

**STATA results for exemplar5**

```

/*-----
> first define the design - only need weighting
> -----*/
. svyset [pweight=weight]
pweight is weight

. /*-----
> now get proportions of various categories
> compared with unweighted tables
> -----*/
. svyprop q85a

```

```

-----
pweight:  weight                Number of obs   =      358
Strata:   <one>                  Number of strata =        1
PSU:     <observations>          Number of PSUs  =      358
                                           Population size = 430.92195
-----

```

Survey proportions estimation

q85a	Obs	Est. Prop.	Std. Err.
never used	210	0.583114	0.027406
tried once or twice	90	0.249705	0.024043
use daily	12	0.034062	0.010116
use weekly	8	0.024087	0.008803
used in last month	11	0.027770	0.008459
used more than a month ago	27	0.081263	0.015942

```
. tabulate q85a
```

q85a	Freq.	Percent	Cum.
never used	210	58.66	58.66
tried once or twice	90	25.14	83.80
use daily	12	3.35	87.15
use weekly	8	2.23	89.39
used in last month	11	3.07	92.46
used more than a month ago	27	7.54	100.00
Total	358	100.00	

```
. svyprop q85b
```

```

-----
pweight:  weight                Number of obs   =      338
Strata:   <one>                  Number of strata =        1
PSU:     <observations>          Number of PSUs  =      338
                                           Population size = 407.87389
-----

```

Survey proportions estimation

q85b	Obs	Est. Prop.	Std. Err.
1	281	0.820471	0.022514
2	32	0.097191	0.016914
3	1	0.002725	0.002726
4	1	0.004927	0.004918
5	5	0.016345	0.007902
6	18	0.058341	0.014174

. tabulate q85b

q85b	Freq.	Percent	Cum.
1	281	83.14	83.14
2	32	9.47	92.60
3	1	0.30	92.90
4	1	0.30	93.20
5	5	1.48	94.67
6	18	5.33	100.00
Total	338	100.00	

. svyprop living

```
-----  
pweight: weight                Number of obs   =    361  
Strata:  <one>                 Number of strata =     1  
PSU:    <observations>        Number of PSUs  =    361  
                                           Population size = 435.17478  
-----
```

Survey proportions estimation

	living	Obs	Est. Prop.	Std. Err.
	single	29	0.086536	0.016202
	sing parent	16	0.050037	0.012584
	couple	52	0.151834	0.020111
	couple w kids	62	0.179171	0.021241
	other	202	0.532421	0.027604

. tabulate living

living	Freq.	Percent	Cum.
single	29	8.03	8.03
sing parent	16	4.43	12.47
couple	52	14.40	26.87
couple w kids	62	17.17	44.04
other	202	55.96	100.00
Total	361	100.00	

. svyprop genhelp

```
-----  
pweight: weight                Number of obs   =    361  
Strata:  <one>                 Number of strata =     1  
PSU:    <observations>        Number of PSUs  =    361  
                                           Population size = 435.17478  
-----
```

Survey proportions estimation

	genhelp	Obs	Est. Prop.	Std. Err.
	excellent	72	0.186660	0.020762
	very good	139	0.377625	0.026673
	good	120	0.343566	0.026397
	fair	24	0.070042	0.014328
	poor	6	0.022107	0.009443

. tabulate genhelp

genhelp	Freq.	Percent	Cum.
excellent	72	19.94	19.94
very good	139	38.50	58.45
good	120	33.24	91.69
fair	24	6.65	98.34
poor	6	1.66	100.00
Total	361	100.00	

```
. svyprop empl
```

```
-----
pweight: weight           Number of obs   =      361
Strata:  <one>           Number of strata =        1
PSU:     <observations>  Number of PSUs  =      361
                               Population size = 435.17478
-----
```

Survey proportions estimation

	empl	Obs	Est. Prop.	Std. Err.
in paid work or self employed - full time		154	0.423934	0.027258
in paid work or self employed - part time		44	0.119879	0.017706
unemployed		22	0.073019	0.015595
intending to look for work but prevented by temp sickness or		11	0.039017	0.012018
looking after the home or family full time		22	0.066568	0.014096
in full time education		108	0.277583	0.024020

```
. tabulate empl
```

empl	Freq.	Percent	Cum.
in paid work or self employed - full ti	154	42.66	42.66
in paid work or self employed - part ti	44	12.19	54.85
unemployed	22	6.09	60.94
intending to look for work but prevente	11	3.05	63.99
looking after the home or family full t	22	6.09	70.08
in full time education	108	29.92	100.00
Total	361	100.00	

```
-----
. /*-----
> some code to recode drug use into scores so that they make
> ordered categories
> -----*/
. recode q85a (1=0) (2=0) (3=1) (4=1) (5=1) (6=0.5) ,gen (canscore)
(358 differences between q85a and canscore)

. recode q85b (3=6) (1=0) (2=0) (3=1) (4=1) (5=1) (6=0.5) ,gen (ampscore)
(338 differences between q85b and ampscore)

. /*-----
> now some mean scores to check design effects
> -----*/
. svymean genhelp sinc sacc
```

Survey mean estimation

```
pweight: weight           Number of obs   =      361
Strata:  <one>           Number of strata =        1
PSU:     <observations>  Number of PSUs  =      361
                               Population size = 435.17478
```

Mean	Estimate	Std. Err.	[95% Conf. Interval]		Deff
genhelp	2.363313	.0527406	2.259594	2.467031	1.141333
sinc	18.71168	.6945963	17.3457	20.07766	1.312113
sacc	-.0675717	.0328835	-.1322396	-.0029038	1.038462

```
-----
. /*-----
> now some regressions to predict general health score
> although this is categorical it is quite legitimate to
> use it in a regression to look for simple associations
> -----*/
. svyregress genhelp sinc canscore ampscore
```

Survey linear regression

```

pweight: weight           Number of obs   =       338
Strata:  <one>           Number of strata =         1
PSU:    <observations>   Number of PSUs  =       338
                               Population size = 407.87389
                               F( 3, 335) = 11.82
                               Prob > F = 0.0000
                               R-squared = 0.0688
    
```

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sinc	.0134715	.0050677	2.66	0.008	.0035033	.0234398
canscore	.5099733	.1471115	3.47	0.001	.2206008	.7993458
ampscore	.2301185	.0971406	2.37	0.018	.0390401	.4211969
_cons	2.015085	.0902006	22.34	0.000	1.837658	2.192512

. regress genhelf sinc canscore amscore

Source	SS	df	MS	Number of obs = 338		
Model	18.0620294	3	6.02067648	F( 3, 334) = 7.77	Prob > F = 0.0000	
Residual	258.757497	334	.774723046	R-squared = 0.0652	Adj R-squared = 0.0569	
Total	276.819527	337	.821422928	Root MSE = .88018		

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sinc	.0130195	.0045219	2.88	0.004	.0041245	.0219145
canscore	.5442357	.1625674	3.35	0.001	.2244507	.8640206
ampscore	.1936597	.1319918	1.47	0.143	-.0659802	.4532997
_cons	2.019351	.0874236	23.10	0.000	1.84738	2.191321

```

. /*-----
> next bit of code gets some stata commands that
> enable you to generate nice output to paste into reports
>
> the findit command gives access to some regression formatting
> commands available from a submission to the stata journal
> FOLLOW THE WEB LINK IF YOU WANT TO INSTALL IT
>
> http://www.ats.ucla.edu/stat/stata/faq/outreg.htm
>
> Results are sent to an external file output.doc
> This was sued to produce the output in the exemplar page
> -----*/
. findit outreg
    
```

. net from http://www.stata.com

-----  
http://www.stata.com/

```
. svyregress genhelf canscore
```

```
Survey linear regression
```

```
pweight: weight          Number of obs   =      358
Strata:  <one>           Number of strata =        1
PSU:    <observations>  Number of PSUs  =      358
                          Population size = 430.92195
                          F( 1, 357) = 9.21
                          Prob > F = 0.0026
                          R-squared = 0.0225
```

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
canscore	.4546734	.1498254	3.03	0.003	.1600221	.7493247
_cons	2.283758	.0566644	40.30	0.000	2.17232	2.395196

```
.   outreg using output.doc, nolabel replace
```

```
. svyregress genhelf ampscore
```

```
Survey linear regression
```

```
pweight: weight          Number of obs   =      338
Strata:  <one>           Number of strata =        1
PSU:    <observations>  Number of PSUs  =      338
                          Population size = 407.87389
                          F( 1, 337) = 3.45
                          Prob > F = 0.0640
                          R-squared = 0.0141
```

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ampscore	.3016327	.1623158	1.86	0.064	-.0176471	.6209125
_cons	2.323339	.0533271	43.57	0.000	2.218443	2.428235

```
.   outreg using output.doc , nolabel append
```

```
. svyregress genhelf sinc
```

```
Survey linear regression
```

```
pweight: weight          Number of obs   =      361
Strata:  <one>           Number of strata =        1
PSU:    <observations>  Number of PSUs  =      361
                          Population size = 435.17478
                          F( 1, 360) = 8.13
                          Prob > F = 0.0046
                          R-squared = 0.0323
```

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sinc	.0146277	.0051302	2.85	0.005	.0045388	.0247167
_cons	2.089603	.0896386	23.31	0.000	1.913322	2.265884

```
.   outreg using output.doc , nolabel append
```

```
. svyregress genhelf canscore sinc
```

```
Survey linear regression
```

```
pweight: weight          Number of obs   =       358
Strata:   <one>          Number of strata =         1
PSU:     <observations> Number of PSUs  =       358
                          Population size = 430.92195
                          F( 2, 356)    =         9.08
                          Prob > F     =         0.0001
                          R-squared     =         0.0475
```

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
canscore	.439577	.1504393	2.92	0.004	.1437184	.7354357
sinc	.01257	.004982	2.52	0.012	.0027723	.0223678
_cons	2.051402	.0879524	23.32	0.000	1.878432	2.224372

```
.      outreg using output.doc , nolabel append
```

```
. svyregress genhelf ampscore sinc
```

```
Survey linear regression
```

```
pweight: weight          Number of obs   =       338
Strata:   <one>          Number of strata =         1
PSU:     <observations> Number of PSUs  =       338
                          Population size = 407.87389
                          F( 2, 336)    =         6.23
                          Prob > F     =         0.0022
                          R-squared     =         0.0426
```

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ampscore	.2850634	.1402797	2.03	0.043	.0091292	.5609975
sinc	.0134487	.0051076	2.63	0.009	.0034019	.0234955
_cons	2.072104	.0915095	22.64	0.000	1.892103	2.252106

```
.      outreg using output.doc , nolabel append
```

```
. svyregress genhelf canscore ampscore
```

```
Survey linear regression
```

```
pweight: weight          Number of obs   =       338
Strata:   <one>          Number of strata =         1
PSU:     <observations> Number of PSUs  =       338
                          Population size = 407.87389
                          F( 2, 336)    =         9.14
                          Prob > F     =         0.0001
                          R-squared     =         0.0402
```

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
canscore	.5090293	.1505447	3.38	0.001	.2129036	.805155
ampscore	.2468177	.1188762	2.08	0.039	.0129848	.4806505
_cons	2.266851	.0580489	39.05	0.000	2.152667	2.381035

```
.      outreg using output.doc , nolabel append
```

```
. svyregress genhelf canscore ampscore sinc
```

Survey linear regression

```

pweight: weight          Number of obs   =      338
Strata:  <one>          Number of strata =       1
PSU:    <observations> Number of PSUs  =     338
                               Population size = 407.87389
                               F( 3, 335) = 11.82
                               Prob > F = 0.0000
                               R-squared = 0.0688
    
```

```

-----
      genhelp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
      canscore |   .5099733   .1471115     3.47  0.001   .2206008   .7993458
      ampscore |   .2301185   .0971406     2.37  0.018   .0390401   .4211969
           sinc |   .0134715   .0050677     2.66  0.008   .0035033   .0234398
           _cons |   2.015085   .0902006    22.34  0.000   1.837658   2.192512
-----
    
```

```

.      outreg using output.doc , nolabel append

. /*-----
> now the same thing for an unweighted regression
> -----*/

. regress genhelp canscore
    
```

```

-----
      Source |      SS      df      MS          Number of obs =      358
-----+-----
      Model |  8.14533065      1  8.14533065      F( 1, 356) = 10.25
      Residual | 282.874222    356  .794590512      Prob > F = 0.0015
-----+-----
      Total | 291.019553    357  .815180821      R-squared = 0.0280
                               Adj R-squared = 0.0253
                               Root MSE = .8914
    
```

```

-----
      genhelp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
      canscore |   .5028069   .157043     3.20  0.001   .1939583   .8116554
           _cons |   2.236383   .0509959    43.85  0.000   2.136092   2.336674
-----
    
```

```

.      outreg using output2.doc, nolabel replace
    
```

```

-----
      Source |      SS      df      MS          Number of obs =      338
-----+-----
      Model |  3.04173794      1  3.04173794      F( 1, 336) = 3.73
      Residual | 273.777789    336  .814814847      Prob > F = 0.0542
-----+-----
      Total | 276.819527    337  .821422928      R-squared = 0.0110
                               Adj R-squared = 0.0080
                               Root MSE = .90267
    
```

```

-----
      genhelp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
      ampscore |   .2594266   .1342712     1.93  0.054   -.0046916   .5235447
           _cons |   2.282698   .0498025    45.84  0.000   2.184734   2.380662
-----
    
```

```

.      outreg using output2.doc , nolabel append
    
```

. regress genhelf sinc

Source	SS	df	MS	Number of obs =	361
Model	8.20930765	1	8.20930765	F( 1, 359) =	9.90
Residual	297.790692	359	.829500536	Prob > F =	0.0018
				R-squared =	0.0268
				Adj R-squared =	0.0241
				Root MSE =	.91077
Total	306	360	.85		

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sinc	.0142075	.0045162	3.15	0.002	.005326	.023089
_cons	2.095164	.0849479	24.66	0.000	1.928106	2.262222

. outreg using output2.doc , nolabel append

. regress genhelf canscore sinc

Source	SS	df	MS	Number of obs =	358
Model	14.2324683	2	7.11623415	F( 2, 355) =	9.13
Residual	276.787085	355	.779681929	Prob > F =	0.0001
				R-squared =	0.0489
				Adj R-squared =	0.0435
				Root MSE =	.883
Total	291.019553	357	.815180821		

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
canscore	.4936069	.1555976	3.17	0.002	.187598	.7996158
sinc	.0123342	.0044143	2.79	0.005	.0036527	.0210157
_cons	2.046646	.084634	24.18	0.000	1.880199	2.213093

. outreg using output2.doc , nolabel append

. regress genhelf ampscore sinc

Source	SS	df	MS	Number of obs =	338
Model	9.37935175	2	4.68967587	F( 2, 335) =	5.87
Residual	267.440175	335	.79832888	Prob > F =	0.0031
				R-squared =	0.0339
				Adj R-squared =	0.0281
				Root MSE =	.89349
Total	276.819527	337	.821422928		

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ampscore	.248102	.1329667	1.87	0.063	-.0134529	.5096569
sinc	.0129331	.0045902	2.82	0.005	.0039039	.0219623
_cons	2.081722	.0867068	24.01	0.000	1.911164	2.252281

. outreg using output2.doc , nolabel append

. regress genhelf canscore ampscore

Source	SS	df	MS	Number of obs =	338
Model	11.6396446	2	5.81982232	F( 2, 335) =	7.35
Residual	265.179882	335	.791581737	Prob > F =	0.0008
				R-squared =	0.0420
				Adj R-squared =	0.0363
				Root MSE =	.88971
Total	276.819527	337	.821422928		

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
canscore	.5415636	.164324	3.30	0.001	.2183267	.8648004
ampscore	.2053269	.1333573	1.54	0.125	-.0569963	.4676501
_cons	2.221969	.052432	42.38	0.000	2.118832	2.325107

. outreg using output2.doc , nolabel append



```
. regress genhelf canscore ampscore sinc
```

Source	SS	df	MS	Number of obs =	338
Model	18.0620294	3	6.02067648	F( 3, 334) =	7.77
Residual	258.757497	334	.774723046	Prob > F =	0.0000
				R-squared =	0.0652
				Adj R-squared =	0.0569
Total	276.819527	337	.821422928	Root MSE =	.88018

genhelf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
canscore	.5442357	.1625674	3.35	0.001	.2244507	.8640206
ampscore	.1936597	.1319918	1.47	0.143	-.0659802	.4532997
sinc	.0130195	.0045219	2.88	0.004	.0041245	.0219145
_cons	2.019351	.0874236	23.10	0.000	1.84738	2.191321

```
. outreg using output2.doc , nolabel append
```

```
. /*-----
> now look at the proportions in health groups
> by the original categories of cannabis use
> and a survey-corrected chi square
> -----*/
```

```
. svytab genhelf q85a,column percent
```

```
pweight: weight                Number of obs    =    358
Strata:  <one>                 Number of strata =     1
PSU:    <observations>        Number of PSUs   =    358
                                           Population size  = 430.92195
```

genhelf	q85a						Total
	never us	tried on	use dail	use week	used in	used mor	
excellen	22.12	19.36	6.093	9.15	0	8.462	18.85
very goo	40.07	35.29	41.22	29.52	16.95	41.52	38.14
good	29.17	37.21	52.68	41.97	54.49	47.69	34.5
fair	6.549	8.149	0	19.36	20.33	2.325	7.073
poor	2.088	0	0	0	8.231	0	1.446
Total	100	100	100	100	100	100	100

```
Key: column percentages
```

```
Pearson:
Uncorrected chi2(20) = 26.0092
Design-based F(18.84, 6725.01)= 1.2200 P = 0.2307
```

```
. /*-----
> now test out the effect of the finite population correction
> The number of women of this age group in the
> population is 29457. This will be set as the population
> size for all units since there is no stratification here
> -----*/
. generate popsize=29457
```

```
. /*-----
> now redo the survey set up putting in the FPC
> this can be either the pop size if > number in strata
> (as here) or the sampling fraction
> GWEIGHT needs to be used since
> we need to have weights that add to population totals
> -----*/
. svyset [pweight=gweight],fpc(popsize)
pweight is gweight
fpc is popsize
```

```

. /*-----
> now rerun one svy mean from above
> it makes little difference
> Though the first run was wrong because it needed
> to have weights that add to population totals
> -----*/
. svymean genhelp sinc sacc

```

Survey mean estimation

```

pweight:  gweight          Number of obs   =      361
Strata:   <one>           Number of strata =        1
PSU:     <observations>   Number of PSUs  =      361
FPC:     popsize          Population size  = 29456.894

```

Mean	Estimate	Std. Err.	[95% Conf. Interval]		Deff
genhelp	2.363313	.0524164	2.260232	2.466393	1.141333
sinc	18.71168	.690327	17.3541	20.06926	1.312113
sacc	-.0675717	.0326814	-.1318421	-.0033013	1.038462

Finite population correction (FPC) assumes simple random sampling without replacement of PSUs within each stratum with no subsampling within PSUs. Weights must represent population totals for deff to be correct when using an FPC. Note: deff is invariant to the scale of weights.

```

.
end of do-file
.

```