Advanced Methods in Health Services Research: Analysis - 309.716
Tuesday and Thursday 9:00-10:20
Instructors: Darrell J. Gaskin, Ph.D.
Roland J. Thorpe, Ph.D.
250 Hampton House
Due: September 15, 2016

Computer Exercise #2: Estimating Impact of Age, Insurance Status and Income on Healthcare Expenditures
Answer Guide

1. What is heteroskedasticity? Why is it a problem? How do you remedy it? List three examples of possible heteroskedastic data from health services research, public health or health policy?
    a. What is heteroskedasticity?
        Heteroskedasticity is the violation of the homoskedasticity assumption $Var(u|x_1, \ldots, x_k) = \sigma^2$, i.e., that the error term has the same variance regardless of the values of the explanatory variables. Heteroskedasticity is exhibited when the variance of the error term varies with different values of the explanatory variables. It occurs whenever the variance of the unobservables varies across different segments of the population, where the segments are determined by the different values of the explanatory variables.
    b. Why is it a problem?
        Heteroskedasticity becomes a problem when one wants to draw statistical inferences from the regression. With heteroskedasticity, the estimators of the variances are biased. Since the OLS standard errors are based the variances, they are no longer valid for constructing t statistics and confidence intervals. Other statistics such as the F and LM statistics are likewise no longer valid in the presence of heteroskedasticity.
    c. How do you remedy it?
        Two well-used remedies are the a) construction of heteroskedasticity-robust standard errors for heteroskedasticity of unknown form, and b) weighted least squares estimation.
    d. List three examples of possible heteroskedastic data from health services research, public health, or health policy.
        Possible heteroskedastic data from HSR, public health, or health policy include a) hospital cost data, b) individual- or patient-level cost data, c) number of hospital days in a year, and d) number of visits to a doctor.

To answer questions 2 and 3, use the MEPS analysis file you built for assignment 1.

2. Estimate a model of total health expenditures for adults excluding expenditures greater than $100,000. Use the following specification – Ln(Total health expenditures+1) = f(health status, age, race/ethnicity, gender, msa, region, poverty status, education, and insurance status). *Use the following reference categories:*

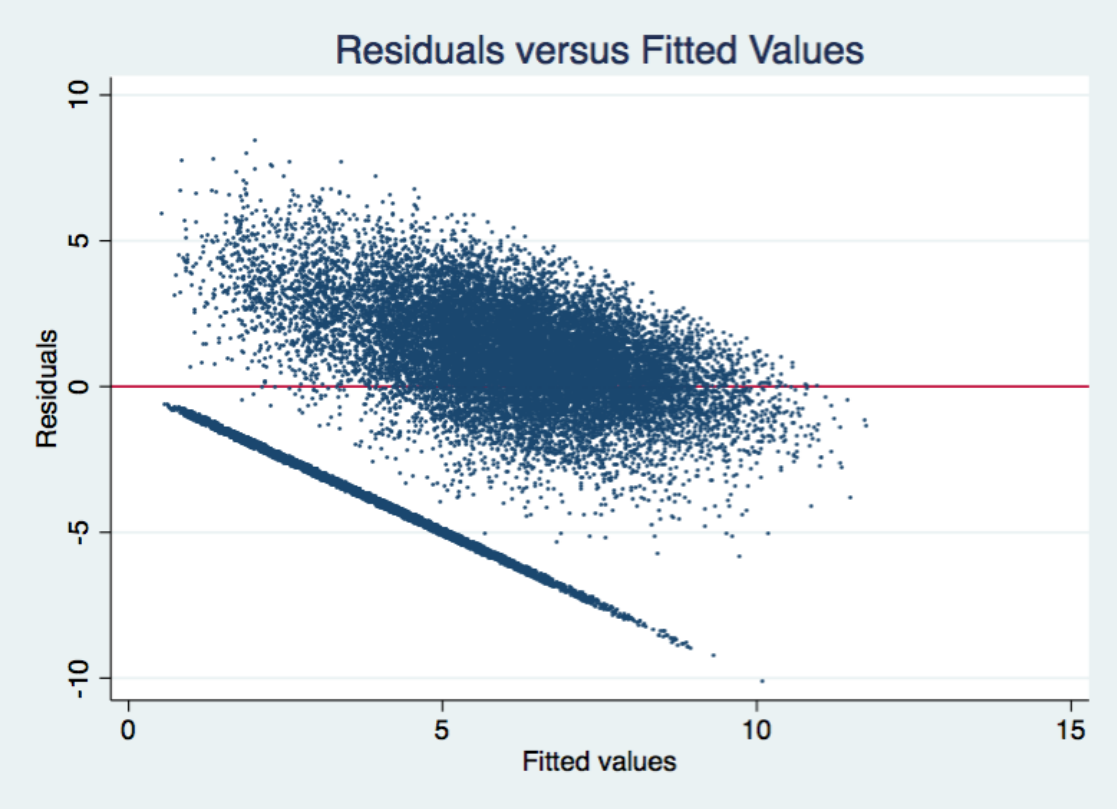
excellent health, age category 25-44, male, white non-Hispanic, HS diploma/GED, high income, privately insured, MSA, East) ***(For the purposes of this question do not use the survey commands.)***

a. Test for heteroskedascity using the White test

```
. estat imtest, white

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

         chi2(393)     =   1730.07
         Prob > chi2   =    0.0000

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |    1730.07    393    0.0000
            Skewness |    2800.08     29    0.0000
            Kurtosis |      16.71      1    0.0000
---------------------+-----------------------------
               Total |    4546.86    423    0.0000
---------------------------------------------------
```

Based on the White test, we reject the null hypothesis of homoscedasticity.

b. Produce a scatterplot of the residuals



The variances of residuals do not appear to be constant. We notice that the variance of the residuals appears to get smaller (i.e., the vertical band of residuals) with higher fitted values. In addition, we observe that the residuals are biased. We also note that the residuals appear to form a line below the main mass of points. This may be due to the high number of zero-expenditure observations in our data.

c. Test whether there is heteroskedascity related to gender using a Goldfeld-Quandt tests comparing males to females.

We show two different ways/procedures:

```
. * c. GQ Test
. * Procedure 1
. qui reg healthexp ib2.agecat female i.race ib3.education ib5.fpl ///
> i.insurance ib1.msa i.region i.healthstatus if msa == 1 & age08x > 17

.
. predict uhat, resid
(6306 missing values generated)

. sdtest uhat, by(female)

Variance ratio test
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    Male |   12811    112.8184     57.8895    6552.263   -.6536725    226.2904
  Female |   13949    23.99063    60.14802    7103.835   -93.90755    141.8888
---------+--------------------------------------------------------------------
combined |   26760    66.51576    41.84582    6845.342   -15.50425    148.5358
------------------------------------------------------------------------------
    ratio = sd(Male) / sd(Female)                                 f =   0.8507
Ho: ratio = 1                              degrees of freedom = 12810, 13948

    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.0000         2*Pr(F < f) = 0.0000           Pr(F > f) = 1.0000

.
. * Procedure 2
. qui reg lnhealthexp i.healthstatus ib2.agecat female i.race ib3.education ib5.fpl ///
> i.insurance i.region if female == 0 & age08x > 17

.
. scalar rss0 = (e(rmse))^2

. scalar df_rss0 = e(df_r)

.
. qui reg lnhealthexp i.healthstatus ib3.agecat female i.race ib3.education ib5.fpl ///
> i.insurance i.region if female == 1 & age08x > 17

.
. scalar rss1 = (e(rmse))^2

. scalar df_rss1 = e(df_r)

.
. scalar FGQ = rss0/rss1

. scalar Fcrit = invFtail(df_rss1, df_rss0, .05)

. scalar pvalue = Ftail(df_rss1, df_rss0, FGQ)

. scalar list FGQ pvalue Fcrit
```

```
      FGQ =  1.2030727
   pvalue =  2.615e-21
    Fcrit =  1.0327935
```

The GQ test suggests that men's health expenditures have a statistically significant higher variance than that of women as the p-values computed from either procedure is less than 0.001.

d. Based on the results from these tests, do you have a problem with heteroskedascity? If so what do you think it is related to?

Based on the three tests, there it appears to be a heteroskedasticity problem. The White test and the plot of residuals against fitted values suggest that heteroscedasticity exists while the GQ test suggests that gender could be a contributor to the heteroscedasticity.

3. Re-estimate the semilog model with 3 different specification for the variable ages : Log (Total health expenditures) = f(health status, age, race, gender, education, poverty status, insurance status and location). *Use the following reference categories: excellent health, male, white non-Hispanic, HS diploma/GED, high income, privately insured, msa, and East.* In this analysis we are going to explore the difference between estimating the association between age and health expenditures using linear, quadratic and spline specifications. ***(For the purposes of this question USE the survey commands.)***

a. Estimate the model 3 different ways:
   i. Model 1 – treat age a strictly linear (use age)
   ii. Model 2 - treat age as quadratic (use age and age squared)
   iii. Model 3 - estimate age as a spline function and interpret coefficients. (Note: use the prior categories as inflection points).

See the full model in the log file.

b. Produce a table displaying only the 2 sets of coefficients for age variables.

TABLE 1. Coefficients for Age Variables

	(1) Linear Age	(2) Quad Age	(3) Spline Age
age	0.0421***	0.0195**	
	(0.00147)	(0.00644)	
agesq		0.000235***	
		(0.0000621)	
age1			-0.0696***
			(0.0208)
age2			0.110***
			(0.0241)
age3			0.0223**
			(0.00771)

age4			-0.0310*
			(0.0134)
age5			-0.0116
			(0.0228)
N	31294	31294	31294
R-sq	0.335	0.336	0.339

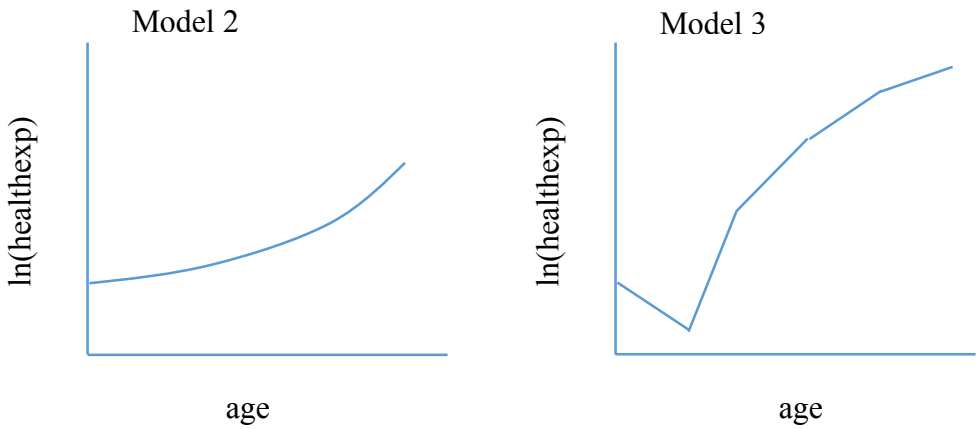
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

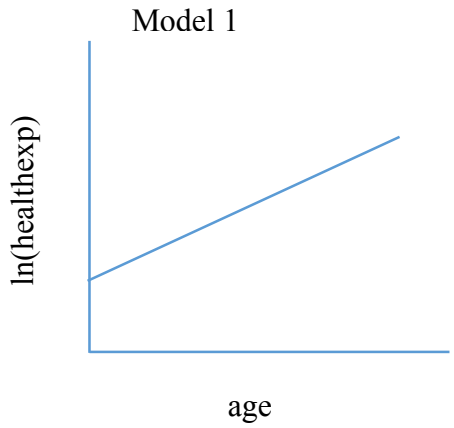
c. You have estimated the model with 3 different specifications of the age variable (linear, quadratic and spline). Examine the coefficients and [hand] draw a picture of the relationship between age and the log of health expenditures. Do these specifications agree or disagree? In your opinion which one is best and why?

The first two model (linear and quadratic age) specifications show that there is a positive and significant relationship between age and the log of health expenditures although they have different magnitudes. In addition, model 2 shows that the positive relationship between age and the log of health expenditures increases with increasing age. These generally disagree with model 3 (age spline) that does not show an increasing relationship between age and the log of health expenditures across all ages. Model 3 shows that there is an initial negative relationship between health expenditures and age but that it then becomes positive from age 25 onwards.

Models 2 and 3 show that there are statistically significant non-linearities with age. These two models may be preferable to our first model where we linearly model age.

Illustrations of the relationship between age and the log of health expenditures are shown below.


Model 2
ln(healthexp)
age
Model 3
ln(healthexp)
age


Model 1
ln(healthexp)
age

```
-------------------------------------------------------------------------------------------------------------
-----
      name:  <unnamed>
       log:  /Users/jojo/Documents/JHU/TA Folder/Advanced HSR Methods -
Analysis/Answers/Assignment2_2016.lo
> g
  log type:  text
 opened on:  16 Sep 2016, 08:21:12

. use "/Users/jojo/Documents/JHU/TA Folder/Advanced HSR Methods - Analysis/meps08.dta"

.
.
. ** Data preparation
. * Trimming/cleaning the data
. summ totexp08

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    totexp08 |     33066    3142.069    9786.619          0     553493

. summ totexp08, detail

                  total health care exp 08
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%            0              0
10%            0              0       Obs               33066
25%           67              0       Sum of Wgt.       33066

50%        528.5                      Mean           3142.069
                        Largest       Std. Dev.      9786.619
75%         2425         238659
90%         7453         264510       Variance       9.58e+07
95%        13582         373799       Skewness       14.40863
99%        40763         553493       Kurtosis        469.447

.
. gen healthexp = totexp08

. replace healthexp = . if totexp08 < 0 /*replaces negative values with missing*/
(0 real changes made)

. replace healthexp = . if totexp08 > 100000 /*replaces values > $100,000 with missing*/
(42 real changes made, 42 to missing)

. summ healthexp

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   healthexp |     33024    2937.216    7412.106          0      99988

. summ healthexp, detail

                          healthexp
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%            0              0
10%            0              0       Obs               33024
25%           66              0       Sum of Wgt.       33024

50%          526                      Mean           2937.216
                        Largest       Std. Dev.      7412.106
75%       2413.5          98210
90%         7353          99251       Variance       5.49e+07
95%        13387          99264       Skewness       5.819768
99%        38550          99988       Kurtosis       48.38959

.
. gen income = ttlp08x

. replace income = . if ttlp08x < 0 /*sets negative values to missing*/
(25 real changes made, 25 to missing)
```

```
. replace income = . if ttlp08x > 170000 /*sets very large values to missing*/
(325 real changes made, 325 to missing)

. /* Note: If you did not exclude those with incomes > 170000, it's fine*/
. summarize income

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      income |     32716    18715.83    24109.08          0     169564

. summarize income, detail

                          income
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%            0              0
10%            0              0       Obs               32716
25%            0              0       Sum of Wgt.       32716

50%        10000                      Mean           18715.83
                        Largest       Std. Dev.      24109.08
75%        28838         164895
90%        52000         164895       Variance       5.81e+08
95%        70000         165391       Skewness       1.717878
99%       103084         169564       Kurtosis       6.147364

.
. summarize age08x

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     age08x |     33066    33.69395    22.38051         -1         85

. gen age = age08x

. replace age = . if age08x < 0 /*sets negative values to missing*/
(251 real changes made, 251 to missing)

. summarize age

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         age |     32815    33.95932    22.25851          0         85

.
. gen agecat = 1 if age08x >=0 & age08x <=24
(20062 missing values generated)

. replace agecat = 2 if age08x >=25 & age08x <=44
(8813 real changes made)

. replace agecat = 3 if age08x >=45 & age08x <=64
(7614 real changes made)

. replace agecat = 4 if age08x >=65 & age08x <=74
(1867 real changes made)

. replace agecat = 5 if age08x >=75
(1517 real changes made)

. label define agecats 1 "18-24" 2  "25-44" 3 "45-64" 4 "65-74" 5 "75+"

. label values agecat agecats /*set label name sexn to the variable sex*/

.
. gen female = 0 if sex == 1
(17181 missing values generated)

. replace female = 1 if sex == 2
(17181 real changes made)

.
. label define sexn 0 "Male" 1 "Female"

. label values female sexn
```

```
.
. label define raceethn 1 "Hispanic" 2 "Black non-Hispanic" 3 "Asian non-Hispanic" 4 "Other race/not
Hispani
> c"

. label values racethnx raceethn

.
. label define racexn 1 "White" 2 "Black" 3 "Amer Indian/Alaska Native" 4 "Asian" 5 "Native
Hawaiian/Pacific
>  Islander" 6 "Multiple races reported"

. label values racex racexn

.
. gen race = 1 if racex == 1 & racethnx != 1
(18775 missing values generated)

. replace race = 2 if racex == 2 & racethnx != 1
(6476 real changes made)

. replace race = 3 if racethnx == 1
(9392 real changes made)

. replace race = 4 if racex == 4 & racethnx != 1
(1997 real changes made)

. replace race = 5 if (racex == 3 | racex == 5 | racex == 6) & racethnx != 1
(910 real changes made)

.
. label define racexn2 1 "White" 2 "Black" 3 "Hispanic" 4 "Asian" 5 "Other"

. label values race racexn2

.
. gen education = 1 if educyr >=0 & educyr <=8
(25465 missing values generated)

. replace education = 2 if educyr >=9 & educyr <=12 & hideg == 1 /*Some High School*/
(3732 real changes made)

. replace education = 2 if educyr >=9 & educyr <=11 & hideg <0 /* Some High School but didn't answer
hideg q
> uestion*/
(3 real changes made)

. replace education = 3 if hideg == 2 | hideg == 3 /*High School*/
(11173 real changes made)

. replace education = 4 if educyr >=13 & educyr <=17 & hideg == 3 /*Some college-In school but only
have hig
> h school diploma*/
(3629 real changes made)

. replace education = 5 if hideg == 4
(3245 real changes made)

. replace education = 6 if hideg == 5 | hideg ==6
(1590 real changes made)

. replace education = 6 if educyr == 17 & hideg == 7 /*Advanced Degree* - implies law degree or similar
type
> */
(66 real changes made)

.
. label define educn 1 "No High School" 2 "Some High" 3 "High School/GED" 4 "Some College/Tech
School/AA deg
> ree" 5 "College"  6 "Advanced Degree"

. label values education educn

.
. gen insurance = 1 if inscov08 ==1
```

```
(14773 missing values generated)

. replace insurance = 2 if inscov08 == 2 & mcrev08 == 2 /*Had public insurance but not medicare*/
(6642 real changes made)

. replace insurance = 3 if inscov08 == 3
(5662 real changes made)

. replace insurance = 4 if inscov08 == 2 & mcrev08 == 1 /* Medicare */
(2469 real changes made)

.
. label define insr 1 "Private Insurance" 2 "Public Insurance" 3 "Uninsured" 4 "Medicare"

. label values insurance insr

.
. gen msa = 0 if msa08 == 0
(28406 missing values generated)

. replace msa = 1 if msa08 == 1
(28155 real changes made)

.
. label define msan 0 "Not in MSA" 1 "In MSA"

. label values msa msan

.
. gen region = 1 if region08 == 1
(28080 missing values generated)

. replace region = 2 if region08 == 3
(12424 real changes made)

. replace region = 3 if region08 == 2
(6499 real changes made)

. replace region = 4 if region08 == 4
(8906 real changes made)

.
. label define regi 1 "North East" 2 "South" 3 "Midwest" 4 "West"

. label values region regi

.
. gen healthstatus = 1 if rthlth42 == 1
(23215 missing values generated)

. replace healthstatus = 2 if rthlth42 == 2
(10188 real changes made)

. replace healthstatus = 3 if rthlth42 == 3
(8628 real changes made)

. replace healthstatus = 4 if rthlth42 == 4
(2777 real changes made)

. replace healthstatus = 5 if rthlth42 == 5
(815 real changes made)

.
. label define health 1 "Excellent" 2 "Very Good" 3 "Good" 4 "Fair" 5 "Poor"

. label values healthstatus health

.
. gen fpl = 1 if povcat08 == 1
(26099 missing values generated)

. replace fpl = 2 if povcat08 == 2
(2171 real changes made)

. replace fpl = 3 if povcat08 == 3
(5667 real changes made)
```

```
. replace fpl = 4 if povcat08 == 4
(9595 real changes made)

. replace fpl = 5 if povcat08 == 5
(8666 real changes made)

.
. label define fplstat 1 "Poor" 2 "Near Poor" 3 "Low Income" 4 "Middle Income" 5 "High Income"

. label values fpl fplstat

.
. * Create the log of health expenditure
. gen lnhealthexp = ln(1+healthexp)
(42 missing values generated)

.
. ** Number 2 **
. * a. White Test *
. set matsize 500

.
. reg lnhealthexp i.healthstatus ib2.agecat female i.race ib3.education ib5.fpl ///
> i.insurance ib1.msa i.region if age08x > 17

      Source |       SS       df       MS              Number of obs =   20959
-------------+------------------------------           F( 29, 20929) =  368.59
       Model |  76918.6534    29  2652.36736           Prob > F      =  0.0000
    Residual |  150606.456 20929  7.19606554           R-squared     =  0.3381
-------------+------------------------------           Adj R-squared =  0.3371
       Total |  227525.109 20958  10.8562415           Root MSE      =  2.6825

------------------------------------------------------------------------------
                   lnhealthexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------------------+----------------------------------------------------------------
                  healthstatus |
                     Very Good |    .558572   .0508435    10.99   0.000     .4589149    .6582292
                          Good |   .9426887    .052604    17.92   0.000     .8395809    1.045797
                          Fair |   2.128763   .0708208    30.06   0.000     1.989949    2.267577
                          Poor |     3.1414   .1113156    28.22   0.000     2.923213    3.359587
                               |
                        agecat |
                         18-24 |  -.2251244   .0593469    -3.79   0.000    -.3414488      -.1088
                         45-64 |   .9534145   .0460207    20.72   0.000     .8632105    1.043619
                         65-74 |    1.68133   .0828887    20.28   0.000     1.518862    1.843798
                           75+ |   2.084888   .0909394    22.93   0.000     1.906639    2.263136
                               |
                        female |   .9988384   .0376714    26.51   0.000     .9249997    1.072677
                               |
                          race |
                         Black |  -.8156551    .054152   -15.06   0.000    -.9217972   -.7095131
                      Hispanic |  -.9807564   .0535761   -18.31   0.000     -1.08577    -.875743
                         Asian |  -1.003312   .0801447   -12.52   0.000    -1.160402   -.8462221
                         Other |   -.268419   .1286007    -2.09   0.037    -.5204863   -.0163518
                               |
                     education |
           No High School      |  -.1972228   .0742023    -2.66   0.008    -.3426651   -.0517805
           Some High           |  -.1886087   .0590976    -3.19   0.001    -.3044446   -.0727728
Some College/Tech School/AA degree |   .3914611   .0553689     7.07   0.000     .2829338    .4999884
           College             |    .620748   .0601121    10.33   0.000     .5029237    .7385722
           Advanced Degree     |   .8005605   .0778214    10.29   0.000     .6480244    .9530965
                               |
                           fpl |
                          Poor |   -.438328   .0685988    -6.39   0.000    -.5727868   -.3038691
                     Near Poor |  -.5005458    .08891    -5.63   0.000    -.6748162   -.3262753
                    Low Income |  -.3906937   .0633345    -6.17   0.000    -.5148343    -.266553
                 Middle Income |  -.3421434   .0503856    -6.79   0.000    -.4409031   -.2433837
                               |
                     insurance |
           Public Insurance    |   .2793372   .0736108     3.79   0.000     .1350544    .4236201
                  Uninsured    |  -2.116063   .0544856   -38.84   0.000    -2.222859   -2.009267
                   Medicare    |   .2240236   .0811173     2.76   0.006     .0650274    .3830198
                               |
                           msa |
                  Not in MSA   |  -.1132069   .0550936    -2.05   0.040    -.2211946   -.0052192
```

```
                            |
                     region |
                      South |  -.0651758   .0571114    -1.14   0.254    -.1771185    .0467669
                    Midwest |   .1210185   .0640297     1.89   0.059    -.0044847    .2465218
                       West |   -.195898   .0605089    -3.24   0.001       -.3145    -.077296
                            |
                      _cons |   4.858846   .0796552    61.00   0.000     4.702716    5.014977
------------------------------------------------------------------------------------------------

.
. imtest, white

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

         chi2(393)    =   1730.07
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |    1730.07    393    0.0000
            Skewness |    2800.08     29    0.0000
            Kurtosis |      16.71      1    0.0000
---------------------+-----------------------------
               Total |    4546.86    423    0.0000
---------------------------------------------------

. estat imtest, white

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

         chi2(393)    =   1730.07
         Prob > chi2  =    0.0000

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |    1730.07    393    0.0000
            Skewness |    2800.08     29    0.0000
            Kurtosis |      16.71      1    0.0000
---------------------+-----------------------------
               Total |    4546.86    423    0.0000
---------------------------------------------------

.
. * b. Scatterplot
. rvfplot, yline(0) msize(tiny) title("Residuals versus Fitted Values") ms(oh) jitter(1)

.
. * c. GQ Test
. * Procedure 1
. qui reg healthexp ib2.agecat female i.race ib3.education ib5.fpl ///
> i.insurance ib1.msa i.region i.healthstatus if msa == 1 & age08x > 17

.
. predict uhat, resid
(6306 missing values generated)

. sdtest uhat, by(female)

Variance ratio test
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    Male |   12811    112.8184     57.8895    6552.263   -.6536725    226.2904
  Female |   13949    23.99063    60.14802    7103.835   -93.90755    141.8888
---------+--------------------------------------------------------------------
combined |   26760    66.51576    41.84582    6845.342   -15.50425    148.5358
------------------------------------------------------------------------------
    ratio = sd(Male) / sd(Female)                              f =   0.8507
```

```
Ho: ratio = 1                              degrees of freedom = 12810, 13948

    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.0000         2*Pr(F < f) = 0.0000           Pr(F > f) = 1.0000

.
. * Procedure 2
. qui reg lnhealthexp i.healthstatus ib2.agecat female i.race ib3.education ib5.fpl ///
> i.insurance i.region if female == 0 & age08x > 17

.
. scalar rss0 = (e(rmse))^2

. scalar df_rss0 = e(df_r)

.
. qui reg lnhealthexp i.healthstatus ib3.agecat female i.race ib3.education ib5.fpl ///
> i.insurance i.region if female == 1 & age08x > 17

.
. scalar rss1 = (e(rmse))^2

. scalar df_rss1 = e(df_r)

.
. scalar FGQ = rss0/rss1

. scalar Fcrit = invFtail(df_rss1, df_rss0, .05)

. scalar pvalue = Ftail(df_rss1, df_rss0, FGQ)

. scalar list FGQ pvalue Fcrit
       FGQ =  1.2030727
    pvalue =  2.615e-21
     Fcrit =  1.0327935

.
.
. ** Number 3 **
. * Account for the survey nature of the data
. svyset varpsu [pweight = perwt08f], strata(varstr)

      pweight: perwt08f
          VCE: linearized
  Single unit: missing
     Strata 1: varstr
         SU 1: varpsu
        FPC 1: <zero>

.
. * a. Model Estimation
. * Linear Age
. xi: svy, subpop(if age08x>17): reg lnhealthexp i.healthstatus age female i.race ib3.education ///
> ib5.fpl i.insurance ib1.msa i.region
i.healthstatus    _Ihealthsta_1-5     (naturally coded; _Ihealthsta_1 omitted)
i.race            _Irace_1-5          (naturally coded; _Irace_1 omitted)
i.insurance       _Iinsurance_1-4     (naturally coded; _Iinsurance_1 omitted)
i.region          _Iregion_1-4        (naturally coded; _Iregion_1 omitted)
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =       165                Number of obs      =      31294
Number of PSUs     =       370                Population size    =  284700522
                                              Subpop. no. of obs =      20143
                                              Subpop. size       =  207688790
                                              Design df          =        205
                                              F(  26,    180)    =     282.90
                                              Prob > F           =     0.0000
                                              R-squared          =     0.3355

------------------------------------------------------------------------------------------------
                             |             Linearized
                  lnhealthexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------------------+-----------------------------------------------------------------
                _Ihealthsta_2 |   .5121225   .0631533     8.11   0.000     .3876093    .6366357
```

```
                    _Ihealthsta_3 |   .8881314   .0666777    13.32   0.000     .7566695    1.019593
                    _Ihealthsta_4 |   1.934006   .0743736    26.00   0.000     1.787371    2.080642
                    _Ihealthsta_5 |    2.87649   .0968474    29.70   0.000     2.685545    3.067434
                              age |   .0420615   .0014748    28.52   0.000     .0391537    .0449692
                           female |   .9257583   .0403825    22.92   0.000       .84614    1.005377
                          _Irace_2 |  -.8989643   .0706027   -12.73   0.000    -1.038165   -.7597638
                          _Irace_3 |  -1.044754   .0759093   -13.76   0.000    -1.194417   -.8950909
                          _Irace_4 |  -1.099867   .1027103   -10.71   0.000    -1.302371    -.897363
                          _Irace_5 |  -.1970117   .2182564    -0.90   0.368    -.6273268    .2333033
                                   |
                         education |
                       No High School  |  -.1716976     .10463    -1.64   0.102    -.3779865    .0345913
                          Some High  |  -.0430418   .0792738    -0.54   0.588    -.1993382    .1132547
Some College/Tech School/AA degree  |   .4077452   .0596733     6.83   0.000     .2900932    .5253973
                            College  |    .541613   .0616112     8.79   0.000     .4201402    .6630859
                    Advanced Degree  |    .635801    .081976     7.76   0.000     .4741768    .7974252
                                   |
                               fpl |
                              Poor  |   -.454575   .0915002    -4.97   0.000    -.6349771   -.2741729
                         Near Poor  |  -.4563592   .1058417    -4.31   0.000    -.6650371   -.2476814
                        Low Income  |  -.3608722   .0871577    -4.14   0.000    -.5327126   -.1890317
                     Middle Income  |  -.2858143    .058817    -4.86   0.000    -.4017782   -.1698504
                                   |
                    _Iinsurance_2 |   .3670992   .1055246     3.48   0.001     .1590465     .575152
                    _Iinsurance_3 |  -2.097906   .0910845   -23.03   0.000    -2.277489   -1.918324
                    _Iinsurance_4 |   .1866515   .0736963     2.53   0.012     .0413515    .3319514
                                   |
                               msa |
                      Not in MSA  |  -.1973134   .0542497    -3.64   0.000    -.3042723   -.0903545
                                   |
                      _Iregion_2 |  -.0389674    .067737    -0.58   0.566    -.1725179    .0945831
                      _Iregion_3 |   .0703673   .0760304     0.93   0.356    -.0795344     .220269
                      _Iregion_4 |  -.1379579   .0695311    -1.98   0.049    -.2750457    -.00087
                           _cons |   3.704508   .1195687    30.98   0.000     3.468766    3.940251
------------------------------------------------------------------------------------------------------

.
. eststo linearage

.
. * Quadratic Age
. gen agesq = age^2
(251 missing values generated)

.
. xi: svy, subpop(if age08x>17): reg lnhealthexp i.healthstatus age agesq female i.race ib3.education
///
> ib5.fpl i.insurance ib1.msa i.region
i.healthstatus    _Ihealthsta_1-5     (naturally coded; _Ihealthsta_1 omitted)
i.race            _Irace_1-5          (naturally coded; _Irace_1 omitted)
i.insurance       _Iinsurance_1-4     (naturally coded; _Iinsurance_1 omitted)
i.region          _Iregion_1-4        (naturally coded; _Iregion_1 omitted)
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =       165                Number of obs      =      31294
Number of PSUs     =       370                Population size    =  284700522
                                              Subpop. no. of obs =      20143
                                              Subpop. size       =  207688790
                                              Design df          =        205
                                              F(  27,    179)    =     271.18
                                              Prob > F           =     0.0000
                                              R-squared          =     0.3360

------------------------------------------------------------------------------------------------------
                                   |             Linearized
                       lnhealthexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------------------------+------------------------------------------------------------------
                    _Ihealthsta_2 |   .5203971   .0629957     8.26   0.000     .3961945    .6445996
                    _Ihealthsta_3 |   .9038524   .0669907    13.49   0.000     .7717732    1.035932
                    _Ihealthsta_4 |   1.957562   .0752026    26.03   0.000     1.809292    2.105832
                    _Ihealthsta_5 |   2.909562   .0982181    29.62   0.000     2.715914    3.103209
                              age |   .0195293   .0064392     3.03   0.003     .0068338    .0322247
                            agesq |    .000235   .0000621     3.79   0.000     .0001126    .0003574
                           female |   .9235824   .0403528    22.89   0.000     .8440227    1.003142
```

```
                        _Irace_2 |  -.8881191   .0707823   -12.55   0.000    -1.027674   -.7485646
                        _Irace_3 |  -1.032648   .0758497   -13.61   0.000    -1.182193   -.8831023
                        _Irace_4 |   -1.09242   .1027163   -10.64   0.000    -1.294936   -.8899038
                        _Irace_5 |  -.1895165   .2178451    -0.87   0.385    -.6190207    .2399877
                                 |
                       education |
                     No High School  |  -.1845797   .1039732    -1.78   0.077    -.3895735    .0204141
                          Some High  |  -.0826262   .0786321    -1.05   0.295    -.2376576    .0724052
Some College/Tech School/AA degree  |   .4037511   .0595959     6.77   0.000     .2862516    .5212505
                           College  |   .5569071   .0618218     9.01   0.000      .435019    .6787952
                   Advanced Degree  |    .656239   .0827451     7.93   0.000     .4930984    .8193796
                                 |
                               fpl |
                              Poor  |  -.4680493   .0916665    -5.11   0.000    -.6487793   -.2873194
                         Near Poor  |  -.4758413   .1067746    -4.46   0.000    -.6863585   -.2653241
                        Low Income  |  -.3748723   .0877005    -4.27   0.000    -.5477829   -.2019617
                     Middle Income  |  -.2942635   .0590667    -4.98   0.000    -.4107197   -.1778073
                                 |
                      _Iinsurance_2 |   .3650266   .1052116     3.47   0.001      .157591    .5724621
                      _Iinsurance_3 |  -2.086658    .090949   -22.94   0.000    -2.265973   -1.907343
                      _Iinsurance_4 |   .1041032    .074704     1.39   0.165    -.0431834    .2513898
                                 |
                               msa |
                        Not in MSA  |  -.1947654   .0542234    -3.59   0.000    -.3016724   -.0878583
                                 |
                        _Iregion_2 |  -.0377599   .0677795    -0.56   0.578    -.1713942    .0958745
                        _Iregion_3 |   .0710205   .0756644     0.94   0.349    -.0781597    .2202006
                        _Iregion_4 |  -.1383062    .069328    -1.99   0.047    -.2749936   -.0016188
                             _cons |   4.168979   .1738293    23.98   0.000     3.826257    4.511702
------------------------------------------------------------------------------------------------------

.
. eststo quadage

.
. * Spline function with agecat
. mkspline age1 25 age2 45 age3 65 age4 75 age5 = age, marginal

.
. xi: svy, subpop(if age08x>17): reg lnhealthexp i.healthstatus age1-age5 female i.race ib3.education
///
> ib5.fpl i.insurance ib1.msa i.region
i.healthstatus    _Ihealthsta_1-5     (naturally coded; _Ihealthsta_1 omitted)
i.race            _Irace_1-5          (naturally coded; _Irace_1 omitted)
i.insurance       _Iinsurance_1-4     (naturally coded; _Iinsurance_1 omitted)
i.region          _Iregion_1-4        (naturally coded; _Iregion_1 omitted)
(running regress on estimation sample)

Survey: Linear regression

Number of strata   =       165                Number of obs      =      31294
Number of PSUs     =       370                Population size    =  284700522
                                              Subpop. no. of obs =      20143
                                              Subpop. size       =  207688790
                                              Design df          =        205
                                              F(  30,    176)    =     242.78
                                              Prob > F           =     0.0000
                                              R-squared          =     0.3389

------------------------------------------------------------------------------------------------------
                                |             Linearized
                     lnhealthexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------------------------+-----------------------------------------------------------------
                      _Ihealthsta_2 |   .5317394   .0632209     8.41   0.000     .4070929    .6563859
                      _Ihealthsta_3 |   .9128047   .0664667    13.73   0.000     .7817587    1.043851
                      _Ihealthsta_4 |   1.978117   .0750548    26.36   0.000     1.830139    2.126095
                      _Ihealthsta_5 |   2.916667    .099027    29.45   0.000     2.721425    3.111909
                               age1 |  -.0695871     .02081    -3.34   0.001    -.1106162   -.0285581
                               age2 |   .1095697   .0240925     4.55   0.000     .0620688    .1570707
                               age3 |   .0223183   .0077093     2.90   0.004     .0071187    .0375179
                               age4 |   -.031034   .0133733    -2.32   0.021    -.0574009   -.0046671
                               age5 |   -.011556   .0227862    -0.51   0.613    -.0564814    .0333694
                             female |   .9309174   .0402365    23.14   0.000     .8515869    1.010248
                          _Irace_2 |  -.8871511   .0713018   -12.44   0.000    -1.02773   -.7465723
                          _Irace_3 |  -1.000685   .0759549   -13.17   0.000    -1.150438   -.8509321
                          _Irace_4 |   -1.08567   .1023312   -10.61   0.000    -1.287427   -.8839139
```

```
                          _Irace_5 |  -.1930708   .2159089    -0.89   0.372    -.6187575    .2326158
                                   |
                         education |
                       No High School  |  -.1878617   .1025405    -1.83   0.068    -.3900309    .0143075
                          Some High  |  -.2141338   .0763699    -2.80   0.006     -.364705   -.0635625
Some College/Tech School/AA degree  |   .4158075   .0598542     6.95   0.000     .2977989    .5338162
                             College  |   .6121623    .063724     9.61   0.000     .4865238    .7378008
                    Advanced Degree  |   .6881745   .0833713     8.25   0.000     .5237993    .8525496
                                   |
                               fpl |
                              Poor  |  -.4257452   .0931445    -4.57   0.000    -.6093891   -.2421012
                         Near Poor  |  -.4191787   .1084071    -3.87   0.000    -.6329145    -.205443
                        Low Income  |  -.3315869   .0890619    -3.72   0.000    -.5071816   -.1559922
                     Middle Income  |  -.2570264   .0593423    -4.33   0.000     -.374026   -.1400269
                                   |
                       _Iinsurance_2 |   .3357545   .1061752     3.16   0.002     .1264192    .5450898
                       _Iinsurance_3 |  -2.074129   .0902156   -22.99   0.000    -2.251998   -1.896259
                       _Iinsurance_4 |   .1296172   .0771944     1.68   0.095    -.0225795     .281814
                                   |
                                  msa |
                         Not in MSA  |  -.2073151   .0539536    -3.84   0.000    -.3136902   -.1009399
                                   |
                          _Iregion_2 |  -.0450062   .0679355    -0.66   0.508    -.1789481    .0889356
                          _Iregion_3 |    .067955   .0757348     0.90   0.371    -.0813639     .217274
                          _Iregion_4 |  -.1467433   .0695944    -2.11   0.036    -.2839559   -.0095307
                               _cons |   6.318828   .4580268    13.80   0.000      5.41578    7.221875
------------------------------------------------------------------------------------------------------

.
. eststo splage

.
. * b. Tables
. esttab linearage quadage splage, se r2 keep(age agesq age1 age2 age3 age4 age5) ///
> mtitles("Linear Age" "Quad Age" "Spline Age") ///
> title(TABLE 1. Coefficients for Age Variables)

TABLE 1. Coefficients for Age Variables
------------------------------------------------------------
                      (1)             (2)             (3)
               Linear Age        Quad Age      Spline Age
------------------------------------------------------------
age               0.0421***       0.0195**
                 (0.00147)       (0.00644)

agesq                             0.000235***
                                 (0.0000621)

age1                                             -0.0696***
                                                 (0.0208)

age2                                               0.110***
                                                 (0.0241)

age3                                              0.0223**
                                                (0.00771)

age4                                             -0.0310*
                                                 (0.0134)

age5                                             -0.0116
                                                 (0.0228)
------------------------------------------------------------
N                   31294           31294           31294
R-sq                0.335           0.336           0.339
------------------------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

.
end of do-file

. log close
      name:  <unnamed>
       log:  /Users/jojo/Documents/JHU/TA Folder/Advanced HSR Methods -
Analysis/Answers/Assignment2_2016.lo
```

```
> g
  log type:  text
 closed on:  16 Sep 2016, 08:22:23
-------------------------------------------------------------------------------------------------------
-----
```