Measurement, Design, and Analytic Techniques in Mental Health and Behavioral Sciences

Lecture 8 (Feb 6, 2007): SAS Proc MI and Proc MiAnalyze

XH Andrew Zhou

azhou@u.washington.edu

Professor, Department of Biostatistics, University of Washington

SAS combining procedures of MI

- The MIanalyze procedure in SAS combines the results of analyses of imputations and generate statistical inferences.
- Syntax

```
PROC MIANALYZE <options>;
 BY variables;
 CLASS variables;
 MODELEFFECTS effects;
 TEST;
 STDERR;
```

Summary of Proc MiAnalyze Options

Specific Input data sets	Options
COV, CORR, or EST type data set	DATA=
parameter estimates and standard errors	DATA=
parameter estimates	PARAMS=
parameter information	PARMINFO=
covariance matrices	COVB=
(XX')-1	XPXI=

Specify statistical analysis	
parameters under the null hypothesis	THETA0=
level for the confidence interval	ALPHA

Descriptions of other commands

- The statement MODELEFFECTS lists the effects to be analyzed. Variables in this statement that are not specified in a CLASS statement are assumed to be continuous.
- The statement TEST can test the hypothesis about linear combinations of parameters. An F statistics is used to test jointly the null hypothesis ($H_0 : L\beta = 0$) in a single TEST statement.

Example 1, using regression analysis with class

- Combine results from a regression model with continuous covariates

```
proc mi data=MonotoneData noprint out=outmi seed=501213;
 class female;
   monotone reg (mh1 mh2 mh3 mh4/details);
   var  female age mh1 mh2 mh3 mh4 ;
   run;
   proc reg data=outmi outest=outreg covout noprint;
   model mh4= age mh1 mh2 mh3;
   by _imputation_;
   run;
   proc mianalyze data=outreg;
   modeleffects Intercept age mh1 mh2 mh3;
   test mh1=mh2=mh3;
   run;
```

Results

Multiple Imputation Parameter Estimates

Parameter	Estimate	Std Error	95% Confidence Limits		DF
Intercept	4.782181	0.514103	3.77355	5.790814	1210.3
age	-0.012707	0.007356	-0.02716	0.001750	432.31
mh1	0.098358	0.044554	0.01092	0.185799	897.68
mh2	0.242225	0.043256	0.15703	0.327418	249.34
mh3	0.339826	0.037684	0.26585	0.413800	788.98

Results, cont

Multiple Imputation Parameter Estimates

Parameter	Minimum	Maximum	Theta0	t for H0: Parameter=Theta0	Pr > |t|
Intercept	4.669552	4.937240	0	9.30	<.0001
age	-0.015818	-0.010066	0	-1.73	0.0848
mh1	0.086720	0.109104	0	2.21	0.0275
mh2	0.220644	0.258681	0	5.60	<.0001
mh3	0.325534	0.348182	0	9.02	<.0001

Results, cont

The MIANALYZE Procedure

Test: Test 1

Test Specification

	L Matrix					
Parameter	Intercept	age	mh1	mh2	mh3	C
TestPrm1	0	0	1.000000	-1.000000	0	0
TestPrm2	0	0	0	1.000000	-1.000000	0

Results, cont

Multiple Imputation Variance Information

Parameter	Between (Variance)	Within (Variance)	Total (Variance)	DF
TestPrm1	0.000473	0.004561	0.005128	326.88
TestPrm2	0.000505	0.004211	0.004817	252.71

Multiple Imputation Variance Information

Parameter	Relative Increase in Variance	Fraction Missing Information	Relative Efficiency
TestPrm1	0.124380	0.116014	0.977323
TestPrm2	0.143919	0.132650	0.974156

SAS Output

Multiple Imputation Parameter Estimates

Parameter	Estimate	Std Error	95% Confidence Limits	
TestPrm1	-0.143867	0.071611	-0.28474	-0.00299
TestPrm2	-0.097602	0.069401	-0.23428	0.03908

Multiple Imputation Parameter Estimates

Parameter	Minimum	Maximum	C	t for H0: Parameter=C	Pr > |t|
TestPrm1	-0.171960	-0.111540	0	-2.01	0.0454
TestPrm2	-0.127537	-0.066854	0	-1.41	0.1609

Mixed-effect regression analysis with class

- Use mixed-effect regression with discrete covariates
- ```
  proc mi data=MonotoneData noprint out=outmi seed=501213;
   class male;
     monotone reg (mh1 mh2 mh3 mh4/details);
     var  male age mh1 mh2 mh3 mh4 ;
     run;
  proc mixed data=outmi;
   class male;
   model  mh4=male age mh1 mh2 mh3 /solution covb;
    by _imputation_;
     ods output solutionf=mxparms CovB=mxcovb;
      run;
  proc mianalyze parms=mxparms;
   class male;
    modeleffects Intercept male age mh1 mh2 mh3;
    run;
  ```
  ```

Results

Multiple Imputation Parameter Estimates

Parameter	male	Estimate	Std Error	95% Confidence Limits		DF
Intercept	.	4.541591	0.548015	3.46576	5.617417	749.76
male	0	0.307126	0.224783	-0.13375	0.747996	1760
male	1.000000	0	.	.	.	.
age	.	-0.012175	0.007360	-0.02664	0.002292	436.68
mh1	.	0.096506	0.044519	0.00914	0.183875	933.56
mh2	.	0.242102	0.043226	0.15697	0.327234	250.3
mh3	.	0.341579	0.037700	0.26757	0.415585	774.07

Results

The MIANALYZE Procedure

Multiple Imputation Parameter Estimates

Parameter	male	Minimum	Maximum
Intercept	.	4.394153	4.691847
male	0	0.232033	0.351137
male	1.000000	0	0
age	.	-0.015273	-0.009524
mh1	.	0.084952	0.107302
mh2	.	0.220599	0.258667
mh3	.	0.326980	0.349701

Results, cont

Multiple Imputation Parameter Estimates

Parameter	male	Theta0	t for H0: Parameter=Theta0	Pr > |t|
Intercept	.	0	8.29	<.0001
male	0	0	1.37	0.1720
male	1.000000	0	.	.
age	.	0	-1.65	0.0988
mh1	.	0	2.17	0.0304
mh2	.	0	5.60	<.0001
mh3	.	0	9.06	<.0001

A generalized linear model (genmod)

- This example illustrates the use of a generalized linear model (normal error and identify link function) to analyze imputed data sets and save parameter estimates and corresponding covariate matrices and then combine them to generate statistical inferences.
- $E(Y) = X'\beta$ and $var(y) = \sigma^2$, where σ is called the scale parameter.
- ```
  proc genmod data=outmi;
      model mh4= age  mh1 mh2 mh3/covb;
      by _Imputation_;
      ods output ParameterEstimates=gmparms  CovB=gmcovb;
   run;
  proc mianalyze parms=gmparms;
   modeleffects   Intercept  age mh1 mh2 mh3;
   run;
  ```
  ```

Results

Multiple Imputation Parameter Estimates

Parameter	Estimate	Std Error	95% Confidence Limits		DF
Intercept	4.782181	0.512725	3.77624	5.788121	1197.4
age	-0.012707	0.007337	-0.02713	0.001713	427.88
mh1	0.098358	0.044435	0.01115	0.185569	888.19
mh2	0.242225	0.043149	0.15724	0.327211	246.87
mh3	0.339826	0.037585	0.26605	0.413605	780.68

Results, cont

Parameter	Minimum	Maximum	Theta0	t for H0: Parameter=Theta0	Pr > |t|
Intercept	4.669552	4.937240	0	9.33	<.0001
age	-0.015818	-0.010066	0	-1.73	0.0840
mh1	0.086720	0.109104	0	2.21	0.0271
mh2	0.220644	0.258681	0	5.61	<.0001
mh3	0.325534	0.348182	0	9.04	<.0001

MI with a general linear model (GLM) model

- This example illustrates the use of a generalized linear model to analyze imputed data sets and save parameter estimates and corresponding covariate matrices and then combine them to generate statistical inferences.
- ```
  proc glm data=outmi;
        model  mh4=age mh1 mh2 mh3/inverse;
        by _Imputation_;
        ods output ParameterEstimates=glmparms
                   InvXPX=glmxpxi;
     run;
  proc mianalyze parms=glmparms;
   modeleffects Intercept  age mh1 mh2 mh3;
   run;
  ```
  ```

output

Multiple Imputation Parameter Estimates

Parameter	Estimate	Std Error	95% Confidence Limits		DF
Intercept	4.782181	0.514103	3.77355	5.790814	1210.3
age	-0.012707	0.007356	-0.02716	0.001750	432.31
mh1	0.098358	0.044554	0.01092	0.185799	897.68
mh2	0.242225	0.043256	0.15703	0.327418	249.34
mh3	0.339826	0.037684	0.26585	0.413800	788.98

Output, cont

Parameter	Minimum	Maximum	Theta0	t for H0: Parameter=Theta0	Pr > |t
Intercept	4.669552	4.937240	0	9.30	<.0001
age	-0.015818	-0.010066	0	-1.73	0.0848
mh1	0.086720	0.109104	0	2.21	0.0275
mh2	0.220644	0.258681	0	5.60	<.0001

An example using a logistic regression

- This example illustrates the use of a logistic regression model to analyze imputed data sets and save parameter estimates and corresponding covariate matrices and then combine them to generate statistical inferences.

SAS Code

```
data exam3;
 set example.education6;
  run;
  proc mi data=exam3
  out=outmi seed=501213;
  class  npcerad ;
   monotone discrim  (npcerad=mmselast npgender educ npdage/details);
   var mmselast npgender educ npdage npcerad ;
 run;
data outmi2;
 set outmi;
 if mmselast  <=24 then cogimpair=1;
 else cogimpair=0;
 run; proc logistic data=outmi2;
 model cogimpair= npgender educ npdage npcerad/covb;
 by _imputation_;
 ods output ParameterEstimates=lgsparms Covb=lgcovb;
 run;  proc mianalyze parms=lgsparms covb=lgcovb;
  modeleffects  npgender educ npdage npcerad;
   run;
```

SAS Output

```
Model Information

PARMS Data Set              WORK.LGSPARMS
COVB Data Set               WORK.LGCOVB
Number of Imputations       5


Multiple Imputation Variance Information

 -----------------Variance-----------------
 Parameter          Between        Within         Total        DF

 npgender          0.009341      0.028195      0.039404    49.431
 educ              0.000126      0.000651      0.000802    112.06
 npdage         0.000010892      0.000111      0.000124    359.96
 npcerad           0.001183      0.004365      0.005784    66.461
```

SAS Output, cont

Multiple Imputation Variance Information

Parameter	Relative Increase in Variance	Fraction Missing Information	Relative Efficiency
npgender	0.397558	0.311760	0.941308
educ	0.232945	0.203033	0.960978
npdage	0.117837	0.110344	0.978408
npcerad	0.325080	0.267058	0.949297

Output, cont

Multiple Imputation Parameter Estimates

Parameter	Estimate	Std Error	95\% Confidence Limits		DF
npgender	-0.393643	0.198504	-0.79246	0.005178	49.431
educ	0.132880	0.028328	0.07675	0.189009	112.06
npdage	0.049236	0.011135	0.02734	0.071133	359.96
npcerad	1.280466	0.076055	1.12864	1.432296	66.461

Output, cont

Multiple Imputation Parameter Estimates

Parameter	Minimum	Maximum	Theta0	t for H0: Parameter=Theta0	Pr > |t
npgender	-0.474383	-0.227524	0	-1.98	0.0529
educ	0.119974	0.150788	0	4.69	<.0001
npdage	0.046631	0.054650	0	4.42	<.0001
npcerad	1.244445	1.326876	0	16.84	<.0001

Example on combining correlation coefficients

- Fisher's z transformation of the sample correlation r is

$$z = \frac{1}{2}\log(\frac{1+r}{1-r}).$$

The statistic z is approximately normal with mean

$$\log(\frac{1+\rho}{1-\rho})$$

and variance $1/(n-3)$. Here ρ is the population correlation, and n is the sample size.

SAS Code

```
proc corr  data=outmi fisher (biasadj=no);
   var mh2 mh3;
   by _imputation_;
   ods output FisherPearsonCorr = outz;
   run;

   data outz;
   set outz;
   StdZ=1./sqrt(Nobs-3);
   run;

   proc mianalyze data=outz;
    ods output ParameterEstimates=parms;
    modeleffects Zval;
    stderr stdZ;
    run;
```