SPSS MULTIPLE IMPUTATION

IMPUTATION ALGORITHM

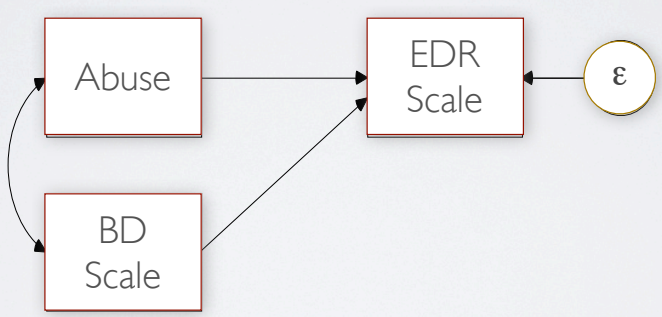
- The SPSS uses an MCMC algorithm known as fully conditional specification (FCS) or chained equations imputation
- The basic idea is to impute incomplete variables one at a time, using the filled-in variable from one step as a predictor in all subsequent steps
- SPSS uses linear regression for continuous variables and logistic regression for categorical variables

EATING DISORDER RISK DATA

- Questionnaire data from a study of eating disorder risk in a sample of 500 college-aged women
- Variables:
- Body mass index (BMI), 7 questionnaire items measuring body dissatisfaction, 6 questionnaire items measuring eating disorder risk, binary indicator of past sexual abuse history (0 = no abuse history, 1 = abuse history)
- All questionnaire items measured on a 7-point Likert scale

ANALYSIS MODEL

- Multiple regression model that predicts eating disorder risk (sum of 6 items) from abuse history and body dissatisfaction (sum of 7 items)



$$\text{edrscale} = B_0 + B_1(\text{abuse}) + B_2(\text{bdscale}) + \varepsilon$$

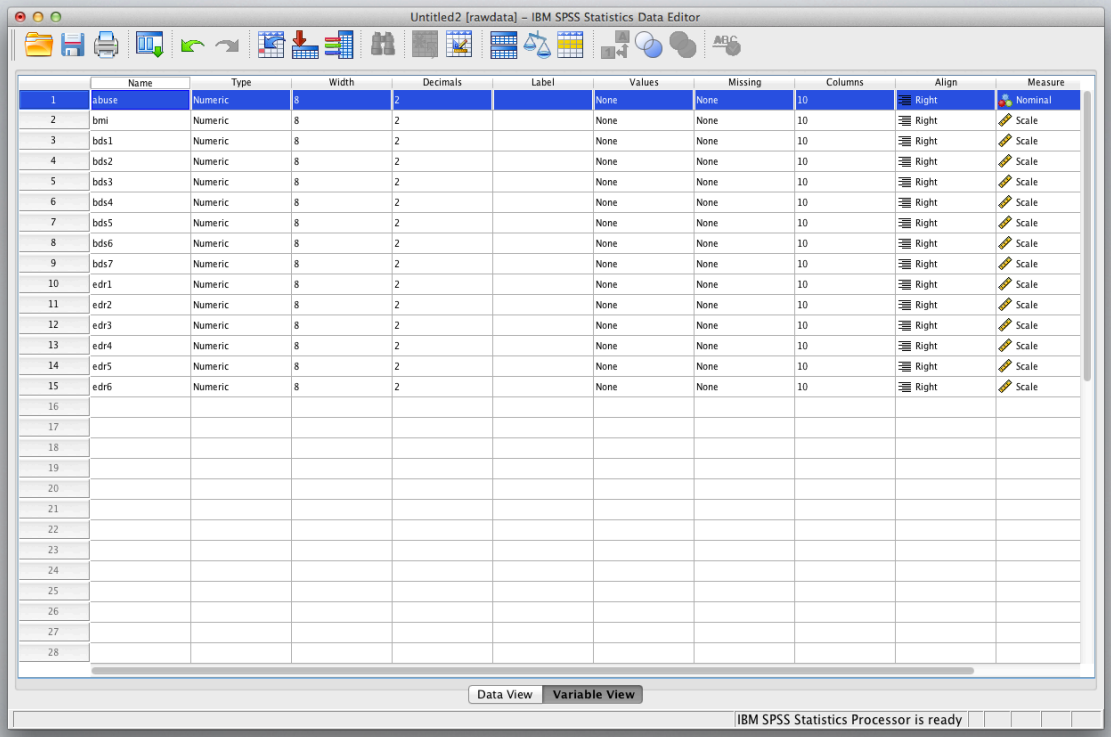
DEFINING VARIABLES

- Incomplete variables must be defined as nominal or scale (i.e., continuous) prior to imputation
- SPSS applies linear imputation to scale variables and logistic (or multinomial logistic) regression to categorical variables
- Define variables in the Variable View tab or with syntax

CATEGORICAL VARIABLES

- A multinomial logistic regression model for a Likert outcome has many parameters
- Imputation can be exceedingly slow (the ordinal imputation model in Mplus is much faster)
- Treating ordinal scales as continuous is often fine, but rounding imputed values to the nearest integer can introduce bias

VARIABLE VIEW TAB

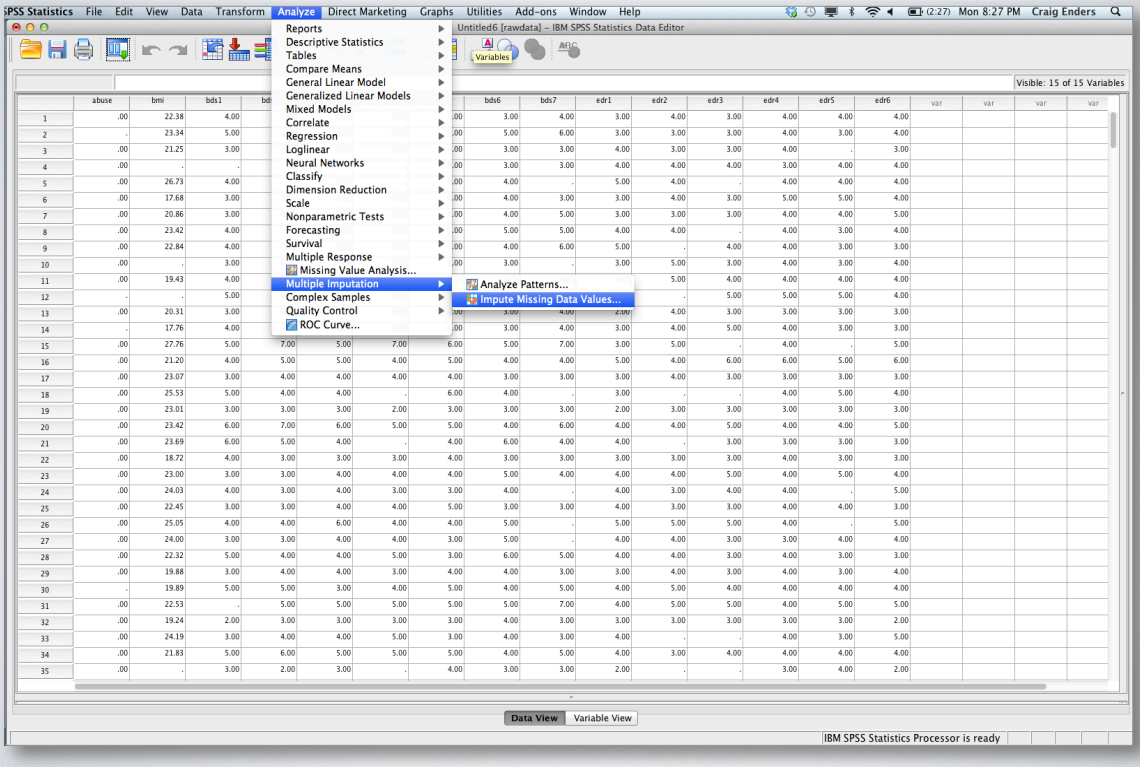


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	abuse	Numeric	8	2		None	None	10	Right	Nominal
2	bmi	Numeric	8	2		None	None	10	Right	Scale
3	bds1	Numeric	8	2		None	None	10	Right	Scale
4	bds2	Numeric	8	2		None	None	10	Right	Scale
5	bds3	Numeric	8	2		None	None	10	Right	Scale
6	bds4	Numeric	8	2		None	None	10	Right	Scale
7	bds5	Numeric	8	2		None	None	10	Right	Scale
8	bds6	Numeric	8	2		None	None	10	Right	Scale
9	bds7	Numeric	8	2		None	None	10	Right	Scale
10	edr1	Numeric	8	2		None	None	10	Right	Scale
11	edr2	Numeric	8	2		None	None	10	Right	Scale
12	edr3	Numeric	8	2		None	None	10	Right	Scale
13	edr4	Numeric	8	2		None	None	10	Right	Scale
14	edr5	Numeric	8	2		None	None	10	Right	Scale
15	edr6	Numeric	8	2		None	None	10	Right	Scale

EXPLORATORY ANALYSIS

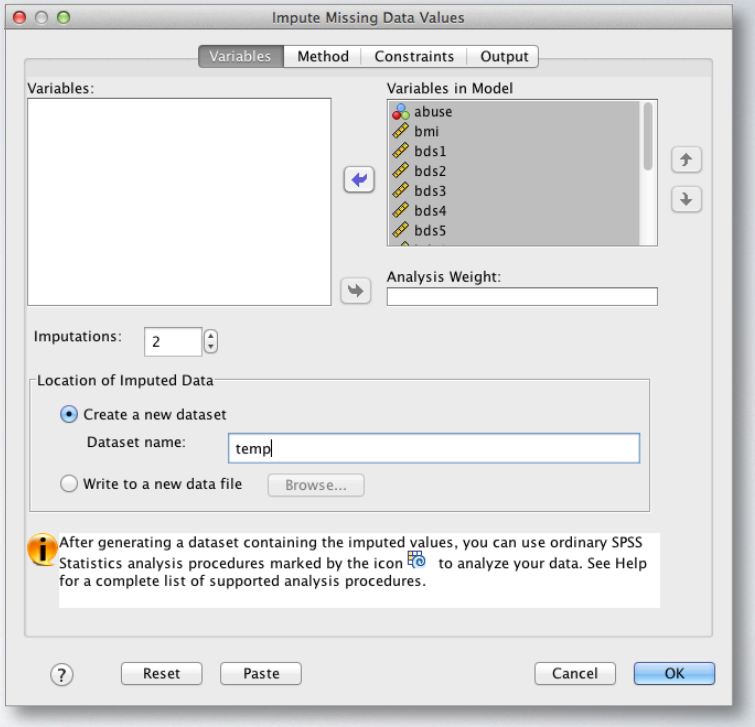
- Prior to imputing the data, run an exploratory analysis to assess MCMC convergence
- SPSS provides limited diagnostic information
- To implement my diagnostic macro program (described later), specify imputations = 2 and iterations = 1000

MULTIPLE IMPUTATION COMMAND



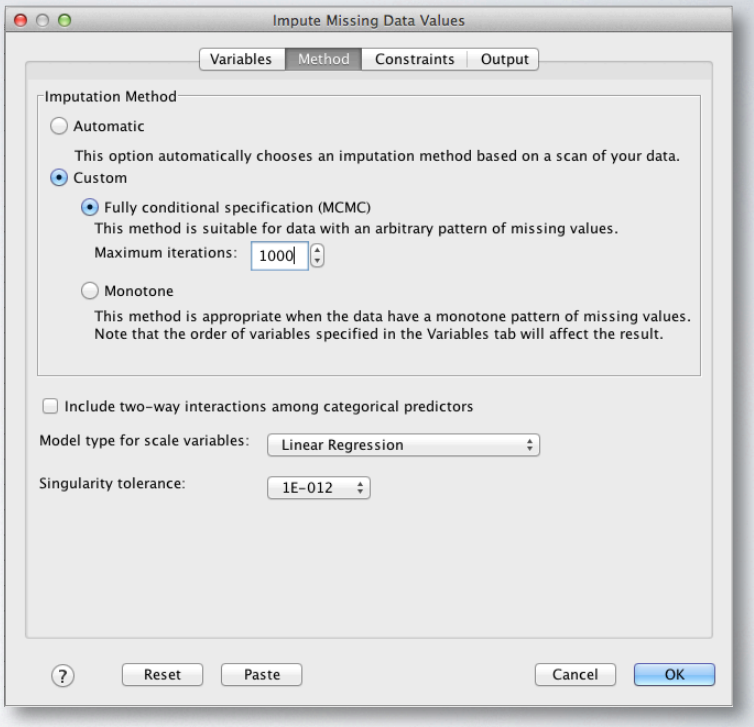
VARIABLES TAB

- Select variables, specify the number of imputed data sets, and specify a file name for the imputed data
- To use the diagnostic macro, specify imputations = 2



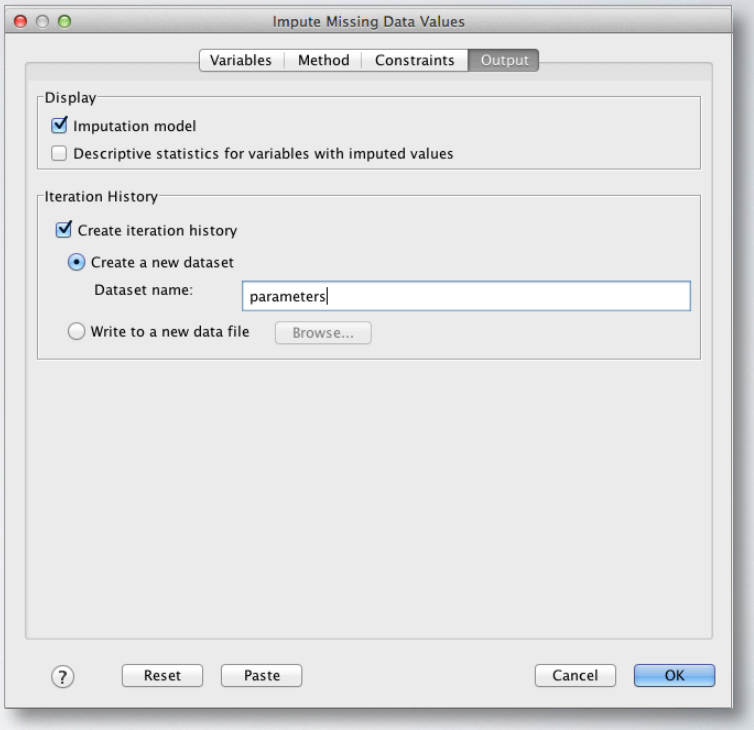
METHOD TAB

- Select the algorithm (fully conditional specification), and specify the number of between-imputation iterations
- To use the diagnostic macro, specify iterations = 1000



OUTPUT TAB

- Checking the Create iteration history button saves P-step means and standard deviations
- To use the diagnostic macro, save the parameters to a file named parameters.sav



SPSS SYNTAX

```
*/ OPEN DATA */.

get file = 'c:\spss ex\eating risk data.sav'.
dataset name rawdata window = front.

*/ DEFINE MEASUREMENT SCALE */.

variable level bmi bds1 to bds7 edr1 to edr6 (scale)
 /abuse (nominal).

*/ PERFORM EXPLORATORY ANALYSIS */.

dataset activate rawdata.

multiple imputation bmi bds1 to bds7 edr1 to edr6 abuse
 /impute method = fcs maxiter = 1000 nimputations = 2
 /outfile imputations = temp fcsiterations = 'c:\spss ex\parameters.sav'.
```

DIAGNOSTIC INFORMATION (OR LACK THEREOF)

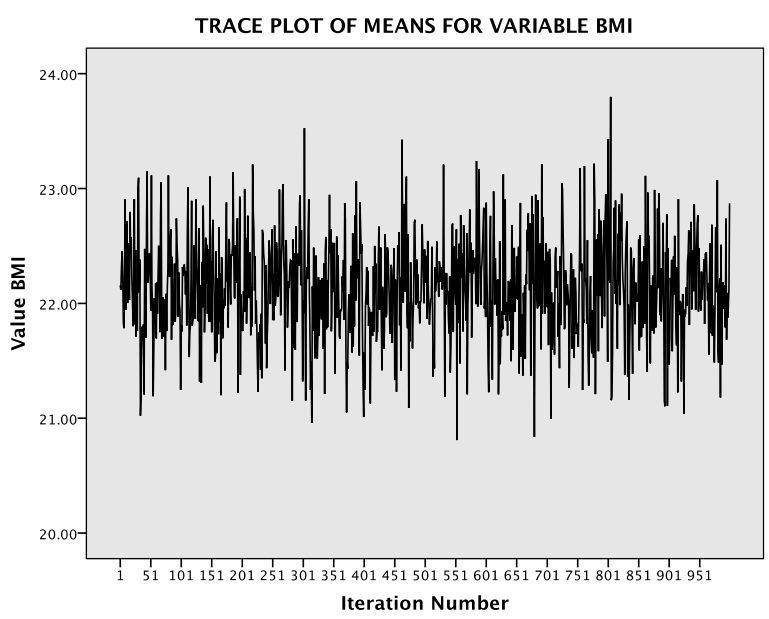
- SPSS provides limited diagnostic information
- The program saves P-step means and standard deviations to a file but does not produce diagnostic information
- I wrote an SPSS macro that creates trace plots and computes the potential scale reduction factor

DIAGNOSTIC MACRO

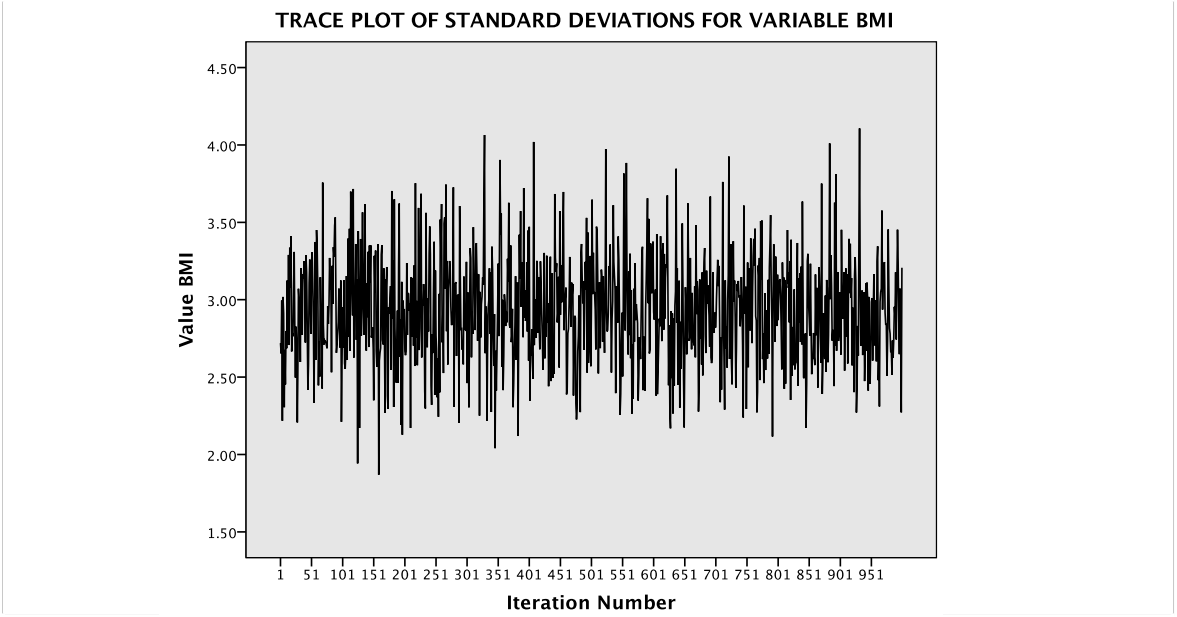
- !let !folder specifies the folder that contains parameters.sav
- !let !vars specifies the variables in parameters.sav

```
define diagnosticmacro ()

*/ !FOLDER = FILE PATH TO THE FOLDER CONTAINING THE PARAMETERS.SAV FILE */.
*/ !VARS = VARIABLES IN THE PARAMETERS.SAV FILE */.

!let !folder =  'c:\spss ex'
!let !vars = 'bmi bds1 bds2 bds4 bds7 edr2 edr3 edr5'
```

TRACE PLOT OF THE BMI MEAN



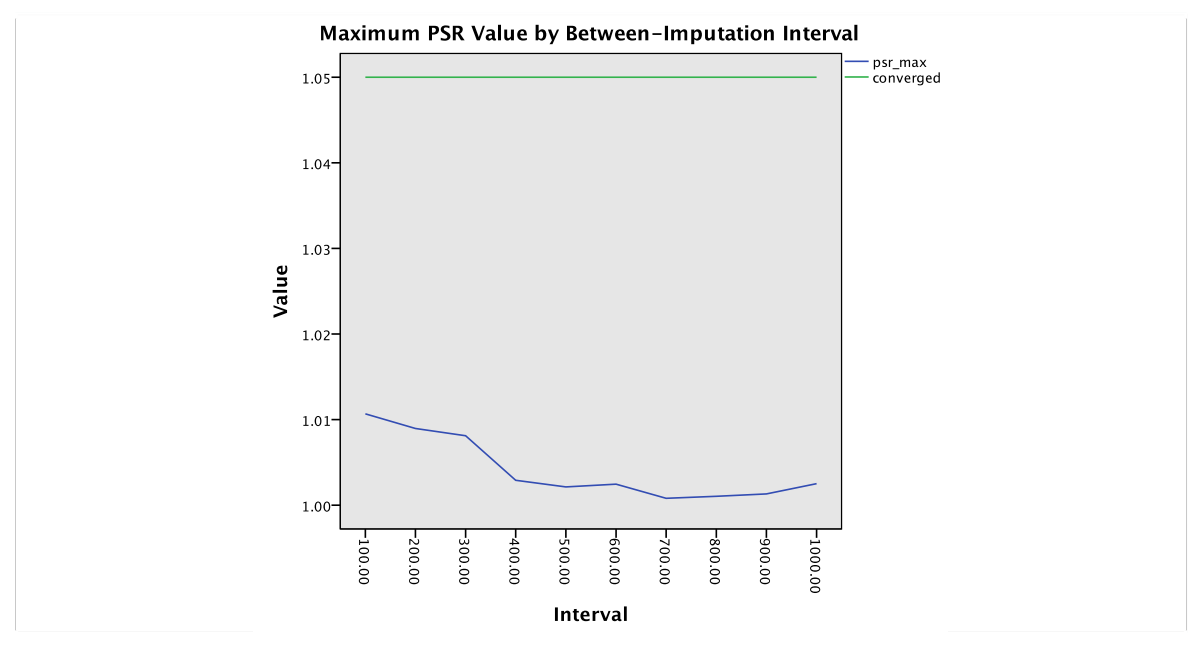
TRACE PLOT OF THE BMI STANDARD DEVIATION



POTENTIAL SCALE REDUCTION

- The macro computes the PSR for the P-step means and standard deviations after every 100 iterations
- These parameters often converge quickly than covariances (which SPSS does not provide)
- Be conservative when choosing the between-imputation interval based on the PSR
- The macro displays the maximum PSR for all parameters in a table and in a line graph

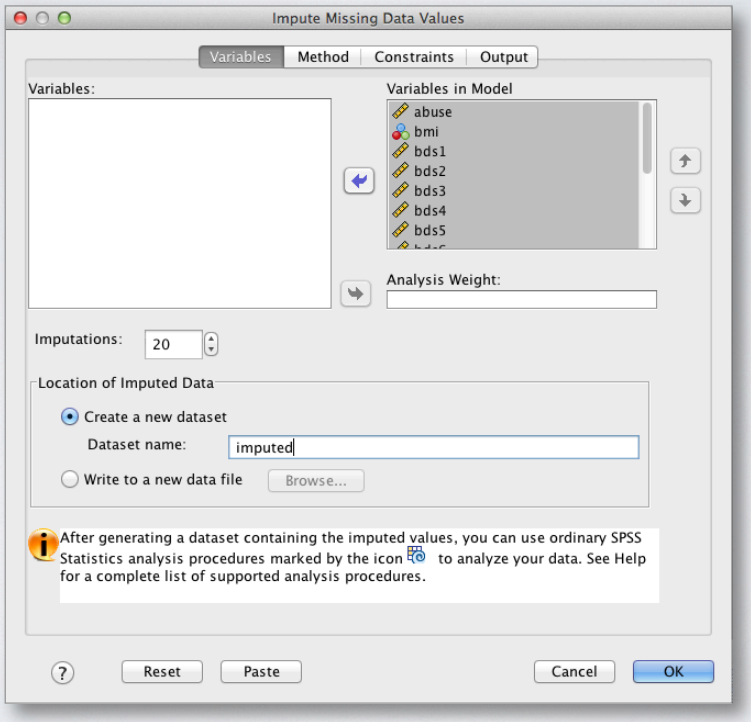
LINE GRAPH OF MAXIMUM PSR VALUES



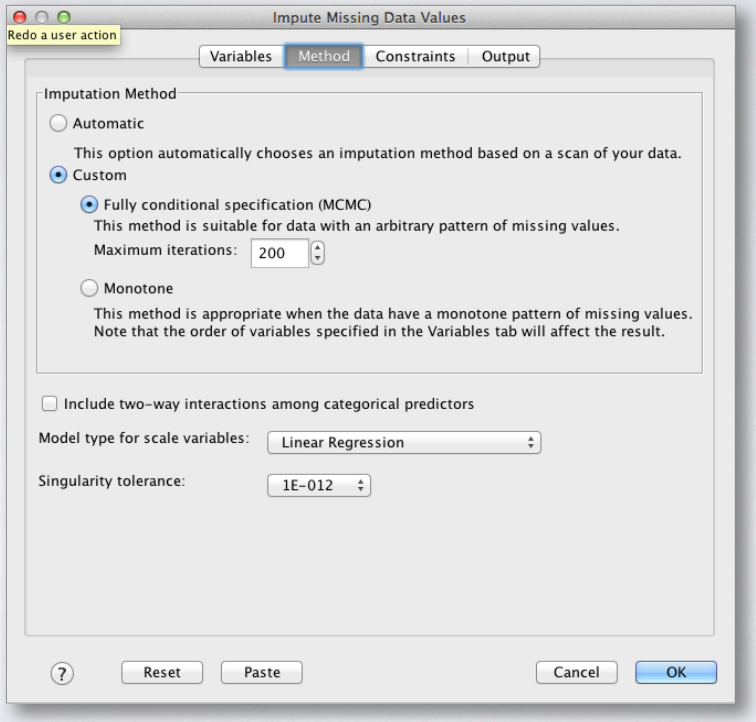
FINAL IMPUTATION RUN

- After establishing convergence, run MCMC a second time to generate the imputed data sets
- Imputation details for this example:
  - 20 imputed data sets
  - 200 iterations separating each imputed data set

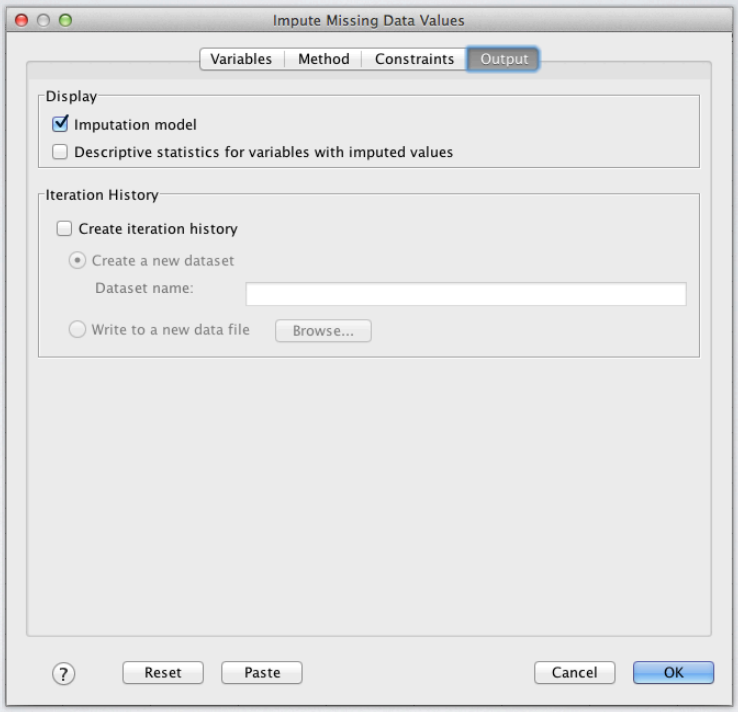
VARIABLES TAB



METHOD TAB



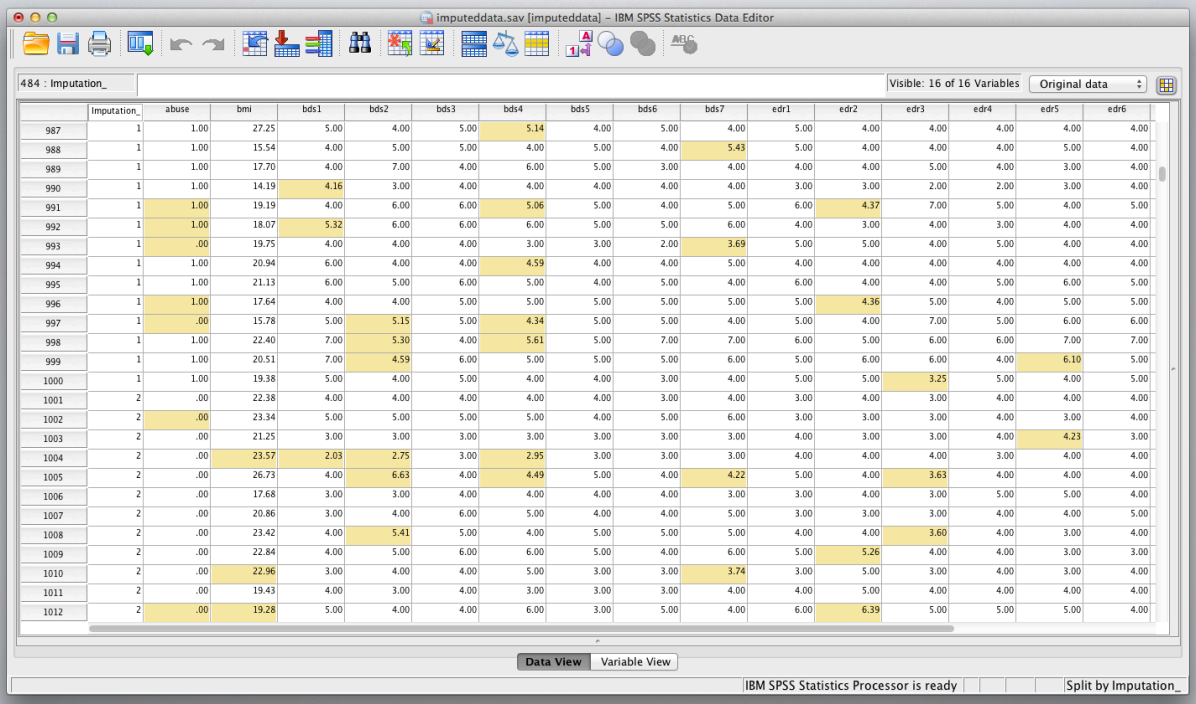
OUTPUT TAB



IMPUTATION OUTPUT

- SPSS stacks the imputed data sets into a single file
- A variable named IMPUTATION_ differentiates the data sets
- The stacked file format is convenient because data manipulation tasks (e.g., computing new variables, recoding, etc.) need only be executed once
- The IMPUTATION_ variable plays an important role in the subsequent analyses …

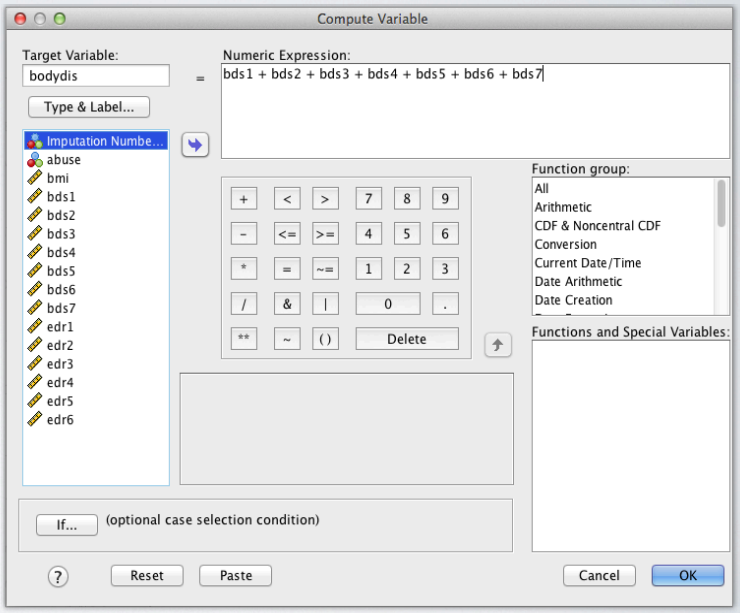
IMPUTED DATA



	Imputation_	abuse	bmi	bds1	bds2	bds3	bds4	bds5	bds6	bds7	edr1	edr2	edr3	edr4	edr5	edr6
987	1	1.00	27.25	5.00	4.00	5.00	5.14	4.00	5.00	4.00	5.00	4.00	4.00	4.00	4.00	4.00
988	1	1.00	15.54	4.00	5.00	5.00	4.00	5.00	4.00	5.43	5.00	4.00	4.00	4.00	5.00	4.00
989	1	1.00	17.70	4.00	7.00	4.00	6.00	5.00	3.00	4.00	4.00	4.00	5.00	4.00	3.00	4.00
990	1	1.00	14.19	4.16	3.00	4.00	4.00	4.00	4.00	4.00	3.00	3.00	2.00	2.00	3.00	4.00
991	1	1.00	19.19	4.00	6.00	6.00	5.06	5.00	4.00	5.00	6.00	4.37	7.00	5.00	4.00	5.00
992	1	1.00	18.07	5.32	6.00	6.00	6.00	5.00	5.00	6.00	4.00	3.00	4.00	3.00	4.00	4.00
993	1	.00	19.75	4.00	4.00	4.00	3.00	3.00	2.00	3.69	5.00	5.00	4.00	5.00	4.00	4.00
994	1	1.00	20.94	6.00	4.00	4.00	4.59	4.00	4.00	5.00	4.00	4.00	4.00	4.00	4.00	4.00
995	1	1.00	21.13	6.00	5.00	6.00	5.00	4.00	5.00	4.00	6.00	4.00	4.00	5.00	6.00	5.00
996	1	1.00	17.64	4.00	4.00	5.00	5.00	5.00	5.00	5.00	5.00	4.36	5.00	4.00	5.00	5.00
997	1	.00	15.78	5.00	5.15	5.00	4.34	5.00	5.00	4.00	5.00	4.00	7.00	5.00	6.00	6.00
998	1	1.00	22.40	7.00	5.30	4.00	5.61	5.00	7.00	7.00	6.00	5.00	6.00	6.00	7.00	7.00
999	1	1.00	20.51	7.00	4.59	6.00	5.00	5.00	5.00	6.00	5.00	6.00	6.00	4.00	6.10	5.00
1000	1	1.00	19.38	5.00	4.00	5.00	4.00	4.00	3.00	4.00	5.00	5.00	3.25	5.00	4.00	5.00
1001	2	.00	22.38	4.00	4.00	4.00	4.00	4.00	3.00	4.00	3.00	4.00	3.00	4.00	4.00	4.00
1002	2	.00	23.34	5.00	5.00	5.00	5.00	4.00	5.00	6.00	3.00	3.00	3.00	4.00	3.00	4.00
1003	2	.00	21.25	3.00	3.00	3.00	3.00	3.00	3.00	3.00	4.00	3.00	3.00	4.00	4.23	3.00
1004	2	.00	23.57	2.03	2.75	3.00	2.95	3.00	3.00	3.00	4.00	4.00	4.00	3.00	4.00	4.00
1005	2	.00	26.73	4.00	6.63	4.00	4.49	5.00	4.00	4.22	5.00	4.00	3.63	4.00	4.00	4.00
1006	2	.00	17.68	3.00	3.00	4.00	4.00	4.00	4.00	3.00	3.00	4.00	3.00	5.00	5.00	4.00
1007	2	.00	20.86	3.00	4.00	6.00	5.00	4.00	4.00	5.00	3.00	3.00	3.00	4.00	4.00	5.00
1008	2	.00	23.42	4.00	5.41	5.00	4.00	5.00	5.00	5.00	4.00	4.00	3.60	4.00	3.00	4.00
1009	2	.00	22.84	4.00	5.00	6.00	6.00	5.00	4.00	6.00	5.00	5.26	4.00	4.00	3.00	3.00
1010	2	.00	22.96	3.00	4.00	4.00	5.00	3.00	3.00	3.74	3.00	5.00	3.00	4.00	3.00	4.00
1011	2	.00	19.43	4.00	3.00	4.00	3.00	3.00	3.00	4.00	4.00	5.00	4.00	4.00	4.00	4.00
1012	2	.00	19.28	5.00	4.00	4.00	6.00	3.00	5.00	4.00	6.00	6.39	5.00	5.00	5.00	4.00

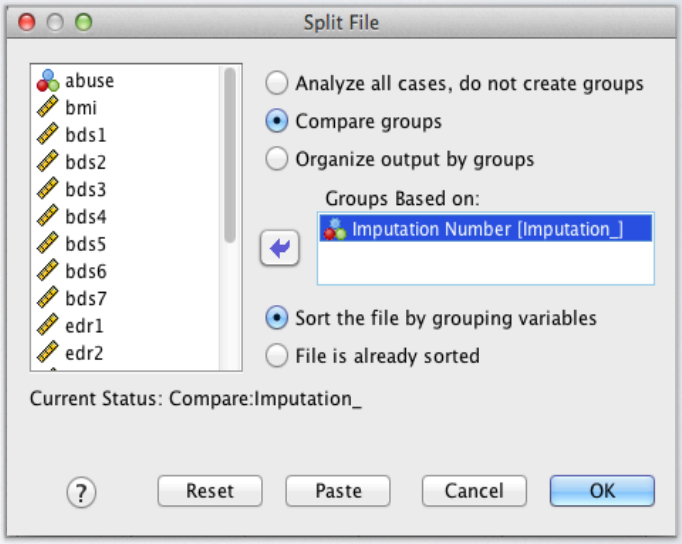
COMPUTING SCALE SCORES

- Use the COMPUTE command once for each scale score
- Transformations are automatically applied to every data set because the imputed files are stacked



SPLIT FILE COMMAND

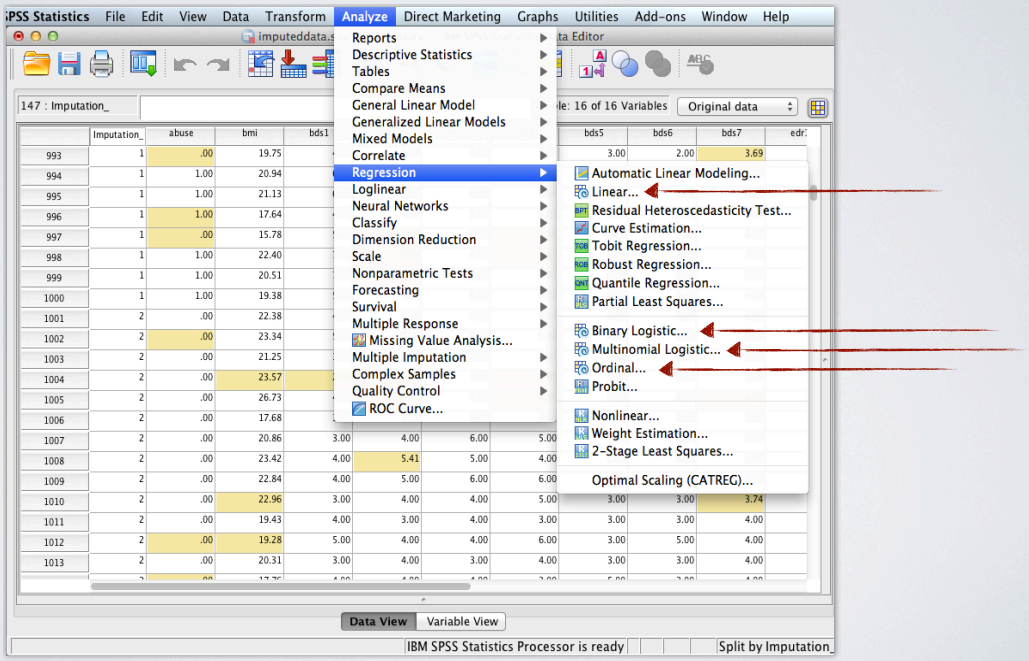
- Splitting the file by the IMPUTATION_ variable invokes the analysis and pooling procedures, if available



ANALYZING MULTIPLY IMPUTED DATA

- Analyze the data as usual
- SPSS pools estimates for many common analyses, but not all
- The program is idiosyncratic in its application of the pooling formulas (e.g., in a regression analysis, SPSS pools the unstandardized coefficients but not the beta weights)
- The icon denotes a procedure that can accommodate multiply imputed data

SUPPORTED PROCEDURES



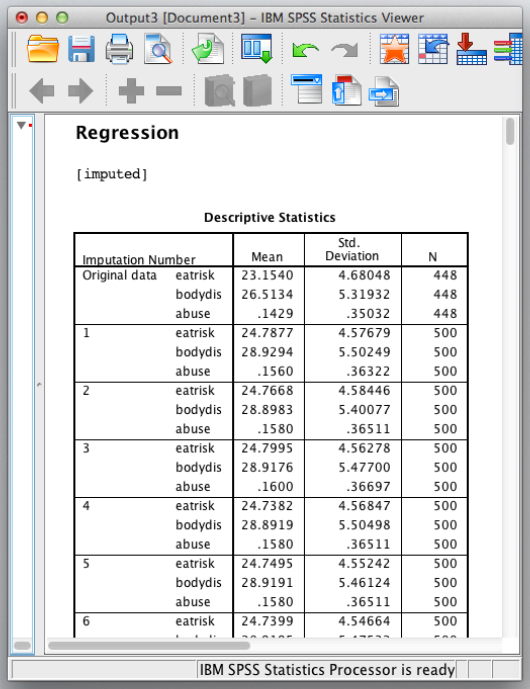
SPSS SYNTAX

```
*/ PERFORM MULTIPLE IMPUTATION ANALYSIS */.

dataset activate rawdata.

multiple imputation bmi bds1 to bds7 edr1 to edr6 abuse
 /impute method = fcs maxiter = 200 nimputations = 20
 /outfile imputations = imputed.

*/ COMPUTE SCALE SCORES WITHIN EACH DATA SET */

dataset activate imputed.
compute bodydis = sum(bds1 to bds7).
compute eatrisk = sum(edr1 to edr6).
exe.

*/ SPLIT THE FILE BY VARIABLE IMPUTATION_ */

sort cases by imputation_.
split file layered by imputation_.
```

SPSS OUTPUT

- SPSS reports the analysis results separately for each imputed data set even when it does not pool the estimates
- The pooled estimates and standard errors appear at the bottom of each table, when available
- Some estimates get pooled, some do not …

DESCRIPTIVES

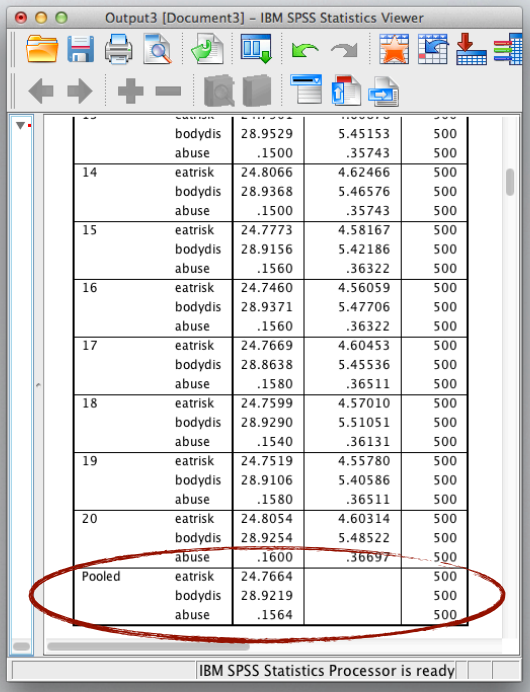


Regression

[imputed]

Descriptive Statistics

Imputation Number		Mean	Std. Deviation	N
Original data	eatrisk	23.1540	4.68048	448
	bodydis	26.5134	5.31932	448
	abuse	.1429	.35032	448
1	eatrisk	24.7877	4.57679	500
	bodydis	28.9294	5.50249	500
	abuse	.1560	.36322	500
2	eatrisk	24.7668	4.58446	500
	bodydis	28.8983	5.40077	500
	abuse	.1580	.36511	500
3	eatrisk	24.7995	4.56278	500
	bodydis	28.9176	5.47700	500
	abuse	.1600	.36697	500
4	eatrisk	24.7382	4.56847	500
	bodydis	28.8919	5.50498	500
	abuse	.1580	.36511	500
5	eatrisk	24.7495	4.55242	500
	bodydis	28.9191	5.46124	500
	abuse	.1580	.36511	500
6	eatrisk	24.7399	4.54664	500



	bodydis	28.9529	5.45153	500
	abuse	.1500	.35743	500
14	eatrisk	24.8066	4.62466	500
	bodydis	28.9368	5.46576	500
	abuse	.1500	.35743	500
15	eatrisk	24.7773	4.58167	500
	bodydis	28.9156	5.42186	500
	abuse	.1560	.36322	500
16	eatrisk	24.7460	4.56059	500
	bodydis	28.9371	5.47706	500
	abuse	.1560	.36322	500
17	eatrisk	24.7669	4.60453	500
	bodydis	28.8638	5.45536	500
	abuse	.1580	.36511	500
18	eatrisk	24.7599	4.57010	500
	bodydis	28.9290	5.51051	500
	abuse	.1540	.36131	500
19	eatrisk	24.7519	4.55780	500
	bodydis	28.9106	5.40586	500
	abuse	.1580	.36511	500
20	eatrisk	24.8054	4.60314	500
	bodydis	28.9254	5.48522	500
	abuse	.1600	.36697	500
Pooled	eatrisk	24.7664		500
	bodydis	28.9219		500
	abuse	.1564		500

CORRELATIONS

Output3 [Document3] – IBM SPSS Statistics Viewer

Correlations

Imputation Number			eatrisk	bodydis	abuse
Original data	Pearson Correlation	eatrisk	1.000	.484	.272
		bodydis	.484	1.000	.246
		abuse	.272	.246	1.000
	Sig. (1-tailed)	eatrisk	.	.000	.000
		bodydis	.000	.	.000
		abuse	.000	.000	.
	N	eatrisk	448	448	448
		bodydis	448	448	448
		abuse	448	448	448
1	Pearson Correlation	eatrisk	1.000	.638	.301
		bodydis	.638	1.000	.338
		abuse	.301	.338	1.000
	Sig. (1-tailed)	eatrisk	.	.000	.000
		bodydis	.000	.	.000
		abuse	.000	.000	.
	N	eatrisk	500	500	500
		bodydis	500	500	500
		abuse	500	500	500
2	Pearson Correlation	eatrisk	1.000	.641	.347
		bodydis	.641	1.000	.318
		abuse	.347	.318	1.000
	Sig. (1-tailed)	eatrisk	.	.000	.000
		bodydis	.000	.	.000
		abuse	.000	.000	.
	N	eatrisk	500	500	500

IBM SPSS Statistics Processor is ready | H: 3835, W: 496 pt.

Output3 [Document3] – IBM SPSS Statistics Viewer

19	Pearson Correlation	eatrisk	1.000	.648	.309
		bodydis	.648	1.000	.348
		abuse	.309	.348	1.000
	Sig. (1-tailed)	eatrisk	.	.000	.000
		bodydis	.000	.	.000
		abuse	.000	.000	.
	N	eatrisk	500	500	500
		bodydis	500	500	500
		abuse	500	500	500
20	Pearson Correlation	eatrisk	1.000	.642	.316
		bodydis	.642	1.000	.345
		abuse	.316	.345	1.000
	Sig. (1-tailed)	eatrisk	.	.000	.000
		bodydis	.000	.	.000
		abuse	.000	.000	.
	N	eatrisk	500	500	500
		bodydis	500	500	500
		abuse	500	500	500
Pooled	Pearson Correlation	eatrisk	1.000	.637	.317
		bodydis	.637	1.000	.328
		abuse	.317	.328	1.000
	Sig. (1-tailed)	eatrisk	.	.000	.000
		bodydis	.000	.	.000
		abuse	.000	.000	.
	N	eatrisk	500	500	500
		bodydis	500	500	500
		abuse	500	500	500

IBM SPSS Statistics Processor is ready | H: 3835, W: 496 pt.

REGRESSION

Output3 [Document3] – IBM SPSS Statistics Viewer

Coefficients[a]

Imputation Number	Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta
Original data	1	(Constant)	12.490	.984	
		bodydis	.391	.037	.444
		abuse	2.170	.563	.162
1	1	(Constant)	10.031	.870	
		bodydis	.504	.030	.605
		abuse	1.211	.459	.096
2	1	(Constant)	9.966	.867	
		bodydis	.501	.030	.590
		abuse	2.001	.447	.159
3	1	(Constant)	10.189	.862	
		bodydis	.496	.030	.595
		abuse	1.681	.449	.135
4	1	(Constant)	10.347	.868	
		bodydis	.490	.030	.590
		abuse	1.547	.456	.124
5	1	(Constant)	9.958	.858	
		bodydis	.503	.030	.604
		abuse	1.502	.446	.120
6	1	(Constant)	9.960	.849	
		bodydis	.503	.029	.605
		abuse	1.643	.454	.128
7	1	(Constant)	10.139	.857	
		bodydis	.499	.030	.604

IBM SPSS Statistics Processor is ready | H: 3835, W: 496 pt.

Output3 [Document3] – IBM SPSS Statistics Viewer

		bodydis	.505	.031	.598
		abuse	1.217	.474	.094
14	1	(Constant)	9.999	.879	
		bodydis	.504	.031	.595
		abuse	1.564	.467	.121
15	1	(Constant)	10.080	.869	
		bodydis	.498	.030	.590
		abuse	1.871	.451	.148
16	1	(Constant)	10.128	.860	
		bodydis	.496	.030	.596
		abuse	1.615	.451	.129
17	1	(Constant)	9.827	.863	
		bodydis	.508	.030	.602
		abuse	1.734	.451	.138
18	1	(Constant)	10.107	.870	
		bodydis	.500	.030	.603
		abuse	1.155	.462	.091
19	1	(Constant)	9.572	.873	
		bodydis	.519	.030	.615
		abuse	1.193	.452	.096
20	1	(Constant)	9.912	.874	
		bodydis	.507	.030	.605
		abuse	1.347	.456	.107
Pooled	1	(Constant)	10.058	.892	
		bodydis	.500	.031	
		abuse	1.507	.516	

a. Dependent Variable: eatrisk

IBM SPSS Statistics Processor is ready | H: 3835, W: 496 pt.