CHAPTER 5 ASDA ANALYSIS EXAMPLES REPLICATION-SAS v9.2

GENERAL NOTES ABOUT ANALYSIS EXAMPLES REPLICATION
These examples are intended to provide guidance on how to use the commands/procedures for analysis of complex sample survey data and assume all data management and other preliminary work is done. The relevant syntax for the procedure of interest is shown first along with the associated output for that procedure(s). In some examples, there may be more than one block of syntax and in this case all syntax is first presented followed by the output produced.

In some software packages certain procedures or options are not available but we have made every attempt to demonstrate how to match the output produced by Stata 10+ in the textbook. Check the ASDA website for updates to the various software tools we cover.

NOTES ABOUT DESCRIPTIVE ANALYSES IN SAS V9.2
SAS does not offer an easy way to do weighted histograms or box plots/bar charts therefore these are not included in this output. SAS PROC SURVEYMEANS can perform nearly all of the descriptive analyses presented in Chapter 5 of ASDA. Some of the fine points of this procedure are the use of a DOMAIN statement for subpopulation analyses, various output statistics specified on the PROC statement, and use of ratio statements for a ratio analysis. Note that the DEFF option for design effects is not available in this particular procedure. Another important point to note is that the DOMAIN statement is not available when requesting quantiles or ratios, hence the use of the BY statement in these examples. Finally, SAS PROC SURVEYMEANS does not include the ability to perform linear contrasts for difference between means but a SAS Institute provided macro called smsub.sas does offer this ability via a macro program, see examples of the use of smsub.sas in this chapter. This macro can be downloaded from the SAS Support website, see the ASDA website for the links to SAS. Also, for more details on the procedures used, see the SAS documentation.

ANALYSIS EXAMPLE 5.1: NOTE: UNWEIGHTED HISTOGRAM IS THE ONLY OPTION VIA PROC UNIVARIATE IN SAS 9.2 (NOTE: NOT INCLUDED)

ANALYSIS EXAMPLE 5.2: NOTE: UNWEIGHTED BOX PLOT IS THE ONLY OPTION IN SAS 9.2 (NOTE: NOT INCLUDED)

```
options ls=125 ps=64 ;
title "Analysis Examples 5.3 : Totals: NCSR data " ;
proc surveymeans data=ncsr nobs sum df stderr clsum ;
strata sestrat ;
cluster seclustr ;
weight ncsrwtsh_pop ;
var mde ;
run ;
```

                          Analysis Examples 5.3 : Totals: NCSR data

                                  The SURVEYMEANS Procedure

                                        Data Summary

                        Number of Strata   42
                        Number of Clusters   84
                        Number of Observations   9282
                        Sum of Weights   209128097

                                         Statistics

                            Std Error
Variable   N   DF   of Mean   Sum   Std Dev   95% CL for Sum
_____

mde   9282   42   0.004877   40092207   2567488   34910806.0 45273607.0
_____

```
proc surveymeans data=ncsr nobs sum df stderr clsum ;
strata sestrat ;
cluster seclustr ;
weight ncsrwtsh_pop  ;
format mar3cat marf. ;
domain mar3cat ;
var mde  ;
run ;
```

The SURVEYMEANS Procedure

             Data Summary

Number of Strata   42
Number of Clusters   84
Number of Observations   9282
Sum of Weights   209128097

                                          Statistics

                                  Std Error
Variable   N   DF   of Mean   Sum   Std Dev   95% CL for Sum
----
mde   9282   42   0.004877   40092207   2567488   34910806.0 45273607.0

                          Domain Analysis: Marital Status-3 categories

    Marital
    Status-3   Std Error
   categories   Variable   N   DF   of Mean   Sum   Std Dev   95% CL for Sum
----
1=Married   mde   5322   42   0.006415   20304191   1584109   17107329.8 23501051.2
2=Previously   mde   2017   42   0.010533   10360671   702622   8942723.0 11778618.3
3=Never Marr   mde   1943   42   0.010247   9427345   773138   7867090.6 10987600.2

```

```sas
title  "Analysis Example 5.4: HRS 2006 HH Level Analysis of Total Assets" ;
proc surveymeans data=hrs nobs df sum clsum ;
strata stratum ;
cluster secu ;
domain kfinr ;
weight kwgthh ;
var h8atota ;
run ;
```

Analysis Example 5.4: HRS 2006 HH Level Analysis of Total Assets

The SURVEYMEANS Procedure

Data Summary

Number of Strata	56
Number of Clusters	112
Number of Observations	18467
Number of Observations Used	17826
Number of Obs with Nonpositive Weights	641
Sum of Weights	82249292

Statistics

Variable	Label	N	DF	Sum	Std Dev	95% CL for Sum	
H8ATOTA	h8atota:w8 total of all assets	17826	56	4.8676892E13	2.8623769E12	4.29429E13	5.44109E13

Domain Analysis: 2006 whether financial respondent

2006 whether financial respondent	Variable	Label	N	DF	Sum	Std Dev	95% CL for Sum	
1	H8ATOTA	h8atota:w8 total of all assets	11942	56	2.8397486E13	1.5955856E12	2.52011E13	3.15938E13
5	H8ATOTA	h8atota:w8 total of all assets	5884	56	2.0279407E13	1.3230999E12	1.76289E13	2.29299E13

```
title "Analysis Example 5.5: NCS-R Data HH Income" ;
proc surveymeans data=ncsr nobs df mean stderr clm ;
strata sestrat ;
cluster seclustr ;
weight ncsrwtlg  ;
var hhinc ;
run ;
```

Analysis Example 5.5: NCS-R Data HH Income

The SURVEYMEANS Procedure

Data Summary

Number of Strata	42
Number of Clusters	84
Number of Observations	9282
Number of Observations Used	5692
Number of Obs with Nonpositive Weights	3590
Sum of Weights	5692.00048

Statistics

Variable	Label	N	DF	Mean	Std Error of Mean	95% CL for Mean
HHINC	Household Income : Topcode	5692	42	59277	1596.342971	56055.5078 62498.6089

```
title "Analysis Example 5.6: NHANES 2005-2006" ;
proc surveymeans data=nhanes0506 nobs df mean stderr clm  ;
strata sdmvstra ;
cluster sdmvpsu ;
weight wtmec2yr ;
domain age18p ;
var bpxsy1 ;
run ;
```

Analysis Example 5.6: NHANES 2005-2006

The SURVEYMEANS Procedure

Data Summary

Number of Strata	15
Number of Clusters	30
Number of Observations	10348
Number of Observations Used	9950
Number of Obs with Nonpositive Weights	398
Sum of Weights	291616892

Statistics

Variable	Label	N	DF	Mean	Std Error of Mean	95% CL for Mean
BPXSY1	Systolic: Blood pres (1st rdg) mm Hg	6668	15	120.655644	0.490947	119.609215 121.702073

Domain Analysis: age18p

age18p	Variable	Label	N	DF	Mean	Std Error of Mean
0	BPXSY1	Systolic: Blood pres (1st rdg) mm Hg	2053	15	107.866079	0.629871
1	BPXSY1	Systolic: Blood pres (1st rdg) mm Hg	4615	15	123.110876	0.541694

Domain Analysis: age18p

age18p	Variable	95% CL for Mean
0	BPXSY1	106.523540 109.208618
1	BPXSY1	121.956283 124.265469

```
title  "Analysis Example 5.7: HRS HH Level Analysis " ;
proc surveymeans data=hrs nobs df mean stderr clm  ;
strata stratum ;
cluster secu ;
weight kwgthh ;
domain kfinr ;
var h8atota ;
run ;
```

Analysis Example 5.7: HRS HH Level Analysis

The SURVEYMEANS Procedure

Data Summary

Number of Strata	56
Number of Clusters	112
Number of Observations	18467
Number of Observations Used	17826
Number of Obs with Nonpositive Weights	641
Sum of Weights	82249292

Statistics

Variable	Label	N	DF	Mean	Std Error of Mean	95% CL for Mean
H8ATOTA	h8atota:w8 total of all assets	17826	56	591821	32992	525729.712 657913.118

Domain Analysis: 2006 whether financial respondent

2006 whether financial respondent	Variable	Label	N	DF	Mean	Std Error of Mean	95% CL for Mean
1	H8ATOTA	h8atota:w8 total of all assets	11942	56	527313	28013	471196.817 583429.505
5	H8ATOTA	h8atota:w8 total of all assets	5884	56	714161	44892	624231.939 804090.369

```
title "Analysis Example 5.8:  HRS HH Level Analysis of Quantiles" ;
proc surveymeans data=hrs q1 median q3 ;
strata stratum ;
cluster secu ;
weight kwgthh ;
where kfinr=1  ;
var h8atota ;
run ;
```

Analysis Example 5.8: HRS HH Level Analysis of Quantiles

The SURVEYMEANS Procedure

Data Summary

Number of Strata	56
Number of Clusters	112
Number of Observations	12558
Number of Observations Used	11942
Number of Obs with Nonpositive Weights	616
Sum of Weights	53853171

Quantiles

Variable	Label	Percentile	Estimate	Std Error	95% Confidence Limits	
H8ATOTA	h8atota:w8 total of all assets	25% Q1	39853	3258.139382	33326.094	46379.769
	h8atota:w8 total of all assets	50% Median	183309	9977.330641	163322.041	203296.031
	h8atota:w8 total of all assets	75% Q3	495931	17394	461086.518	530776.236

```sas
title  "Analysis Example 5.9: NHANES Data for Ratios " ;
* Note: domain analysis not available with ratio, use where statement instead ;
proc surveymeans data=nhanes0506 nobs df mean df stderr clm ratio ;
strata sdmvstra ;
cluster sdmvpsu ;
weight wtmec2yr ;
where age18p=1 ;
var lbdhdd lbxtc ;
ratio lbdhdd/lbxtc ;
run ;
```

Analysis Example 5.9: NHANES Data for Ratios

The SURVEYMEANS Procedure

Data Summary

Number of Strata	15
Number of Clusters	30
Number of Observations	5563
Number of Observations Used	5334
Number of Obs with Nonpositive Weights	229
Sum of Weights	217700471

Statistics

Variable	Label	N	DF	Mean	Std Error of Mean	95% CL for Mean	
LBDHDD	Direct HDL-Cholesterol (mg/dL)	4996	15	54.530873	0.332403	53.822373	55.239373
LBXTC	Total Cholesterol(mg/dL)	4996	15	198.038723	0.778063	196.380322	199.697125

Ratio Analysis

Numerator	Denominator	N	DF	Ratio	Std Err	95% CL for Ratio	
LBDHDD	LBXTC	4996	15	0.275355	0.002254	0.27054956	0.28015963

```
title  "Analysis Example 5.10: Proportions in SubGroups: HRS" ;
proc format ;
value genf 1='Male' 2='Female' ;
proc surveymeans data=hrs mean clm df ;
strata stratum ;
cluster secu ;
weight kwgtr ;
domain age70*gender ;
var diabetes;
format gender genf. ;
run ;
```

Analysis Example 5.10: Proportions in SubGroups: HRS

The SURVEYMEANS Procedure

Data Summary

Number of Strata	56
Number of Clusters	112
Number of Observations	18467
Number of Observations Used	16954
Number of Obs with Nonpositive Weights	1513
Sum of Weights	76540667

Statistics

Variable	DF	Mean	Std Error of Mean	95% CL for Mean	
DIABETES	56	0.183412	0.003657	0.17608639	0.19073733

Domain Analysis: age70*gender

age70	gender	Variable	DF	Mean	Std Error of Mean	95% CL for Mean	
0	Male	DIABETES	56	0.181131	0.006595	0.16791945	0.19434347
	Female	DIABETES	56	0.165275	0.005961	0.15333403	0.17721670
1	Male	DIABETES	56	0.235374	0.008317	0.21871368	0.25203506
	Female	DIABETES	56	0.183865	0.008531	0.16677582	0.20095508

```
title  "Analysis Example 5.11: Means by Subgroups: NHANES " ;
proc format ;
value femf 0='Male' 1='Female' ;
proc surveymeans data=nhanes0506 nobs mean df clm ;
strata sdmvstra ;
cluster sdmvpsu ;
weight wtmec2yr ;
domain age45 * female ;
var bpxsy1 ;
format female femf. ;
run ;
```

Analysis Example 5.11: Means by Subgroups: NHANES

The SURVEYMEANS Procedure

Data Summary

Number of Strata	15
Number of Clusters	30
Number of Observations	10348
Number of Observations Used	9950
Number of Obs with Nonpositive Weights	398
Sum of Weights	291616892

Statistics

Variable	Label	N	DF	Mean	Std Error of Mean	95% CL for Mean	
BPXSY1	Systolic: Blood pres (1st rdg) mm Hg	6668	15	120.655644	0.490947	119.609215	121.702073

Domain Analysis: age45*female

age45	female	Variable	Label	N	DF	Mean	Std Error of Mean
0	Male	BPXSY1	Systolic: Blood pres (1st rdg) mm Hg	2148	15	117.564830	0.552835
	Female	BPXSY1	Systolic: Blood pres (1st rdg) mm Hg	2426	15	110.836053	0.392178
1	Male	BPXSY1	Systolic: Blood pres (1st rdg) mm Hg	1093	15	128.962932	0.756667
	Female	BPXSY1	Systolic: Blood pres (1st rdg) mm Hg	1001	15	132.087263	1.064538

Domain Analysis: age45*female

age45	female	Variable	95% CL for Mean	
0	Male	BPXSY1	116.386491	118.743170
	Female	BPXSY1	110.000145	111.671962
1	Male	BPXSY1	127.350135	130.575728
	Female	BPXSY1	129.818253	134.356272

```
proc format ;
value edf 1='0-11 Years' 2='12 Years' 3='13-15 Years' 4='16+ Years' ;

title  "Analysis Example 5.12: Proportions in SubGroups: HRS" ;
proc surveymeans data=hrs mean clm ;
strata stratum ;
cluster secu ;
weight kwgthh ;
domain kfinr*edcat ;
var h8atota ;
format edcat edf. ;
run ;
```

The SURVEYMEANS Procedure

Data Summary

Number of Strata	56
Number of Clusters	112
Number of Observations	18467
Number of Observations Used	17826
Number of Obs with Nonpositive Weights	641
Sum of Weights	82249292

Statistics

Variable	Label	Mean	Std Error of Mean	95% CL for Mean	
H8ATOTA	h8atota:w8 total of all assets	591821	32992	525729.712	657913.118

Domain Analysis: 2006 whether financial respondent*EDCAT

2006 whether financial respondent	EDCAT	Variable	Label	Mean	Std Error of Mean	95% CL for Mean	
1	0-11 Years	H8ATOTA	h8atota:w8 total of all assets	178386	24561	129184.215	227587.89
	12 Years	H8ATOTA	h8atota:w8 total of all assets	328392	17083	294171.161	362612.75
	13-15 Years	H8ATOTA	h8atota:w8 total of all assets	455458	27000	401369.445	509545.77
	16+ Years	H8ATOTA	h8atota:w8 total of all assets	1107204	102114	902646.143	1311762.06
5	0-11 Years	H8ATOTA	h8atota:w8 total of all assets	255742	21602	212468.405	299016.23
	12 Years	H8ATOTA	h8atota:w8 total of all assets	479708	35778	408035.852	551379.19
	13-15 Years	H8ATOTA	h8atota:w8 total of all assets	860052	126500	606643.316	1113461.39
	16+ Years	H8ATOTA	h8atota:w8 total of all assets	1258384	147710	962485.829	1554282.96

```sas
%inc 'f:\applied_analysis_book\smsub.sas' ;
%smsub (
data=hrs,
statistics = mean semean clmean ,
strata=stratum,
cluster=secu,
weight= kwgthh,
subpop = kfinr=1,
var= h8atota,
contrast= 'edcat: contrasts 1 v. 4 ' edcat 1 0 0 -1 ,
title= 'Example of Difference of Means with smsub.sas code' ) ;
run ;
```

Example of Difference of Means with smsub.sas code

The SURVEYMEANS SUBGROUP Macro

                    Data Summary

Subpopulation: kfinr=1

Number of Observations Read: 18467
Number of Observations Used: 17826   Weight Sum: 82249292
Number of Observations Deleted: 641
  (Nonpositive Weights)
Subpopulation Observations: 11942   Weight Sum: 53853171

Number of Strata: 56
Number of PSUs: 112
Denominator Degrees of Freedom: 56

                    Contrast Vector(s)

            Label   Variable   Values   Coefficients

edcat: contrasts 1 v. 4   edcat   1   1
                                               2   0
                                               3   0
                                               4   -1

                    Table: Overall *Contrast

Analysis Variable: h8atota

|By: Overall, Contrast | | Std Error | Lower 95% | Upper 95% |
| | Mean | of Mean |CL for Mean |CL for Mean |
|------------------------+------------+------------+------------+------------|
Total				
edcat: contrasts 1				
v. 4	-928818.0495	108250.0782	-1145669.014	-711967.0851

```

Analysis Example 5.13 HRS Data

```
*data management first, as example of combining two data sets for analysis ;
data ex5_13 ;
set d.d2004_2006_dec2008 ;
if year=2004 then weight=jwgthh ; else weight=kwgthh ;
finr0406 =jfinr ;
if year=2006 then finr0406=kfinr ;

proc surveymeans ;
strata stratum ;
cluster secu ;
weight weight ;
domain finr0406 * year ;
var totassets ;
run ;
* obtain difference in means between 2004 and 2006 for financial respondents ;

%inc 'f:\applied_analysis_book\smsub.sas' ;
%smsub (
data=ex5_13,
statistics = mean semean clmean ,
strata=stratum,
cluster=secu,
weight= weight,
subpop = finr0406=1,
var= totassets,
contrast= 'total assets 2004 v 2006' year 1 -1 ,
title= 'Example 5.13: HRS Total Assets Difference of Means with smsub.sas code' ) ;
run ;
```

The SURVEYMEANS Procedure

                    Data Summary

Number of Strata   56
Number of Clusters   112
Number of Observations   36288
Number of Observations Used   35268
Number of Obs with Nonpositive Weights   1020
Sum of Weights   159934747

                        Statistics

                                      Std Error
Variable   N   Mean   of Mean   95% CL for Mean
----
totassets   35268   528257   26508   475155.266 581358.607

                    Domain Analysis: finr0406*year

                                              Std Error
finr0406   year Variable   N   Mean   of Mean   95% CL for Mean
----
       1   2004 totassets   11563   411786   20640 370440.065 453132.704
            2006 totassets   11942   527313   28013 471196.817 583429.505
       5   2004 totassets   5879   556185   31626 492830.882 619538.329
            2006 totassets   5884   714161   44892 624231.939 804090.369

Example 5.13: HRS Total Assets Difference of Means with smsub.sas code

The SURVEYMEANS SUBGROUP Macro

                Data Summary

Subpopulation: finr0406=1

Number of Observations Read: 36288
Number of Observations Used: 35268   Weight Sum: 159934747
Number of Observations Deleted: 1020
 (Nonpositive Weights)
Subpopulation Observations: 23505   Weight Sum: 105084460

Number of Strata: 56
Number of PSUs: 112
Denominator Degrees of Freedom: 56

                Contrast Vector(s)

         Label   Variable   Values   Coefficients

total assets 2004 v 2006   year   2004   1
                                           2006   -1

                Table: Overall *Contrast

Analysis Variable: totassets

|By: Overall, Contrast | | Std Error | Lower 95% | Upper 95% |
| | Mean | of Mean |CL for Mean |CL for Mean |
|------------------------+------------+------------+------------+------------|
Total				
total assets 2004 v				
2006	-115526.7767	20025.4113	-155642.4960	-75411.0574