Statistics/Data Analysis

```
      log:  /scratch/OhioFamilyHealthSurvey/analysislogStrataCluster.smcl
  log type:  smcl
 opened on:  16 Jul 2008, 15:38:53

1 . do /tmp/SD27137.000000

2 . /*************************************************************/
3 . /* Estimation with STRATA                                    */
4 . /*************************************************************/
5 .
6 .
7 . /*****************************************/
8 . /*create a pseudo-SRS from Mahoning County*/
9 .
10. set memory 1g  /*increase available memory, since the data is huge*/
  (1048576k)

11. use "ofhs_virgin_file.dta"   /*read in the data*/

12.
13. /*recode some missing data*/
14. replace s10 = . if s10>9  /*maximum recorded adults = 9 */
  (0 real changes made)

15. replace s12 = . if s12>12 /*maximum number of children = 12 */
  (32 real changes made, 32 to missing)

16.
17. /*create a new variable - yes/no for children in hh*/
18. gen kids = 0

19. replace kids = 1 if s12>0
  (16375 real changes made)

20.
21. keep if stratum==50  /*keep only one county - this is the coded number for Mahoning*/
  (38927 observations deleted)

22. keep if s10!=.      /*keep only observations with no missing values - DON'T DO THIS IN PRACTICE!*/
  (1 observation deleted)

23. keep if s12!=.
  (2 observations deleted)

24. gen N=102587

25. gen n=1023

26.
27. save "/scratch/OhioFamilyHealthSurvey/MahoningCounty.dta", replace  /*save the reduced data set*/
  file /scratch/OhioFamilyHealthSurvey/MahoningCounty.dta saved

28. clear

29.
  end of do-file

30. do /tmp/SD27137.000000
```

```
31. /*****************************************/
32. /* Append the two county data sets       */
33.
34. use "MahoningCounty.dta"

35. append using "ButlerCounty.dta"

36.
37. set more off

38. tab stratum
```

design stratum for variance computations	Freq.	Percent	Cum.
9	**1,082**	**51.40**	**51.40**
50	**1,023**	**48.60**	**100.00**
Total	**2,105**	**100.00**	

```
39.
40.
41. /*****************************************/
42. /* Calculate a Mean by hand              */
43.
44. su s10 if stratum==9           /*calculate the mean and standard error*/
```

Variable	Obs	Mean	Std. Dev.	Min	Max
s10	**1082**	**1.892791**	**.7643561**	**1**	**6**

```
45.
46. scalar define N9 = 123082

47. scalar define n9 = r(N)  /* sample size*/

48. scalar define mean9 = r(mean)  /*mean */

49. scalar define stdev9 = r(sd)    /* variance */

50.
51. su s10 if stratum==50           /*calculate the mean and standard error*/
```

Variable	Obs	Mean	Std. Dev.	Min	Max
s10	**1023**	**1.848485**	**.8620705**	**1**	**7**

```
52.
53. scalar define N50 = 102587

54. scalar define n50 = r(N)  /* sample size*/

55. scalar define mean50 = r(mean)  /*mean */

56. scalar define stdev50 = r(sd)    /* variance */

57.
```

```
58. scalar define Ntot=N9+N50

59.
60. /*stratified mean*/
61. scalar define meanstr = N9/(Ntot)*mean9 + N50/(Ntot)*mean50

62. display meanstr
  1.8726499

63.
64. /*stratified variance*/
65. scalar define varstr = N9^2/Ntot^2*(1-n9/N9)*stdev9^2/n9 + N50^2/Ntot^2*(1-n50/N50)*stdev50^2/n50

66. display sqrt(varstr)
  .01754534

67.
68. scalar define varstrnofpc = (N9^2)/(Ntot^2)*stdev9^2/n9 + (N50^2)/(Ntot^2)*stdev50^2/n50

69. display sqrt(varstrnofpc)
  .01762804

70.
71.
  end of do-file

72. do /tmp/SD27137.000000

73. /*****************************************/
74. /* Calculate a Mean using weights          */
75.
76.
77. /*make a weight variable*/
78. gen mywt = N/n

79.
80. svyset, clear

81. svyset masterid [pweight=mywt], strata(stratum) fpc(N)

        pweight: mywt
            VCE: linearized
    Single unit: missing
       Strata 1: stratum
           SU 1: masterid
          FPC 1: N

82.
83. svy: mean s10
  (running mean on estimation sample)

  Survey: Mean estimation

  Number of strata =       2          Number of obs    =    2105
  Number of PSUs   =    2105          Population size  =  225605
                                      Design df        =    2103

------------------------------------------------------------------------------
             |             Linearized
             |        Mean   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         s10 |    1.872644   .0175456      1.838236    1.907053
------------------------------------------------------------------------------
```

```
84.
85. /* compare to the ‘‘by hand" results*/
86. display meanstr
  1.8726499

87. display sqrt(varstr)
  .01754534

88.
89.
90. /*******No FPC********/
91. svyset, clear

92. svyset masterid [pweight=mywt], strata(stratum)

        pweight: mywt
            VCE: linearized
    Single unit: missing
       Strata 1: stratum
           SU 1: masterid
         FPC 1: <zero>

93.
94. svy: mean s10
  (running mean on estimation sample)

  Survey: Mean estimation

  Number of strata =       2          Number of obs    =    2105
  Number of PSUs   =    2105          Population size  =  225605
                                      Design df        =    2103

  ------------------------------------------------------------------------------
               |             Linearized
               |        Mean   Std. Err.     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
           s10 |    1.872644   .0176283      1.838073    1.907215
  ------------------------------------------------------------------------------

95.
96. /* compare to the ‘‘by hand" results*/
97. display meanstr
  1.8726499

98. display sqrt(varstrnofpc)
  .01762804

99.
  end of do-file

100 do /tmp/SD27137.000000

101 /****************************************/
102 /* Now include all counties as strata      */
103
104
105 clear

106
107 set memory 1g  /*increase available memory, since the data is huge*/
  (1048576k)
```

```
108 use "ofhs_virgin_file.dta"   /*read in the data*/

109
110 /*recode some missing data*/
111 replace s10 = . if s10>9  /*maximum recorded adults = 9 */
  (0 real changes made)

112 replace s12 = . if s12>12 /*maximum number of children = 12 */
  (32 real changes made, 32 to missing)

113
114 /*create a new variable − yes/no for children in hh*/
115 gen kids = 0

116 replace kids = 1 if s12>0
  (16375 real changes made)

117
118 /*create a new variable − total number of people in a household */
119 capture gen tothhmembers = s10 + s12

120
121
122 keep if stratum<=88  /*keep only the counties (not the oversample)*/
  (1684 observations deleted)

123 keep if s10!=.       /*keep only observations with no missing values − DON'T DO THIS IN PRACTICE!*
  (163 observations deleted)

124 keep if s12!=.
  (32 observations deleted)

125
126
127 run StrWts.do /*this file assigns naive weights to units in each stratum to a variable called myst

128
129
130 svyset masterid [pweight=mystrwt], strata(stratum)

        pweight: mystrwt
            VCE: linearized
    Single unit: missing
       Strata 1: stratum
           SU 1: masterid
          FPC 1: <zero>

131
132 /*DESCRIBE THE SURVEY DESIGN*/
133 set more off

134 svydes

  Survey: Describing stage 1 sampling units

        pweight: mystrwt
            VCE: linearized
    Single unit: missing
       Strata 1: stratum
           SU 1: masterid
          FPC 1: <zero>
```

			#Obs per Unit		
Stratum	#Units	#Obs	min	mean	max
1	234	234	1	1.0	1
2	294	294	1	1.0	1
3	157	157	1	1.0	1
4	334	334	1	1.0	1
5	222	222	1	1.0	1
6	106	106	1	1.0	1
7	239	239	1	1.0	1
8	556	556	1	1.0	1
9	1082	1082	1	1.0	1
10	125	125	1	1.0	1
11	119	119	1	1.0	1
12	328	328	1	1.0	1
13	798	798	1	1.0	1
14	146	146	1	1.0	1
15	391	391	1	1.0	1
16	157	157	1	1.0	1
17	165	165	1	1.0	1
18	2973	2973	1	1.0	1
19	340	340	1	1.0	1
20	256	256	1	1.0	1
21	197	197	1	1.0	1
22	405	405	1	1.0	1
23	200	200	1	1.0	1
24	102	102	1	1.0	1
25	3114	3114	1	1.0	1
26	250	250	1	1.0	1
27	147	147	1	1.0	1
28	109	109	1	1.0	1
29	253	253	1	1.0	1
30	145	145	1	1.0	1
31	1589	1589	1	1.0	1
32	275	275	1	1.0	1
33	95	95	1	1.0	1
34	42	42	1	1.0	1
35	211	211	1	1.0	1
36	633	633	1	1.0	1
37	85	85	1	1.0	1
38	919	919	1	1.0	1
39	307	307	1	1.0	1
40	148	148	1	1.0	1
41	268	268	1	1.0	1
42	180	180	1	1.0	1
43	298	298	1	1.0	1
44	291	291	1	1.0	1
45	245	245	1	1.0	1
46	157	157	1	1.0	1
47	1370	1370	1	1.0	1
48	1920	1920	1	1.0	1
49	86	86	1	1.0	1
50	1023	1023	1	1.0	1
51	219	219	1	1.0	1
52	203	203	1	1.0	1
53	387	387	1	1.0	1
54	272	272	1	1.0	1
55	258	258	1	1.0	1
56	86	86	1	1.0	1
57	1649	1649	1	1.0	1
58	379	379	1	1.0	1
59	111	111	1	1.0	1
60	292	292	1	1.0	1
61	67	67	1	1.0	1
62	134	134	1	1.0	1
63	141	141	1	1.0	1
64	88	88	1	1.0	1
65	93	93	1	1.0	1
66	106	106	1	1.0	1
67	162	162	1	1.0	1
68	303	303	1	1.0	1

69	**137**	**137**	**1**	**1.0**	**1**
70	**243**	**243**	**1**	**1.0**	**1**
71	**200**	**200**	**1**	**1.0**	**1**
72	**322**	**322**	**1**	**1.0**	**1**
73	**404**	**404**	**1**	**1.0**	**1**
74	**317**	**317**	**1**	**1.0**	**1**
75	**190**	**190**	**1**	**1.0**	**1**
76	**785**	**785**	**1**	**1.0**	**1**
77	**2700**	**2700**	**1**	**1.0**	**1**
78	**554**	**554**	**1**	**1.0**	**1**
79	**482**	**482**	**1**	**1.0**	**1**
80	**109**	**109**	**1**	**1.0**	**1**
81	**116**	**116**	**1**	**1.0**	**1**
82	**60**	**60**	**1**	**1.0**	**1**
83	**684**	**684**	**1**	**1.0**	**1**
84	**159**	**159**	**1**	**1.0**	**1**
85	**522**	**522**	**1**	**1.0**	**1**
86	**273**	**273**	**1**	**1.0**	**1**
87	**694**	**694**	**1**	**1.0**	**1**
88	**87**	**87**	**1**	**1.0**	**1**
88	**38074**	**38074**	**1**	**1.0**	**1**

```
135
136 svy: mean s10 s12
  (running mean on estimation sample)

  Survey: Mean estimation

  Number of strata =      88          Number of obs    =   38074
  Number of PSUs   =   38074          Population size  = 1.6e+06
                                      Design df        =   37986
```

	Mean	Linearized Std. Err.	[95% Conf.	Interval]
s10	**1.885946**	**.0057589**	**1.874658**	**1.897234**
s12	**.7498394**	**.0086419**	**.7329011**	**.7667777**

```
137 svy: ratio s12 / tothhmembers
  (running ratio on estimation sample)

  Survey: Ratio estimation

  Number of strata =      88          Number of obs    =   38074
  Number of PSUs   =   38074          Population size  = 1.6e+06
                                      Design df        =   37986

       _ratio_1: s12/tothhmembers
```

	Ratio	Linearized Std. Err.	[95% Conf.	Interval]
_ratio_1	**.2844842**	**.0023404**	**.279897**	**.2890715**

```
138
  end of do-file
```

```
139 do /tmp/SD27137.000000

140 /*************************************************************/
141 /* Estimation with CLUSTERS                                  */
142 /*************************************************************/
143
144 set more off

145
146 clear

147
148 set memory 1g
  (1048576k)

149 fdause "NIS05.xpt", novallabels

150 sort seqnumhh

151
152 /*Look only at Franklin County == approximately an SRS of phone numbers */
153 keep if estiap == 43
  (27345 observations deleted)

154
155 /*Look at polio vaccines*/
156 tab full_pol, missing  /*this variable containes yes/no polio series*/
```

HH REPORT OF 3+ POLIO-CONTAINING SHOTS	Freq.	Percent	Cum.
1	**132**	**46.81**	**46.81**
2	**64**	**22.70**	**69.50**
77	**86**	**30.50**	**100.00**
Total	**282**	**100.00**	

```
157
158 drop if full_pol > 2 /*cheat by dropping the missing data*/
  (86 observations deleted)

159 replace full_pol = 0 if full_pol == 2 /*recode the nos to '0'
  >                              so that the proportions work out*/
  (64 real changes made)

160
161 /*Look at the households */
162
163 tab seqnumhh /* there are 4 households with 2 children, rest have only one */
```

UNIQUE HOUSEHOLD IDENTIFIER	Freq.	Percent	Cum.
00151	**1**	**0.51**	**0.51**
00314	**1**	**0.51**	**1.02**
00820	**1**	**0.51**	**1.53**
00920	**1**	**0.51**	**2.04**
01056	**1**	**0.51**	**2.55**
01232	**1**	**0.51**	**3.06**
01567	**1**	**0.51**	**3.57**
01651	**1**	**0.51**	**4.08**
01868	**1**	**0.51**	**4.59**
01892	**1**	**0.51**	**5.10**
02949	**1**	**0.51**	**5.61**
03167	**1**	**0.51**	**6.12**
03340	**1**	**0.51**	**6.63**
03386	**1**	**0.51**	**7.14**
03396	**1**	**0.51**	**7.65**
03487	**1**	**0.51**	**8.16**

03504	1	0.51	8.67
03543	1	0.51	9.18
03633	1	0.51	9.69
03652	2	1.02	10.71
03708	1	0.51	11.22
04065	1	0.51	11.73
04288	1	0.51	12.24
04404	1	0.51	12.76
04698	1	0.51	13.27
05107	1	0.51	13.78
06025	1	0.51	14.29
06457	1	0.51	14.80
06463	1	0.51	15.31
06481	1	0.51	15.82
06680	1	0.51	16.33
06852	1	0.51	16.84
07061	1	0.51	17.35
07085	1	0.51	17.86
07091	1	0.51	18.37
07132	1	0.51	18.88
07305	1	0.51	19.39
07349	1	0.51	19.90
07455	1	0.51	20.41
07553	1	0.51	20.92
07806	1	0.51	21.43
07960	1	0.51	21.94
08071	1	0.51	22.45
08654	1	0.51	22.96
08656	1	0.51	23.47
08668	1	0.51	23.98
08678	1	0.51	24.49
08743	1	0.51	25.00
08750	1	0.51	25.51
09103	1	0.51	26.02
09339	1	0.51	26.53
09429	1	0.51	27.04
09778	1	0.51	27.55
09941	1	0.51	28.06
09944	1	0.51	28.57
10160	1	0.51	29.08
11071	1	0.51	29.59
11570	1	0.51	30.10
11885	1	0.51	30.61
15765	1	0.51	31.12
15789	1	0.51	31.63
15929	1	0.51	32.14
16196	1	0.51	32.65
16256	1	0.51	33.16
16277	1	0.51	33.67
16361	1	0.51	34.18
16386	1	0.51	34.69
16390	1	0.51	35.20
16401	1	0.51	35.71
16473	1	0.51	36.22
16502	1	0.51	36.73
16630	1	0.51	37.24
16727	1	0.51	37.76
16826	1	0.51	38.27
16828	1	0.51	38.78
16910	1	0.51	39.29
16981	1	0.51	39.80
17014	1	0.51	40.31
17103	1	0.51	40.82
17306	1	0.51	41.33
17433	1	0.51	41.84
17763	1	0.51	42.35
18004	1	0.51	42.86
18006	1	0.51	43.37
18187	1	0.51	43.88
18334	2	1.02	44.90
18336	1	0.51	45.41
18447	1	0.51	45.92

18597	**1**	**0.51**	**46.43**
18640	**1**	**0.51**	**46.94**
18683	**1**	**0.51**	**47.45**
18691	**1**	**0.51**	**47.96**
18734	**1**	**0.51**	**48.47**
18738	**1**	**0.51**	**48.98**
18781	**1**	**0.51**	**49.49**
18823	**2**	**1.02**	**50.51**
18862	**1**	**0.51**	**51.02**
18936	**1**	**0.51**	**51.53**
18980	**1**	**0.51**	**52.04**
18987	**1**	**0.51**	**52.55**
19006	**1**	**0.51**	**53.06**
19174	**1**	**0.51**	**53.57**
19247	**1**	**0.51**	**54.08**
19353	**1**	**0.51**	**54.59**
19438	**1**	**0.51**	**55.10**
19574	**1**	**0.51**	**55.61**
19606	**1**	**0.51**	**56.12**
19631	**1**	**0.51**	**56.63**
19638	**1**	**0.51**	**57.14**
19737	**1**	**0.51**	**57.65**
19811	**1**	**0.51**	**58.16**
19901	**1**	**0.51**	**58.67**
20197	**1**	**0.51**	**59.18**
20217	**1**	**0.51**	**59.69**
20381	**1**	**0.51**	**60.20**
20507	**1**	**0.51**	**60.71**
20525	**1**	**0.51**	**61.22**
20599	**1**	**0.51**	**61.73**
20607	**1**	**0.51**	**62.24**
20626	**1**	**0.51**	**62.76**
20793	**1**	**0.51**	**63.27**
20825	**1**	**0.51**	**63.78**
20832	**1**	**0.51**	**64.29**
20847	**1**	**0.51**	**64.80**
20872	**1**	**0.51**	**65.31**
20895	**1**	**0.51**	**65.82**
20911	**1**	**0.51**	**66.33**
21155	**1**	**0.51**	**66.84**
21265	**1**	**0.51**	**67.35**
21419	**1**	**0.51**	**67.86**
21558	**1**	**0.51**	**68.37**
21773	**1**	**0.51**	**68.88**
21902	**1**	**0.51**	**69.39**
22025	**1**	**0.51**	**69.90**
22058	**1**	**0.51**	**70.41**
22102	**1**	**0.51**	**70.92**
22136	**1**	**0.51**	**71.43**
22168	**1**	**0.51**	**71.94**
22297	**1**	**0.51**	**72.45**
22323	**1**	**0.51**	**72.96**
22413	**1**	**0.51**	**73.47**
22516	**1**	**0.51**	**73.98**
22638	**1**	**0.51**	**74.49**
22644	**1**	**0.51**	**75.00**
22727	**1**	**0.51**	**75.51**
22793	**1**	**0.51**	**76.02**
22832	**1**	**0.51**	**76.53**
22966	**1**	**0.51**	**77.04**
23113	**1**	**0.51**	**77.55**
23167	**1**	**0.51**	**78.06**
23176	**1**	**0.51**	**78.57**
23218	**1**	**0.51**	**79.08**
23571	**1**	**0.51**	**79.59**
23602	**1**	**0.51**	**80.10**
23704	**1**	**0.51**	**80.61**
23767	**1**	**0.51**	**81.12**
23819	**1**	**0.51**	**81.63**
24265	**1**	**0.51**	**82.14**
24350	**1**	**0.51**	**82.65**
24383	**1**	**0.51**	**83.16**

24430	**1**	**0.51**	**83.67**
24453	**1**	**0.51**	**84.18**
24483	**1**	**0.51**	**84.69**
24496	**1**	**0.51**	**85.20**
24501	**1**	**0.51**	**85.71**
24816	**1**	**0.51**	**86.22**
24881	**1**	**0.51**	**86.73**
24890	**2**	**1.02**	**87.76**
25094	**1**	**0.51**	**88.27**
25154	**1**	**0.51**	**88.78**
25169	**1**	**0.51**	**89.29**
25201	**1**	**0.51**	**89.80**
25228	**1**	**0.51**	**90.31**
25418	**1**	**0.51**	**90.82**
25480	**1**	**0.51**	**91.33**
25489	**1**	**0.51**	**91.84**
25500	**1**	**0.51**	**92.35**
25567	**1**	**0.51**	**92.86**
25626	**1**	**0.51**	**93.37**
25654	**1**	**0.51**	**93.88**
25666	**1**	**0.51**	**94.39**
25673	**1**	**0.51**	**94.90**
25718	**1**	**0.51**	**95.41**
25779	**1**	**0.51**	**95.92**
25864	**1**	**0.51**	**96.43**
25919	**1**	**0.51**	**96.94**
25935	**1**	**0.51**	**97.45**
25977	**1**	**0.51**	**97.96**
26137	**1**	**0.51**	**98.47**
26331	**1**	**0.51**	**98.98**
26373	**1**	**0.51**	**99.49**
26418	**1**	**0.51**	**100.00**
Total	**196**	**100.00**	

```
164 return list  /* 196-192 confirms this */

  scalars:
                  r(N) =  196
                  r(r) =  192

165 duplicates report seqnumhh /*yet another way to get this*/

  Duplicates in terms of seqnumhh
```

copies	observations	surplus
1	**188**	**0**
2	**8**	**4**

```
166
167 /*identify houshohld with more than one child*/
168 duplicates tag seqnumhh, generate(numkids)

  Duplicates in terms of seqnumhh

169 tab numkids /*look at the generated variable*/
```

numkids	Freq.	Percent	Cum.
0	**188**	**95.92**	**95.92**
1	**8**	**4.08**	**100.00**
Total	**196**	**100.00**	

```
170 replace numkids = numkids + 1 /*this is the number of kids in the child's household*/
  (196 real changes made)

171
172 /*look at the polio of the kids in multi-kid households*/
173 list seqnumhh full_pol if numkids==2
```

	seqnumhh	full_pol
20.	**03652**	**1**
21.	**03652**	**1**
87.	**18334**	**1**
88.	**18334**	**1**
98.	**18823**	**0**
99.	**18823**	**0**
171.	**24890**	**0**
172.	**24890**	**1**

```
174
175 /* record the number of households in the sample*/
176 scalar nn = 192 /*we use nn because 'n' is ambiguous*/

177
178 /*Cheat and use the weights to find the number of households*/
179 total rddwt if numkids==1

  Total estimation                    Number of obs    =     188
```

	Total	Std. Err.	[95% Conf.	Interval]
rddwt	**16903.73**	**545.3352**	**15827.93**	**17979.53**

```
180 matrix define totonekid = e(b)

181 total rddwt if numkids==2

  Total estimation                    Number of obs    =       8
```

	Total	Std. Err.	[95% Conf.	Interval]
rddwt	**670.1883**	**82.01259**	**476.2594**	**864.1173**

```
182 matrix define tottwokids = e(b)

183 scalar N_est = round(totonekid[1,1] + tottwokids[1,1]/2)

184         /*This estimates the number of households*/
185 gen N = N_est

186 gen KK = N_est * 1.05

187
188 /*generate the number of children per household (in the sample)*/
```

```
189 generate M = numkids

190
191
192 /*******************************************************/
193 /* to do estimation, it's easier to collapse the data set*/
194 /* into totals with one row per household                */
195
196 by seqnumhh: egen tot_pol = total(full_pol) /* sum the polio for the household */

197 list seqnumhh full_pol tot_pol if numkids==2
```

	seqnumhh	full_pol	tot_pol
20.	03652	1	2
21.	03652	1	2
87.	18334	1	2
88.	18334	1	2
98.	18823	0	0
99.	18823	0	0
171.	24890	0	1
172.	24890	1	1

```
198
199 /*save the current state first!*/
200 save "/scratch/OhioFamilyHealthSurvey/NIS05.dta", replace
  file /scratch/OhioFamilyHealthSurvey/NIS05.dta saved

201
202 /***now collapse***/
203 duplicates drop seqnumhh, force

  Duplicates in terms of seqnumhh

  (4 observations deleted)

204
205
206 /******unbiased estimate******/
207
208 duplicates drop seqnumhh, force

  Duplicates in terms of seqnumhh

  (0 observations are duplicates)

209
210 display N_est
  17239

211 display nn
  192

212
213 summarize tot_pol
```

Variable	Obs	Mean	Std. Dev.	Min	Max
tot_pol	192	.6875	.486735	0	2

```
214 scalar t_bar = r(mean)

215 scalar t_hat_unb = N_est * t_bar

216 scalar st_sq = r(Var)

217 scalar t_unb_var = N_est^2*(1-nn/N)*st_sq/nn

218 display t_hat_unb
  11851.813

219 display sqrt(t_unb_var)
  602.17401

220
221 scalar ybar_unb = t_hat_unb / KK

222 scalar ybar_unb_var = t_unb_var / (KK^2)

223 display ybar_unb
  .65476193

224 display sqrt(ybar_unb_var)
  .03326754

225
226 /*******ratio estimate*******/
227
228 total tot_pol M

  Total estimation                     Number of obs    =     192

------------------------------------------------------------------------------
             |      Total   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     tot_pol |        132   6.744398      118.6969    145.3031
           M |        196   1.984231      192.0862    199.9138
------------------------------------------------------------------------------

229 matrix totests = e(b)

230 scalar ybar_r = totests[1,1]/totests[1,2]

231 display ybar_r
  .67346939

232
233 summarize M

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           M |        192    1.020833    .1431995          1          2

234 scalar mbar_U = r(mean)

235 gen ybar_i = tot_pol/M

236 gen numerator = M^2 * (ybar_i - ybar_r)^2

237 summarize numerator

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   numerator |        192     .229145    .1992247   .1066222   1.814244
```

```
238 scalar nummean=r(mean)

239 scalar ybar_r_var = (1- nn/N) * 1 / (mbar_U^2) / nn * nummean * nn / (nn-1)

240 display ybar_r
  .67346939

241 display sqrt(ybar_r_var)
  .03374051

242
243 /******Use built-in Functions*****/
244
245 clear

246 use "/scratch/OhioFamilyHealthSurvey/NIS05.dta"

247
248 svyset seqnumhh, fpc(N)

        pweight: <none>
            VCE: linearized
    Single unit: missing
       Strata 1: <one>
           SU 1: seqnumhh
          FPC 1: N

249 svy: mean full_pol
  (running mean on estimation sample)

  Survey: Mean estimation

  Number of strata =       1          Number of obs    =     196
  Number of PSUs   =     192          Population size  =     196
                                      Design df        =     191

-------------------------------------------------------------------------------
             |             Linearized
             |        Mean   Std. Err.     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
    full_pol |   .6734694   .0337405      .6069175    .7400213
-------------------------------------------------------------------------------

250
251
252 /******Use Weights ******/
253
254 gen clustwt = N_est / nn

255 svyset seqnumhh [pweight = clustwt], fpc(N)

        pweight: clustwt
            VCE: linearized
    Single unit: missing
       Strata 1: <one>
           SU 1: seqnumhh
          FPC 1: N

256 svy: total full_pol
  (running total on estimation sample)

  Survey: Total estimation

  Number of strata =       1          Number of obs    =     196
  Number of PSUs   =     192          Population size  = 17598.1
                                      Design df        =     191
```

	Total	Linearized Std. Err.	[95% Conf.	Interval]
full_pol	**11851.81**	**602.174**	**10664.05**	**13039.58**

```
257
258 svy: mean full_pol
  (running mean on estimation sample)

  Survey: Mean estimation

  Number of strata =       1          Number of obs    =     196
  Number of PSUs   =     192          Population size  = 17598.1
                                      Design df        =     191
```

	Mean	Linearized Std. Err.	[95% Conf.	Interval]
full_pol	**.6734694**	**.0337405**	**.6069175**	**.7400213**

```
259 display ybar_r
  .67346939

260 display sqrt(ybar_r_var)
  .03374051

261
262
263
264 clear

265
  end of do-file

266 do /tmp/SD27137.000000

267
268 /*************************************************************/
269 /**  Back to the OFHS                                       **/
270 /*************************************************************/
271
272 set memory 8000    /*don't need as much memory for this reduced data set*/
  (8000k)

273 use "/scratch/OhioFamilyHealthSurvey/ButlerCounty.dta"  /*read in the Butler County data set*/

274
275 capture gen N=123082             /*make a variable to hold the population size*/

276 capture gen n=1082               /*make a variable to hold the sample size*/

277
278 /* Weight the data using the simplistic 2-stage sample */
279
280 gen p1 = n/N    /*probability of household being selected*/
```

```
281 gen p2 = 1/s10  /*probability of adult being selected */

282
283 gen totprob = p1 * p2

284 gen twostagewt = 1/totprob

285
286 /* Define the design [two stage cluster]*/
287
288 svyset masterid [pweight = twostagewt], fpc(N)

        pweight: twostagewt
            VCE: linearized
    Single unit: missing
       Strata 1: <one>
           SU 1: masterid
          FPC 1: N

289
290 svy: prop a1  /* a1 is "do you have insurance" */
  (running proportion on estimation sample)

  Survey: Proportion estimation

  Number of strata =       1          Number of obs    =    1082
  Number of PSUs   =    1082          Population size  =  232847
                                      Design df        =    1081

  --------------------------------------------------------------
                  |             Linearized
                  | Proportion   Std. Err.     [95% Conf. Interval]
  ----------------+---------------------------------------------
  a1              |
               1  |   .8901367   .0104946      .8695446    .9107288
               2  |   .1074219   .0104167      .0869826    .1278612
              98  |   .0024414   .0014577     -.0004188    .0053016
  --------------------------------------------------------------

291
292
293 clear

294
295
  end of do-file

296 log close
        log:  /scratch/OhioFamilyHealthSurvey/analysislogStrataCluster.smcl
   log type:  smcl
  closed on:  16 Jul 2008, 15:40:20
```