

## Chapter 8 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 8. 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.

All analysis examples presented can be done in Stata 10.1 and are included in this chapter's output.

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 8 ANALYSIS EXAMPLES REPLICATION STATA 10

\*example 8.1 Bivariate analyses for predicting MDE NCS-R DATA

svy: tab ag4cat mde, row se  
(running tabulate on estimation sample)

Number of strata	=	42	Number of obs	=	5692
Number of PSUs	=	84	Population size	=	5692.0005
			Design df	=	42

ag4cat	MajorDepEpisode		Total
	0	1	
<=29	.816 (.0089)	.184 (.0089)	1
30-44	.7712 (.011)	.2288 (.011)	1
45-59	.7767 (.0126)	.2233 (.0126)	1
>=60	.8894 (.0096)	.1106 (.0096)	1
Total	.8082 (.0064)	.1918 (.0064)	1

Key: row proportions  
(linearized standard errors of row proportions)

Pearson:  
Uncorrected chi2(3) = 75.9697  
Design-based F(2.76, 115.97) = 26.3902 P = 0.0000

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

Number of strata	=	42	Number of obs	=	5692
Number of PSUs	=	84	Population size	=	5692.0005
			Design df	=	42

sex	MajorDepEpisode		Total
	0	1	
Male	.8471 (.0091)	.1529 (.0091)	1
Female	.7738 (.0067)	.2262 (.0067)	1
Total	.8082 (.0064)	.1918 (.0064)	1

Key: row proportions  
(linearized standard errors of row proportions)

Pearson:  
Uncorrected chi2(1) = 49.1166  
Design-based F(1, 42) = 44.8339 P = 0.0000

. svy: tab ald mde, row se  
 (running tabulate on estimation sample)

Number of strata	=	42	Number of obs	=	5692
Number of PSUs	=	84	Population size	=	5692.0005
			Design df	=	42

AlcDep	MajorDepEpisode		Total
	0	1	
0	.8231 (.0065)	.1769 (.0065)	1
1	.5484 (.029)	.4516 (.029)	1
Total	.8082 (.0064)	.1918 (.0064)	1

Key: row proportions  
 (linearized standard errors of row proportions)

Pearson:  
 Uncorrected chi2(1) = 141.7044  
 Design-based F(1, 42) = 120.0282 P = 0.0000

. svy: tab ed4cat mde, row se  
 (running tabulate on estimation sample)

Number of strata	=	42	Number of obs	=	5692
Number of PSUs	=	84	Population size	=	5692.0005
			Design df	=	42

years of education -4 categor ies	MajorDepEpisode		Total
	0	1	
0-11	.8369 (.0121)	.1631 (.0121)	1
12	.8145 (.0083)	.1855 (.0083)	1
13-15	.7875 (.0104)	.2125 (.0104)	1
16+	.8033 (.0109)	.1967 (.0109)	1
Total	.8082 (.0064)	.1918 (.0064)	1

Key: row proportions  
 (linearized standard errors of row proportions)

Pearson:  
 Uncorrected chi2(3) = 10.0806  
 Design-based F(2.90, 121.93) = 4.3043 P = 0.0069

. svy: tab mar3cat mde, row se  
 (running tabulate on estimation sample)

Number of strata = 42  
 Number of PSUs = 84

Number of obs = 5692  
 Population size = 5692.0005  
 Design df = 42

marital status-3 categor ies	MajorDepEpisode		Total
	0	1	
married	.8267 (.0074)	.1733 (.0074)	1
sep/wid/	.761 (.0145)	.239 (.0145)	1
never ma	.806 (.0115)	.194 (.0115)	1
Total	.8082 (.0064)	.1918 (.0064)	1

Key: row proportions  
 (linearized standard errors of row proportions)

Pearson:  
 Uncorrected chi2(2) = 24.1420  
 Design-based F(1.90, 79.74) = 11.0849 P = 0.0001

\*stage 2 model estimation

. char sex[omit] 2

. 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>

. xi: svy: logistic mde i.ag4cat i.sex ald i.ed4cat i.mar3cat  
i.ag4cat            \_Iag4cat\_1-4           (naturally coded; \_Iag4cat\_1 omitted)  
i.sex               \_Isex\_1-2             (naturally coded; \_Isex\_2 omitted)  
i.ed4cat           \_Ied4cat\_1-4           (naturally coded; \_Ied4cat\_1 omitted)  
i.mar3cat           \_Imar3cat\_1-3         (naturally coded; \_Imar3cat\_1 omitted)  
(running logistic on estimation sample)

Survey: Logistic regression

Number of strata	=	42	Number of obs	=	5692
Number of PSUs	=	84	Population size	=	5692.0005
			Design df	=	42
			F( 10, 33)	=	28.07
			Prob > F	=	0.0000

mde	Odds Ratio	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
_Iag4cat_2	1.29126	.1218742	2.71	0.010	1.067313	1.562197
_Iag4cat_3	1.229302	.112517	2.26	0.029	1.021971	1.478695
_Iag4cat_4	.5087563	.0718888	-4.78	0.000	.3825303	.6766338
_Isex_1	.5613867	.0433476	-7.48	0.000	.4803828	.6560499
ald	4.152358	.6401312	9.24	0.000	3.042157	5.667712
_Ied4cat_2	1.08248	.1048871	0.82	0.418	.8902193	1.316264
_Ied4cat_3	1.259243	.1171994	2.48	0.017	1.04361	1.519432
_Ied4cat_4	1.176949	.1301881	1.47	0.148	.9414778	1.471313
_Imar3cat_2	1.626487	.1389338	5.69	0.000	1.368943	1.932484
_Imar3cat_3	1.122524	.1210876	1.07	0.290	.9029277	1.395526

. test \_Iag4cat\_2 \_Iag4cat\_3 \_Iag4cat\_4

Adjusted Wald test

( 1) \_Iag4cat\_2 = 0  
( 2) \_Iag4cat\_3 = 0  
( 3) \_Iag4cat\_4 = 0

F( 3, 40) = 19.03  
Prob > F = 0.0000

. test \_Ied4cat\_2 \_Ied4cat\_3 \_Ied4cat\_4

Adjusted Wald test

( 1) \_Ied4cat\_2 = 0  
( 2) \_Ied4cat\_3 = 0  
( 3) \_Ied4cat\_4 = 0

F( 3, 40) = 2.13  
Prob > F = 0.1116

. test \_Imar3cat\_2 \_Imar3cat\_3

Adjusted Wald test

( 1) \_Imar3cat\_2 = 0  
( 2) \_Imar3cat\_3 = 0

F( 2, 41) = 16.60  
Prob > F = 0.0000

```

xi: svy: logistic mde i.ag4cat i.sex ald i.ed4cat i.mar3cat , coef
i.ag4cat      _Iag4cat_1-4      (naturally coded; _Iag4cat_1 omitted)
i.sex         _Isex_1-2        (naturally coded; _Isex_2 omitted)
i.ed4cat      _Ied4cat_1-4      (naturally coded; _Ied4cat_1 omitted)
i.mar3cat     _Imar3cat_1-3     (naturally coded; _Imar3cat_1 omitted)
(running logistic on estimation sample)

```

Survey: Logistic regression

```

Number of strata =      42      Number of obs      =      5692
Number of PSUs  =      84      Population size    = 5692.0005
Design df       =              =      42
F( 10,          33) =          =      28.07
Prob > F        =              =      0.0000

```

mde	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
_Iag4cat_2	.2556185	.0943839	2.71	0.010	.0651441	.4460929
_Iag4cat_3	.2064465	.0915292	2.26	0.029	.0217332	.3911598
_Iag4cat_4	-.6757863	.1413031	-4.78	0.000	-.9609475	-.390625
_Isex_1	-.5773452	.0772153	-7.48	0.000	-.733172	-.4215184
ald	1.423676	.1541609	9.24	0.000	1.112567	1.734786
_Ied4cat_2	.079255	.0968952	0.82	0.418	-.1162874	.2747974
_Ied4cat_3	.2305111	.0930713	2.48	0.017	.0426856	.4183365
_Ied4cat_4	.1629254	.1106149	1.47	0.148	-.0603045	.3861553
_Imar3cat_2	.4864225	.0854196	5.69	0.000	.3140388	.6588062
_Imar3cat_3	.1155794	.1078708	1.07	0.290	-.1021128	.3332716
_cons	-1.583077	.1206597	-13.12	0.000	-1.826578	-1.339575

. \* test of goodness of fit using svylogitgof

```

xi: svy: logistic mde i.ag4cat i.sex ald i.ed4cat i.mar3cat , coef
i.ag4cat      _Iag4cat_1-4      (naturally coded; _Iag4cat_1 omitted)
i.sex         _Isex_1-2        (naturally coded; _Isex_2 omitted)
i.ed4cat      _Ied4cat_1-4      (naturally coded; _Ied4cat_1 omitted)
i.mar3cat     _Imar3cat_1-3     (naturally coded; _Imar3cat_1 omitted)
(running logistic on estimation sample)

```

Survey: Logistic regression

```

Number of strata =      42      Number of obs      =      5692
Number of PSUs  =      84      Population size    = 5692.0005
Design df       =              =      42
F( 10,          33) =          =      28.07
Prob > F        =              =      0.0000

```

mde	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
_Iag4cat_2	.2556185	.0943839	2.71	0.010	.0651441	.4460929
_Iag4cat_3	.2064465	.0915292	2.26	0.029	.0217332	.3911598
_Iag4cat_4	-.6757863	.1413031	-4.78	0.000	-.9609475	-.390625
_Isex_1	-.5773452	.0772153	-7.48	0.000	-.733172	-.4215184
ald	1.423676	.1541609	9.24	0.000	1.112567	1.734786
_Ied4cat_2	.079255	.0968952	0.82	0.418	-.1162874	.2747974
_Ied4cat_3	.2305111	.0930713	2.48	0.017	.0426856	.4183365
_Ied4cat_4	.1629254	.1106149	1.47	0.148	-.0603045	.3861553
_Imar3cat_2	.4864225	.0854196	5.69	0.000	.3140388	.6588062
_Imar3cat_3	.1155794	.1078708	1.07	0.290	-.1021128	.3332716
_cons	-1.583077	.1206597	-13.12	0.000	-1.826578	-1.339575

. svylogitgof

```

F-adjusted test statistic = 1.2292479
p-value                   = .31029551

```

\*interaction testing for sex\*each other predictor group in model

```
. xi: svy: logistic mde i.ag4cat*i.sex i.sex*ald i.ed4cat*i.sex i.mar3cat*i.sex , coef
i.ag4cat      _Iag4cat_1-4      (naturally coded; _Iag4cat_1 omitted)
i.sex         _Isex_1-2        (naturally coded; _Isex_2 omitted)
i.ag4~t*i.sex _Iag4Xsex_#_#        (coded as above)
i.sex*ald     _IsexXald_#      (coded as above)
i.ed4cat      _Ied4cat_1-4      (naturally coded; _Ied4cat_1 omitted)
i.ed4~t*i.sex _Ied4Xsex_#_#        (coded as above)
i.mar3cat     _Imar3cat_1-3     (naturally coded; _Imar3cat_1 omitted)
i.mar~t*i.sex _ImarXsex_#_#      (coded as above)
```

(running logistic on estimation sample)  
note: \_Isex\_1 dropped because of collinearity  
note: \_Isex\_1 dropped because of collinearity  
note: \_Isex\_1 dropped because of collinearity

```
Survey: Logistic regression
Number of strata =      42      Number of obs      =      5692
Number of PSUs  =      84      Population size    = 5692.0005
                                          Design df        =      42
                                          F( 19, 24)      =      17.15
                                          Prob > F        =      0.0000
```

mde	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
_Iag4cat_2	.2204041	.1137606	1.94	0.059	-.0091742	.4499823
_Iag4cat_3	.214641	.1024942	2.09	0.042	.0077993	.4214827
_Iag4cat_4	-.6455558	.1751924	-3.68	0.001	-.9991084	-.2920033
_Isex_1	-.5464415	.3571641	-1.53	0.134	-1.267228	.1743449
_Iag4Xse~2_1	.0967431	.2008412	0.48	0.633	-.3085708	.502057
_Iag4Xse~3_1	.002637	.212819	0.01	0.990	-.4268491	.432123
_Iag4Xse~4_1	-.0378099	.3020415	-0.13	0.901	-.6473543	.5717345
ald	1.55314	.2110231	7.36	0.000	1.127278	1.979002
_IsexXald_1	-.2004168	.2422447	-0.83	0.413	-.6892865	.2884528
_Ied4cat_2	.1305184	.0837105	1.56	0.126	-.0384162	.299453
_Ied4cat_3	.2973241	.1170419	2.54	0.015	.0611241	.5335242
_Ied4cat_4	.2422184	.1518702	1.59	0.118	-.0642682	.5487049
_Ied4Xse~2_1	-.1377802	.2710163	-0.51	0.614	-.6847131	.4091528
_Ied4Xse~3_1	-.1687904	.2693435	-0.63	0.534	-.7123476	.3747667
_Ied4Xse~4_1	-.1940178	.3441702	-0.56	0.576	-.8885814	.5005457
_Imar3cat_2	.4177856	.1105251	3.78	0.000	.194737	.6408342
_Imar3cat_3	.017337	.1297818	0.13	0.894	-.2445733	.2792472
_ImarXse~2_1	.182504	.2078787	0.88	0.385	-.2370122	.6020202
_ImarXse~3_1	.2318977	.2119616	1.09	0.280	-.1958581	.6596536
_cons	-1.599889	.1340095	-11.94	0.000	-1.870331	-1.329447

```
. test _Iag4Xsex_2_1 _Iag4Xsex_3_1 _Iag4Xsex_4_1
```

```
Adjusted Wald test
( 1) _Iag4Xsex_2_1 = 0
( 2) _Iag4Xsex_3_1 = 0
( 3) _Iag4Xsex_4_1 = 0
      F( 3, 40) = 0.25
      Prob > F = 0.8626
```

```
. test _IsexXald_1
```

```
Adjusted Wald test
( 1) _IsexXald_1 = 0
      F( 1, 42) = 0.68
      Prob > F = 0.4127
```

```
. test _Ied4Xsex_2_1 _Ied4Xsex_3_1 _Ied4Xsex_4_1
```

```
Adjusted Wald test
( 1) _Ied4Xsex_2_1 = 0
( 2) _Ied4Xsex_3_1 = 0
( 3) _Ied4Xsex_4_1 = 0
      F( 3, 40) = 0.13
      Prob > F = 0.9444
```

```
. test _ImarXsex_2_1 _ImarXsex_3_1
```

```
Adjusted Wald test
( 1) _ImarXsex_2_1 = 0
( 2) _ImarXsex_3_1 = 0
      F( 2, 41) = 0.77
      Prob > F = 0.4718
```

\*comparison of logit, probit and cloglog

. char sex[omit] 2

. xi: svy: logit ald i.ag4cat i.sex i.ed4cat i.mar3cat  
i.ag4cat \_Iag4cat\_1-4 (naturally coded; \_Iag4cat\_1 omitted)  
i.sex \_Isex\_1-2 (naturally coded; \_Isex\_2 omitted)  
i.ed4cat \_Ied4cat\_1-4 (naturally coded; \_Ied4cat\_1 omitted)  
i.mar3cat \_Imar3cat\_1-3 (naturally coded; \_Imar3cat\_1 omitted)  
(running logit on estimation sample)

Survey: Logistic regression

Number of strata	=	42	Number of obs	=	5692
Number of PSUs	=	84	Population size	=	5692.0005
			Design df	=	42
			F( 9, 34)	=	23.48
			Prob > F	=	0.0000

ald	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
_Iag4cat_2	.1462759	.1781311	0.82	0.416	-.2132072	.505759
_Iag4cat_3	-.050707	.1439205	-0.35	0.726	-.3411503	.2397363
_Iag4cat_4	-1.120335	.2124836	-5.27	0.000	-1.549144	-.6915259
_Isex_1	.9979891	.1191044	8.38	0.000	.7576268	1.238351
_Ied4cat_2	-.2684385	.1937281	-1.39	0.173	-.6593975	.1225206
_Ied4cat_3	-.2644809	.1761355	-1.50	0.141	-.6199367	.0909749
_Ied4cat_4	-.7362285	.1971872	-3.73	0.001	-1.134168	-.3382886
_Imar3cat_2	.5178315	.1420638	3.65	0.001	.2311352	.8045277
_Imar3cat_3	.065315	.1687073	0.39	0.701	-.27515	.4057801
_cons	-3.124321	.2252724	-13.87	0.000	-3.578939	-2.669703

. test \_Iag4cat\_2 \_Iag4cat\_3 \_Iag4cat\_4

Adjusted Wald test

- ( 1) \_Iag4cat\_2 = 0
- ( 2) \_Iag4cat\_3 = 0
- ( 3) \_Iag4cat\_4 = 0

F( 3, 40) = 12.06  
Prob > F = 0.0000

. test \_Ied4cat\_2 \_Ied4cat\_3 \_Ied4cat\_4

Adjusted Wald test

- ( 1) \_Ied4cat\_2 = 0
- ( 2) \_Ied4cat\_3 = 0
- ( 3) \_Ied4cat\_4 = 0

F( 3, 40) = 4.80  
Prob > F = 0.0060

. test \_Imar3cat\_2 \_Imar3cat\_3

Adjusted Wald test

- ( 1) \_Imar3cat\_2 = 0
- ( 2) \_Imar3cat\_3 = 0

F( 2, 41) = 6.54  
Prob > F = 0.0034



```

. xi: svy: probit ald i.ag4cat i.sex i.ed4cat i.mar3cat
i.ag4cat      _Iag4cat_1-4 (naturally coded; _Iag4cat_1 omitted)
i.sex         _Isex_1-2   (naturally coded; _Isex_2 omitted)
i.ed4cat      _Ied4cat_1-4 (naturally coded; _Ied4cat_1 omitted)
i.mar3cat     _Imar3cat_1-3 (naturally coded; _Imar3cat_1 omitted)
(running probit on estimation sample)

```

Survey: Probit regression

```

Number of strata =      42      Number of obs      =      5692
Number of PSUs  =      84      Population size    = 5692.0005
                                           Design df        =      42
                                           F( 9, 34)        =      21.17
                                           Prob > F         =      0.0000

```

ald	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
_Iag4cat_2	.0653189	.0846188	0.77	0.444	-.1054487	.2360865
_Iag4cat_3	-.0344724	.0669712	-0.51	0.609	-.1696258	.100681
_Iag4cat_4	-.5312854	.0933035	-5.69	0.000	-.7195794	-.3429913
_Isex_1	.4708444	.0563443	8.36	0.000	.357137	.5845517
_Ied4cat_2	-.1237841	.0950781	-1.30	0.200	-.3156595	.0680912
_Ied4cat_3	-.1243791	.0851594	-1.46	0.152	-.2962377	.0474795
_Ied4cat_4	-.3395563	.0924728	-3.67	0.001	-.5261739	-.1529386
_Imar3cat_2	.254783	.0697718	3.65	0.001	.1139778	.3955882
_Imar3cat_3	.038951	.0770503	0.51	0.616	-.1165427	.1944448
_cons	-1.719442	.1053591	-16.32	0.000	-1.932066	-1.506819

```

. test _Iag4cat_2 _Iag4cat_3 _Iag4cat_4

```

Adjusted Wald test

- ( 1) \_Iag4cat\_2 = 0
- ( 2) \_Iag4cat\_3 = 0
- ( 3) \_Iag4cat\_4 = 0

```

F( 3, 40) = 15.26
Prob > F = 0.0000

```

```

. test _Ied4cat_2 _Ied4cat_3 _Ied4cat_4

```

Adjusted Wald test

- ( 1) \_Ied4cat\_2 = 0
- ( 2) \_Ied4cat\_3 = 0
- ( 3) \_Ied4cat\_4 = 0

```

F( 3, 40) = 4.79
Prob > F = 0.0061

```

```

. test _Imar3cat_2 _Imar3cat_3

```

Adjusted Wald test

- ( 1) \_Imar3cat\_2 = 0
- ( 2) \_Imar3cat\_3 = 0

```

F( 2, 41) = 6.66
Prob > F = 0.0031

```

```
. xi: svy: cloglog ald i.ag4cat i.sex i.ed4cat i.mar3cat
i.ag4cat      _Iag4cat_1-4      (naturally coded; _Iag4cat_1 omitted)
i.sex         _Isex_1-2        (naturally coded; _Isex_2 omitted)
i.ed4cat      _Ied4cat_1-4      (naturally coded; _Ied4cat_1 omitted)
i.mar3cat     _Imar3cat_1-3     (naturally coded; _Imar3cat_1 omitted)
(running cloglog on estimation sample)
```

Survey: Complementary log-log regression

```
Number of strata =      42      Number of obs      =      5692
Number of PSUs  =      84      Population size     = 5692.0005
Design df       =      42
F( 9, 34)      =      24.23
Prob > F       =      0.0000
```

ald	Coef.	Linearized Std. Err.	t	P> t	[95% Conf. Interval]	
_Iag4cat_2	.1430431	.1710567	0.84	0.408	-.2021633	.4882495
_Iag4cat_3	-.0452402	.1396156	-0.32	0.748	-.3269959	.2365155
_Iag4cat_4	-1.082867	.2085918	-5.19	0.000	-1.503822	-.661912
_Isex_1	.9652235	.1150985	8.39	0.000	.7329453	1.197502
_Ied4cat_2	-.2600972	.1849546	-1.41	0.167	-.6333507	.1131562
_Ied4cat_3	-.2555628	.1685477	-1.52	0.137	-.5957057	.0845802
_Ied4cat_4	-.7126507	.1904971	-3.74	0.001	-1.097089	-.3282121
_Imar3cat_2	.4935371	.1356554	3.64	0.001	.2197735	.7673007
_Imar3cat_3	.0604917	.163547	0.37	0.713	-.2695596	.390543
_cons	-3.148401	.2176024	-14.47	0.000	-3.587541	-2.709262

```
. test _Iag4cat_2 _Iag4cat_3 _Iag4cat_4
```

Adjusted Wald test

- ( 1) [ald]\_Iag4cat\_2 = 0
- ( 2) [ald]\_Iag4cat\_3 = 0
- ( 3) [ald]\_Iag4cat\_4 = 0

```
F( 3, 40) = 11.52
Prob > F = 0.0000
```

```
. test _Ied4cat_2 _Ied4cat_3 _Ied4cat_4
```

Adjusted Wald test

- ( 1) [ald]\_Ied4cat\_2 = 0
- ( 2) [ald]\_Ied4cat\_3 = 0
- ( 3) [ald]\_Ied4cat\_4 = 0

```
F( 3, 40) = 4.77
Prob > F = 0.0062
```

```
. test _Imar3cat_2 _Imar3cat_3
```

Adjusted Wald test

- ( 1) [ald]\_Imar3cat\_2 = 0
- ( 2) [ald]\_Imar3cat\_3 = 0

```
F( 2, 41) = 6.50
Prob > F = 0.0035
```