

Survey Data Analysis in Stata 10: Accessible and Comprehensive

Christine Wells

Statistical Consulting Group
Academic Technology Services
University of California, Los Angeles

Thursday, October 25, 2007

Table of contents

Part 1: Who, when and what

Part 2: Accessible

Part 3: Putting it together

Survey Data Analysis Programs

- ▶ Stata
- ▶ SUDAAN
- ▶ WesVar
- ▶ SAS
- ▶ SPSS
- ▶ Others, including R, AM, IVEware, Mplus, etc.

A Timeline - the '70s and '80s

- 1970:** RTI develops a general proc called stderr
- 1976:** RTI develops 4 routines as SAS procs
- Mid 1980s:** RTI SAS procs become SUDAAN
- 1985:** Westat develops proc wesvar, wesreg and weslog for the mainframe

A Timeline - the '90s

- 1991:** SUDAAN 5
- 1992:** Stata 3: pweights introduced
- 1996:** SUDAAN 7: survival, multilog, catan and hystest
- 1996:** WesVarPC: multidimensional tables, create replicate weights, medians and quantiles
- 1996:** Stata 5: introduces 10 survey commands including mean, total, reg and logit
- 1997:** WesVarPC 2.1
- 1998:** WesVar 3
- 1999:** Stata 6: adds 7 more survey commands, including tab, olog, mlog, pois, ivreg
- 1999:** SAS 8: select, means, reg

A Timeline - 2000 to present

- 2000:** WesVar 4.0: use mi data sets, multinomial logistic, standardized reg coeffs
- 2001:** SUDAAN 8: replicate weights, loglink (poisson)
- 2003:** Stata 8: data must be svyset, 4 new commands
- 2003:** SPSS 12: select, plan, descript, tab
- 2003:** SAS 9.1: freq, logistic
- 2004:** SPSS 14: glm, logistic
- 2004:** SUDAAN 9: replicate weights for survival, use mi data sets, kapmeier
- 2005:** Stata 9: changes svyset command, adds more features to svyset
- 2006:** SPSS 15: ordinal
- 2007:** Stata 10: adds 27 more svy: procedures, including survival (total of 48)
- 2007:** WesVar 5.1: descriptive stats
- 2007:** SPSS 16: cox

Definitions

- ▶ Sampling plan or design - departure from SRS

Definitions

- ▶ Sampling plan or design - departure from SRS
- ▶ Survey procedures (e.g., `svy: tabulate`, `svy: regress`, `svy: logistic`, etc.)

Definitions

- ▶ Sampling plan or design - departure from SRS
- ▶ Survey procedures (e.g., `svy: tabulate`, `svy: regress`, `svy: logistic`, etc.)
- ▶ Elements of the sampling design

Definitions

- ▶ Sampling plan or design - departure from SRS
- ▶ Survey procedures (e.g., `svy: tabulate`, `svy: regress`, `svy: logistic`, etc.)
- ▶ Elements of the sampling design
 - ▶ Sampling or probability weights (AKA `pweights`)

Definitions

- ▶ Sampling plan or design - departure from SRS
- ▶ Survey procedures (e.g., `svy: tabulate`, `svy: regress`, `svy: logistic`, etc.)
- ▶ Elements of the sampling design
 - ▶ Sampling or probability weights (AKA `pweights`)
 - ▶ Stratification (and post-stratification)

Definitions

- ▶ Sampling plan or design - departure from SRS
- ▶ Survey procedures (e.g., `svy: tabulate`, `svy: regress`, `svy: logistic`, etc.)
- ▶ Elements of the sampling design
 - ▶ Sampling or probability weights (AKA `pweights`)
 - ▶ Stratification (and post-stratification)
 - ▶ Primary sampling units (PSUs and maybe SSUs, TSUs, etc.)

Definitions

- ▶ Sampling plan or design - departure from SRS
- ▶ Survey procedures (e.g., `svy: tabulate`, `svy: regress`, `svy: logistic`, etc.)
- ▶ Elements of the sampling design
 - ▶ Sampling or probability weights (AKA `pweights`)
 - ▶ Stratification (and post-stratification)
 - ▶ Primary sampling units (PSUs and maybe SSUs, TSUs, etc.)
 - ▶ Replicate weights (jackknife and BRR)

Comparison of packages

Features of sampling plans accommodated by the various packages

Sampling design feature	Stata	SUDAAN	WesVar	SAS	SPSS
probability weight	✓	✓	✓	✓	✓
stratification	✓	✓	✓	✓	✓
PSUs	✓	✓	✓	✓	✓
levels of sampling	multi	two	one	one	three
post-stratification	✓	✓	✓		
replicate weights	✓	✓	✓		
subpop for all procs	✓	✓	✓		✓

Comparison of packages

Procedures available in current version of packages

Procedure	Stata 10	SUDAAN 9	WesVar 5.1	SAS 9.1	SPSS 16
frequencies and crosstabs	✓	✓	✓	✓	✓
3-way crosstabs		✓	✓		✓
means, totals and ratios	✓	✓	✓	✓	✓
medians and quantiles		✓	✓		
correlations	*		✓		
OLS regression/GLM	✓	✓	✓	✓	✓
logistic	✓	✓	✓	✓	✓
ordered logit	✓	✓		✓	✓
multinomial logit	✓	✓	✓	✓	
survival (Cox regression)	✓	✓			✓
select sample	✓			✓	✓
create replicate weights	*		✓		
handle mi data sets	*	✓	✓	✓	

* user-written .ado available

Three things that make Stata's survey offerings comprehensive

- ▶ sampling plans (Stata 9)
- ▶ procedures (Stata 10, especially survival analysis)
- ▶ computational options (e.g., standard errors and singleton PSUs)

Commonly requested by clients

- ▶ 3-way tabulations (over option for svy: tabulate?)
- ▶ correlation
- ▶ median / quantiles

Commonly requested by clients

- ▶ 3-way tabulations (over option for svy: tabulate?)
- ▶ correlation
- ▶ median / quantiles
- ▶ create replicate weights
(may become more important as more researchers make their data sets available on the web)

Commonly requested by clients

- ▶ 3-way tabulations (over option for svy: tabulate?)
- ▶ correlation
- ▶ median / quantiles
- ▶ create replicate weights
(may become more important as more researchers make their data sets available on the web)
- ▶ create multiply imputed data sets
(many requests from clients for this)

Accessibility

▶ svyset

```
svyset my_psu [pweight=final_pw], strat(strat_var)
```

Accessibility

- ▶ svyset

```
svyset my_psu [pweight=final_pw], strat(strat_var)
```

- ▶ svy: prefix - don't have to learn new procedures

```
regress write read math female
```

```
svy: regress write read math female
```

Accessibility

- ▶ svyset

```
svyset my_psu [pweight=final_pw], strat(strat_var)
```

- ▶ svy: prefix - don't have to learn new procedures

```
regress write read math female
```

```
svy: regress write read math female
```

- ▶ svy: procedures now in reference manuals

Accessibility

- ▶ svyset

```
svyset my_psu [pweight=final_pw], strat(strat_var)
```

- ▶ svy: prefix - don't have to learn new procedures

```
regress write read math female  
svy: regress write read math female
```

- ▶ svy: procedures now in reference manuals

- ▶ good documentation with helpful examples

Accessibility

- ▶ svyset

```
svyset my_psu [pweight=final_pw], strat(strat_var)
```

- ▶ svy: prefix - don't have to learn new procedures

```
regress write read math female  
svy: regress write read math female
```

- ▶ svy: procedures now in reference manuals

- ▶ good documentation with helpful examples

- ▶ cross platform

Commonly needed by clients

- ▶ Stata NetCourse (or 2)

Commonly needed by clients

- ▶ Stata NetCourse (or 2)
- ▶ User-written books:
 - ▶ applied
 - ▶ syntax
 - ▶ both basic and advanced examples
 - ▶ details regarding when to use certain options and why

The consequences

- ▶ raise the standard - we have been seeing this trend in many areas of statistics

The consequences

- ▶ raise the standard - we have been seeing this trend in many areas of statistics
- ▶ survey designers can use more not-so-standard designs, knowing that easy-to-use software is readily available that will handle the sampling design correctly

The consequences

- ▶ raise the standard - we have been seeing this trend in many areas of statistics
- ▶ survey designers can use more not-so-standard designs, knowing that easy-to-use software is readily available that will handle the sampling design correctly
- ▶ increasing interest from researchers to "do things right", continued desire for even more procedures and options

The consequences

- ▶ raise the standard - we have been seeing this trend in many areas of statistics
- ▶ survey designers can use more not-so-standard designs, knowing that easy-to-use software is readily available that will handle the sampling design correctly
- ▶ increasing interest from researchers to "do things right", continued desire for even more procedures and options
- ▶ many clients simply expect that they can do anything with survey data that they can do with data collected via SRS