

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 MULTIPLE IMPUTATION USING IVEware

Special note: IVEware MUST BE RUN IN THE REGULAR PROGRAM EDITOR IF RUNNING UNDER SAS!! THE ENHANCED EDITOR DOES NOT WORK WITH SAS BASED IVEware!!

IVEware %impute performs multiple imputation using the Sequential Regression method and offers the ability to impute missing data and analyze the imputed data sets taking the complex sample design and the variability introduced by imputation into account (%describe and %regress). This program is very easy to use and provides a convenient way to impute and analyze data correctly.

Note that some of the post-estimation statistics such as within, between and total variance that are available in SAS and Stata are not included in the IVEware output.

These examples use a subset of the NHANES 2005-2006 data for adults (18+ years of age) who had a non-zero weight on the WTMEC2YR variable (n=5334). There is some missing data on a few of the variables included in this subset and a means analysis of diastolic blood pressure and a linear regression analysis are included in the chapter examples.

```

data c11 ;
set d.impute_subset_nhanes ;
where age18p=1 and wtmec2yr > 0 ;

* examine missing data ;
proc means nmiss mean min max ;
run ;

```

The SAS System

The MEANS Procedure

Variable	Label	N Miss	Mean	Minimum	Maximum
SEQN	respondent sequence number	0	36358.80	31130.00	41473.00
RIAGENDR	gender - adjudicated	0	1.5198725	1.0000000	2.0000000
RIDRETH1	1=mex 2=oth hisp 3=white 4=black 5=other	0	2.8710161	1.0000000	5.0000000
INDFMPIR	family pir	268	2.5923431	0	5.0000000
WTMEC2YR	full sample 2 year mec exam weight	0	40813.74	1363.17	156152.18
SDMVPSU	masked variance pseudo-psu	0	1.5050619	1.0000000	2.0000000
SDMVSTRA	masked variance pseudo-stratum	0	50.5466817	44.0000000	58.0000000
MARCAT	1=married 2=prev married 3=never married	5	1.6595984	1.0000000	3.0000000
BPXDI1_1		753	68.9945427	4.0000000	124.0000000
AGE18P		0	1.0000000	1.0000000	1.0000000
AGEC		0	-0.5079417	-27.5979309	39.4020691
AGECSQ		0	406.0373060	0.1616603	1552.52
BMXBMI	body mass index (kg/m**2)	97	28.5119572	13.3599997	130.2100067

```
* use IVEware %impute to multiply impute missing data for chapter 11 exercises ;
%impute (name=11_21, setup=new, dir=. ) ;
title "Analysis Example for Chapter 11 Imputation: NHANES" ;
datain C11 ;
dataout c11_imp all ;
categorical marcat riagendr ridreth1 ;
transfer sdmvstra sdmvpsu wtmecc2yr age18p agecsq seqn ;
bounds indfmpir (>=0, <=5) bmx bmi (>=13.0, <=130) ;
multiples 5 ;
seed 67 ;
run ;
```

IVEware Setup Checker, Fri Mar 12 13:43:17 2010

1

Setup listing:

```
title "Analysis Example for Chapter 11 Imputation: NHANES" ;
datain C11 ;
dataout c11_imp all ;
categorical marcat riagendr ridreth1 ;
transfer sdmvstra sdmvpsu wtmecc2yr age18p agecsq seqn ;
bounds indfmpir (>=0, <=5) bmx bmi (>=13.0, <=130) ;
multiples 5 ;
seed 67 ;
run ;
```

"Analysis Example for Chapter 11 Imputation: NHANES"

Imputation 1

Variable	Observed	Imputed	Double counted
RIAGENDR	5334	0	0
RIDRETH1	5334	0	0
INDFMPIR	5066	268	0
MARCAT	5329	5	0
BPXDI1_1	4581	753	0
AGEC	5334	0	0
BMXBMI	5237	97	0

Variable	Observed	Imputed	Combined
INDFMPIR			
Number	5066	268	5334
Minimum	0	0.00040815	0
Maximum	5	4.94199	5
Mean	2.59234	2.44446	2.58491
Std Dev	1.60717	1.23838	1.59092

Variable	Observed		Imputed		Combined		
MARCAT	Code	Freq	Per	Freq	Per	Freq	Per
	1	3055	57.33	3	60.00	3058	57.33
	2	1033	19.38	2	40.00	1035	19.40
	3	1241	23.29	0	0.00	1241	23.27
	Total	5329	100.00	5	100.00	5334	100.00

Variable	Observed	Imputed	Combined
BPXDI1_1			
Number	4581	753	5334
Minimum	4	28.5706	4
Maximum	124	109.703	124
Mean	68.9945	68.9528	68.9886
Std Dev	12.8696	13.2026	12.9159

Variable	Observed	Imputed	Combined
BMXBMI			
Number	5237	97	5334
Minimum	13.36	13.9507	13.36
Maximum	130.21	42.7459	130.21
Mean	28.512	29.503	28.53
Std Dev	6.7813	6.3041	6.77367

"Analysis Example for Chapter 11 Imputation: NHANES"

Imputation 2

Variable	Observed	Imputed	Double counted
RIAGENDR	5334	0	0
RIDRETH1	5334	0	0
INDFMPIR	5066	268	0
MARCAT	5329	5	0
BPXDI1_1	4581	753	0
AGEC	5334	0	0
BMXBMI	5237	97	0

Variable	Observed	Imputed	Combined
INDFMPIR			
Number	5066	268	5334
Minimum	0	0.00481076	0
Maximum	5	4.9265	5
Mean	2.59234	2.38196	2.58177
Std Dev	1.60717	1.23641	1.59118

Variable	Observed		Imputed		Combined		
MARCAT	Code	Freq	Per	Freq	Per	Freq	Per
	1	3055	57.33	5	100.00	3060	57.37
	2	1033	19.38	0	0.00	1033	19.37
	3	1241	23.29	0	0.00	1241	23.27
	Total	5329	100.00	5	100.00	5334	100.00

Variable	Observed	Imputed	Combined
BPXDI1_1			
Number	4581	753	5334
Minimum	4	27.2559	4
Maximum	124	114.162	124
Mean	68.9945	69.1957	69.0229
Std Dev	12.8696	13.5095	12.9608

Variable	Observed	Imputed	Combined
BMXBMI			
Number	5237	97	5334
Minimum	13.36	14.6106	13.36
Maximum	130.21	48.3718	130.21
Mean	28.512	29.9377	28.5379
Std Dev	6.7813	6.82258	6.78409

"Analysis Example for Chapter 11 Imputation: NHANES"

Imputation 3

Variable	Observed	Imputed	Double counted
RIAGENDR	5334	0	0
RIDRETH1	5334	0	0
INDFMPIR	5066	268	0
MARCAT	5329	5	0
BPXDI1_1	4581	753	0
AGEC	5334	0	0
BMXBMI	5237	97	0

Variable	Observed	Imputed	Combined
INDFMPIR			
Number	5066	268	5334
Minimum	0	0.0234594	0
Maximum	5	4.96166	5
Mean	2.59234	2.4347	2.58442
Std Dev	1.60717	1.20235	1.58958

Variable	Observed		Imputed		Combined		
MARCAT	Code	Freq	Per	Freq	Per	Freq	Per
	1	3055	57.33	4	80.00	3059	57.35
	2	1033	19.38	1	20.00	1034	19.39
	3	1241	23.29	0	0.00	1241	23.27
	Total	5329	100.00	5	100.00	5334	100.00

Variable	Observed	Imputed	Combined
BPXDI1_1			
Number	4581	753	5334
Minimum	4	30.7565	4
Maximum	124	110.225	124
Mean	68.9945	68.9007	68.9813
Std Dev	12.8696	12.8812	12.8701

Variable	Observed	Imputed	Combined
BMXBMI			
Number	5237	97	5334
Minimum	13.36	13.2616	13.2616
Maximum	130.21	48.8765	130.21
Mean	28.512	27.6117	28.4956
Std Dev	6.7813	7.0303	6.7863

"Analysis Example for Chapter 11 Imputation: NHANES"

Imputation 4

Variable	Observed	Imputed	Double counted
RIAGENDR	5334	0	0
RIDRETH1	5334	0	0
INDFMPIR	5066	268	0
MARCAT	5329	5	0
BPXDI1_1	4581	753	0
AGEC	5334	0	0
BMXBMI	5237	97	0

Variable	Observed	Imputed	Combined
INDFMPIR			
Number	5066	268	5334
Minimum	0	0.0619482	0
Maximum	5	4.99557	5
Mean	2.59234	2.32427	2.57887
Std Dev	1.60717	1.21912	1.59092

Variable	Observed		Imputed		Combined		
MARCAT	Code	Freq	Per	Freq	Per	Freq	Per
	1	3055	57.33	3	60.00	3058	57.33
	2	1033	19.38	2	40.00	1035	19.40
	3	1241	23.29	0	0.00	1241	23.27
	Total	5329	100.00	5	100.00	5334	100.00

Variable	Observed	Imputed	Combined
BPXDI1_1			
Number	4581	753	5334
Minimum	4	23.6411	4
Maximum	124	114.412	124
Mean	68.9945	69.4968	69.0654
Std Dev	12.8696	13.3281	12.9352

Variable	Observed	Imputed	Combined
BMXBMI			
Number	5237	97	5334
Minimum	13.36	14.8793	13.36
Maximum	130.21	44.2298	130.21
Mean	28.512	28.5294	28.5123
Std Dev	6.7813	6.69916	6.7792

"Analysis Example for Chapter 11 Imputation: NHANES"

Imputation 5

Variable	Observed	Imputed	Double counted
RIAGENDR	5334	0	0
RIDRETH1	5334	0	0
INDFMPIR	5066	268	0
MARCAT	5329	5	0
BPXDI1_1	4581	753	0
AGEC	5334	0	0
BMXBMI	5237	97	0

Variable	Observed	Imputed	Combined
INDFMPIR			
Number	5066	268	5334
Minimum	0	0.0136309	0
Maximum	5	4.99698	5
Mean	2.59234	2.49119	2.58726
Std Dev	1.60717	1.20282	1.58938

Variable	Observed		Imputed		Combined		
MARCAT	Code	Freq	Per	Freq	Per	Freq	Per
	1	3055	57.33	5	100.00	3060	57.37
	2	1033	19.38	0	0.00	1033	19.37
	3	1241	23.29	0	0.00	1241	23.27
	Total	5329	100.00	5	100.00	5334	100.00

Variable	Observed	Imputed	Combined
BPXDI1_1			
Number	4581	753	5334
Minimum	4	27.246	4
Maximum	124	105.761	124
Mean	68.9945	68.9914	68.9941
Std Dev	12.8696	13.4443	12.951

Variable	Observed	Imputed	Combined
BMXBMI			
Number	5237	97	5334
Minimum	13.36	15.0277	13.36
Maximum	130.21	42.6221	130.21
Mean	28.512	28.3038	28.5082
Std Dev	6.7813	6.16486	6.77012


```
%describe (name=ex11_1, setup=new, dir=. ) ;
title "Analysis Example 11.1 Means Analysis of Imputed Data Sets Diastolic BP: NHANES " ;
datain c11_imp ;
stratum sdmvstra ;
cluster sdmvpsu ;
weight wtmec2yr ;
mean bpxdi1_1 ;
run ;
```

IVEware Setup Checker, Fri Mar 12 13:44:29 2010

1

Setup listing:

```
title "Analysis Example 11.1 Means Analysis of Imputed Data Sets Diastolic BP:
NHANES " ;
datain c11_imp ;
stratum sdmvstra ;
cluster sdmvpsu ;
weight wtmec2yr ;
mean bpxdi1_1 ;
run ;
```

"Analysis Example 11.1 Means Analysis of Imputed Data Sets Diastolic BP: NHANES "

Stratum variable: SDMVSTRA masked variance pseudo-stratum
Cluster variable: SDMVPSU masked variance pseudo-psu
Weight variable: WTMEC2YR full sample 2 year mec exam weight

Analysis description:

4 Variables
15 Strata
30 Secus

Strata Model

15 Multiple PSU
0 Paired Selection
0 Successive Differences

26670 Cases Read

"Analysis Example 11.1 Means Analysis of Imputed Data Sets Diastolic BP: NHANES "

Problem 1

Degrees of freedom
15

Factor Covariance of denominator
None 0.06040

Mean	Number of	Sum of	Weighted	Standard
BPXDI1_1	Cases	Weights	Mean	Error
	26670	1.088502e+009	70.43194	0.2928074
	Lower	Upper	T Test	Prob > T
	Bound	Bound		
	69.80783	71.05604	240.54016	0.00000
	Unweighted	Bias	Design	
	Mean		Effect	
	69.01048	-2.01820	14.72664	

```

data c11_imp ;
set c11_imp ;
*CREATE INDICATOR VARIABLES FOR MODEL ;
mex=(ridreth1=1) ;
othhis=(ridreth1=2) ;
white=(ridreth1=3) ;
black=(ridreth1=4) ;
other=(ridreth1=5) ;

prevmar=(marcat=2) ;
nevmar=(marcat=3) ;

female=(riagendr=2) ;

%regress (name=ex11_1, setup=new, dir=. ) ;
title "Analysis Example 11.1 Regression Analysis of Imputed Data Sets : NHANES " ;
datain c11_imp ;
stratum sdmvstra ;
cluster sdmvpsu ;
weight wtmecl2yr ;
predictor othhis white black other female prevmar nevmar agec agecsq bmx bmi indfmpir ;
dependent bpxdi1_1 ;
run ;

```

IVAware Setup Checker, Fri Mar 12 13:45:44 2010

1

Setup listing:

```

title "Analysis Example 11.1 Regression Analysis of Imputed Data Sets : NHANES "
;
datain c11_imp ;
stratum sdmvstra ;
cluster sdmvpsu ;
weight wtmecl2yr ;
predictor othhis white black other female prevmar nevmar agec agecsq bmx bmi
indfmpir ;
dependent bpxdi1_1 ;
run ;

```

"Analysis Example 11.1 Regression Analysis of Imputed Data Sets: NHANES "

Regression type: Linear
 Dependent variable: BPXDI1_1
 Predictors: othhis
 white
 black
 other
 female
 prevmar
 nevmar
 AGEC
 AGECSQ
 BMXBMI body mass index (kg/m**2)
 INDFMPIR family pir
 Stratum variable: SDMVSTRA masked variance pseudo-stratum
 Cluster variable: SDMVPSU masked variance pseudo-psu
 Weight variable: WTMEC2YR full sample 2 year mec exam weight

Valid cases 26670
 Sum weights 1088502356
 Replicates 15
 Degr freedom 15

Sum of squares:
 Model 2.081852077e+010
 Error 1.481853413e+011
 Total 1.690038621e+011
 R-square 0.12318
 F-value 0.17561
 P-value 0.99784

Variable	Estimate	Std Error	T Test	Prob > T
Intercept	67.1927807	1.0050059	66.85809	0.00000
othhis	1.4182722	1.0650997	1.33159	0.20288
white	2.1653758	0.5706060	3.79487	0.00176
black	3.5424413	0.6732177	5.26196	0.00010
other	1.5387342	0.7448995	2.06569	0.05658
female	-2.8675012	0.3337574	-8.59157	0.00000
prevmar	0.8052565	0.6022942	1.33698	0.20115
nevmar	-1.1132055	0.6115019	-1.82044	0.08870
AGEC	0.1001993	0.0136785	7.32533	0.00000
AGECSQ	-0.0100079	0.0007710	-12.98066	0.00000
BMXBMI	0.2020072	0.0285515	7.07518	0.00000
INDFMPIR	0.0042498	0.1128745	0.03765	0.97046

Variable	Estimate	95% Confidence Interval	
		Lower	Upper
Intercept	67.1927807	65.0506669	69.3348944
othhis	1.4182722	-0.8519281	3.6884725
white	2.1653758	0.9491611	3.3815905
black	3.5424413	2.1075156	4.9773671

"Analysis Example 11.1 Regression Analysis of Imputed Data Sets : NHANES "

Variable	Estimate	95% Confidence Interval	
		Lower	Upper
other	1.5387342	-0.0489773	3.1264456
female	-2.8675012	-3.5788865	-2.1561159
prevmar	0.8052565	-0.4784998	2.0890129
nevmar	-1.1132055	-2.4165876	0.1901766
AGEC	0.1001993	0.0710444	0.1293542
AGECSQ	-0.0100079	-0.0116512	-0.0083646
BMXBMI	0.2020072	0.1411513	0.2628632
INDFMPIR	0.0042498	-0.2363359	0.2448355

Variable	Design Effect	SRS	% Diff
		Estimate	SRS v Est
Intercept	6.09904	68.1070100	1.36061
othhis	5.64080	1.6946150	19.48447
white	7.67251	1.9254734	-11.07902
black	9.03091	3.7137882	4.83697
other	3.50998	2.0444889	32.86823
female	4.99210	-3.3383885	16.42152
prevmar	8.29364	1.2606433	56.55176
nevmar	7.81081	-0.4762582	-57.21740
AGEC	7.53182	0.1240846	23.83778
AGECSQ	11.94389	-0.0096245	-3.83154
BMXBMI	6.55060	0.1600714	-20.75959
INDFMPIR	4.98973	-0.0149131	-450.91452

```

* complete case analysis for comparison to imputed regression;
data c11_r ;
set c11 ;
where age18p=1 and wtmec2yr > 0 ;
mex=(ridreth1=1) ;
othhis=(ridreth1=2) ;
white=(ridreth1=3) ;
black=(ridreth1=4) ;
other=(ridreth1=5) ;

married=. ; prevmar=. ; nevmar=. ;
if marcat=1 then married=1 ; else if marcat ne 1 and marcat ne . then married=0 ;
if marcat=2 then prevmar=1 ; else if marcat ne 2 and marcat ne . then prevmar=0 ;
if marcat=3 then nevmar=1 ; else if marcat ne 3 and marcat ne . then nevmar=0 ;

female=(riagendr=2) ;

proc means n nmiss mean ;
run ;

%regress (name=ex11_2, setup=new, dir=. ) ;
title "Analysis Example 11.1 Regression Analysis of Imputed Data Sets : NHANES " ;
datain c11_r ;
stratum sdmvstra ;
cluster sdmvpsu ;
weight wtmec2yr ;
predictor othhis white black other female prevmar nevmar agec agecsq bmx bmi indfmpir ;
dependent bpxdi1_1 ;
run ;

```

The SAS System

The MEANS Procedure

Variable	Label	N		Mean
		N	Miss	
SEQN	respondent sequence number	5334	0	36358.80
RIAGENDR	gender - adjudicated	5334	0	1.5198725
RIDRETH1	1=mex 2=oth hisp 3=white 4=black 5=other	5334	0	2.8710161
INDFMPIR	family pir	5066	268	2.5923431
WTMEC2YR	full sample 2 year mec exam weight	5334	0	40813.74
SDMVPSU	masked variance pseudo-psu	5334	0	1.5050619
SDMVSTRA	masked variance pseudo-stratum	5334	0	50.5466817
MARCAT	1=married 2=prev married 3=never married	5329	5	1.6595984
BPXDI1_1		4581	753	68.9945427
AGE18P		5334	0	1.0000000
AGEC		5334	0	-0.5079417
AGECSQ		5334	0	406.0373060
BMXBMI	body mass index (kg/m**2)	5237	97	28.5119572
mex		5334	0	0.2124109
othhis		5334	0	0.0307462
white		5334	0	0.4716910
black		5334	0	0.2437195
other		5334	0	0.0414323
married		5329	5	0.5732783
prevmar		5329	5	0.1938450
nevmar		5329	5	0.2328767
female		5334	0	0.5198725

Setup listing:

```
title "Analysis Example 11.1 Regression Analysis of Imputed Data Sets : NHANES "  
;  
datain c11_r ;  
stratum sdmvstra ;  
cluster sdmvpsu ;  
weight wtmecl2yr ;  
predictor othhis white black other female prevmar nevmar agec agecsq bmxbmi  
indfmpir ;  
dependent bpxdi1_1 ;  
run ;
```


"Analysis Example 11.1 Regression Analysis of Imputed Data Sets : NHANES "

Regression type: Linear
 Dependent variable: BPXDI1_1
 Predictors: othhis
 white
 black
 other
 female
 prevmar
 nevmar
 AGEC
 AGECSQ
 BMXBMI body mass index (kg/m**2)
 INDFMPIR family pir
 Stratum variable: SDMVSTRA masked variance pseudo-stratum
 Cluster variable: SDMVPSU masked variance pseudo-psu
 Weight variable: WTMEC2YR full sample 2 year mec exam weight

Valid cases 4308
 Sum weights 181480255.5
 Replicates 15
 Degr freedom 15

Sum of squares:
 Model 4077289536
 Error 2.326985597e+010
 Total 2.734714551e+010
 R-square 0.14909
 F-value 0.21902
 P-value 0.99422

Variable	Estimate	Std Error	T Test	Prob > T
Intercept	68.2871395	1.4304690	47.73759	0.00000
othhis	1.6527524	1.1891522	1.38986	0.18486
white	2.1453575	0.6725533	3.18987	0.00609
black	3.3105810	0.8854629	3.73881	0.00198
other	1.6638912	0.9250625	1.79868	0.09222
female	-2.6992622	0.3941630	-6.84809	0.00001
prevmar	1.0339497	0.6909995	1.49631	0.15532
nevmar	-0.3729060	0.5540424	-0.67306	0.51114
AGEC	0.1193541	0.0154403	7.73003	0.00000
AGECSQ	-0.0123063	0.0008644	-14.23624	0.00000
BMXBMI	0.1973282	0.0380274	5.18910	0.00011
INDFMPIR	-0.1076662	0.1369220	-0.78633	0.44392

Variable	Estimate	95% Confidence Interval	
		Lower	Upper
Intercept	68.2871395	65.2381751	71.3361040
othhis	1.6527524	-0.8818588	4.1873635
white	2.1453575	0.7118479	3.5788671
black	3.3105810	1.4232664	5.1978955

"Analysis Example 11.1 Regression Analysis of Imputed Data Sets : NHANES "

Variable	Estimate	95% Confidence Interval	
		Lower	Upper
other	1.6638912	-0.3078278	3.6356101
female	-2.6992622	-3.5393986	-1.8591258
prevmar	1.0339497	-0.4388769	2.5067763
nevmar	-0.3729060	-1.5538163	0.8080043
AGEC	0.1193541	0.0864439	0.1522642
AGECSQ	-0.0123063	-0.0141488	-0.0104638
BMXBMI	0.1973282	0.1162749	0.2783815
INDFMPIR	-0.1076662	-0.3995076	0.1841753

Variable	Design Effect	SRS	% Diff
		Estimate	SRS v Est
Intercept	2.02960	69.5859438	1.90197
othhis	1.19478	1.6219303	-1.86490
white	1.78792	1.9335364	-9.87347
black	2.55682	3.4338559	3.72367
other	0.90874	2.0327620	22.16917
female	1.18260	-3.1977188	18.46640
prevmar	1.81874	1.3875763	34.20153
nevmar	1.10300	0.2365145	-163.42470
AGEC	1.61578	0.1434284	20.17052
AGECSQ	2.40534	-0.0119425	-2.95614
BMXBMI	1.88126	0.1407308	-28.68184
INDFMPIR	1.26008	-0.1079249	0.24036