

## STATA RESULTS FOR EXEMPLAR 3

Black code is comments

Red code commands

Blue code results

```

. /*-----
> first set up the survey design and view its properties with svydes
> note that we need strata within regions (regstrat)
> -----*/

. svyset [pwei=weighta],psu(psu) strata(regstrat)
pweight is weighta
strata is regstrat
psu is psu

. svydes

pweight:  weighta
Strata:   regstrat
PSU:     psu

                #Obs per PSU
Strata
regstrat  #PSUs  #Obs   min   mean   max
-----
      101      2    48    23   24.0   25
      102      2    45    21   22.5   24
      103      2    58    18   29.0   40
      104      2    65    26   32.5   39
lines missed out here
      415      3    91    24   30.3   37
lines missed out here
      718      2    57    24   28.5   33
      719      2    51    24   25.5   27
-----
      154     312   9047   12   29.0   43

. /*-----
> you should find that you have strata each with 2 or (in a few cases)
> 3 PSUs
>
> Now get proportions in cigarette smoking categories and their standard errors
> -----*/

. svyprop cigst1

-----
pweight:  weighta                Number of obs   =    9047
Strata:   regstrat              Number of strata =     154
PSU:     psu                    Number of PSUs  =     312
                                           Population size = 9006.178
-----

```

## Survey proportions estimation

```

+-----+
|               cigst1   Obs   Est. Prop.   Std. Err. |
+-----+

```

Refused/Not answered	14	0.001534	0.000491
Dont know	16	0.002548	0.000716
schedule not obtained	3	0.000601	0.000371
not applicable	3	0.000230	0.000133
Never smoked cigarettes at all	3711	0.436668	0.006127
-----			
Used to smoke cigarettes occasionally	269	0.030702	0.002425
Used to smoke cigarettes regularly	1895	0.196159	0.004592
Current cigarette smoker	3136	0.331559	0.005967

```

. /*-----
> svyprop does not give design effects or confidence intervals
> to get these for smokers you need to recode
> to a 0/1 variable and get its mean value
> -----*/
. recode cigst1 (-9 -8 -6 =.) (-1 1 2 3=0) (4=1),gen(smoker)
(5911 differences between cigst1 and smoker)

. svyprop smoker

```

```

-----
pweight:  weighta          Number of obs      =      9014
Strata:   regstrat        Number of strata   =       154
PSU:      psu             Number of PSUs    =       312
                               Population size     = 8964.0037
-----

```

Survey proportions estimation

smoker	Obs	Est. Prop.	Std. Err.
0	5878	0.666881	0.006010
1	3136	0.333119	0.006010

```

. svymean smoker,deff deff ci
. /*-----
> To investigate the effect of other survey designs
> one can redo the svyset command
> BUT before rerunning we need to clear previous settings
> -----*/
. /*-----first just weights-----*/
. svyset, clear(all)
no variables are set

. svyset [pwei=weighta]

```

pweight is weighta

```

. svymean smoker,deff deff

```

## Survey mean estimation

```

pweight:  weighta                Number of obs    =      9014
Strata:   <one>                  Number of strata =         1
PSU:      <observations>         Number of PSUs   =      9014
                                           Population size  = 8964.0037

```

```

-----+-----
      Mean |      Estimate      Std. Err.      Deff      Deft
-----+-----
smoker |      .3331194      .0057008      1.318523      1.14827
-----+-----

```

```
. /*-----then add strata-----*/
```

```
. svyset, clear(all)
```

```
no variables are set
```

```
. svyset [pwei=weighta],strata(regstrat)
```

```
pweight is weighta
```

```
strata is regstrat
```

```
. svymean smoker,deff deft
```

## Survey mean estimation

```

pweight:  weighta                Number of obs    =      9014
Strata:   regstrat              Number of strata =      154
PSU:      <observations>         Number of PSUs   =      9014
                                           Population size  = 8964.0037

```

```

-----+-----
      Mean |      Estimate      Std. Err.      Deff      Deft
-----+-----
smoker |      .3331194      .0056322      1.286988      1.134455
-----+-----

```

```
. /*-----now the full design as before-----*/
```

```
. svyset, clear(all)
```

```
no variables are set
```

```
. svyset [pwei=weighta],strata(regstrat) psu(psu)
```

```
pweight is weighta
```

```
strata is regstrat
```

```
psu is psu
```

```
. svymean smoker,deff deft
```

```
.
```

```
. /*-----
```

```
> now looking at rates by sex
```

```
> -----*/
```

```
. svymean smoker, by(sex)
```

## Survey mean estimation

```
pweight:  weighta      Number of obs   =   9014
Strata:   regstrat    Number of strata =   154
PSU:     psu          Number of PSUs  =   312
                          Population size = 8964.0037
```

Mean	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
smoker						
	male	.3419507	.0084952	.3251718	.3587296	1.419651
	female	.3245964	.0078806	.3090315	.3401614	1.29928

```
. /*----- to get a test of differences by sex use lincom
>           for linear combinations-----*/
. lincom [smoker]male-[smoker]female
```

( 1) [smoker]male - [smoker]female = 0

Mean	Estimate	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	.0173543	.0111258	1.56	0.121	-.0046203	.0393288

```
. /*-----
> and by adults in the household
> -----*/
. svymean smoker, by(nofad)
```

Survey mean estimation

```
pweight:  weighta      Number of obs   =   9014
Strata:   regstrat    Number of strata =   154
PSU:     psu          Number of PSUs  =   312
                          Population size = 8964.0037
```

Mean	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
smoker						
	nofad==1	.4408193	.0099747	.4211183	.4605202	.6521785
	nofad==2	.3189418	.0073425	.3044398	.3334438	1.219789
	nofad==3	.2887937	.0145437	.2600685	.3175188	1.607311
	nofad==4	.2894893	.0283638	.2334682	.3455104	2.801671
	nofad==5	.3479849	.0740851	.2016601	.4943097	3.931923
	nofad==6	0	0	0	0	.
	nofad==7	0	0	0	0	.
	nofad==8	.617777	.3339362	-.0417778	1.277332	6.100542
	nofad==9	0	0	0	0	.

```
. /*-----and compare nofad=1 with nofad=2-----*/
. lincom [smoker]1-[smoker]2
```

( 1) [smoker]1 - [smoker]2 = 0

Mean	Estimate	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	.1218775	.012969	9.40	0.000	.0962625	.1474925

```
. /*-----
> smoking rates by region or health board are also easily calculated
> and lincom can give the comparisons between any pair
> or other combination
>
> -----*/
. svymean smoker, by(region)
```

Survey mean estimation

```
pweight:  weighta      Number of obs   =      9014
Strata:   regstrat     Number of strata =      154
PSU:     psu           Number of PSUs  =      312
                          Population size = 8964.0037
```

Mean	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
smoker						
	Highland	.3278274	.0250704	.2783111	.3773436	1.375797
	Grampian	.3267013	.0160343	.295032	.3583705	1.900358
	Lothian_	.3213779	.0114178	.2988266	.3439292	1.198679
	Borders,	.2893518	.0228964	.2441293	.3345743	1.128549
	Glagow	.3633258	.0164955	.3307457	.395906	1.845421
	Lanarksh	.3425938	.0131008	.3167184	.3684691	1.256116
	Forth_Va	.3273929	.0154412	.296895	.3578907	1.342229

```
. svymean smoker, by(hboard)
```

Survey mean estimation

```
pweight:  weighta      Number of obs   =      9014
Strata:   regstrat     Number of strata =      154
PSU:     psu           Number of PSUs  =      312
```

Population size = 8964.0037

Mean	Subpop.	Estimate	Std. Err.	[95% Conf. Interval]		Deff
smoker						
	Ayreshir	.3405227	.0249012	.2913406	.3897047	2.091303
	Borders	.2751782	.0339687	.2080869	.3422696	1.216756
	Argyll_& Fife	.3377994	.0226662	.2930315	.3825673	1.405142
	Greater_ Highland	.3514359	.0179934	.3158974	.3869745	1.048564
	Lanarksh	.3633258	.0164955	.3307457	.395906	1.845421
	Grampian	.3594827	.0311305	.2979971	.4209684	1.563275
	Orkney	.3443544	.0187207	.3073794	.3813295	1.382976
	Lothian	.2982887	.0189685	.2608242	.3357531	1.442587
	Tayside	.2323456	.0107839	.2110463	.2536448	.0225086
	Forth_Va	.3038667	.0152057	.2738341	.3338993	1.385
	Western_ Dumfries	.3570112	.024537	.3085483	.4054741	2.063279
	Shetland	.317252	.0228504	.2721205	.3623836	1.513442
		.2505246	.0263358	.198509	.3025402	.1420666
		.3021828	.0269545	.2489452	.3554204	.8004963
		.1831689	.0670901	.0506597	.3156781	1.141379

```
. lincom [smoker]Fife-[smoker]Lothian
```

( 1) [smoker]Fife - [smoker]Lothian = 0

Mean	Estimate	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	.0475692	.0237606	2.00	0.047	.0006399	.0944985

```
. lincom [smoker]Lanarksh-[smoker]Ayreshir
```

( 1) - [smoker]Ayreshir + [smoker]Lanarksh = 0

Mean	Estimate	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	.0038318	.0349296	0.11	0.913	-.0651573	.0728209

```
. /*----- sorry about spelling mistake - in original file-----*/
. /*-----
> now logistic regressions to predict smoking
>
```

> To use categorical variables you must first generate a set of dummy variables  
 > here for number of adults

```
> -----*/
. tabulate nofad,generate(nofad)
```

Number of adults.	Freq.	Percent	Cum.
1	3,046	33.67	33.67
2	4,613	50.99	84.66
3	992	10.96	95.62
4	330	3.65	99.27
5	56	0.62	99.89
6	6	0.07	99.96
7	1	0.01	99.97
8	2	0.02	99.99
9	1	0.01	100.00
Total	9,047	100.00	

```
. /*-----*/
> check the data set to see the new variables
> as there are so few households of more than 5
> it seems sensible to group them together
> and then to carry out the regression
> -----*/
```

```
. replace nofad5=1 if nofad>5
(10 real changes made)
```

```
. /*---regressions include the comparisons with nofad1 only-----*/
. svylogit smoker nofad2 nofad3 nofad4
```

Survey logistic regression

```
pweight:  weighta          Number of obs   =      9014
Strata:   regstrat        Number of strata =      154
PSU:     psu              Number of PSUs  =      312
                          Population size = 8964.0037
                          F( 3, 156) = 25.41
                          Prob > F = 0.0000
```

smoker	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
nofad2	-.4625912	.0598125	-7.73	0.000	-.5807264 - .3444561
nofad3	-.6052021	.0829158	-7.30	0.000	-.7689683 - .4414358
nofad4	-.6018177	.147868	-4.07	0.000	-.8938706 - .3097647
_cons	-.296048	.0465228	-6.36	0.000	-.3879349 - .2041611

```
. /*----- we can compare with simple logistic regression-----
> -----use coef to get comaparable results to the svy command----*/
. logistic smoker nofad2 nofad3 nofad4,coef
```

```

Logistic regression                               Number of obs   =       9014
                                                  LR chi2(3)      =       143.44
                                                  Prob > chi2     =       0.0000
Log likelihood =  -5752.579                    Pseudo R2      =       0.0123

```

```

-----+-----
smoker |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
nofad2 |   -0.5172813  .0482623   -10.72  0.000   -0.6118737   -0.4226889
nofad3 |   -0.6473839  .0795895    -8.13  0.000   -0.8033764   -0.4913914
nofad4 |   -0.6380871  .1279033    -4.99  0.000   -0.8887729   -0.3874012
_cons  |   -0.2803519  .0362628    -7.73  0.000   -0.3514257   -0.2092781
-----+-----

```

```

. /*----- and we can get more complicated models
>           looking at joint effect of age group sex
>           and number of adults
> Test commands can be used to check if variables are significant in
> the larger models
> -----*/
. tabulate hboard,generate(hboard)

```

```

-----+-----
Health Board |      Freq.   Percent   Cum.
-----+-----
Ayresshire & Arran |      744     8.22     8.22
  Borders |      388     4.29    12.51
Argyll & Clyde |      614     6.79    19.30
  Fife |      662     7.32    26.62
Greater Glasgow |     1,294    14.30    40.92
  Highland |      681     7.53    48.45
  Lanarkshire |      871     9.63    58.07
  Grampian |      726     8.02    66.10
  Orkney |         63     0.70    66.80
  Lothian |     1,174    12.98    79.77
  Tayside |      725     8.01    87.79
  Forth Valley |      509     5.63    93.41
  Western Isles |         98     1.08    94.50
Dumfries & Galloway |      438     4.84    99.34
  Shetland |         60     0.66   100.00
-----+-----
Total |     9,047   100.00

```

```

. tabulate ageg,generate(ageg)

```

```

-----+-----
ageg |      Freq.   Percent   Cum.
-----+-----
16-19 |      391     4.32     4.32
25-29 |      536     5.92    10.25
35-39 |      765     8.46    18.70
45-49 |      973    10.75    29.46
55-59 |      984    10.88    40.33
65-69 |      852     9.42    49.75
70-74 |      759     8.39    58.14
60-64 |      831     9.19    67.33
50-54 |      742     8.20    75.53

```

40-44	750	8.29	83.82
30-34	760	8.40	92.22
20-24	704	7.78	100.00
-----			
Total	9,047	100.00	

```
. tabulate sex,generate(sex)
```

Sex of respondent from household grid. 0	Freq.	Percent	Cum.
male	3,941	43.56	43.56
female	5,106	56.44	100.00
-----			
Total	9,047	100.00	

```
. svylogit smoker nofad2 nofad3 nofad4 sex2 ageg2-ageg12 hboard2-hboard15
```

Survey logistic regression

```
pweight:  weighta      Number of obs   =      9014
Strata:   regstrat     Number of strata =      154
PSU:     psu           Number of PSUs  =      312
                               Population size = 8964.0037
                               F( 29, 130) = 11.94
                               Prob > F = 0.0000
```

smoker	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
nofad2	-.5612387	.0609933	-9.20	0.000	-.681706 - .4407714
nofad3	-.7485357	.0909213	-8.23	0.000	-.9281135 - .5689578
nofad4	-.7860213	.1571321	-5.00	0.000	-1.096372 - .475671
sex2	-.1193292	.0521901	-2.29	0.024	-.2224095 - .0162489
ageg2	.4888972	.1821246	2.68	0.008	.1291843 .8486101
ageg3	.2063644	.1559656	1.32	0.188	-.1016821 .5144109
ageg4	.3205004	.1600491	2.00	0.047	.0043888 .6366121
ageg5	.1117259	.1457159	0.77	0.444	-.1760764 .3995282
ageg6	.2468631	.1576949	1.57	0.119	-.064599 .5583251
ageg7	.1650847	.1686705	0.98	0.329	-.1680551 .4982244
ageg8	.1918504	.1501647	1.28	0.203	-.1047388 .4884395
ageg9	.1454189	.1645114	0.88	0.378	-.1795063 .4703441
ageg10	-.1560555	.1598613	-0.98	0.330	-.4717963 .1596854
ageg11	-.4393657	.1757703	-2.50	0.013	-.7865283 -.0922032
ageg12	-.7704922	.1644425	-4.69	0.000	-1.095281 -.4457032
hboard2	-.2949776	.203661	-1.45	0.149	-.6972269 .1072717
hboard3	-.0308084	.145793	-0.21	0.833	-.318763 .2571462
hboard4	.017459	.1351148	0.13	0.897	-.2494052 .2843233
hboard5	.0521162	.1321553	0.39	0.694	-.2089026 .3131351
hboard6	.0988347	.1723596	0.57	0.567	-.2415914 .4392609
hboard7	.0257908	.1519292	0.17	0.865	-.2742833 .3258648

hboard8		-.2064544	.1482992	-1.39	0.166	-.4993589	.0864502
hboard9		-.5131615	.1361878	-3.77	0.000	-.7821449	-.2441781
hboard10		-.2055899	.1291238	-1.59	0.113	-.4606214	.0494415
hboard11		.0477432	.1538141	0.31	0.757	-.2560537	.3515401
hboard12		-.1433606	.1576678	-0.91	0.365	-.4547691	.1680478
hboard13		-.4630294	.1779328	-2.60	0.010	-.814463	-.1115958
hboard14		-.1637861	.1662252	-0.99	0.326	-.4920962	.164524
hboard15		-.8051732	.4804435	-1.68	0.096	-1.754093	.1437469
_cons		-.196425	.1790967	-1.10	0.274	-.5501576	.1573076

---

```
. test sex2
```

Adjusted Wald test

```
( 1) sex2 = 0
```

```

F( 1, 158) = 1.03
Prob > F = 0.3127

```

```
. test ageg2 ageg3 ageg4 ageg5 ageg6 ageg7 ageg8 ageg9 ageg10 ageg11 ageg12
```

Adjusted Wald test

```

( 1) ageg2 = 0
( 2) ageg3 = 0
( 3) ageg4 = 0
( 4) ageg5 = 0
( 5) ageg6 = 0
( 6) ageg7 = 0
( 7) ageg8 = 0
( 8) ageg9 = 0
( 9) ageg10 = 0
(10) ageg11 = 0
(11) ageg12 = 0

```

```

F( 11, 148) = 6.42
Prob > F = 0.0000

```

```

. /*-----
> get dummies for the age sex interaction
> -----*/
. generate ageg2s=ageg2*(sex==1)
. generate ageg3s=ageg3*(sex==1)
. generate ageg4s=ageg4*(sex==1)
. generate ageg5s=ageg5*(sex==1)
. generate ageg6s=ageg6*(sex==1)

```

```
. generate ageg7s=ageg7*(sex==1)
. generate ageg8s=ageg8*(sex==1)
. generate ageg9s=ageg9*(sex==1)
. generate ageg10s=ageg10*(sex==1)
. generate ageg11s=ageg11*(sex==1)
. generate ageg12s=ageg12*(sex==1)
```

```
svylogit smoker nofad2 nofad3 nofad4 sex2 ageg2-ageg12 hboard2-hboard15 ageg2s-
ageg12s
```

```
Survey logistic regression
```

```
pweight:  weighta          Number of obs   =       9014
Strata:   regstrat        Number of strata =        154
PSU:     psu              Number of PSUs  =         312
                               Population size = 8964.0037
                               F( 40, 119) =       7.71
                               Prob > F      =       0.0000
```

smoker	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
nofad2	-.5649544	.0608856	-9.28	0.000	-.6852091 -.4446997
nofad3	-.7536616	.0926128	-8.14	0.000	-.9365805 -.5707427
nofad4	-.8098296	.1577345	-5.13	0.000	-1.12137 -.4982894
sex2	.2744983	.2710147	1.01	0.313	-.2607807 .8097773
ageg2	.0620744	.2462163	0.25	0.801	-.4242254 .5483743
ageg3	.0229399	.2295454	0.10	0.921	-.4304334 .4763131
ageg4	-.0247639	.2006926	-0.12	0.902	-.4211503 .3716225
ageg5	-.1542414	.1982643	-0.78	0.438	-.5458315 .2373488
ageg6	.0865896	.2108846	0.41	0.682	-.3299268 .5031061
ageg7	-.0092496	.224278	-0.04	0.967	-.4522192 .4337201
ageg8	.0533689	.1981527	0.27	0.788	-.3380009 .4447388
ageg9	.1510162	.2365119	0.64	0.524	-.3161166 .618149
ageg10	-.5590369	.2192275	-2.55	0.012	-.9920313 -.1260425
ageg11	-.5204729	.2332741	-2.23	0.027	-.9812107 -.0597351
ageg12	-.8953448	.2270752	-3.94	0.000	-1.343839 -.4468504
hboard2	-.3028365	.207801	-1.46	0.147	-.7132627 .1075896
hboard3	-.0337248	.1478331	-0.23	0.820	-.3257087 .2582592
hboard4	.0032226	.1385716	0.02	0.981	-.2704691 .2769143
hboard5	.0453948	.1320654	0.34	0.732	-.2154465 .306236
hboard6	.0938063	.1739568	0.54	0.590	-.2497744 .437387
hboard7	.0224807	.1529716	0.15	0.883	-.2796522 .3246136
hboard8	-.2122447	.1511458	-1.40	0.162	-.5107715 .0862821
hboard9	-.5275344	.1574306	-3.35	0.001	-.8384743 -.2165945
hboard10	-.2124226	.1302862	-1.63	0.105	-.4697497 .0449046
hboard11	.0445227	.1562242	0.28	0.776	-.2640345 .35308
hboard12	-.1588901	.1584109	-1.00	0.317	-.4717662 .153986
hboard13	-.4628882	.1811377	-2.56	0.012	-.8206519 -.1051246
hboard14	-.1826604	.1682906	-1.09	0.279	-.5150499 .1497291
hboard15	-.8062307	.4707845	-1.71	0.089	-1.736073 .1236121

ageg2s		.824665	.3715844	2.22	0.028	.0907517	1.558578
ageg3s		.3504848	.3230779	1.08	0.280	-.2876237	.9885933
ageg4s		.6673566	.2912065	2.29	0.023	.092197	1.242516
ageg5s		.5117199	.2978448	1.72	0.088	-.076551	1.099991
ageg6s		.3084944	.3122679	0.99	0.325	-.3082634	.9252521
ageg7s		.3364531	.3147512	1.07	0.287	-.2852094	.9581156
ageg8s		.2614396	.3404926	0.77	0.444	-.4110647	.9339438
ageg9s		-.0546145	.3289639	-0.17	0.868	-.7043484	.5951195
ageg10s		.7881987	.3413227	2.31	0.022	.114055	1.462342
ageg11s		.1063126	.3237929	0.33	0.743	-.5332082	.7458333
ageg12s		.1880666	.3350127	0.56	0.575	-.4736143	.8497476
_cons		-.3746603	.2330058	-1.61	0.110	-.8348683	.0855476

```
-----
. test ageg2s ageg3s ageg4s ageg5s ageg6s ageg7s ageg8s ageg9s ageg10s ageg11s
ageg12s
```

```
Adjusted Wald test
```

```
( 1) ageg2s = 0
( 2) ageg3s = 0
( 3) ageg4s = 0
( 4) ageg5s = 0
( 5) ageg6s = 0
( 6) ageg7s = 0
( 7) ageg8s = 0
( 8) ageg9s = 0
( 9) ageg10s = 0
(10) ageg11s = 0
(11) ageg12s = 0
      F( 11, 148) = 2.02
      Prob > F = 0.0303
```

```
/*-----
```

```
Shows little evidence of any difference in pattern by age for men and women
once adjusted for no of adults and health board
```

```
-----*/
```