    [99% Conf. Interval]
--------+-----------------------------------------------
m47   |  1.491152   .5222973     2.85   0.004     .1452928    2.837012
obamacare |  5.184225   .1894643
 teaparty | -6.736499   .5415667   -12.44   0.000    -8.132012   -5.340986
black  |   2.9218    .6054786     4.83   0.000     1.361598    4.482001
 _cons  |   47.65426   .7544787    63.16   0.000     45.71011    49.59841
------------------------------------------------------------------------------

. test teaparty

( 1)  teaparty = 0

F( 1, 5027) = [5]  Prob > F = 0.0000

. test obamacare = 5

( 1)  obamacare = 5

F( 1, 5027) = 0.95  Prob > F = 0.3309

. pcorr biden obamacare m47 teaparty black
(obs=5032)

Partial and semipartial correlations of biden with

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>obamacare</td>
<td>0.3600</td>
<td>0.3313</td>
<td>0.1296</td>
<td>0.1097</td>
<td>0.0000</td>
</tr>
<tr>
<td>m47</td>
<td>0.0402</td>
<td>0.0346</td>
<td>0.0016</td>
<td>0.0002</td>
<td>0.0043</td>
</tr>
<tr>
<td>teaparty</td>
<td>-0.1728</td>
<td>-0.1506</td>
<td>0.0299</td>
<td>0.0227</td>
<td>0.0000</td>
</tr>
<tr>
<td>black</td>
<td>0.0679</td>
<td>0.0584</td>
<td>0.0046</td>
<td>0.0034</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

. estat hettest

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

chi2(1) = 0.22
Prob > chi2 = 0.6421
. estat imtest

Cameron & Trivedi's decomposition of IM-test

<table>
<thead>
<tr>
<th>Source</th>
<th>chi2</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heteroskedasticity</td>
<td>82.26</td>
<td>11</td>
<td>0.0000</td>
</tr>
<tr>
<td>Skewness</td>
<td>159.73</td>
<td>4</td>
<td>0.0000</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>34.13</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>276.12</td>
<td>16</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

a) (10 pts) Fill in the missing quantities [1] – [5]. (A few other values may 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. Which groups or types of individuals are most supportive of Biden and which are least supportive?

2. Based on the results of the `pcorr` command, one analyst thinks m47 should be dropped from the model. Explain what you think her reasoning is and why you agree or disagree.

3. There was concern that the variables teaparty, m47 and black would be highly collinear. Do you think that fear was justified? Would you recommend combining the items into a scale?

4. The party leaders are upset because they thought the analysis revealed a clear violation of OLS assumptions but nothing was done about it. Why did they feel that way?

5. Previous studies had found that the slope coefficient for obamacare was 6. The Republican leaders wanted to see if that had changed, so, using the .01 level of significance, they wanted to test the hypothesis

\[ H_0: \beta_{\text{obamacare}} = 6 \]
\[ H_A: \beta_{\text{obamacare}} \neq 6 \]

Unfortunately Morris thought they said 5, not 6, so the wrong test was conducted. Explain to the party leaders why they still have the information they need to reject the null hypothesis.