

```
name: <unnamed>
log: M:\Data\T2000\resurvey\koulutus_2013_06\t2011.smc1
log type: smc1
opened on: 15 Jun 2013, 18:55:17
```

```
1 .
2 . // Load the data:
3 . use "M:\Data\T2000\resurvey\koulutus_2013_06\t0011b.dta", clear

4 .
5 . // variable name variable label
6 . // -----
7 . // havtun Individual ID
8 . // vuosi Study year
9 . // osite Stratum
10 . // ryvas PSU in Health 2000
11 . // all_analysis_w Weight for correcting nonresponse
12 . // and varying sampling probabilities
13 . // sukupuoli Gender (1=M,2=F)
14 . // ika Age at measurement,
15 . // rounded to min. of 5-year-intervals
16 . // ika_51k Age group (5 classes)
17 . // 30-44, 45-54, 55-64, 65-74, 75+
18 . // r_aa01 Marital status
19 . // r_bmi BMI
20 . // r_fs_ko1 fs-ko1, mmol/l
21 . // r_systbp2 RR: syst,mittaus 2
22 . // r_systbp2_01 Syst.bp (0=85-140,1=141-232)
23 . // r_systbp2_123 Syst.bp (1=85-120,2=121-160,3=161-232)
24 .
25 . // Define some new variables.
26 . // Create variable for baseline age:
27 . sort havtun vuosi

28 . g ika_bl=ika

29 . replace ika_bl=ika_bl-11 if vuosi==2011
(1720 real changes made)

30 . // and create also corresponding categorical age variable:
31 . g ika_51k_bl=ika_bl

32 . recode ika_51k_bl (min/44=1) (45/54=2) (55/64=3) (65/74=4) (75/max=5)
(ika_51k_bl: 3248 changes made)

33 .
34 . // Define the sampling design:
35 . // The Health 2000 sampling design (PSU=ryvas):
36 . //svyset ryvas [pw=all_analysis_w], strata(osite)
37 . // PSU is individual:
38 . svyset havtun [pw=all_analysis_w], strata(osite)

pweight: all_analysis_w
VCE: linearized
Single unit: missing
Strata 1: osite
SU 1: havtun
FPC 1: <zero>
```

```
39 .
40 . // Check data:
41 . svydescribe
```

Survey: Describing stage 1 sampling units

```
pweight: all_analysis_w
VCE: linearized
Single unit: missing
Strata 1: osite
SU 1: havtun
FPC 1: <zero>
```

Stratum	#Units	#Obs	#Obs per Unit		
			min	mean	max
1	226	373	1	1.7	2
2	162	267	1	1.6	2
3	267	436	1	1.6	2
4	277	464	1	1.7	2
5	228	379	1	1.7	2
124	84	135	1	1.6	2
152	232	360	1	1.6	2
153	64	92	1	1.4	2
181	14	21	1	1.5	2
192	37	58	1	1.6	2
270	24	40	1	1.7	2
280	36	59	1	1.6	2
299	39	65	1	1.7	2
305	30	48	1	1.6	2
390	46	76	1	1.7	2
424	26	45	1	1.7	2
532	84	136	1	1.6	2
545	83	128	1	1.5	2
565	24	39	1	1.6	2
635	17	27	1	1.6	2
20	2000	3248	1	1.6	2

```
42 .
43 . // Simple statistics:
44 . svy: mean r_bmi r_fs_ko1 r_systbp2
(running mean on estimation sample)
```

Survey: ~~Mean estimation~~

Number of strata =	20	Number of obs =	2876
Number of PSUs =	1877	Population size =	2278.23
		Design df =	1857

	Mean	Linearized Std. Err.	[95% Conf. Interval]	
r_bmi	26.82604	.119118	26.59243	27.05966
r_fs_ko1	5.544204	.0248405	5.495486	5.592923
r_systbp2	132.8246	.4661689	131.9103	133.7388

```
45 .
46 . // Print design effects:
47 . estat effects
```

	Mean	Linearized Std. Err.	DEFF	DEFT
r_bmi	26.82604	.119118	1.53807	1.24019
r_fs_ko1	5.544204	.0248405	1.2573	1.1213
r_systbp2	132.8246	.4661689	1.25463	1.1201

```
48 . // (DEFT expresses, how much larger the standard error is,
49 . // when the sampling design has been accounted for.
50 . // '1' means that there is no difference.)
51 .
```

```
52 . // Means in subgroups:
53 . svy: mean r_bmi r_fs_ko1 r_systbp2, over(vuosi ika_51k)
(running mean on estimation sample)
```

Survey: Mean estimation

Number of strata = 20 Number of obs = 2876
 Number of PSUs = 1877 Population size = 2278.23
 Design df = 1857

- Over: vuosi ika_51k
- _subpop_1: 2000 1
- _subpop_2: 2000 2
- _subpop_3: 2000 3
- _subpop_4: 2000 4
- _subpop_5: 2000 5
- _subpop_6: 2011 1
- _subpop_7: 2011 2
- _subpop_8: 2011 3
- _subpop_9: 2011 4
- _subpop_10: 2011 5

Over	Mean	Linearized Std. Err.	[95% Conf. Interval]	
r_bmi				
_subpop_1	25.43596	.1985189	25.04662	25.82531
_subpop_2	27.41306	.2464322	26.92974	27.89637
_subpop_3	27.42958	.3048862	26.83163	28.02754
_subpop_4	28.21287	.3498487	27.52673	28.89901
_subpop_5	27.4077	.3504255	26.72043	28.09497
_subpop_6	25.36666	.3023903	24.7736	25.95972
_subpop_7	26.6737	.3151941	26.05553	27.29187
_subpop_8	28.07558	.3496211	27.38989	28.76127
_subpop_9	27.09918	.30723	26.49662	27.70173
_subpop_10	27.14575	.4370427	26.2886	28.0029
r_fs_ko1				
_subpop_1	5.332064	.047422	5.239058	5.425069
_subpop_2	5.78677	.0545912	5.679703	5.893836
_subpop_3	5.912476	.0677893	5.779525	6.045428
_subpop_4	5.854731	.0838152	5.690349	6.019112
_subpop_5	5.683559	.0756005	5.535288	5.83183
_subpop_6	5.149644	.0719128	5.008606	5.290683
_subpop_7	5.466661	.0672216	5.334823	5.598499
_subpop_8	5.626499	.0721291	5.485037	5.767962
_subpop_9	5.348159	.0802014	5.190865	5.505453
_subpop_10	5.120947	.1205515	4.884517	5.357378
r_systbp2				
_subpop_1	122.448	.7137873	121.0481	123.8479
_subpop_2	132.5338	.944439	130.6815	134.3861
_subpop_3	139.5595	1.302878	137.0042	142.1148
_subpop_4	147.5556	1.584719	144.4476	150.6637

_subpop_5	148.8314	1.85888	145.1857	152.4771
_subpop_6	121.2528	.9136839	119.4609	123.0448
_subpop_7	127.5088	1.084819	125.3812	129.6364
_subpop_8	136.2299	1.138911	133.9962	138.4636
_subpop_9	138.617	1.471898	135.7303	141.5038
_subpop_10	138.9783	2.218131	134.628	143.3286

```
54 . //test _b[r_fs_ko1:_subpop_2] = _b[r_fs_ko1:_subpop_1]
55 .
56 . // Test the average difference of the two measurements:
57 . // 1. Ignore sampling design and assume independency (WRONG!)
58 . ttest r_systbp2 if osite <=5 & ika >= 30, by(vuosi)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
2000	872	136.6904	.771037	22.76845	135.1771	138.2037
2011	601	133.8619	.7785062	19.08531	132.333	135.3908
combined	1473	135.5363	.5570932	21.38106	134.4435	136.6291
diff		2.82847	1.13152		.6089055	5.048035

diff = mean(2000) - mean(2011) t = 2.4997
 Ho: diff = 0 degrees of freedom = 1471
 Ha: diff < 0 Pr(T < t) = 0.9937
 Ha: diff != 0 Pr(|T| > |t|) = 0.0125
 Ha: diff > 0 Pr(T > t) = 0.0063

```
59 .
60 . // 2. Use the sampling design:
61 . svy, subpop(if osite<=5 & ika >= 30): mean r_systbp2, over(vuosi)
    (running mean on estimation sample)
```

Survey: Mean estimation

Number of strata = 5 Number of obs = 1759
 Number of PSUs = 1110 Population size = 1420.61
 Subpop. no. obs = 1462
 Subpop. size = 1420.61
 Design df = 1105

2000: vuosi = 2000
 2011: vuosi = 2011

over	Mean	Linearized Std. Err.	[95% Conf. Interval]	
r_systbp2				
2000	136.0245	.7302797	134.5916	137.4574
2011	134.0304	.7955502	132.4694	135.5914

Note: 15 strata omitted because they contain no subpopulation members.

```
62 . test _b[2000] = _b[2011]
```

Adjusted Wald test

(1) [r_systbp2]2000 - [r_systbp2]2011 = 0

F(1, 1105) = 4.88
 Prob > F = 0.0274

```
63 .
64 . // 3. Use the old sampling design.
65 . // Compare results with the original Health 2000 survey:
66 . svyset ryvas [pw=all_analysis_w], strata(osite)
```

```

pweight: all_analysis_w
VCE: linearized
Single unit: missing
Strata 1: osite
SU 1: ryvas
FPC 1: <zero>
```

```
67 . svy, subpop(if osite<=5 & ika >= 30): mean r_systbp2, over(vuosi)
(running mean on estimation sample)
```

Survey: Mean estimation

```

Number of strata =      5      Number of obs   =    1759
Number of PSUs  =     65      Population size =  1420.61
                                          Subpop. no. obs =   1462
                                          Subpop. size   =  1420.61
                                          Design df      =     60
```

```

2000: vuosi = 2000
2011: vuosi = 2011
```

Over	Mean	Linearized Std. Err.	[95% Conf. Interval]	
r_systbp2				
2000	136.0245	.9346708	134.1548	137.8941
2011	134.0304	.6754929	132.6792	135.3816

Note: 15 strata omitted because they contain no subpopulation members.

```
68 . test _b[2000] = _b[2011]
```

Adjusted Wald test

(1) [r_systbp2]2000 - [r_systbp2]2011 = 0

```

F( 1, 60) = 3.43
Prob > F = 0.0689
```

```
69 .
70 . // Restore the sampling design:
71 . svyset havtun [pw=all_analysis_w], strata(osite)
```

```

pweight: all_analysis_w
VCE: linearized
Single unit: missing
Strata 1: osite
SU 1: havtun
FPC 1: <zero>
```

```
72 .
73 .
74 . // Proportions:
```

75 . **svy: proportion r_systbp2_01**
 (running proportion on estimation sample)

Survey: Proportion estimation

Number of strata = 20 Number of obs = 2966
 Number of PSUs = 1907 Population size = 2360.4
 Design df = 1887

	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
r_systbp2_01				
0	.7032997	.0102619	.6831739	.7234255
1	.2967003	.0102619	.2765745	.3168261

76 .
 77 . // Tables:
 78 . **svy: tabulate r_systbp2_123 ika_51k**
 (running tabulate on estimation sample)

Number of strata = 20 Number of obs = 2458
 Number of PSUs = 1618 Population size = 2360.3955
 Design df = 1598

Syst.vp (3-luok.)	Ikäryhmä					Total
	1	2	3	4	5	
1	.1588	.0709	.0358	.0165	.0164	.2985
2	.1392	.1641	.1365	.1013	.0691	.6102
3	.0047	.0103	.0239	.0279	.0244	.0913
Total	.3027	.2454	.1963	.1457	.11	1

Key: cell proportions

Pearson:
 Uncorrected chi2(8) = 409.2967
 Design-based F(7.92, 12655.73) = 51.6573 P = 0.0000

79 .
 80 .
 81 . // ESTIMATE CHANGES IN THE POPULATION LEVEL:
 82 . // Basic regression models:
 83 . **svy: regress r_systbp2(ika) i.sukupuoli i.vuosi**
 (running regress on estimation sample)

Survey: Linear regression

Number of strata = 20 Number of obs = 2966
 Number of PSUs = 1907 Population size = 2360.3955
 Design df = 1887
 F(3, 1885) = 152.95
 Prob > F = 0.0000
 R-squared = 0.1841

r_systbp2	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika	.575147	.0273405	21.04	0.000	.5215263	.6287677
2.sukupuoli	-3.289542	.835901	-3.94	0.000	-4.928929	-1.650154
vuosi 2011	-4.598484	.652792	-7.04	0.000	-5.878754	-3.318214
_cons	106.421	1.379779	77.13	0.000	103.7149	109.1271

```
84 .
85 . // Compare results with the old sampling design
86 . // (PSU is the same as in the Health 2000 survey):
87 . svyset ryvas [pw=all_analysis_w], strata(osite)
```

```
    pweight: all_analysis_w
           VCE: linearized
Single unit: missing
  Strata 1: osite
        SU 1: ryvas
        FPC 1: <zero>
```

```
88 . svy: regress r_systbp2 ika i.sukupuoli i.vuosi
    (running regress on estimation sample)
```

Survey: Linear regression

```
Number of strata =      20
Number of PSUs  =     862
Number of obs   =     2966
Population size = 2360.3955
Design df       =      842
F( 3, 840)      =     137.77
Prob > F        =     0.0000
R-squared       =     0.1841
```

r_systbp2	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika	.575147	.0287327	20.02	0.000	.518751	.6315431
2.sukupuoli	-3.289542	.8449197	-3.89	0.000	-4.947938	-1.631146
vuosi 2011	-4.598484	.7516414	-6.12	0.000	-6.073795	-3.123174
_cons	106.421	1.571835	67.70	0.000	103.3358	109.5062

```
89 . // (Note that standard error of 'vuosi' is now larger,
90 . // because intraclass correlation is much stronger
91 . // in the individual level than on the in
92 . // the health center district level.)
93 . // Assume independency (ignore clustering):
94 . svyset _n [pw=all_analysis_w], strata(osite)
```

```
    pweight: all_analysis_w
           VCE: linearized
Single unit: missing
  Strata 1: osite
        SU 1: <observations>
        FPC 1: <zero>
```

```
95 . svy: regress r_systbp2 ika i.sukupuoli i.vuosi
    (running regress on estimation sample)
```

Survey: Linear regression

```
Number of strata =      20
Number of PSUs  =     2966
Number of obs   =     2966
Population size = 2360.3955
Design df       =     2946
F( 3, 2944)     =     169.65
Prob > F        =     0.0000
R-squared       =     0.1841
```

r_systbp2	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika	.575147	.0260877	22.05	0.000	.5239952	.6262989
2.sukupuoli	-3.289542	.748434	-4.40	0.000	-4.757048	-1.822035
vuosi 2011	-4.598484	.7498633	-6.13	0.000	-6.068793	-3.128175
_cons	106.421	1.330463	79.99	0.000	103.8123	109.0297

```
96 . // Reset the sampling design:
97 . svyset havtun [pw=all_analysis_w], strata(osite)
```

```

pweight: all_analysis_w
VCE: linearized
Single unit: missing
Strata 1: osite
SU 1: havtun
FPC 1: <zero>
```

```
98 .
99 . // Estimate changes in population between 2000 and 2011
100 . // Average change:
101 . svy: regress r_systbp2 i.ika_51k i.sukupuoli i.vuosi
(running regress on estimation sample)
```

Survey: Linear regression

```

Number of strata = 20
Number of PSUs = 1907
Number of obs = 2966
Population size = 2360.3955
Design df = 1887
F( 6, 1882) = 86.10
Prob > F = 0.0000
R-squared = 0.1875
```

r_systbp2	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika_51k						
2	9.322023	.8654977	10.77	0.000	7.62459	11.01946
3	16.58889	1.047596	15.84	0.000	14.53432	18.64345
4	22.11816	1.235845	17.90	0.000	19.69439	24.54192
5	22.41873	1.501475	14.93	0.000	19.47401	25.36346
2.sukupuoli	-3.030155	.8379372	-3.62	0.000	-4.673536	-1.386775
vuosi 2011	-4.640027	.6483908	-7.16	0.000	-5.911665	-3.368389
_cons	125.0817	.7241962	172.72	0.000	123.6614	126.502

```
102 .
103 . // Change by age group (interactions):
104 . svy: regress r_systbp2 i.sukupuoli i.vuosi##i.ika_51k
(running regress on estimation sample)
```

Survey: Linear regression

```

Number of strata = 20
Number of PSUs = 1907
Number of obs = 2966
Population size = 2360.3955
Design df = 1887
F( 10, 1878) = 54.21
Prob > F = 0.0000
R-squared = 0.1924
```


r_systbp2	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
2.sukupuoli	-3.098597	.837812	-3.70	0.000	-4.741732	-1.455461
vuosi 2011	-1.164259	1.078904	-1.08	0.281	-3.28023	.9517119
ika_51k 2	10.37638	1.190225	8.72	0.000	8.042085	12.71068
3	17.23193	1.482525	11.62	0.000	14.32437	20.13949
4	25.14089	1.723884	14.58	0.000	21.75997	28.52181
5	25.63396	2.036039	12.59	0.000	21.64084	29.62709
vuosi# ika_51k 2011 2	-3.433677	1.907241	-1.80	0.072	-7.174201	.3068466
2011 3	-2.504025	2.017743	-1.24	0.215	-6.461267	1.453217
2011 4	-7.597391	2.411241	-3.15	0.002	-12.32637	-2.868412
2011 5	-7.997606	2.968342	-2.69	0.007	-13.81918	-2.176029
_cons	124.0467	.8129386	152.59	0.000	122.4524	125.6411

```
105 .
106 . // Predictive margins (model adjusted means):
107 . // 1. Calculate predictive means of SystBP assuming that the
108 . // age distributions are the same for both
109 . // men and women (i.e. the same as in the full data (2000+2011):
110 . margins sukupuoli, vce(unconditional)
```

```
Predictive margins                    Number of obs   =   2966
Subpop. no. of obs =   2432
```

Expression : Linear prediction, predict()

	Margin	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
sukupuoli						
1	134.551	.6049319	222.42	0.000	133.3644	135.7375
2	131.4524	.6341813	207.28	0.000	130.2085	132.6963

```
111 .
112 . // 2. Note that the population distributions do not correspond to
113 . // the population distribution at any calendar year.
114 . // Therefore it is more reasonable to calculate the predicted
115 . // margins in a subgroup defined by the calendar year:
116 . margins sukupuoli, vce(unconditional) subpop(if vuosi==2011)
```

```
Predictive margins                    Number of obs   =   2966
Subpop. no. of obs =   980
```

Expression : Linear prediction, predict()

	Margin	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
sukupuoli						
1	133.2177	.72496	183.76	0.000	131.7957	134.6397
2	130.1191	.71851	181.10	0.000	128.7098	131.5284


```

130 . // (Note that in this case the age distributions are the same,
131 . // that is, of the year 2011, thus the figures are comparable.)
132 .
133 . // 5. Parameters can be estimated in a subpopulation, and the
134 . // predicted margins can be calculated in another subpopulation,
135 . // e.g. in years 2000 and 2011, respectively.
136 . // First, run regression analysis in Health 2000 data:
137 . // svy, subpop(if vuosi==2000): //
> regress r_systbp2 i.ika_5lk i.sukupuoli
(running regress on estimation sample)
    
```

Survey: Linear regression

Number of strata	=	20	Number of obs	=	3172
Number of PSUs	=	1965	Population size	=	2563.1074
			Subpop. no. of obs	=	1452
			Subpop. size	=	1446.343
			Design df	=	1945
			F(5, 1941)	=	79.47
			Prob > F	=	0.0000
			R-squared	=	0.2081

r_systbp2	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika_5lk						
2	10.37437	1.188032	8.73	0.000	8.044415	12.70432
3	17.26102	1.483127	11.64	0.000	14.35234	20.16971
4	25.1481	1.725204	14.58	0.000	21.76465	28.53154
5	25.72199	2.044454	12.58	0.000	21.71244	29.73154
2.sukupuoli	-3.593345	1.000473	-3.59	0.000	-5.55458	-1.631232
_cons	124.2921	.8599577	144.53	0.000	122.6056	125.9787

```

138 . // Then calculate the predicted margins in Health 2011 data
139 . // ("standard population"):
140 . margins sukupuoli, vce(unconditional) subpop(if vuosi==2011) noesample
    
```

Predictive margins	Number of obs	=	3172
	Subpop. no. of obs	=	1186

Expression : Linear prediction, predict()

	Margin	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
sukupuoli						
1	137.8415	.7962611	173.11	0.000	136.2799	139.4031
2	134.2482	.7841146	171.21	0.000	132.7104	135.786

```

141 .
142 .
143 .
144 . // ESTIMATE INDIVIDUAL CHANGES:
145 . // Recall the basic regression models for estimating
146 . // population level changes:
    
```

```
147 . svy: regress r_systbp2 ika i.sukupuoli i.vuosi
(running regress on estimation sample)
```

Survey: Linear regression

```
Number of strata = 20
Number of PSUs = 1907
Number of obs = 2966
Population size = 2360.3955
Design df = 1887
F( 3, 1885) = 152.95
Prob > F = 0.0000
R-squared = 0.1841
```

r_systbp2	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika	.575147	.0273405	21.04	0.000	.5215263	.6287677
2.sukupuoli	-3.289542	.835901	-3.94	0.000	-4.928929	-1.650154
vuosi 2011	-4.598484	.652792	-7.04	0.000	-5.878754	-3.318214
_cons	106.421	1.379779	77.13	0.000	103.7149	109.1271

```
148 .
149 . // Individual-level changes:
150 . // The baseline/background factors are usually assumed to be fixed
151 . // --> we use the baseline age (ika_b1) instead of
152 . // the age at the time of the measurement (ika)
153 . svy: regress r_systbp2 ika_b1 i.sukupuoli i.vuosi
(running regress on estimation sample)
```

Survey: Linear regression

```
Number of strata = 20
Number of PSUs = 1907
Number of obs = 2966
Population size = 2360.3955
Design df = 1887
F( 3, 1885) = 152.95
Prob > F = 0.0000
R-squared = 0.1841
```

r_systbp2	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika_b1	.575147	.0273405	21.04	0.000	.5215263	.6287677
2.sukupuoli	-3.289542	.835901	-3.94	0.000	-4.928929	-1.650154
vuosi 2011	1.728133	.6626119	2.61	0.009	.4286042	3.027662
_cons	106.421	1.379779	77.13	0.000	103.7149	109.1271

```
154 .
155 . // Estimate the average individual changes in age groups:
156 . svy: regress r_systbp2 i.ika_51k_b1 i.sukupuoli i.ika_51k_b1#i.vuosi
(running regress on estimation sample)
```

Survey: Linear regression

```
Number of strata = 20
Number of PSUs = 1907
Number of obs = 2966
Population size = 2360.3955
Design df = 1887
F( 10, 1878) = 43.56
Prob > F = 0.0000
R-squared = 0.1786
```

r_systbp2	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika_51k_b1						
2	10.376	1.18977	8.72	0.000	8.042601	12.70941
3	17.23736	1.48235	11.63	0.000	14.33014	20.14457
4	25.14223	1.724099	14.58	0.000	21.76089	28.52357
5	25.65038	2.036892	12.59	0.000	21.65558	29.64517
2.sukupuoli	-3.19084	.8457367	-3.77	0.000	-4.849517	-1.532162
ika_51k_b1#						
vuosi						
1 2011	4.186589	.838473	4.99	0.000	2.542158	5.831021
2 2011	3.54209	1.353297	2.62	0.009	.8879741	6.196205
3 2011	1.609031	1.990591	0.81	0.419	-2.29496	5.513022
4 2011	-9.942111	3.047133	-3.26	0.001	-15.91822	-3.966006
5 2011	-13.29627	5.450561	-2.44	0.015	-23.98603	-2.606508
_cons	124.0925	.8156831	152.13	0.000	122.4927	125.6922

```

157 .
158 .
159 . // HIERARCHICAL MODELS:
160 . // Define sampling weights (pweight) for all levels of a
161 . // hierarchical model:
162 . g w1=all_analysis_w // for residual
163 . g w2=1 // for havtun
164 . g w3=1 // for hcd
165 . g w4=1 // for osite
166 . // For hierarchical models the original PSU variable
167 . // (ryvas) is better be replaced by HCD:
168 . g hcd=osite
169 . replace hcd=ryvas if osite<=5
    (1919 real changes made)
170 .
171 . // Compare results with a hierarchical (random effects) model
172 . // with 3 levels of hierarchy (strata, HCD and individual)
173 . xtmixed r_systbp2 ika i.sukupuoli i.vuosi || osite: || hcd: || havtun:

```

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0: log restricted-likelihood = -10591.564
 Iteration 1: log restricted-likelihood = -10590.634
 Iteration 2: log restricted-likelihood = -10590.634

Computing standard errors:

Mixed-effects REML regression Number of obs = 2458

Group Variable	No. of Groups	Observations per Group		
		Minimum	Average	Maximum
osite	20	17	122.9	372
hcd	80	6	30.7	261
havg	1618	1	1.5	2

Log restricted-likelihood = -10590.634 wald chi2(3) = 477.98
 Prob > chi2 = 0.0000

r_systbp2	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ika	.554682	.0255974	21.67	0.000	.5045119	.604852
2.sukupuoli	-2.621764	.8353977	-3.14	0.002	-4.259114	-.9844147
vuosi						
2011	-4.514285	.6481715	-6.96	0.000	-5.784678	-3.243892
_cons	107.2449	1.611681	66.54	0.000	104.0861	110.4038

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
osite: Identity				
sd(_cons)	2.340458	.9447642	1.060961	5.163002
hcd: Identity				
sd(_cons)	2.491846	.862459	1.264473	4.910578
havgun: Identity				
sd(_cons)	11.21702	.5415067	10.20435	12.33018
sd(Residual)	14.48449	.3548962	13.80535	15.19705

LR test vs. linear regression: chi2(3) = 159.78 Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

174 . // (weights are not available in this procedure.)

175 . margins sukupuoli#vuosi

Predictive margins Number of obs = 2458

Expression : Linear prediction, fixed portion, predict()

	Margin	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
sukupuoli#						
vuosi						
1 2000	136.6817	.9974228	137.03	0.000	134.7268	138.6366
1 2011	132.1674	1.040262	127.05	0.000	130.1285	134.2063
2 2000	134.0599	.947779	141.45	0.000	132.2023	135.9175
2 2011	129.5456	1.000251	129.51	0.000	127.5852	131.5061

176 . // (Note that no weighting is available here, thus the results

177 . // based on margins do not represent the population.)

178 .

179 . // In Stata version 13 multilevel generalized linear models

180 . // (including e.g. logistic regression) are available, but still

181 . // weights are not available.

182 .

183 . // Generalized linear mixed effects (GLME) model using

184 . // the GLLAMM procedure. For details, see

185 . // <http://www.gllamm.org/>

186 . // and for installation, type command (in Stata)

```

187 . // ssc install gllamm
188 . // then run a linear mixed effects model:
189 . xi: gllamm r_systbp2 i.iika_51k i.sukupuoli i.vuosi, ///
> i(havtun hcd) pweight(w) adapt
i.iika_51k      _Iika_51k_1-5      (naturally coded; _Iika_51k_1 omitted)
i.sukupuoli    _Isukupuoli_1-2    (naturally coded; _Isukupuoli_1 omitted)
i.vuosi        _Ivuosi_2000-2011  (naturally coded; _Ivuosi_2000 omitted)

```

```

Running adaptive quadrature
Iteration 0:    log likelihood = -10198.028
Iteration 1:    log likelihood = -10184.209
Iteration 2:    log likelihood = -10181.115
Iteration 3:    log likelihood = -10138.833
Iteration 4:    log likelihood = -10135.581
Iteration 5:    log likelihood = -10135.536
Iteration 6:    log likelihood = -10135.536

```

```

Adaptive quadrature has converged, running Newton-Raphson
Iteration 0:    log likelihood = -10135.536
Iteration 1:    log likelihood = -10135.536 (backed up)
Iteration 2:    log likelihood = -10135.507
Iteration 3:    log likelihood = -10135.507

```

```

number of level 1 units = 2458
number of level 2 units = 1618
number of level 3 units = 80

```

Condition Number = 92.686508

gllamm model

log likelihood = -10135.507

Robust standard errors

r_systbp2	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Iika_51k_2	9.394758	.7778282	12.08	0.000	7.870243	10.91927
_Iika_51k_3	16.52562	1.264607	13.07	0.000	14.04703	19.0042
_Iika_51k_4	22.07778	1.40326	15.73	0.000	19.32744	24.82812
_Iika_51k_5	21.6883	1.655185	13.10	0.000	18.44419	24.9324
_Isukupuoli_2	-2.479792	.7890602	-3.14	0.002	-4.026322	-.9332626
_Ivuosi_2011	-4.608423	.8990806	-5.13	0.000	-6.370588	-2.846257
_cons	125.8513	.8594459	146.43	0.000	124.1668	127.5358

Variance at level 1

Residual variance

211.15966 (12.574905)

Variances and covariances of random effects

***level 2 (havtun)

var(1): 110.34527 (12.24775)

***level 3 (hcd)

Hierarchical model

var(1): 11.354356 (3.8900085)

```

190 . // (Strata (osite) has been excluded due to
191 . // slow or no convergence.)
192 . // Compare results with design-based analysis:
193 . svy: regress r_systbp2 i.ika_51k i.sukupuoli##i.vuosi
    (running regress on estimation sample)
    
```

Survey: Linear regression

```

Number of strata   =      20           Number of obs     =    2966
Number of PSUs    =    1907           Population size   =  2360.3955
                                           Design df        =    1887
                                           F( 7, 1881)     =    73.82
                                           Prob > F        =    0.0000
                                           R-squared       =    0.1877
    
```

r_systbp2	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika_51k						
2	9.325894	.8650177	10.78	0.000	7.629402	11.02239
3	16.61437	1.048563	15.84	0.000	14.5579	18.67083
4	22.12834	1.234795	17.92	0.000	19.70663	24.55005
5	22.44556	1.50206	14.94	0.000	19.49968	25.39143
2.sukupuoli	-3.448542	.9978742	-3.46	0.001	-5.405595	-1.49149
vuosi						
2011	-5.208987	.9607986	-5.42	0.000	-7.093326	-3.324648
sukupuoli#						
vuosi						
2 2011	1.077662	1.267199	0.85	0.395	-1.407596	3.562921
_cons	125.2916	.7731318	162.06	0.000	123.7753	126.8079

```

194 .
195 . // Predictive margins not are available with gllamm!
196 . // Use the design-based methods (svy: ...) with the margins command.
197 . margins sukupuoli#vuosi
    
```

```

Predictive margins                               Number of obs   =    2966
Model VCE    : Linearized
Expression   : Linear prediction, predict()
    
```

	Margin	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
sukupuoli#						
vuosi						
1 2000	136.5322	.7115169	191.89	0.000	135.1376	137.9267
1 2011	131.3232	.8324231	157.76	0.000	129.6917	132.9547
2 2000	133.0836	.7075411	188.09	0.000	131.6969	134.4704
2 2011	128.9523	.7146619	180.44	0.000	127.5516	130.353

198 .


```

199 . // Try also a logistic regression model for a binary response:
200 . xi: gllamm r_systbp2_01 i.iika_51k i.sukupuoli i.vuosi, ///
> i(havtun hcd) pweight(w) adapt ///
> link(logit) fam(binom) eform
i.iika_51k      _Iika_51k_1-5      (naturally coded; _Iika_51k_1 omitted)
i.sukupuoli    _Isukupuoli_1-2    (naturally coded; _Isukupuoli_1 omitted)
i.vuosi         _Ivuosi_2000-2011   (naturally coded; _Ivuosi_2000 omitted)

```

Running adaptive quadrature
Iteration 0: log likelihood = -1238.2158
Iteration 1: log likelihood = -1230.8731
Iteration 2: log likelihood = -1221.7653
Iteration 3: log likelihood = -1221.4309
Iteration 4: log likelihood = -1221.43

Adaptive quadrature has converged, running Newton-Raphson
Iteration 0: log likelihood = -1221.43
Iteration 1: log likelihood = -1221.43 (backed up)
Iteration 2: log likelihood = -1221.4298
Iteration 3: log likelihood = -1221.4298

number of level 1 units = 2458
number of level 2 units = 1618
number of level 3 units = 80

Condition Number = 8.1113483

gllamm model

log likelihood = -1221.4298

Robust standard errors

r_systbp2_01	exp(b)	Std. Err.	z	P> z	[95% Conf. Interval]
_Iika_51k_2	6.155594	1.205726	9.28	0.000	4.193156 9.036471
_Iika_51k_3	13.49301	3.24265	10.83	0.000	8.424561 21.61077
_Iika_51k_4	30.97228	8.588389	12.38	0.000	17.98627 53.33413
_Iika_51k_5	25.97294	6.745387	12.54	0.000	15.61194 43.21009
_Isukupuoli~2	.8065265	.1125031	-1.54	0.123	.6135978 1.060116
_Ivuosi_2011	.4145293	.0574713	-6.35	0.000	.315895 .543961

Variances and covariances of random effects

***level 2 (havtun)

var(1): 1.9780901 (.43517351)

***level 3 (hcd)

var(1): .13297526 (.05645376)

```

201 . // (Note that these were the 'conditional effect' OR's.)
202 .
203 . // Compare with design based model:
204 . // svy: logit r_systbp2_01 i.ika_51k i.sukupuoli i.vuosi
205 . svy: logistic r_systbp2_01 i.ika_51k i.sukupuoli i.vuosi
    (running logistic on estimation sample)
    
```

Survey: Logistic regression

```

Number of strata =      20
Number of PSUs  =     1907
Number of obs   =      2966
Population size =  2360.3955
Design df       =      1887
F( 6, 1882)     =      45.66
Prob > F        =      0.0000
    
```

r_systbp2_01	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika_51k						
2	4.214346	.6936257	8.74	0.000	3.051714	5.819915
3	7.342483	1.244595	11.76	0.000	5.265827	10.2381
4	13.62897	2.419992	14.71	0.000	9.621098	19.3064
5	11.97465	2.197747	13.53	0.000	8.354853	17.16275
2.sukupuoli	.8495689	.09091	-1.52	0.128	.6887403	1.047953
vuosi						
2011	.5188582	.0488269	-6.97	0.000	.4314151	.6240251

```

206 . // (Note that these were the 'marginal effect' OR's,
207 . // which are closer to 1.)
208 .
209 . // Calculate predicted margins to get model adjusted means:
210 . margins sukupuoli#vuosi, vce(unconditional)
    
```

```

Predictive margins
Subpop. no. of obs =      2432
Number of obs      =      2966
    
```

Expression : Pr(r_systbp2_01), predict()

	Margin	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
sukupuoli#vuosi						
1 2000	.3606513	.0169831	21.24	0.000	.3273437	.3939588
1 2011	.2430149	.0162245	14.98	0.000	.2111951	.2748347
2 2000	.3296006	.0154144	21.38	0.000	.2993695	.3598316
2 2011	.2174365	.0146026	14.89	0.000	.1887976	.2460753

```

211 . // Calculate differences and rates:
212 . margins sukupuoli#vuosi, vce(unconditional) post coeflegend
    
```

```

Predictive margins
Subpop. no. of obs =      2432
Number of obs      =      2966
    
```

Expression : Pr(r_systbp2_01), predict()

	Margin	Legend
sukupuoli# vuosi		
1 2000	.3606513	_b[1bn.sukupuoli#2000bn.vuosi]
1 2011	.2430149	_b[1bn.sukupuoli#2011.vuosi]
2 2000	.3296006	_b[2.sukupuoli#2000bn.vuosi]
2 2011	.2174365	_b[2.sukupuoli#2011.vuosi]

```

213 . // ('coeflegend' shows how to refer to the predicted margins.)
214 nlcom (difference: _b[1bn.sukupuoli#2011.vuosi] - _b[1bn.sukupuoli#2000bn.vuo
> si]) ///
> (ratio: _b[1bn.sukupuoli#2011.vuosi] / _b[1bn.sukupuoli#2000bn.vuosi])

```

difference: _b[1bn.sukupuoli#2011.vuosi] - _b[1bn.sukupuoli#2000bn.vuosi]
ratio: _b[1bn.sukupuoli#2011.vuosi] / _b[1bn.sukupuoli#2000bn.vuosi]

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
difference	-.1176364	.0161302	-7.29	0.000	-.1492712	-.0860015
ratio	.6738224	.0389951	17.28	0.000	.5973443	.7503005

```

215 .
216 . // Test if the HCD-level random effects are needed:
217 . // Drop the hcd in the hierarchy and rerun the GLME model:
218 . xi: gllamm r_systbp2_01 i.ika_51k i.sukupuoli i.vuosi, ///
> i(havtun) pweight(w) adapt ///
> link(logit) fam(binom) eform
i.ika_51k _Iika_51k_1-5 (naturally coded; _Iika_51k_1 omitted)
i.sukupuoli _Isukupuoli_1-2 (naturally coded; _Isukupuoli_1 omitted)
i.vuosi _Ivuosi_2000-2011 (naturally coded; _Ivuosi_2000 omitted)

```

Running adaptive quadrature

```

Iteration 0: log likelihood = -1242.9411
Iteration 1: log likelihood = -1237.4151
Iteration 2: log likelihood = -1225.9445
Iteration 3: log likelihood = -1225.4533
Iteration 4: log likelihood = -1225.452
Iteration 5: log likelihood = -1225.4518

```

Adaptive quadrature has converged, running Newton-Raphson

```

Iteration 0: log likelihood = -1225.4518
Iteration 1: log likelihood = -1225.4518 (backed up)
Iteration 2: log likelihood = -1225.4517

```

number of level 1 units = 2458
number of level 2 units = 1618

Condition Number = 9.5451216

gllamm model

log likelihood = -1225.4517

Robust standard errors

r_systbp2_01	exp(b)	Std. Err.	z	P> z	[95% Conf. Interval]	
_Iika_51k_2	6.245191	1.377402	8.31	0.000	4.05331	9.622362
_Iika_51k_3	13.86792	3.442627	10.59	0.000	8.525222	22.55885
_Iika_51k_4	32.32683	9.022913	12.45	0.000	18.70595	55.86586
_Iika_51k_5	27.60137	7.775122	11.78	0.000	15.89106	47.94115
_Isukupuoli~2	.7787546	.111247	-1.75	0.080	.5885785	1.030379
_Ivuosi_2011	.410384	.0547173	-6.68	0.000	.316008	.5329454

Variances and covariances of random effects

***level 2 (hvtun)

var(1): 2.1575989 (.4741731)

```

219 . // Perform LR test:
220 . // Find the 'log likelihood' values of in the above two glme results:
221 . // -1221.4298 and -1225.4517
222 . // Calculate 2 * (-1221.4298 - -1225.4517) = 8.0438
223 . // Compare with the chiSq (df=1) distribution --> p-value is 0.0016684
224 . disp 1-chi2tail(2 * (-1221.4298 - -1225.4517), 1)
    .0016684

225 . // (The HCD level variation is statistically significant and
226 . // should be kept in the model.)
227 .
228 . // For other design-based regression analyses, use for example
229 . // Multinomial logistic regression analysis:
230 . svy: mlogit r_systbp2_123 i.ika_51k i.sukupuoli i.vuosi
    (running mlogit on estimation sample)
    
```

Survey: Multinomial logistic regression

Number of strata	=	20	Number of obs	=	2966
Number of PSUs	=	1907	Population size	=	2360.3955
			Design df	=	1887
			F(12, 1876)	=	30.40
			Prob > F	=	0.0000

r_systbp~123	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
1						
ika_51k						
2	-1.00128	.1154779	-8.67	0.000	-1.227758	-.7748024
3	-1.544841	.1514835	-10.20	0.000	-1.841933	-1.247748
4	-2.023007	.1930036	-10.48	0.000	-2.40153	-1.644484
5	-1.711898	.1974047	-8.67	0.000	-2.099053	-1.324744
2.sukupuoli	.6640344	.1105431	6.01	0.000	.4472347	.880834
vuosi						
2011	.1636671	.0955536	1.71	0.087	-.0237348	.3510689
_cons	-.2526809	.0999523	-2.53	0.012	-.4487095	-.0566523
2	(base outcome)					
3						
ika_51k						
2	.6467188	.3764094	1.72	0.086	-.0915036	1.384941
3	1.738035	.3395164	5.12	0.000	1.072168	2.403902
4	2.232169	.3379807	6.60	0.000	1.569314	2.895024
5	2.455822	.3359382	7.31	0.000	1.796972	3.114671
2.sukupuoli	.2545041	.1592488	1.60	0.110	-.0578182	.5668264
vuosi						
2011	-.8446029	.1532073	-5.51	0.000	-1.145076	-.5441294
_cons	-3.31179	.3227269	-10.26	0.000	-3.944729	-2.678851

```
231 . // Ordinal logistic regression analysis:
232 . svy: ologit r_systbp2_123 i.ika_51k i.sukupuoli i.vuosi
(running ologit on estimation sample)
```

Survey: Ordered logistic regression

```
Number of strata = 20
Number of PSUs = 1907
Number of obs = 2966
Population size = 2360.3955
Design df = 1887
F( 6, 1882) = 54.90
Prob > F = 0.0000
```

r_systbp~123	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
ika_51k						
2	.9822239	.1056043	9.30	0.000	.7751106	1.189337
3	1.75698	.1349019	13.02	0.000	1.492407	2.021552
4	2.324685	.1509542	15.40	0.000	2.02863	2.62074
5	2.375727	.1811681	13.11	0.000	2.020416	2.731038
2.sukupuoli	-.4448969	.0962525	-4.62	0.000	-.6336694	-.2561244
vuosi 2011	-.4019241	.0803586	-5.00	0.000	-.5595252	-.2443231
/cut1	-.2320729	.0923404	-2.51	0.012	-.4131728	-.050973
/cut2	3.403021	.1141118	29.82	0.000	3.179222	3.62682

```
233 .
234 . log close
name: <unnamed>
log: M:\Data\T2000\resurvey\koulutus_2013_06\t2011.smcl
log type: smcl
closed on: 15 Jun 2013, 19:36:32
```