



```
log: /scratch/OhioFamilyHealthSurvey/analysislog-Variance-PostStratify.smcl
log type: smcl
opened on: 17 Jul 2008, 17:17:16

1 . do /tmp/SD02215.000000

2 .
3 . /*****/
4 . /** Variance Estimation Using NHANES 99 **/
5 . /*****/
6 .
7 . set more off

8 . clear

9 .
10. set memory lg /*increase available memory, since the data is huge*/
    (1048576k)

11.
12. /**READ IN DEMOGRAPHIC AND SURVEY DESIGN INFO**/
13. fdause "/scratch/NHANES99/demo.xpt"

14. sort seqn /*sort by unique ID*/

15. save "/scratch/NHANES99/demo", replace
    file /scratch/NHANES99/demo.dta saved

16.
17. clear

18.
19. /**READ IN HEALTH INSURANCE QUESTIONS**/
20. fdause "/scratch/NHANES99/hiq.xpt"

21. sort seqn /*sort by unique ID*/

22. save "/scratch/NHANES99/hiq", replace
    file /scratch/NHANES99/hiq.dta saved

23.
24. /**MERGE THE TWO DATA SETS TOGETHER BASED ON THE UNIQUE ID*/
25. merge seqn using "/scratch/NHANES99/demo"

26.
27. /*READ THE DOCUMENTATION FOR THIS SURVEY - IT'S COMPLICATED!*/
28. /*Question = Proportion of Americans Covered by Health Insurance*/
29. /*HID010 : Health Insurance? 1=y, 2=n, 7=refuse, 9=DK*/
30. gen insuranceMAR = hid010
    (161 missing values generated)

31. replace insuranceMAR = . if hid010==7 | hid010==9
    (30 real changes made, 30 to missing)

32. replace insuranceMAR = 0 if hid010==2
    (2038 real changes made)

33.
34. /*wtint2yr = The appropriate weights for this data set, and this question */
```

```

35. /*sdmvpisu = The 'pseudo' PSUs (clusters) provided for this data */
36. /*sdmvstra = The 'pseudo' Strata provided for this data */
37.
38.
39. /******
40. /* Design Effects - no design info, but published DEFF values */
41.
42.
43. /*HERE I AM PRETENDING TO BE THE AGENCY WITH THE DESIGN INFORMATION*/
44. /*THIS CODE CALCULATES THE DEFF*/
45. svyset, clear

46. svyset sdmvpisu [pweight=wtint2yr], strata(sdmvstra)

```

```

    pweight: wtint2yr
           VCE: linearized
Single unit: missing
  Strata 1: sdmvstra
    SU 1: sdmvpisu
    FPC 1: <zero>

```

```

47. svy: mean insuranceMAR
   (running mean on estimation sample)

```

Survey: Mean estimation

```

Number of strata =      13      Number of obs   =    9774
Number of PSUs  =      27      Population size =  2.7e+08
                                   Design df       =    14

```

	Mean	Linearized Std. Err.	[95% Conf. Interval]	
insuranceMAR	.8298242	.0130441	.8018474	.857801

```

48. estat effects, deff

```

	Mean	Linearized Std. Err.	DEFF
insuranceMAR	.8298242	.0130441	11.7753

```

49. /*PUBLISH THE DESIGN EFFECT = 11.78*/
50. /*NOW FORGET THE DESIGN INFORMATION*/
51. svyset, clear

52.
53.
54. /*NOW WE GET THE SURVEY WITHOUT THE DESIGN INFORMATION*/
55. /*we still need the weights!*/
56. svyset, clear

57. svyset [pweight=wtint2yr]

```

```

    pweight: wtint2yr
           VCE: linearized
Single unit: missing
  Strata 1: <one>
    SU 1: <observations>
    FPC 1: <zero>

```

```
58.
59. /*****estimate using assumption of SRS*****/
60. svy: mean insuranceMAR
    (running mean on estimation sample)
```

Survey: Mean estimation

```
Number of strata =      1          Number of obs   =   9774
Number of PSUs   =   9774          Population size = 2.7e+08
                                   Design df       =   9773
```

	Mean	Linearized Std. Err.	[95% Conf. Interval]	
insuranceMAR	.8298242	.0056605	.8187284	.84092

```
61.
62. /*Find the appropriate t critical value*/
63. scalar define tval = abs(invttail(e(df_r), 0.025)) /*find the critical value*/

64.
65. /*unadjusted CIs*/
66. matrix define lower = e(b) - tval * cholesky(e(V)) /*cholesky finds the square root*/
67. matrix define upper = e(b) + tval * cholesky(e(V))
68. matrix define se = cholesky(e(V))

69.
70. /*adjusted CIs*/
71. matrix define lowerdeff = e(b) - tval * cholesky(e(V)*11.78)
72. matrix define upperdeff = e(b) + tval * cholesky(e(V)*11.78)
73. matrix define sedeff = cholesky(e(V)*11.78)

74.
75. matrix define info = (se, lower, upper) \ (sedeff, lowerdeff, upperdeff)
76. matrix colnames info = Standard_Error LowerCI UpperCI
77. matrix rownames info = SRS Deff_adjusted

78.
79. matrix list info

    info[2,3]
           Standard_Error      LowerCI      UpperCI
    SRS      .00566054      .81872841      .84092004
    Deff_adjusted .01942809      .79174115      .8679073

80.
81.
82. /*****
83. /* Linearization - known design information */
84.
85. svyset, clear
```

86. svyset sdmvpsu [pweight=wtint2yr], strata(sdmvstra)

```

pweight: wtint2yr
VCE: linearized
Single unit: missing
Strata 1: sdmvstra
SU 1: sdmvpsu
FPC 1: <zero>

```

87. svydes

Survey: Describing stage 1 sampling units

```

pweight: wtint2yr
VCE: linearized
Single unit: missing
Strata 1: sdmvstra
SU 1: sdmvpsu
FPC 1: <zero>

```

Stratum	#Units	#Obs	#Obs per Unit		
			min	mean	max
1	3	1065	324	355.0	374
2	2	694	295	347.0	399
3	2	787	376	393.5	411
4	2	858	363	429.0	495
5	2	749	372	374.5	377
6	2	730	346	365.0	384
7	2	859	397	429.5	462
8	2	749	338	374.5	411
9	2	690	345	345.0	345
10	2	646	304	323.0	342
11	2	737	325	368.5	412
12	2	815	384	407.5	431
13	2	586	290	293.0	296
13	27	9965	290	369.1	495

88.

89. svy: mean insuranceMAR
(running mean on estimation sample)

Survey: Mean estimation

```

Number of strata = 13      Number of obs = 9774
Number of PSUs = 27      Population size = 2.7e+08
                          Design df = 14

```

	Mean	Linearized Std. Err.	[95% Conf. Interval]	
insuranceMAR	.8298242	.0130441	.8018474	.857801

90.

91. /*save the CIs, and compare to before*/

92. scalar define tval = abs(invttail(e(df_r), 0.025)) /*find the critical value*/

```

93. matrix define cislinear = cholesky(e(V)), e(b) - tval * cholesky(e(V)), e(b) + tval * cholesky(e(V))
94. matrix define info = (se, lower, upper) \ (seff, lowerdeff, upperdeff) \ cislinear
95. matrix colnames info = Standard_Error LowerCI UpperCI
96. matrix rownames info = SRS Deff_adjusted Linearization
97.
98. matrix list info

```

```

info[3,3]

```

	Standard_Error	LowerCI	UpperCI
SRS	.00566054	.81872841	.84092004
Deff_adjusted	.01942809	.79174115	.8679073
Linearization	.0130441	.80184742	.85780103

```

99.
100
101 /*****
102 /* Jackknife - have resampling weights */
103
104 svyset, clear

```

```

105 svyset sdmvpsu [pweight=wtint2yr], strata(sdmvstra) vce(jackknife) jkrweight(wtirep01-wtirep52)

    pweight: wtint2yr
           VCE: jackknife
           MSE: off
    jkrweight: wtirep01 wtirep02 wtirep03 wtirep04 wtirep05 wtirep06 wtirep07 wtirep08 wtirep09 wtirep10
                wtirep15 wtirep16 wtirep17 wtirep18 wtirep19 wtirep20 wtirep21 wtirep22 wtirep23 wtirep24
                wtirep29 wtirep30 wtirep31 wtirep32 wtirep33 wtirep34 wtirep35 wtirep36 wtirep37 wtirep38
                wtirep43 wtirep44 wtirep45 wtirep46 wtirep47 wtirep48 wtirep49 wtirep50 wtirep51 wtirep52
Single unit: missing
Strata 1: sdmvstra
SU 1: sdmvpsu
FPC 1: <zero>

```

```

106 svyset sdmvpsu [pweight=wtint2yr], strata(sdmvstra) vce(jackknife) jkrweight(wtirep*) /*same thi

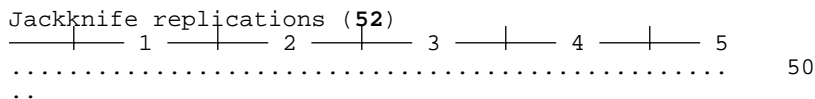
    pweight: wtint2yr
           VCE: jackknife
           MSE: off
    jkrweight: wtirep01 wtirep02 wtirep03 wtirep04 wtirep05 wtirep06 wtirep07 wtirep08 wtirep09 wtirep10
                wtirep15 wtirep16 wtirep17 wtirep18 wtirep19 wtirep20 wtirep21 wtirep22 wtirep23 wtirep24
                wtirep29 wtirep30 wtirep31 wtirep32 wtirep33 wtirep34 wtirep35 wtirep36 wtirep37 wtirep38
                wtirep43 wtirep44 wtirep45 wtirep46 wtirep47 wtirep48 wtirep49 wtirep50 wtirep51 wtirep52
Single unit: missing
Strata 1: sdmvstra
SU 1: sdmvpsu
FPC 1: <zero>

```

```

107
108 svy: mean insuranceMAR
    (running mean on estimation sample)

```



Survey: Mean estimation

```

Number of strata =      13
Number of obs    =      9774
Population size  = 2.7e+08
Replications     =       52
Design df       =       51

```

	Mean	Jackknife Std. Err.	[95% Conf. Interval]	
insuranceMAR	.8298242	.012272	.8051872	.8544612

```

109
110 /*save the CIs, and compare to before*/
111 scalar define tval = abs(invttail(e(df_r), 0.025)) /*find the critical value*/
112 matrix define cisjack = cholesky(e(V)), e(b) - tval * cholesky(e(V)), e(b) + tval * cholesky(e(V))
113 matrix define info = (se, lower, upper) \ (sedeff, lowerdeff, upperdeff) \ cislinear \ cisjack
114 matrix colnames info = Standard_Error LowerCI UpperCI
115 matrix rownames info = SRS Deff_adjusted Linearization Jackknife
116
117 matrix list info

```

```

info[4,3]

```

	Standard_Error	LowerCI	UpperCI
SRS	.00566054	.81872841	.84092004
Deff_adjusted	.01942809	.79174115	.8679073
Linearization	.0130441	.80184742	.85780103
Jackknife	.01227195	.80518725	.8544612

```

118
119 /****Jackknife "guts"****/
120
121 /*estimate using the first weights*/
122 svyset, clear
123 svyset sdmvpsu [pweight=wtirep01], strata(sdmvstra)

```

pweight: **wtirep01**
 VCE: **linearized**
 Single unit: **missing**
 Strata 1: **sdmvstra**
 SU 1: **sdmvpsu**
 FPC 1: <zero>

```

124 svy: mean insuranceMAR
(running mean on estimation sample)

```

Survey: Mean estimation

```

Number of strata =      13          Number of obs   =      9774
Number of PSUs   =      27          Population size =  2.7e+08
                                           Design df      =       14

```

	Mean	Linearized Std. Err.	[95% Conf. Interval]	
insuranceMAR	.8290786	.0133284	.8004921	.8576651

```

125     matrix define j_estst = e(b)

126
127 /*estimate using the other weights, store values in j_estst -- remove the 'quietly's to see what's
128 foreach var of varlist wtirep02-wtirep52 {
129     2.     quietly: svyset sdmvpsu [pweight='var'], strata(sdmvstra)
130     3.     quietly: svy: mean insuranceMAR
131     4.     quietly: matrix define j_estst = j_estst \ e(b)
132     5. }

129
130 matrix list j_estst

j_estst[52,1]
      insuranceMAR
y1    .82907859
y1    .83153158
y1    .82981416
y1    .8307719
y1    .8303635
y1    .82944021
y1    .82919259
y1    .82832447
y1    .8302309
y1    .82924844
y1    .83011964
y1    .83085106
y1    .83063922
y1    .82731126
y1    .82802445
y1    .8280289
y1    .83015251
y1    .83096734
y1    .8287149
y1    .82980588
y1    .83229749
y1    .83020352
y1    .82924821
y1    .82842002
y1    .82967531
y1    .82956954
y1    .82704329
y1    .82761683
y1    .82996439
y1    .83147435
y1    .82964092
y1    .83551917
y1    .82832537
y1    .82984635
y1    .83150068
y1    .82719148
y1    .82939286
y1    .82933456
y1    .82881311
y1    .82948823
y1    .82982709
y1    .82746759
y1    .8294832
y1    .83023257
y1    .83526988
y1    .82666181
y1    .83174158
y1    .82899259
y1    .83075907
y1    .82996428
y1    .83242763
y1    .83073233

```


159 total wtint2yr /*too small*/

Total estimation Number of obs = 9965

	Total	Std. Err.	[95% Conf. Interval]	
wtint2yr	2.72e+08	3000291	2.66e+08	2.78e+08

160

161 /*adjust for total population*/

162 sum wtint2yr

Variable	Obs	Mean	Std. Dev.	Min	Max
wtint2yr	9965	27311.27	30055.55	974.6653	236864.1

163 generate adjwt = wtint2yr * 281400000 / r(sum)

164 /*check*/

165 total adjwt

Total estimation Number of obs = 9965

	Total	Std. Err.	[95% Conf. Interval]	
adjwt	2.81e+08	3102189	2.75e+08	2.87e+08

166

167 /*Population Values are Male=136,100,000 Female = 145,300,000 */

168 /*first see if there's a problem*/

169

170 total adjwt, over(male) /*too little wt for females, too much for males*/

Total estimation Number of obs = 9965

0: male = 0
1: male = 1

Over	Total	Std. Err.	[95% Conf. Interval]	
adjwt				
0	1.44e+08	2237678	1.39e+08	1.48e+08
1	1.38e+08	2148795	1.33e+08	1.42e+08

171 matrix totals = e(b)

172 scalar totalfem = totals[1,1]

173 scalar totalmale = totals[1,2]

174 display totalfem

1.437e+08

175 display totalmale

1.377e+08

```

176
177 /*****TWO WAYS TO POSTSTRATIFY*****/
178
179 /*****/
180 /*By Hand*/
181 gen genderadjwt = adjwt

182 replace genderadjwt = genderadjwt * 136100000 / totalmale if male == 1
    (4883 real changes made)

183 replace genderadjwt = genderadjwt * 145300000 / totalfem if male == 0
    (5082 real changes made)

184
185 total genderadjwt, over(male) /*Just Right!*/

```

```

Total estimation                Number of obs    =    9965

      0: male = 0
      1: male = 1

```

	Over	Total	Std. Err.	[95% Conf. Interval]	
genderadjwt					
	0	1.45e+08	2262428	1.41e+08	1.50e+08
	1	1.36e+08	2123989	1.32e+08	1.40e+08

```

186
187 /*****/
188 /*"Automatically"*/
189 /*create variables specifying the target*/
190 gen malepopsize = 145300000

191 replace malepopsize = 136100000 if male == 1
    (4883 real changes made)

192
193 svyset [pweight = adjwt], poststrata(male) postweight(malepopsize)

      pweight: adjwt
      VCE: linearized
      Poststrata: male
      Postweight: malepopsize
      Single unit: missing
      Strata 1: <one>
      SU 1: <observations>
      FPC 1: <zero>

194 svy: total male
    (running total on estimation sample)

```

```

Survey: Total estimation

Number of strata =      1          Number of obs    =    9965
Number of PSUs  =    9965          Population size = 2.8e+08
N. of poststrata =      2          Design df      =    9964

```

	Total	Linearized Std. Err.	[95% Conf. Interval]	
male	1.36e+08	0	.	.

```

195
196
197 /***** RAKING *****/
198 /*this is in the "svr" package that you must download and install into Stata*/
199 /*this is easy to do via the help menus, if you just search for svr*/
200
201 /*We also need a second variable by which to adjust*/
202 /*I'll use over/under 18 years old */
203 generate kid = 0

204 replace kid = 1 if ridageyr < 18
    (4517 real changes made)

205
206 /*we again, need variables that store the "true" values */
207 generate kidpopsize = 70000000

208 replace kidpopsize = 281400000 - 70000000 if kid == 0
    (5448 real changes made)

209
210 /*see if our unadjusted weights are off*/
211 total adjwt, over(kid)

```

```

Total estimation                Number of obs    =    9965
                                0: kid = 0
                                1: kid = 1

```

	Over	Total	Std. Err.	[95% Conf. Interval]	
adjwt	0	2.06e+08	2603534	2.01e+08	2.12e+08
	1	7.50e+07	1314015	7.24e+07	7.76e+07

```

212
213 /*perform the raking*/
214 survwgt rake adjwt, by(male kid) totvar(malepopsize kidpopsize) generate(rakewt)

215
216 /*check the results*/
217 total rakewt, over(male)

```

```

Total estimation                Number of obs    =    9965
                                0: male = 0
                                1: male = 1

```

	Over	Total	Std. Err.	[95% Conf. Interval]	
rakewt	0	1.45e+08	2300547	1.41e+08	1.50e+08
	1	1.36e+08	2160046	1.32e+08	1.40e+08

```

218 total rakewt, over(kid)

```

```

Total estimation                Number of obs    =    9965
                                0: kid = 0
                                1: kid = 1

```

	Over	Total	Std. Err.	[95% Conf. Interval]	
rakewt					
	0	2.11e+08	2667235	2.06e+08	2.17e+08
	1	7.00e+07	1226261	6.76e+07	7.24e+07

219
220
221
222
223

end of do-file

224 log close

log: /scratch/OhioFamilyHealthSurvey/analysislog-Variance-PostStratify.smcl

log type: smcl

closed on: 17 Jul 2008, 17:17:40