CHAPTER 11 ASDA ANALYSIS EXAMPLES REPLICATION-SUDAAN 10.0.1

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 IMPUTATION AND ANALYZING IMPUTED DATA USING SUDAAN 10.0.1

The analysis replication examples were all run using SAS-callable SUDAAN version 10.0.1. There are very few differences between SAS-callable and stand-alone SUDAAN with the exception of the names of the procedures are sometimes slightly different as to avoid confusion with SAS procedures.

SUDAAN does not offer the ability to perform graphical analyses within the program therefore are not included in this output however output data sets can be saved and used in other software packages.

Nearly all SUDAAN procedures can perform analysis of multiply imputed data sets (produced by some other software package). This example demonstrates use of an imputed data set imputed by SAS PROC MI and then analyzed using SUDAAN proc regress for linear regression.

Some of the fine points of this approach: use of a MI_COUNT statement on the procedure statement and statement of just the first imputed data set. The program will automatically recognize each of the imputed data sets on the MI_COUNT statement provided they are named in a consecutive manner. Another feature used is the EFFECTS statement for a test of the race and marital status variables as a group (testing if they are significantly different than 0), note that this program defaults to a Wald F test, not the adjusted Wald F test. Please see the Language and Examples Guides for additional detail on programming and analysis of imputed data sets.

```
*EXAMPLE OF HOW TO CREATE 5 OUTPUT DATA SETS FROM ONE "LONG" FILE OF IMPUTED DATA PRODUCED BY SAS ;
*NOTE THAT EACH FILE ALSO NEEDS TO BE SORTED FOR SUDAAN ;

data c11_m1 c11_m2 c11_m3 c11_m4 c11_m5  ;
set d.c11_imputed ;
if _imputation_=1  then output c11_m1 ;
if _imputation_=2  then output c11_m2 ;
if _imputation_=3  then output c11_m3 ;
if _imputation_=4  then output c11_m4 ;
if _imputation_=5  then output c11_m5 ;

proc sort data=c11_m1 ;
       by sdmvstra sdmvpsu ;
run ;
proc sort data=c11_m2 ;
       by sdmvstra sdmvpsu ;
run ;
proc sort data=c11_m3 ;
       by sdmvstra sdmvpsu ;
run ;
proc sort data=c11_m4 ;
       by sdmvstra sdmvpsu ;
run ;
proc sort data=c11_m5 ;
       by sdmvstra sdmvpsu ;
run ;


* regression example using imputed data sets (from SAS) ;
options ls=120 ;
proc regress data=c11_m1 filetype=sas mi_count=5 ;
rtitle "Analysis Example Replication Chapter 11: Sudaan Regression using Fully Imputed Data Set Produced by SAS" ;
nest sdmvstra sdmvpsu  ;
weight wtmec2yr  ;
class marcat ridreth1  riagendr / nofreq ;
reflevel marcat=1 ridreth1=1 riagendr=1 ;
model  bpxdi1_1 = ridreth1 marcat riagendr agec agecsq bmxbmi indfmpir  ;
setenv decwidth=3 ;
effects marcat ridreth1 / name="Test Race and Marital Status " ;
print / style=nchs ;
run ;
```

Analysis Example Replication Chapter 11: Sudaan Regression using Fully Imputed Data Set Produced by SAS

S U D A A N
Software for the Statistical Analysis of Correlated Data
Copyright Research Triangle Institute January 2009
Release 10.0.1

DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method, Assuming a With Replacement (WR) Design
Sample Weight: WTMEC2YR
Stratification Variables(s): SDMVSTRA
Primary Sampling Unit: SDMVPSU

Processing data for set 1 of imputed variables:

Processing data for set 2 of imputed variables:

Processing data for set 3 of imputed variables:

Processing data for set 4 of imputed variables:

Processing data for set 5 of imputed variables:

Processing data for set 1 of imputed variables:

Number of observations read : 5334 Weighted count:217700471
Observations used in the analysis : 5334 Weighted count:217700471
Denominator degrees of freedom : 15

Maximum number of estimable parameters for the model is 12

File C11_M1 contains 30 Clusters
30 clusters were used to fit the model
Maximum cluster size is 274 records
Minimum cluster size is 128 records

Weighted mean response is 70.516998

Multiple R-Square for the dependent variable BPXDI1_1: 0.121957

Processing data for set 2 of imputed variables:

Number of observations read : 5334 Weighted count:217700471
Observations used in the analysis : 5334 Weighted count:217700471
Denominator degrees of freedom : 15

Maximum number of estimable parameters for the model is 12

File C11_M2 contains 30 Clusters
30 clusters were used to fit the model
Maximum cluster size is 274 records
Minimum cluster size is 128 records

Weighted mean response is 70.435860

Multiple R-Square for the dependent variable BPXDI1_1: 0.124054

```
Processing data for set 3 of imputed variables:

Number of observations read       :   5334    Weighted count:217700471
Observations used in the analysis :   5334    Weighted count:217700471
Denominator degrees of freedom    :     15


Maximum number of estimable parameters for the model is 12

File C11_M3 contains   30 Clusters
  30 clusters were used to fit the model
Maximum cluster size is 274 records
Minimum cluster size is 128 records

Weighted mean response is 70.486586


Multiple R-Square for the dependent variable BPXDI1_1: 0.120386


Processing data for set 4 of imputed variables:

Number of observations read       :   5334    Weighted count:217700471
Observations used in the analysis :   5334    Weighted count:217700471
Denominator degrees of freedom    :     15


Maximum number of estimable parameters for the model is 12

File C11_M4 contains   30 Clusters
  30 clusters were used to fit the model
Maximum cluster size is 274 records
Minimum cluster size is 128 records

Weighted mean response is 70.491619


Multiple R-Square for the dependent variable BPXDI1_1: 0.127576

Processing data for set 5 of imputed variables:

Number of observations read       :   5334    Weighted count:217700471
Observations used in the analysis :   5334    Weighted count:217700471
Denominator degrees of freedom    :     15


Maximum number of estimable parameters for the model is 12

File C11_M5 contains   30 Clusters
  30 clusters were used to fit the model
Maximum cluster size is 274 records
Minimum cluster size is 128 records

Weighted mean response is 70.384540


Multiple R-Square for the dependent variable BPXDI1_1: 0.121190

Overall degrees of freedom (Rubin):  5.18
```

Variance Estimation Method: Taylor Series (WR) Using Multiply Imputed Data
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Identity
Response variable BPXDI1_1: BPXDI1_1
Analysis Example Replication Chapter 11: Sudaan Regression using Fully Imputed Data Set Produced by SAS
Results for Summary Over All Imputations
by: Independent Variables and Effects.

Independent Variables and Effects	Beta Coeff.	SE Beta	Lower 95% Limit Beta	Upper 95% Limit Beta	T-Test B=0	P-value T-Test B=0	DDF Beta
Intercept	67.434	1.349	64.395	70.473	49.971	0.000	9.267
1=married 2=prev married 3=never married							
1	0.000	0.000	.	.	.	.	.
2	0.624	0.639	-0.760	2.008	0.976	0.347	12.763
3	-0.651	0.805	-2.698	1.396	-0.809	0.454	5.179
1=mex 2=oth hisp 3=white 4=black 5=other							
1	0.000	0.000	.	.	.	.	.
2	1.397	1.144	-1.206	4.000	1.221	0.254	8.670
3	2.205	0.584	0.936	3.474	3.773	0.003	12.396
4	3.209	0.887	1.055	5.363	3.619	0.011	6.186
5	2.142	1.130	-0.725	5.010	1.895	0.114	5.228
gender - adjudicated							
1	0.000	0.000	.	.	.	.	.
2	-2.865	0.432	-3.909	-1.820	-6.629	0.000	6.325
AGEC	0.108	0.016	0.072	0.143	6.815	0.000	8.938
AGECSQ	-0.010	0.001	-0.012	-0.008	-11.300	0.000	10.583
body mass index (kg/m**2)	0.196	0.037	0.115	0.277	5.358	0.000	10.499
family pir	-0.021	0.148	-0.356	0.315	-0.139	0.892	8.834

Variance Estimation Method: Taylor Series (WR) Using Multiply Imputed Data
SE Method: Robust (Binder, 1983)
Working Correlations: Independent
Link Function: Identity
Response variable BPXDI1_1: BPXDI1_1
Analysis Example Replication Chapter 11: Sudaan Regression using Fully Imputed Data Set Produced by SAS
Results for Summary Over All Imputations
by: Contrast.

Contrast	Degrees of Freedom	Wald F	P-value Wald F
OVERALL MODEL	12.000	8373.028	0.000
MODEL MINUS INTERCEPT	11.000	107.437	0.000
INTERCEPT	.	.	.
MARCAT	2.000	1.022	0.423
RIDRETH1	4.000	4.664	0.058
RIAGENDR	1.000	43.937	0.001
AGEC	1.000	46.440	0.001
AGECSQ	1.000	127.681	0.000
BMXBMI	1.000	28.708	0.003
INDFMPIR	1.000	0.019	0.894
Test Race and Marital Status	6.000	4.771	0.051