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 8, 2016

Homework Exercise #1
Answer Guide

1. List the five assumptions of the Classical Linear Model. Describe them briefly. Give a real world example of a violation of two of these assumptions.
    a. Linearity in parameters.

        The dependent variable can be calculated as a linear function of a specific set of independent variables, plus a disturbance term; a linear relationship exists between the dependent variable and the regressors. The unknown coefficients of this linear function form the vector $\boldsymbol{\beta}$ and are assumed to be constants.
    b. Full rank or no perfect collinearity.

        This implies that there are no exact linear relationships between the independent variables and that the number of observations is greater than the number of independent variables.

        In mathematical terms, each of the rows and columns of the regression matrix are linearly independent. This implies that none of the variables are multiples of each other, i.e., they have no exact linear relationships.
    c. Exogeneity of the independent variables: Expected Value of Error Term is Zero. $E[\varepsilon_i \mid \boldsymbol{X}] = 0$.

        *This states that the expected value of the disturbance term or error is zero. This assumption also implies that none of the independent variables in **X** is systematically related to ε_i. Thus, none of the independent variables give information on what ε_i would be. This is also the zero conditional mean assumption.*
    d. Homoscedasticity and nonautocorrelation.

        *Each error term ε_i has the same variance σ^2 (homoscedasticity) and does not depend on any independent variable in **X**. In addition, each error term ε_i is not correlated with other error terms ε_j (nonautocorrelation); error terms are not correlated with each other.*
    e. Exogenously generated data.

        This draws on the assumptions in (c) and (d). The data may be any mixture of constants and random variables as long as they are generated in a manner that is not systematically correlated to the disturbance or error term.

2. Review the attached regression output and answer the question below. This is a model where the dependent variable is total health expenditures for adults using data from the MEPS.
    a. How much the variation does the model explain?

        The model explains 20.5% of the variation since the R-squared is 0.2050.
    b. What is the association uninsured status and health expenditures?

        Individuals who are uninsured have $1618.61 lower annual health expenditures compared to individuals who are privately insured, holding all other things constant. This is statistically significant.

c. What is the association between being female and health expenditures?

Individuals who are female have $709.55 higher annual health expenditures compared to individuals who are male, holding all other things constant. This is also statistically significant.

d. Describe the race/ethnic disparity in health expenditures.

Compared to white individuals, the following racial/ethnic groups have less health expenditures: black, Hispanic, and Asian. Specifically, individuals who are black, Hispanic, Asian, and others have $942.06, $893.20, and $1967.14 lower health expenditures compared to white individuals. Those who belong to other non-black, non-Hispanic, or non-Asian racial/ethnic groups on the other hand have $380.66 higher health expenditures compared to white individuals. All are statistically significant.

e. Interpret the joint significance tests for health status and poverty status.

The joint significance test for health status tests the hypothesis that all the coefficients are zero, i.e., that the collective contribution of the coefficients for health status is statistically significant. We see that the p-value is $p < 0.001$ and thus we reject the null hypothesis that all the coefficients are zero. Thus, the model as a whole with health status is statistically significant. On the other hand, the p-value for the F-test involving poverty status is $0.5670 > 0.05$. This implies that the collective contribution of the coefficients of the poverty status to the model is not significant, i.e., the model as a whole with only poverty status is not statistically significant.

3. Estimate a model of health expenditures using the log of health expenditures as the dependent variable. Include only adults, persons age 18 and over in the sample. Exclude persons with expenditures that are less the zero and greater than $100,000. Note you will have to create a variable that identifies the adult sample. Use the subpop command with the svy procedures to estimate the models correctly in Stata. Use the following model specification:

   a. Log(Total health expenditures+1) = f(age, race/ethnicity, gender, msa, region, education, poverty status, insurance status, and health status) note for the categorical variables you must designate reference categories for the categorical variables. *Use the following reference categories: age category 25-44, male, white non-Hispanic, HS diploma/GED, high income, privately insured, MSA, East, excellent health)*
   b. Produce 2 tables
      i. Report the means and standard deviations of the dependent and independent variables.

	Mean (SE)
lnhealthexp	6.224 (0.035)
	Proportion
Age Category	
18-24	0.12
25-44	0.36
45-64	0.35

65-74	0.09
75+	0.08
Gender	
Male	0.48
Female	0.52
Race	
White	0.68
Black	0.12
Hispanic	0.14
Asian	0.04
Other	0.02
Education	
No HS	0.05
Some HS	0.12
HS	0.36
Some College	0.18
College	0.18
Advanced	0.1
Insurance	
Private	0.68
Public	0.06
Uninsured	0.15
Medicare	0.1
MSA	
Not in MSA	0.16
In MSA	0.84
Region	
North East	0.18
South	0.37
Midwest	0.22
West	0.23
Health Status	
Excellent	0.25
Very Good	0.34
Good	0.28
Fair	0.1
Poor	0.03
FPL	
Poor	0.11

Near Poor	0.04
Low Income	0.13
Middle Income	0.3
High Income	0.41

ii. Report the coefficients, standard errors, and p-values.

	est1		
	b	se	p
_Iagecat_1	-.1757938	.0829758	.0353271
_Iagecat_3	.939058	.0455355	6.79e-52
_Iagecat_4	1.642337	.08392	8.54e-49
_Iagecat_5	1.991141	.0790059	1.03e-64
female	.9314282	.0404729	9.97e-59
_Irace_2	-.9058272	.0705141	4.26e-28
_Irace_3	-1.039848	.0750837	3.18e-31
_Irace_4	-1.121758	.1026713	3.40e-22
_Irace_5	-.2412625	.2172839	.2681462
_Ieducatio~1	-.1468646	.1026121	.1538784
_Ieducatio~2	-.1499418	.0777491	.0551699
_Ieducatio~4	.3887705	.0602822	7.93e-10
_Ieducatio~5	.5896828	.0632943	1.94e-17
_Ieducatio~6	.6938485	.0853632	4.07e-14
_Iinsuranc~2	.3123708	.1040915	.0030256
_Iinsuranc~3	-2.121578	.0912869	2.32e-59
_Iinsuranc~4	.1976065	.0807481	.0152391
_Ifpl_1	-.4593343	.0921327	1.31e-06
_Ifpl_2	-.4543905	.1069581	.0000327
_Ifpl_3	-.3603304	.0876661	.0000571
_Ifpl_4	-.2829634	.058658	2.74e-06
msa	.1847822	.0542062	.0007853
_Iregion_2	-.0420186	.0668074	.530081
_Iregion_3	.0594148	.0753337	.4312051
_Iregion_4	-.144537	.0696169	.0391246
_Ihealthst~2	.5423581	.0635283	3.08e-15
_Ihealthst~3	.9417198	.0669778	6.97e-32
_Ihealthst~4	2.005296	.0748377	6.63e-69
_Ihealthst~5	2.970561	.0982248	1.64e-77
_cons	4.829568	.1103436	5.7e-106

- Note: Categories defined/assigned as in the provided base program.

c. What is the estimated relationship between health expenditures and insurance status and race/ethnicity?

Based on the regression, the total health expenditures of uninsured individuals in 2008 is 212% less than that of privately insured individuals, holding all other things constant. The

coefficients for public insurance and Medicare (4.insurance) can be interpreted the same way – that their expenditures are 31.2% and 19.8% higher than privately insured individuals. All are found to be statistically significant.

For race, the coefficients on "Black", "Hispanic", and "Asian" are statistically significant. Compared to white/Caucasian individuals, Black, Hispanic, and Asian individuals have 90.6%, 104%, and 112.2% less total health expenditures in 2008, respectively, again holding all other things constant.

d. Compute and interpret an F-test for insurance status and race/ethnicity.

```
. test _Iinsurance_2 _Iinsurance_3 _Iinsurance_4

Adjusted Wald test

 ( 1)  _Iinsurance_2 = 0
 ( 2)  _Iinsurance_3 = 0
 ( 3)  _Iinsurance_4 = 0

       F(  3,   203) =  219.86
            Prob > F =    0.0000

. test _Irace_2 _Irace_3 _Irace_4 _Irace_5

Adjusted Wald test

 ( 1)  _Irace_2 = 0
 ( 2)  _Irace_3 = 0
 ( 3)  _Irace_4 = 0
 ( 4)  _Irace_5 = 0

       F(  4,   202) =   84.39
            Prob > F =    0.0000
```

We see that the F-statistics for both tests are statistically significant. Thus we reject the null hypotheses that a) the coefficients on health insurance status other than private insurance are all zero, and that b) the coefficients on race other than white/Caucasian are all zero. The collective contributions of the coefficients on health status and poverty status are statistically significant. Thus the models as a whole with insurance status and race are statistically significant.

e. Why did we add $1 to total health expenditures in the log model?

Our expenditure data contains $0 entries and the log 0 is undefined. Adding $1 to expenditures allows us to make calculations that include data from individuals who report no/zero expenditures.

f. In your opinion, which model better explains the health expenditures, my model using health expenditures or your model using the log of health expenditures? Explain?

The model using the log of health expenditures seems more appropriate our setting. Our data follows a right-skewed distribution and the log transformation allows our data to be

more normally approximated. In addition, the log expenditure model explains more of the variation compared to the level expenditure model as seen by its higher R-squared.

In a pdf format, in addition to your tables and answers to please hand in a clean versus of your program and log file.

Survey: Linear regression

Number of strata	= 165	Number of obs	= 31155
Number of PSUs	= 370	Population size	= 283098674
		Subpop. no. of obs	= 20004
		Subpop. size	= 206086942
		Design df	= 205
		F(29, 177)	= 62.87
		Prob > F	= 0.0000
		R-squared	= 0.2050

Linearized	Coef.	Std. Err.	t	P>t
Age (25-44 ref)				
18-24	-9.36	124.56	-0.08	0.94
45-64	1163.68	112.15	10.38	0.00
65-74	2768.71	283.56	9.76	0.00
over 75	4429.24	367.20	12.06	0.00
Female	709.55	109.24	6.50	0.00
Race (White ref)				
Black	-942.06	139.52	-6.75	0.00
Hispanic	-893.20	142.49	-6.27	0.00
Asian	-1967.14	143.23	-13.73	0.00
Other	380.66	550.64	0.69	0.49
Education (HS ref)				
No High School	-546.60	227.06	-2.41	0.02
Some High School	-238.16	202.24	-1.18	0.24
Some College and AA Degree	477.57	170.40	2.80	0.01
College Degree	694.50	185.61	3.74	0.00
Advanced Degree	904.44	223.24	4.05	0.00
Insurance (Private ref)				
Medicaid or Other Public	404.84	256.23	1.58	0.12
Uninsured	-1618.61	108.61	-14.90	0.00
Medicare	1290.76	341.42	3.78	0.00

Poverty (High Income ref)				
Poor	-282.86	217.36	-1.30	0.20
Near Poor	95.53	350.19	0.27	0.79
Low Income	-236.80	203.85	-1.16	0.25
Middle Income	-160.14	137.39	-1.17	0.25
MSA	612.69	130.84	4.68	0.00
Region (NE ref)				
South	-263.57	202.07	-1.30	0.19
MidWest	-10.41	216.52	-0.05	0.96
West	-360.08	207.02	-1.74	0.08
Health Status (Excellent ref)				
Very Good	698.75	112.51	6.21	0.00
Good	1857.17	115.38	16.10	0.00
Fair	4399.07	249.36	17.64	0.00
Poor	9690.40	526.69	18.40	0.00
_cons	745.41	232.44	3.21	0.00

. test _Ihealthsta_2 _Ihealthsta_3 _Ihealthsta_4 _Ihealthsta_5

Adjusted Wald test

(1) _Ihealthsta_2 = 0 (very good)
(2) _Ihealthsta_3 = 0 (good)
(3) _Ihealthsta_4 = 0 (fair)
(4) _Ihealthsta_5 = 0 (poor)

F(4, 202) = 198.68
Prob > F = 0.0000

. test _Ifpl_1 _Ifpl_2 _Ifpl_3 _Ifpl_4

Adjusted Wald test

(1) _Ifpl_1 = 0 (poor)
(2) _Ifpl_2 = 0 (near poor)
(3) _Ifpl_3 = 0 (low income)
(4) _Ifpl_4 = 0 (middle income)

F(4, 202) = 0.74
Prob > F = 0.5670

Log file

```
-------------------------------------------------------------------------------
----------------------
      name:  <unnamed>
       log:  /Users/jojo/Documents/JHU/TA Folder/Advanced HSR Methods -
Analysis/Answers/Assignment1_2016.lo
> g
  log type:  text
 opened on:   8 Sep 2016, 08:41:10

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

.
.
. ** Data preparation
. * 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>

.
. * 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 "0-24" 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

.
. ** 3a, 3b, & 3c **
. gen lnhealthexp = ln(1+healthexp)
(42 missing values generated)

.
. char agecat[omit] 2

. char race[omit] 1

. char healthstatus[omit] 1

. char education[omit] 3

. char insurance[omit] 1

. char region[omit] 1

. char fpl[omit] 5

.
. eststo: xi: svy, subpop(if age08x>17): reg lnhealthexp i.agecat female i.race
i.education i.insurance i.fp
> l msa i.region i.healthstatus
i.agecat          _Iagecat_1-5        (naturally coded; _Iagecat_2 omitted)
i.race            _Irace_1-5          (naturally coded; _Irace_1 omitted)
i.education       _Ieducation_1-6     (naturally coded; _Ieducation_3 omitted)
i.insurance       _Iinsurance_1-4     (naturally coded; _Iinsurance_1 omitted)
i.fpl             _Ifpl_1-5           (naturally coded; _Ifpl_5 omitted)
i.region          _Iregion_1-4        (naturally coded; _Iregion_1 omitted)
i.healthstatus    _Ihealthsta_1-5     (naturally coded; _Ihealthsta_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(  29,    177)    =     252.34
                                              Prob > F           =     0.0000
                                              R-squared          =     0.3309
```

```
------------------------------------------------------------------------------
             |             Linearized
  lnhealthexp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _Iagecat_1 |  -.1757938   .0829758    -2.12   0.035    -.3393892   -.0121983
   _Iagecat_3 |    .939058   .0455355    20.62   0.000     .8492801    1.028836
   _Iagecat_4 |   1.642337     .08392    19.57   0.000      1.47688    1.807794
   _Iagecat_5 |   1.991141   .0790059    25.20   0.000     1.835373    2.146909
       female |   .9314282   .0404729    23.01   0.000     .8516316    1.011225
     _Irace_2 |  -.9058272   .0705141   -12.85   0.000    -1.044853   -.7668013
     _Irace_3 |  -1.039848   .0750837   -13.85   0.000    -1.187883    -.891813
     _Irace_4 |  -1.121758   .1026713   -10.93   0.000    -1.324185   -.9193307
     _Irace_5 |  -.2412625   .2172839    -1.11   0.268    -.6696601    .1871351
_Ieducation_1 |  -.1468646   .1026121    -1.43   0.154     -.349175    .0554458
_Ieducation_2 |  -.1499418   .0777491    -1.93   0.055    -.3032321    .0033486
_Ieducation_4 |   .3887705   .0602822     6.45   0.000     .2699179    .5076231
_Ieducation_5 |   .5896828   .0632943     9.32   0.000     .4648915    .7144742
_Ieducation_6 |   .6938485   .0853632     8.13   0.000     .5255462    .8621508
_Iinsurance_2 |   .3123708   .1040915     3.00   0.003     .1071436     .517598
_Iinsurance_3 |  -2.121578   .0912869   -23.24   0.000     -2.30156   -1.941597
_Iinsurance_4 |   .1976065   .0807481     2.45   0.015     .0384032    .3568098
      _Ifpl_1 |  -.4593343   .0921327    -4.99   0.000    -.6409835   -.2776852
      _Ifpl_2 |  -.4543905   .1069581    -4.25   0.000    -.6652694   -.2435116
      _Ifpl_3 |  -.3603304   .0876661    -4.11   0.000    -.5331731   -.1874877
      _Ifpl_4 |  -.2829634    .058658    -4.82   0.000    -.3986138    -.167313
          msa |   .1847822   .0542062     3.41   0.001      .077909    .2916554
   _Iregion_2 |  -.0420186   .0668074    -0.63   0.530    -.1737363    .0896991
   _Iregion_3 |   .0594148   .0753337     0.79   0.431    -.0891133    .2079429
   _Iregion_4 |   -.144537   .0696169    -2.08   0.039    -.2817938   -.0072802
_Ihealthsta_2 |   .5423581   .0635283     8.54   0.000     .4171055    .6676107
_Ihealthsta_3 |   .9417198   .0669778    14.06   0.000     .8096662    1.073773
_Ihealthsta_4 |   2.005296   .0748377    26.80   0.000     1.857745    2.152846
_Ihealthsta_5 |   2.970561   .0982248    30.24   0.000     2.776901    3.164222
        _cons |   4.829568   .1103436    43.77   0.000     4.612014    5.047122
------------------------------------------------------------------------------
(est1 stored)

. estout, cells("b se p")

----------------------------------------------------
                      est1
                         b           se           p
----------------------------------------------------
_Iagecat_1        -.1757938    .0829758    .0353271
_Iagecat_3          .939058    .0455355    6.79e-52
_Iagecat_4         1.642337      .08392    8.54e-49
_Iagecat_5         1.991141    .0790059    1.03e-64
female             .9314282    .0404729    9.97e-59
_Irace_2          -.9058272    .0705141    4.26e-28
_Irace_3          -1.039848    .0750837    3.18e-31
_Irace_4          -1.121758    .1026713    3.40e-22
_Irace_5          -.2412625    .2172839    .2681462
_Ieducatio~1      -.1468646    .1026121    .1538784
_Ieducatio~2      -.1499418    .0777491    .0551699
_Ieducatio~4       .3887705    .0602822    7.93e-10
_Ieducatio~5       .5896828    .0632943    1.94e-17
_Ieducatio~6       .6938485    .0853632    4.07e-14
_Iinsuranc~2       .3123708    .1040915    .0030256
_Iinsuranc~3      -2.121578    .0912869    2.32e-59
_Iinsuranc~4       .1976065    .0807481    .0152391
_Ifpl_1           -.4593343    .0921327    1.31e-06
_Ifpl_2           -.4543905    .1069581    .0000327
_Ifpl_3           -.3603304    .0876661    .0000571
```

```
_Ifpl_4         -.2829634      .058658     2.74e-06
msa              .1847822     .0542062     .0007853
_Iregion_2      -.0420186     .0668074      .530081
_Iregion_3       .0594148     .0753337     .4312051
_Iregion_4       -.144537     .0696169     .0391246
_Ihealthst~2     .5423581     .0635283     3.08e-15
_Ihealthst~3     .9417198     .0669778     6.97e-32
_Ihealthst~4     2.005296     .0748377     6.63e-69
_Ihealthst~5     2.970561     .0982248     1.64e-77
_cons            4.829568     .1103436     5.7e-106
------------------------------------------------------

.
. test _Iinsurance_2 _Iinsurance_3 _Iinsurance_4

Adjusted Wald test

 ( 1)  _Iinsurance_2 = 0
 ( 2)  _Iinsurance_3 = 0
 ( 3)  _Iinsurance_4 = 0

       F(  3,   203) =  219.86
            Prob > F =    0.0000

. test _Irace_2 _Irace_3 _Irace_4 _Irace_5

Adjusted Wald test

 ( 1)  _Irace_2 = 0
 ( 2)  _Irace_3 = 0
 ( 3)  _Irace_4 = 0
 ( 4)  _Irace_5 = 0

       F(  4,   202) =   84.39
            Prob > F =    0.0000

.
. * Means of dependent and independent variables
. svy, subpop(if age08x>17): mean lnhealthexp
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =     165       Number of obs    =      33034
Number of PSUs   =     370       Population size  =  303998491
                                 Subpop. no. obs  =      21883
                                 Subpop. size     =  226986759
                                 Design df        =        205

--------------------------------------------------------------
             |             Linearized
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
 lnhealthexp |   6.223972   .0353434      6.154289    6.293656
--------------------------------------------------------------

. svy, subpop(if age08x>17): tabulate agecat, count cell format(%11.0g) /*format
intented only for count sin
> ce it was displaying things in scientific notation*/
(running tabulate on estimation sample)

Number of strata   =       165                 Number of obs      =      33066
Number of PSUs     =       370                 Population size    =  304375942
```

```
                                                Subpop. no. of obs =      21915
                                                Subpop. size       =  227364210
                                                Design df          =        205

------------------------------------
   agecat |       count  proportions
----------+-------------------------
     0-24 |  28348199.6   .124681891
    25-44 |  82027707.1    .36077669
    45-64 |  78804391.5   .346599808
    65-74 |  20217269.8   .088920195
      75+ |    17966642   .079021417
          |
    Total |   227364210           1
------------------------------------
  Key:  count     =  weighted counts
        propor~s  =  cell proportions

. svy, subpop(if age08x>17): tabulate female, count cell format(%11.0g)
(running tabulate on estimation sample)

Number of strata   =       165                 Number of obs      =      33066
Number of PSUs     =       370                 Population size    =  304375942
                                                Subpop. no. of obs =      21915
                                                Subpop. size       =  227364210
                                                Design df          =        205

------------------------------------
   female |       count  proportions
----------+-------------------------
     Male |   110026855   .483923369
   Female |   117337355   .516076631
          |
    Total |   227364210           1
------------------------------------
  Key:  count     =  weighted counts
        propor~s  =  cell proportions

. svy, subpop(if age08x>17): tabulate race, count cell format(%11.0g)
(running tabulate on estimation sample)

Number of strata   =       165                 Number of obs      =      33066
Number of PSUs     =       370                 Population size    =  304375942
                                                Subpop. no. of obs =      21915
                                                Subpop. size       =  227364210
                                                Design df          =        205

------------------------------------
     race |       count  proportions
----------+-------------------------
    White |   155201311      .682611
    Black |  26181246.4   .115151133
 Hispanic |  31119179.4   .136869296
    Asian |  10001745.3   .043989972
    Other |  4860728.17   .021378599
          |
    Total |   227364210           1
------------------------------------
  Key:  count     =  weighted counts
        propor~s  =  cell proportions

. svy, subpop(if age08x>17): tabulate education, count cell format(%11.0g)
(running tabulate on estimation sample)
```

```
Number of strata   =       165                Number of obs      =      31118
Number of PSUs     =       370                Population size    =  285212314
                                              Subpop. no. of obs =      20192
                                              Subpop. size       =  208200582
                                              Design df          =        205

------------------------------------
education |       count  proportions
----------+-------------------------
  No High |  11416488.2   .054834084
 Some Hig |  24807306.3   .119150994
 High Sch |  74554393.6   .358089266
 Some Col |  37948072.5   .182266889
  College |  38450429.5    .18467974
 Advanced |  21023892.1   .100979026
          |
    Total |   208200582           1
------------------------------------
  Key:  count     =  weighted counts
        propor~s  =  cell proportions

. svy, subpop(if age08x>17): tabulate insurance, count cell format(%11.0g)
(running tabulate on estimation sample)

Number of strata   =       165                Number of obs      =      33066
Number of PSUs     =       370                Population size    =  304375942
                                              Subpop. no. of obs =      21915
                                              Subpop. size       =  227364210
                                              Design df          =        205

------------------------------------
insurance |       count  proportions
----------+-------------------------
  Private |   155497516   .683913778
 Public I |  14425814.7   .063448045
 Uninsure |  34793985.4   .153031937
 Medicare |    22646894    .09960624
          |
    Total |   227364210           1
------------------------------------
  Key:  count     =  weighted counts
        propor~s  =  cell proportions

. svy, subpop(if age08x>17): tabulate msa, count cell format(%11.0g)
(running tabulate on estimation sample)

Number of strata   =       165                Number of obs      =      33066
Number of PSUs     =       370                Population size    =  304375942
                                              Subpop. no. of obs =      21915
                                              Subpop. size       =  227364210
                                              Design df          =        205

------------------------------------
      msa |       count  proportions
----------+-------------------------
 Not in M |    36490058   .160491654
   In MSA |   190874152   .839508346
          |
    Total |   227364210           1
------------------------------------
  Key:  count     =  weighted counts
        propor~s  =  cell proportions
```

```
. svy, subpop(if age08x>17): tabulate region, count cell format(%11.0g)
(running tabulate on estimation sample)

Number of strata   =       165                Number of obs      =      33066
Number of PSUs     =       370                Population size    =  304375942
                                              Subpop. no. of obs =      21915
                                              Subpop. size       =  227364210
                                              Design df          =        205

------------------------------------
   region |       count  proportions
----------+-------------------------
 North Ea |    41800038   .183846165
    South |  83005841.2   .365078748
  Midwest |  49882055.9   .219392735
     West |  52676274.8   .231682352
          |
    Total |   227364210           1
------------------------------------
  Key:  count     =  weighted counts
        propor~s  =  cell proportions

. svy, subpop(if age08x>17): tabulate healthstatus, count cell format(%11.0g)
(running tabulate on estimation sample)

Number of strata   =       165                Number of obs      =      32719
Number of PSUs     =       370                Population size    =  304183512
                                              Subpop. no. of obs =      21894
                                              Subpop. size       =  227171780
                                              Design df          =        205

------------------------------------
healthsta |
tus       |       count  proportions
----------+-------------------------
 Excellen |  55929350.4   .246198495
 Very Goo |  76791482.8   .338032667
     Good |  64406560.7   .283514795
     Fair |  22764011.7    .10020616
     Poor |  7280374.71   .032047883
          |
    Total |   227171780           1
------------------------------------
  Key:  count     =  weighted counts
        propor~s  =  cell proportions

. svy, subpop(if age08x>17): tabulate fpl, count cell format(%11.0g)
(running tabulate on estimation sample)

Number of strata   =       165                Number of obs      =      33066
Number of PSUs     =       370                Population size    =  304375942
                                              Subpop. no. of obs =      21915
                                              Subpop. size       =  227364210
                                              Design df          =        205

------------------------------------
      fpl |       count  proportions
----------+-------------------------
     Poor |  25687590.2    .11297992
 Near Poo |  9831209.62   .043239917
 Low Inco |  30180744.3   .132741843
 Middle I |  68854477.4   .302837801
```

```
 High Inc |  92810188.5   .408200519
          |
    Total |   227364210            1
-----------------------------------
  Key:  count     =  weighted counts
        propor~s  =  cell proportions

.
.
end of do-file
```