

Chapter 6 Stata v10.1 Analysis Examples Syntax and Output

General Notes on Stata 10.1

Given that this tool is used throughout the ASDA textbook this chapter includes only the syntax and output for the analysis examples provided in Chapter 6. Stata 10.1 is an excellent tool for survey data analysis as well as graphing and related data management tasks. It offers a very comprehensive set of svy commands as well as weighted graphics and convenient syntax and data management abilities. For these reasons, we use Stata as the primary software for the ASDA text.

The examples and syntax presented here assume that all data management including variable construction, labels for variable values and other preparation steps are complete. See the Stata documentation for assistance with these issues.

One analysis included in this chapter cannot be run in Stata 10.1. For example, the CMH trends test (example 6.1) is run in sudaan rather than Stata.

Please check the Stata documentation and also the ASDA web site for updates to Stata as new versions are released. For example, we have already included an example of how to use Stata 11.0 with the new "factor" variable features/syntax and compared this to the older "xi" type of syntax for including categorical variables in data analysis.

```
. * 6.1 NHANES DATA
. svy, subpop(age18p): tab irregular, se ci col deff
(running tabulate on estimation sample)
```

```
Number of strata = 15      Number of obs = 9906
Number of PSUs  = 30      Population size = 284231791
                                   Subpop. no. of obs = 5121
                                   Subpop. size = 210315370
                                   Design df = 15
```

```
-----
```

1=yes 0=no	column	se	lb	ub	deff
0	.9705	.0067	.9525	.9818	11.33
1	.0295	.0067	.0182	.0475	11.33
Total	1				

```
-----
```

```
Key: column = column proportions
      se      = linearized standard errors of column proportions
      lb      = lower 95% confidence bounds for column proportions
      ub      = upper 95% confidence bounds for column proportions
      deff    = deff for variances of column proportions
```

```
. svy, subpop(age18p): prop irregular
(running proportion on estimation sample)
```

Survey: Proportion estimation

```
Number of strata = 15      Number of obs = 9906
Number of PSUs  = 30      Population size = 284231791
                                   Subpop. no. obs = 5121
                                   Subpop. size = 210315370
                                   Design df = 15
```

```
-----
```

	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
irregular				
0	.970469	.0066546	.956285	.9846529
1	.029531	.0066546	.0153471	.043715

```
-----
```

```
. estat effects
```

```
-----
```

	Proportion	Linearized Std. Err.	DEFF	DEFT
irregular				
0	.970469	.0066546	11.325	3.36526
1	.029531	.0066546	11.325	3.36526

```
-----
```

```
. svy: mean irregular, subpop(age18p)
(running mean on estimation sample)
```

Survey: Mean estimation

```
Number of strata = 15      Number of obs = 9906
Number of PSUs  = 30      Population size = 284231791
                                   Subpop. no. obs = 5121
                                   Subpop. size = 210315370
                                   Design df = 15
```

```
-----
```

	Mean	Linearized Std. Err.	[95% Conf. Interval]	
irregular				
	.029531	.0066546	.0153471	.043715

```
-----
```

```
. estat effects
```

```
-----
```

	Mean	Linearized Std. Err.	DEFF	DEFT
irregular				
	.029531	.0066546	11.325	3.36526

```
-----
```

```

. * example 6.2
. * add a table of race/eth from NHANES file
. svy: proportion ridreth1, subpop(age18p)
(running proportion on estimation sample)

```

Survey: Proportion estimation

```

Number of strata =      15      Number of obs   =      10348
Number of PSUs   =      30      Population size = 291616892
                                   Subpop. no. obs  =      5334
                                   Subpop. size    = 217700471
                                   Design df       =         15

```

```

-----
|          |          |          |          |          |
|          | Proportion | Linearized |          | [95% Conf. Interval] |
|          |            | Std. Err.  |          |                       |
-----+-----
ridreth1 |          |            |          |                       |
  1       | .0807834 | .0100534   | .059355 | .1022117              |
  2       | .0337852 | .0074214   | .0179669 | .0496035              |
  3       | .7141428 | .0276981   | .6551057 | .77318                 |
  4       | .1172628 | .0198491   | .0749555 | .1595701              |
  5       | .0540257 | .005825    | .04161   | .0664414              |
-----

```

```

. estat effects

```

```

-----
|          |          |          |          |          |
|          | Proportion | Linearized |          |          |
|          |            | Std. Err.  | DEFF    | DEFT    |
-----+-----
ridreth1 |          |            |          |          |
  1       | .0807834 | .0100534   | 10.2808 | 3.20638 |
  2       | .0337852 | .0074214   | 12.7441 | 3.56989 |
  3       | .7141428 | .0276981   | 28.3863 | 5.32788 |
  4       | .1172628 | .0198491   | 28.7495 | 5.36185 |
  5       | .0540257 | .005825    | 5.01483 | 2.23938 |
-----

```

```

. tab ridreth1 if age18p==1

```

```

1=mex 2=oth |
  hisp      |
  3=white   |
  4=black   |
  5=other   |
-----+-----
          | Freq.  | Percent | Cum.  |
-----+-----
  1       | 1,185  | 21.30   | 21.30 |
  2       | 171    | 3.07    | 24.38 |
  3       | 2,633  | 47.33   | 71.71 |
  4       | 1,341  | 24.11   | 95.81 |
  5       | 233    | 4.19    | 100.00 |
-----+-----
Total    | 5,563  | 100.00  |

```

* example 6.3

. svy: tab bp_cat , subpop(age18p) obs se ci
(running tabulate on estimation sample)

Number of strata	=	15	Number of obs	=	9842
Number of PSUs	=	30	Population size	=	281904771
			Subpop. no. of obs	=	5057
			Subpop. size	=	207988351
			Design df	=	15

bp_cat	proportions	se	lb	ub	obs
normal	.4711	.0111	.4475	.4948	2441
pre-hype	.4185	.0118	.3937	.4439	1988
stage 1	.0864	.0062	.0741	.1006	470
stage 2	.024	.0024	.0193	.0297	158
Total	1				5057

Key: proportions = cell proportions
se = linearized standard errors of cell proportions
lb = lower 95% confidence bounds for cell proportions
ub = upper 95% confidence bounds for cell proportions
obs = number of observations

. svy: tab bp_cat , subpop (age18p) deff
(running tabulate on estimation sample)

Number of strata	=	15	Number of obs	=	9842
Number of PSUs	=	30	Population size	=	281904771
			Subpop. no. of obs	=	5057
			Subpop. size	=	207988351
			Design df	=	15

bp_cat	proportions	deff
normal	.4711	3.581
pre-hype	.4185	4.145
stage 1	.0864	3.544
stage 2	.024	1.794
Total	1	

Key: proportions = cell proportions
deff = deff for variances of cell proportions

```

. * example 6.4
. * use 4 category blood pressure variable to examine multinomial gof tool
. * note bp_cat 1=normal 2=pre-hypertensive 3=stage 1 high blood pressure 4=stage 2 high blood pressure
.
. svy: tab bp_cat, subpop(age18p) se ci col deff
(running tabulate on estimation sample)

```

```

Number of strata =      15      Number of obs      =      9842
Number of PSUs   =      30      Population size    = 281904771
                                           Subpop. no. of obs =      5057
                                           Subpop. size      = 207988351
                                           Design df         =      15

```

bp_cat	column	se	lb	ub	deff
1	.4711	.0111	.4475	.4948	3.581
2	.4185	.0118	.3937	.4439	4.145
3	.0864	.0062	.0741	.1006	3.544
4	.024	.0024	.0193	.0297	1.794
Total	1				

```

Key: column = column proportions
se = linearized standard errors of column proportions
lb = lower 95% confidence bounds for column proportions
ub = upper 95% confidence bounds for column proportions
deff = deff for variances of column proportions

```

```

. recode bp_cat (1=.5) (2=.3) (3=.15) (4=.05), generate (p)
(7359 differences between bp_cat and p)

```

```

. mgof bp_cat=p if age18p==1, svy

```

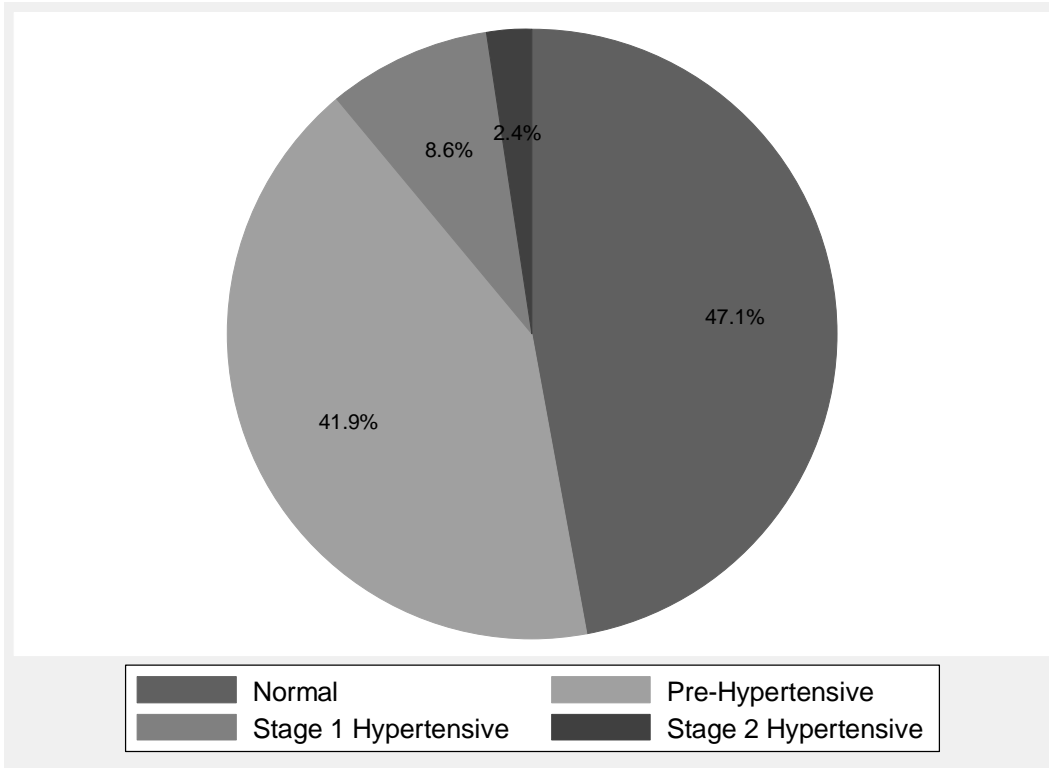
```

Number of strata =      15      Number of obs =      5057
Number of PSUs   =      30      Pop size      = 2.1e+08
                                           Design df    =      15
                                           N of outcomes =      4
                                           F df1       = 2.62707
                                           F df2       = 39.406

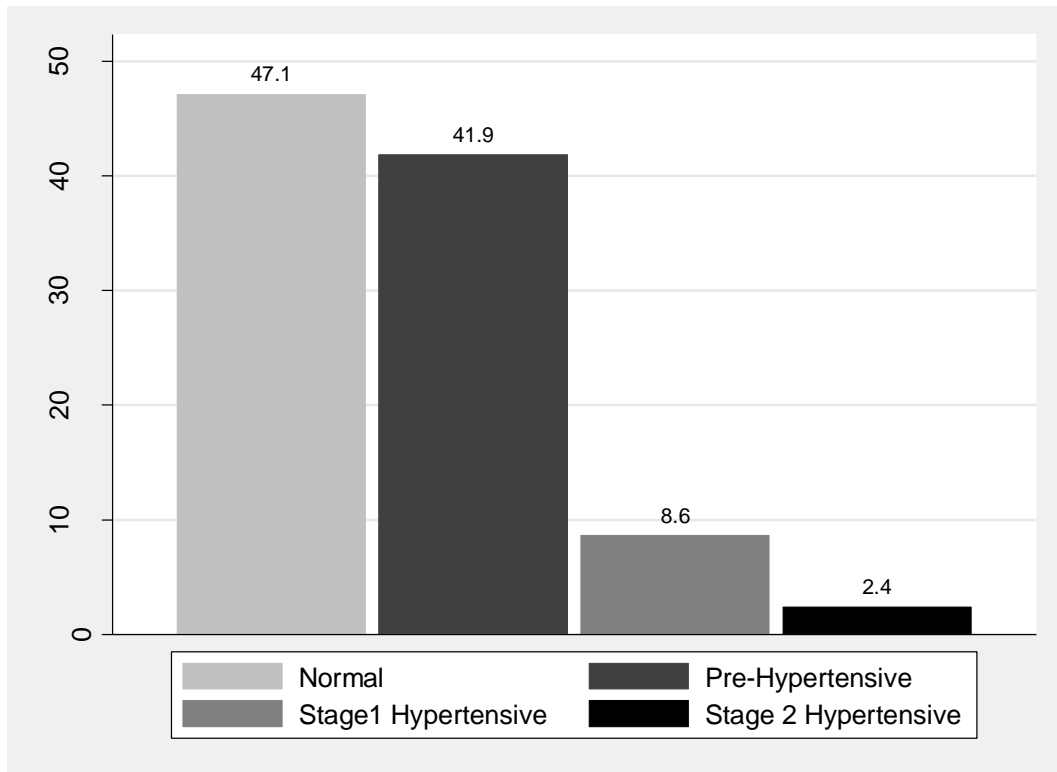
```

Goodness-of-fit	Coef.	F-value	P-value
Pearson's X2	450.1971	69.5663	0.0000
Log likelihood ratio	465.482	71.9281	0.0000

```
* figure 6.3
. graph pie bp_cat1 bp_cat2 bp_cat3 bp_cat4 [pweight=wtmec2yr] if age18p==1, plabel(_all percent,
format(%9.1f)) ///
> scheme(s2mono) ///
> legend (label(1 "Normal")label(2 "Pre-Hypertensive") label(3 "Stage 1 Hypertensive") label (4 "Stage
2 Hypertensive"))
```



```
. * figure 6.4
graph bar (mean) bp_cat1 bp_cat2 bp_cat3 bp_cat4 [pweight=wtmec2yr] if age18p==1 , percentages ///
> bar(1,color(gs12)) bar(2,color(gs4)) bar(3,color(gs8)) bar(4,color(black)) ///
> blabel(bar, format(%5.1f)) bargap(7) scheme(s2mono) ///
> legend (label(1 "Normal") label(2 "Pre-Hypertensive") label(3 "Stage1 Hypertensive") label(4 "Stage 2
Hypertensive")) ///
> ytitle ("Percentage")
```



Example 6.6 NCS-R DATA

```
. svyset seclustr [pweight=ncsrwtsh], strata(sestrat)
  pweight: ncsrwtsh
  VCE: linearized
  Single unit: missing
  Strata 1: sestrat
  SU 1: seclustr
  FPC 1: <zero>
```

```
. svy: tab sex mde , se ci deff
(running tabulate on estimation sample)
```

```
Number of strata = 42          Number of obs = 9282
Number of PSUs  = 84          Population size = 9282.0002
                                   Design df = 42
```

sex	MajorDepEpisode		Total
	0	1	
Male	.4066 (.007) [.3926, .4208] 1.874	.0722 (.0034) [.0656, .0795] 1.637	.4789 (.0053) [.4681, .4896] 1.051
Female	.4016 (.0054) [.3909, .4125] 1.11	.1195 (.003) [.1135, .1258] .8086	.5211 (.0053) [.5104, .5319] 1.051
Total	.8083 (.0049) [.7983, .8179] 1.424	.1917 (.0049) [.1821, .2017] 1.424	1

Key: cell proportions
(linearized standard errors of cell proportions)
[95% confidence intervals for cell proportions]
deff for variances of cell proportions

Pearson:
Uncorrected chi2(1) = 92.1499
Design-based F(1, 42) = 57.9784 P = 0.0000

Mean generalized deff = 1.3727
CV of generalized deffs = 0.0000

```
. svy: tab sex mde , row se ci deff
(running tabulate on estimation sample)
```

```
Number of strata = 42          Number of obs = 9282
Number of PSUs  = 84          Population size = 9282.0002
                                   Design df = 42
```

sex	MajorDepEpisode		Total
	0	1	
Male	.8492 (.0077) [.8329, .8642] 2.083	.1508 (.0077) [.1358, .1671] 2.083	1
Female	.7707 (.0056) [.7591, .7819] .8728	.2293 (.0056) [.2181, .2409] .8728	1
Total	.8083 (.0049) [.7983, .8179] .3639	.1917 (.0049) [.1821, .2017] .3639	1

Key: row proportions
(linearized standard errors of row proportions)
[95% confidence intervals for row proportions]
deff for variances of row proportions

Pearson:
Uncorrected chi2(1) = 92.1499
Design-based F(1, 42) = 57.9784 P = 0.0000

Mean generalized deff = 1.3727
CV of generalized deffs = 0.0000

* example 6.7 mde with over statement for sex NCS-R DATA

. svy: prop mde
(running proportion on estimation sample)

Survey: Proportion estimation

Number of strata = 42 Number of obs = 9282
Number of PSUS = 84 Population size = 9282
Design df = 42

	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
mde				
0	.8082888	.0048768	.798447	.8181306
1	.1917112	.0048768	.1818694	.201553

. svy: proportion mde, over(sex)
(running proportion on estimation sample)

Survey: Proportion estimation

Number of strata = 42 Number of obs = 9282
Number of PSUS = 84 Population size = 9282
Design df = 42

_prop_1: mde = 0
_prop_2: mde = 1

Male: sex = Male
Female: sex = Female

Over	Proportion	Linearized Std. Err.	[95% Conf. Interval]	
_prop_1				
Male	.8492067	.0077478	.8335709	.8648424
Female	.7706917	.0056473	.7592951	.7820883
_prop_2				
Male	.1507933	.0077478	.1351576	.1664291
Female	.2293083	.0056473	.2179117	.2407049

. lincom [_prop_2]Male-[_prop_2]Female
(1) [_prop_2]Male - [_prop_2]Female = 0

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	-.078515	.0095518	-8.22	0.000	-.0977914	-.0592386

```
. * example 6.8 use part 2 weight for alcohol dependence NCS-R DATA
. svyset seclustr [pweight=ncsrwtlg], strata(sestrat) vce(linearized) singleunit(missing)
```

```
    pweight: ncsrwtlg
          VCE: linearized
Single unit: missing
  Strata 1: sestrat
        SU 1: seclustr
        FPC 1: <zero>
```

```
. svy: tab ed4cat ald , subpop (if age < 29 ) se ci row deff
(running tabulate on estimation sample)
```

```
Number of strata =      42          Number of obs      =      5692
Number of PSUs  =      84          Population size   = 5692.0005
                                          Subpop. no. of obs =      1275
                                          Subpop. size     = 1266.5565
                                          Design df       =       42
```

years of education -4 categor ies	AlcDep		Total
	0	1	
0-11	.9087 (.0294) [.8297, .9531] 2.3	.0913 (.0294) [.0469, .1703] 2.3	1
12	.9514 (.0135) [.9159, .9724] 1.585	.0486 (.0135) [.0276, .0841] 1.585	1
13-15	.951 (.01) [.9263, .9678] .9407	.049 (.01) [.0322, .0737] .9407	1
16+	.931 (.0136) [.8978, .9539] .5983	.069 (.0136) [.0461, .1022] .5983	1
Total	.9405 (.0088) [.92, .956] .0855	.0595 (.0088) [.044, .08] .0855	1

```
Key: row proportions
      (linearized standard errors of row proportions)
      [95% confidence intervals for row proportions]
      deff for variances of row proportions
```

```
Pearson:
  Uncorrected chi2(3) = 27.2130
  Design-based F(2.75, 115.53) = 1.6498    P = 0.1857

Mean generalized deff = 6.6250
CV of generalized deffs = 0.5594
```

```

. * example 6.9 simple logistic regression of mde on male part 1 weight NCS-R DATA
. svyset seclustr [pweight=ncsrwtsh], strata(sestrat) vce(linearized) singleunit(missing)

    pweight: ncsrwtsh
      VCE: linearized
Single unit: missing
  Strata 1: sestrat
    SU 1: seclustr
    FPC 1: <zero>

. svy: logistic mde sexm
(running logistic on estimation sample)

```

Survey: Logistic regression

```

Number of strata =      42      Number of obs      =      9282
Number of PSUS  =      84      Population size    = 9282.0002
                                   Design df              =      42
                                   F( 1, 42)                =     57.28
                                   Prob > F                 =     0.0000

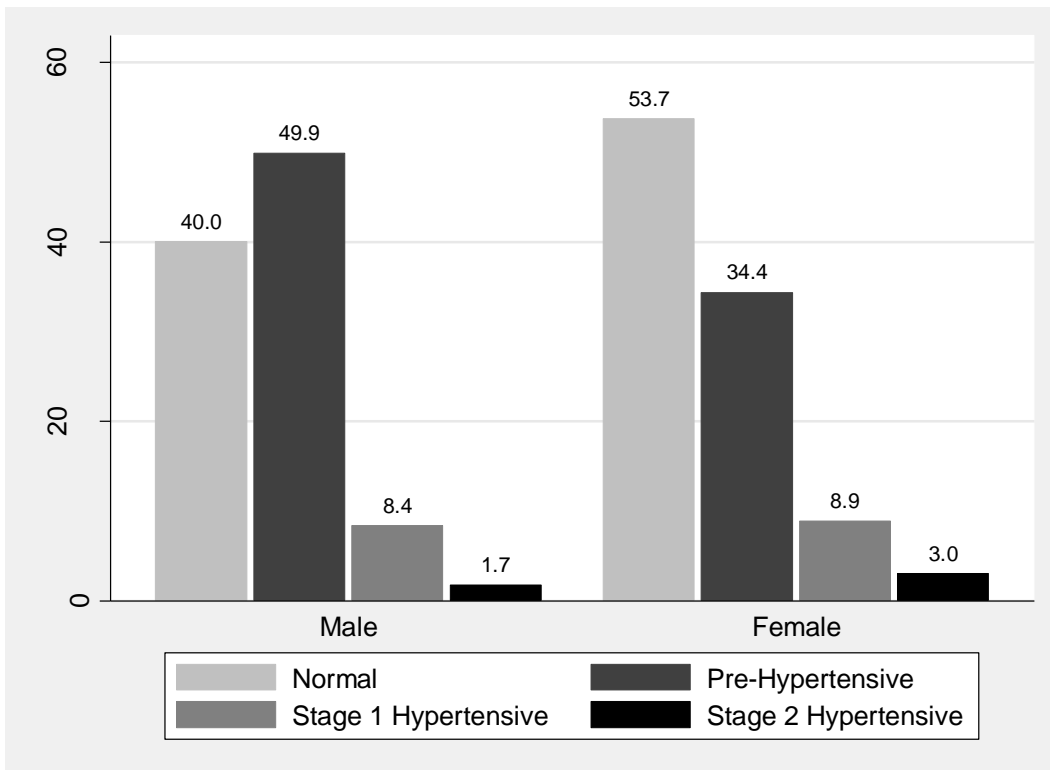
```

mde	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
sexm	.5968012	.0407038	-7.57	0.000	.5200601	.6848663

```

. * figure 6.8
. graph bar (mean) bp_cat1 bp_cat2 bp_cat3 bp_cat4 [pweight=wtmec2yr] if age18p==1, blabel(bar,
format(%9.1f) color(none)) //
> /
> bar(1,color(gs12)) bar(2,color(gs4)) bar(3,color(gs8)) bar(4,color(black)) ///
> bargap(7) scheme(s2mono) over(riagendr) percentages ///
> legend (label(1 "Normal")label(2 "Pre-Hypertensive") label(3 "Stage 1 Hypertensive") label (4 "Stage
2 Hypertensive")) ///
> ytitle ("Percentage")

```



*example 6.10 done in Sudaan as Stata does not offer a complex design corrected CMH trends test