

## Chapter 5 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 5. 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 cannot be run in Stata 10.1. For example, the quantile example (example 5.8) 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.

CHAPTER 5 ASDA ANALYSIS EXAMPLES REPLICATION IN STATA

\*example 5.1 and 5.2 graphs NHANES DATA

\*example 5.1

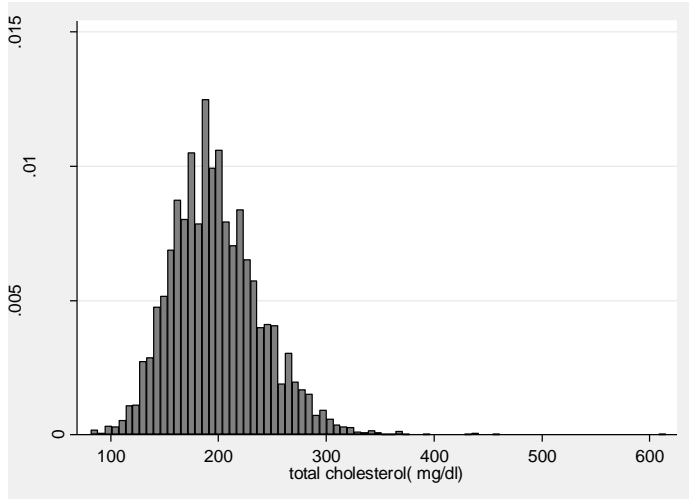
\* generate an integer version of the wtmec2yr for fweight in histogram

```
generate int_wtmec2yr=int(wtmec2yr)
```

```
gen wtmec2yr10000=wtmec2yr*10000
```

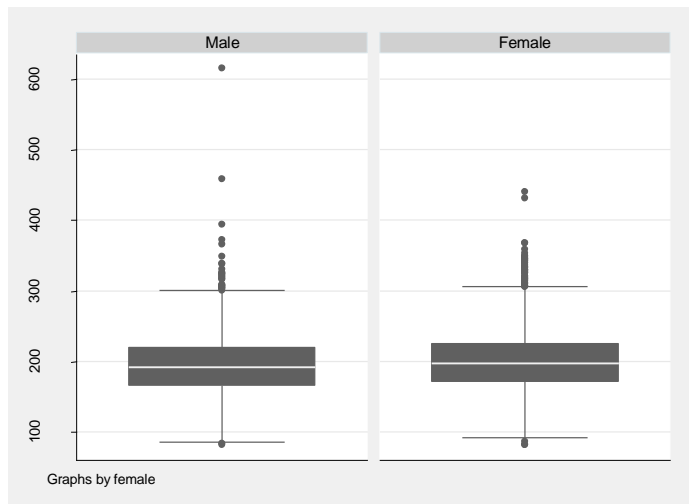
```
replace wtmec2yr1000=round(wtmec2yr)
```

```
histogram lbxtc if age18p [fweight=wtmec2yr1000]
```



\* example 5.2 weighted boxplot by gender

```
graph box lbxtc [pweight = wtmec2yr] if age18p==1, by(female)
```



\*example 5.3 NCSR DATA  
 . svy: tab mde , format(%16.1g) count se ci  
 (running tabulate on estimation sample)

Number of strata	=	42	Number of obs	=	9282
Number of PSUs	=	84	Population size	=	209128097
			Design df	=	42

MajorDepE pisode	count	se	lb	ub
0	169035891	7876170	153141136	184930645
1	40092206	2567488	34910806	45273607
Total	209128097			

Key: count = weighted counts  
 se = linearized standard errors of weighted counts  
 lb = lower 95% confidence bounds for weighted counts  
 ub = upper 95% confidence bounds for weighted counts

. svy : total mde  
 (running total on estimation sample)

Survey: Total estimation

Number of strata	=	42	Number of obs	=	9282
Number of PSUs	=	84	Population size	=	209128097
			Design df	=	42

	Total	Linearized Std. Err.	[95% Conf. Interval]	
mde	4.01e+07	2567488	3.49e+07	4.53e+07

. estat effects

	Total	Linearized Std. Err.	DEFF	DEFT
mde	4.01e+07	2567488	9.02761	3.0046

. svy : total mde , over (mar3cat)  
 (running total on estimation sample)

Survey: Total estimation

Number of strata = 42      Number of obs = 9282  
 Number of PSUs = 84      Population size = 209128097  
                          Design df = 42

married: mar3cat = married  
 \_subpop\_2: mar3cat = sep/wid/div  
 \_subpop\_3: mar3cat = never married

Over	Total	Linearized Std. Err.	[95% Conf. Interval]	
mde				
married	2.03e+07	1584109	1.71e+07	2.35e+07
_subpop_2	1.04e+07	702621.5	8942723	1.18e+07
_subpop_3	9427345	773137.6	7867091	1.10e+07

. estat size

married: mar3cat = married  
 \_subpop\_2: mar3cat = sep/wid/div  
 \_subpop\_3: mar3cat = never married

Over	Total	Linearized Std. Err.	Obs	Size
mde				
married	2.03e+07	1584109	5322	116763777.511
_subpop_2	1.04e+07	702621.5	2017	42747375.875
_subpop_3	9427345	773137.6	1943	49616943.9053

. estat effects

married: mar3cat = married  
 \_subpop\_2: mar3cat = sep/wid/div  
 \_subpop\_3: mar3cat = never married

Over	Total	Linearized Std. Err.	DEFF	DEFT
mde				
married	2.03e+07	1584109	6.07466	2.46468
_subpop_2	1.04e+07	702621.5	2.22487	1.4916
_subpop_3	9427345	773137.6	2.94672	1.7166

\*example 5.4 HRS DATA

```
. svy, subpop(finr): total h8atota  
(running total on estimation sample)
```

Survey: Total estimation

```
Number of strata =    56      Number of obs   =    18467  
Number of PSUS  =    112      Population size = 82249292  
Subpop. no. obs =    11942  
Subpop. size    = 53853171  
Design df      =         56
```

```
-----  
|          Total      Linearized  
|          |          Std. Err.      [95% Conf. Interval]  
-----+-----  
h8atota | 2.84e+13  1.60e+12      2.52e+13  3.16e+13  
-----
```

. \* example 5.5 NCSR DATA

```
. svy : mean hhinc  
(running mean on estimation sample)
```

Survey: Mean estimation

```
Number of strata =    42      Number of obs   =    5692  
Number of PSUS  =    84      Population size =    5692  
Design df      =         42
```

```
-----  
|          Mean      Linearized  
|          |          Std. Err.      [95% Conf. Interval]  
-----+-----  
hhinc | 59277.06  1596.343      56055.51  62498.61  
-----
```

. estat vce

Covariance matrix of coefficients of mean model

```
      e(v) |      hhinc  
-----+-----  
hhinc | 2548310.9
```

. estat effects

```
-----  
|          Mean      Linearized  
|          |          Std. Err.      DEFF      DEFT  
-----+-----  
hhinc | 59277.06  1596.343      6.09381  2.46856  
-----
```

\*example 5.6 NHANES DATA

. svy, subpop (age18p) : mean bpxsy1  
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =	15	Number of obs =	9400
Number of PSUs =	30	Population size =	264962444
		Subpop. no. obs =	4615
		Subpop. size =	191046023
		Design df =	15

	Mean	Linearized Std. Err.	[95% Conf. Interval]	
bpxsy1	123.1109	.5416936	121.9563	124.2655

. estat effects

	Mean	Linearized Std. Err.	DEFF	DEFT
bpxsy1	123.1109	.5416936	5.7922	2.4067

---

Example 5.7 HRS DATA

. svy, subpop(finr): mean h8atota  
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =	56	Number of obs =	18467
Number of PSUs =	112	Population size =	82249292
		Subpop. no. obs =	11942
		Subpop. size =	53853171
		Design df =	56

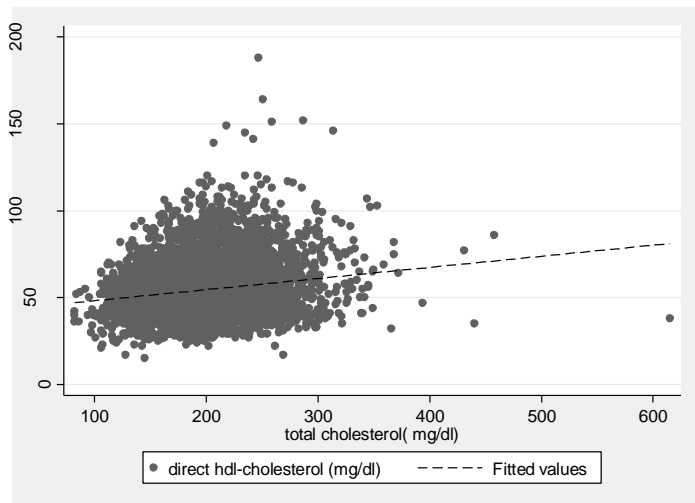
	Mean	Linearized Std. Err.	[95% Conf. Interval]	
h8atota	527313.2	28012.78	471196.8	583429.5

. estat effects

	Mean	Linearized Std. Err.	DEFF	DEFT
h8atota	527313.2	28012.78	1.5274	1.23588

Example 5.8 Quantiles not available in Stata 10

```
*Scatter Plot of direct HDL and total cholesterol
twoway (scatter lbdhdd lbxtc if age18p==1 ) (lfit lbdhdd lbxtc if age18p==1 [pweight=wtmec2yr])
```



```
*example 5.9 NHANES DATA
. svy, subpop(age18p): ratio (lbdhdd/lbxtc)
(running ratio on estimation sample)
```

Survey: Ratio estimation

```
Number of strata =      15      Number of obs   =      9781
Number of PSUs   =      30      Population size = 279216604
Subpop. no. obs  =      4996
Subpop. size     = 205300183
Design df        =           15
```

\_ratio\_1: lbdhdd/lbxtc

	Ratio	Linearized Std. Err.	[95% Conf. Interval]	
_ratio_1	.2753546	.0022544	.2705496	.2801596

\*example 5.10 HRS DATA

```

. * example 5.10
. svyset secu [pweight=kwgtr], strata(stratum) vce(linearized) singleunit(missing)

      pweight: kwgtr
          VCE: linearized
Single unit: missing
  Strata 1: stratum
    SU 1: secu
    FPC 1: <zero>

```

. svy: mean diabetes, subpop(if kage > 70) over (gender)  
(running mean on estimation sample)

Survey: Mean estimation

```

Number of strata =      52      Number of obs   =      18181
Number of PSUs   =      104     Population size =  74816616
                               Subpop. no. obs   =      7185
                               Subpop. size      =  25032597
                               Design df         =         52

```

Male: gender = Male  
Female: gender = Female

Over	Mean	Linearized Std. Err.	[95% Conf. Interval]	
diabetes				
Male	.2353744	.0083169	.2186854	.2520634
Female	.1838655	.008531	.1667468	.2009841

\*example 5.11 NHANES DATA

\* 5.11 Subpopulation analysis

. svy: mean bpxsy1 , subpop (if age > 45) over (gender)  
(running mean on estimation sample)

Survey: Mean estimation

```

Number of strata =      15      Number of obs   =      9915
Number of PSUs   =      30     Population size = 278949498
                               Subpop. no. obs   =      2094
                               Subpop. size      =  90292383.4
                               Design df         =         15

```

1: gender = 1  
2: gender = 2

Over	Mean	Linearized Std. Err.	[95% Conf. Interval]	
bpxsy1				
1	128.9629	.7566666	127.3501	130.5757
2	132.0873	1.064538	129.8183	134.3563

. estat effects

1: gender = 1  
2: gender = 2

Over	Mean	Linearized Std. Err.	DEFF	DEFT
bpxsy1				
1	128.9629	.7566666	2.60336	1.61349
2	132.0873	1.064538	3.62241	1.90326



\*example 5.12 HRS DATA

```
. svy, subpop(finr): mean h8atota, over(edcat)
(running mean on estimation sample)
```

Survey: Mean estimation

```
Number of strata =      56      Number of obs   =      18430
Number of PSUs   =     112      Population size = 82044060
                                   Subpop. no. obs  =     11914
                                   Subpop. size     = 53647939
                                   Design df        =         56
```

```
_subpop_1: edcat = 0-11
_subpop_2: edcat = 12
_subpop_3: edcat = 13-15
_subpop_4: edcat = 16+
```

Over	Mean	Linearized Std. Err.	[95% Conf. Interval]	
h8atota				
_subpop_1	178386.1	24561.12	129184.2	227587.9
_subpop_2	328392	17082.72	294171.2	362612.7
_subpop_3	455457.6	27000.33	401369.4	509545.8
_subpop_4	1107204	102113.5	902646.1	1311762

```
. lincom [h8atota]_subpop_1 - [h8atota]_subpop_4
```

```
( 1) [h8atota]_subpop_1 - [h8atota]_subpop_4 = 0
```

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	-928818	108250.1	-8.58	0.000	-1145669	-711967.1

\*example 5.13 HRS DATA

```
. svy, subpop(finr0406): mean totassets, over (year)
(running mean on estimation sample)
```

Survey: Mean estimation

```
Number of strata =      56      Number of obs   =      24486
Number of PSUs   =     112      Population size = 105084460
                                   Subpop. no. obs  =     23505
                                   Subpop. size     = 105084460
                                   Design df        =         56
```

```
2004: year = 2004
2006: year = 2006
```

Over	Mean	Linearized Std. Err.	[95% Conf. Interval]	
totassets				
2004	411786.4	20639.72	370440.1	453132.7
2006	527313.2	28012.78	471196.8	583429.5

```
. lincom [totassets]2004 - [totassets]2006
```

```
( 1) [totassets]2004 - [totassets]2006 = 0
```

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	-115526.8	20025.41	-5.77	0.000	-155642.5	-75411.06