2014 ■ vol. 10 ■ no. 2

DOI: 10.20982/tqmp.10.2.p143

Exploratory factor analysis and reliability analysis with missing data: A simple method for SPSS users

Bruce Weaver [a], Hillary Maxwell [b]

[a] Human Sciences Division, Northern Ontario School of Medicine; and Centre for Research on Safe Driving, Lakehead University

[b] Centre for Research on Safe Driving, Lakehead University

Abstract ■ Missing data is a frequent problem for researchers conducting exploratory factor analysis (EFA) or reliability analysis. The SPSS FACTOR procedure allows users to select listwise deletion, pairwise deletion or mean substitution as a method for dealing with missing data. The shortcomings of these methods are well-known. Graham (2009) argues that a much better way to deal with missing data in this context is to use a matrix of expectation maximization (EM) covariances (or correlations) as input for the analysis. SPSS users who have the Missing Values Analysis add-on module can obtain vectors of EM means and standard deviations plus EM correlation and covariance matrices via the MVA procedure. But unfortunately, MVA has no /MATRIX sub-command, and therefore cannot write the EM correlations directly to a matrix dataset of the type needed as input to the FACTOR and RELIABILITY procedures. We describe two macros that (in conjunction with an intervening MVA command) carry out the data management steps needed to create two matrix datasets, one containing EM correlations and the other EM covariances. Either of those matrix datasets can then be used as input to the FACTOR procedure, and the EM correlations can also be used as input to RELIABILITY. We provide an example that illustrates the use of the two macros to generate the matrix datasets and how to use those datasets as input to the FACTOR and RELIABILITY procedures. We hope that this simple method for handling missing data will prove useful to both students and researchers who are conducting EFA or reliability analysis.

Keywords ■ Exploratory factor analysis, reliability analysis, missing data, expectation maximization, SPSS

bweaver@lakeheadu.ca

Introduction

The SPSS MVA procedure (IBM, 2011a), which was introduced in release 12 as part of the *Missing Values Analysis* add-on module, includes four methods for dealing with missing values: listwise and pairwise deletion, single imputation via regression, and expectation maximization (EM). Prior to release 12, several procedures had listwise and pairwise deletion available as options on a /MISSING sub-command (e.g., CORRELATIONS, FACTOR, NONPAR CORR, REGRESSION), but single imputation via regression and EM were not available prior to the introduction of the MVA procedure.

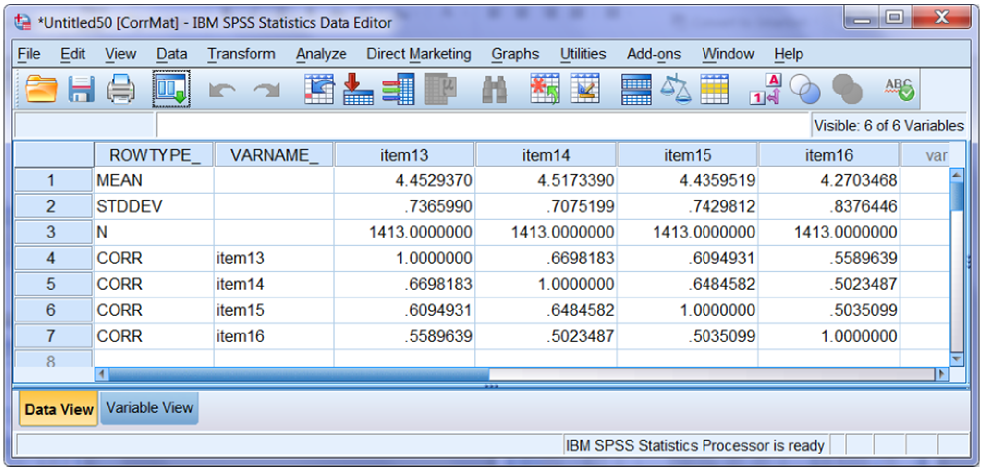
In his review of MVA, von Hippel (2004) pointed out the well-known limitations of listwise deletion, pairwise deletion and single imputation methods (see also Acock, 2005; Donders et al., 2006; Schaefer & Graham, 2002). Regarding EM, however, von Hippel (p. 164) said this:

> *The one bright spot in MVA is its implementation of the EM method, which can produce maximum likelihood point estimates of means, variances, and covariances. The EM method can also be used to impute missing values. Unfortunately, MVA imputes these values without residual variation. Analyses based on the EM imputations can therefore be biased.*

Unfortunately, it is very easy to become fixated on the last part of that excerpt, and to mistakenly conclude that the MVA procedure provides nothing useful (by modern standards) for the treatment of missing data. While it is true that analyses of *raw data* that include the EM imputed values can produce biased estimates, we must not lose sight of the fact that MVA's EM sub-command does provide perfectly good maximum likelihood (ML) estimates of the means, standard deviations, covariances and correlations, and that *these* can be used either descriptively or as input for certain types of analyses. Graham (2009, p. 556) makes this point very nicely as follows:

> *Although the EM algorithm provides excellent parameter estimates, the lack of convenient standard errors means that EM is not particularly*

A.



B.

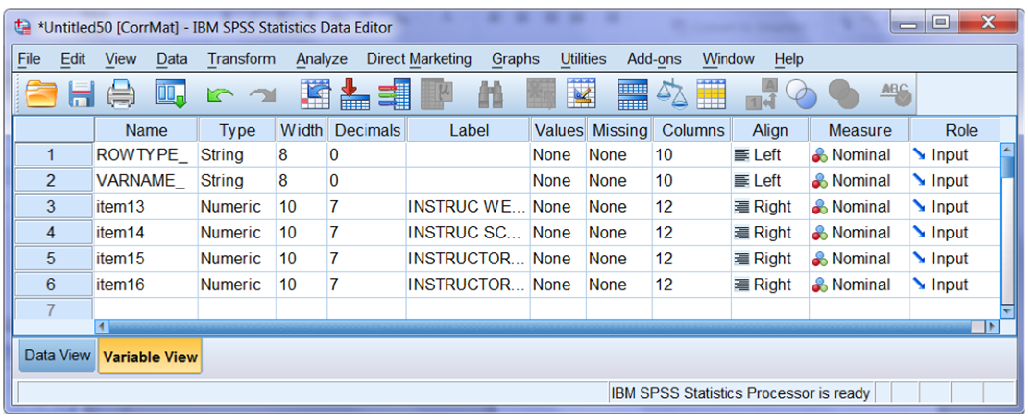


Figure 1 ■ A correlation matrix dataset in the format needed as input to the FACTOR and RELIABILITY procedures. Panels A and B show the Data View and Variable View windows respectively.

good for hypothesis testing. On the other hand, several important analyses, often preliminary analyses, don't use standard errors anyway, so the EM estimates are very useful. First, it is often desirable to report means, standard deviations, and sometimes a correlation matrix in one's paper. I would argue that the best estimates for these quantities are the ML estimates provided by EM. Second, data quality analyses, for example, coefficient alpha analyses, because they typically do not involve standard errors, can easily be based on the EM covariance matrix (e.g., see Enders 2003; Graham et al. 2002, 2003). The EM covariance matrix is also an excellent basis for exploratory factor analysis with missing data.

Contrary to what Graham states above, Truxillo (2005) has suggested that the EM covariance matrix and vector of means can *also* be used as input for procedures that entail inference, but that one must then use a nominal sample size that properly accounts for the fact that some data are missing. Allison (2012, p. 15), on the other hand, argues that there is no single sample size that would yield correct results for all coefficients in a regression model. That debate, while interesting, is beyond the scope of this article. We focus exclusively on how to use a matrix dataset of EM correlations as input for exploratory factor analysis (EFA) and reliability analysis, two "good uses of the EM

Univariate Statistics

	N	Mean	Std. Deviation	Missing Count	Missing Percent	No. of Extremes[a] Low	No. of Extremes[a] High
item13	1419	4.45	.737	9	.6	33	0
item14	1424	4.52	.709	4	.3	28	0
item15	1424	4.43	.748	4	.3	33	0
item16	1420	4.27	.839	8	.6	69	0

a. Number of cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR).

Summary of Estimated Means

	item13	item14	item15	item16
All Values	4.45	4.52	4.43	4.27
EM	4.45	4.52	4.43	4.27

Summary of Estimated Standard Deviations

	item13	item14	item15	item16
All Values	.737	.709	.748	.839
EM	.738	.709	.748	.840

EM Means[a]

item13	item14	item15	item16
4.45	4.52	4.43	4.27

a. Little's MCAR test: Chi-Square = 22.857, DF = 15, Sig. = .087

EM Covariances[a]

	item13	item14	item15	item16
item13	.544			
item14	.349	.503		
item15	.336	.346	.559	
item16	.345	.300	.318	.705

a. Little's MCAR test: Chi-Square = 22.857, DF = 15, Sig. = .087

EM Correlations[a]

	item13	item14	item15	item16
item13	1			
item14	.668	1		
item15	.609	.653	1	
item16	.558	.504	.507	1

a. Little's MCAR test: Chi-Square = 22.857, DF = 15, Sig. = .087

Figure 2 ■ Example of output from the MVA procedure with an /EM sub-command.

algorithm", as Graham (2009) puts it.

In SPSS, EFA and reliability analysis are carried out via the FACTOR and RELIABILITY commands respectively. Both of those procedures have a /MATRIX sub-command that allows users to specify a matrix dataset (rather than a raw data file) as input. For FACTOR, the matrix dataset can be either a correlation matrix or a covariance matrix; for RELIABILITY, it must be a correlation matrix.

Figure 1 shows the required format of a correlation matrix dataset. The data are from an example provided by the UCLA Statistical Consulting Group (n.d.). The analyses to be performed later use variables item13 to item 24, but here (and in some other figures) we use only item13 to item16 in order to avoid having output tables that are too large to display easily. Panels A and B show the Data View and Variable View windows respectively. The first two variables in the dataset are ROWTYPE_ and VARNAME_. Both of these are short string variables (i.e., string variables with maximum length of 8 characters). As its name suggests, ROWTYPE_ indicates what type of information is held in that row. In a correlation matrix dataset, the possible values of ROWTYPE_ are MEAN, STDDEV, N and CORR. For a covariance matrix dataset, everything is the same except that CORR is replaced with COV (and a covariance matrix replaces the correlation matrix).[1]

As noted earlier, the MVA procedure computes EM means, standard deviations (SDs), covariances and correlations, and displays them in the output viewer (see Figure 2). But unfortunately, MVA has no /MATRIX sub-command, and therefore cannot write the EM means and correlations (or covariances) directly to a matrix dataset formatted as required for input to the FACTOR or RELIABLILITY procedures. Thus, anyone wishing to use EM correlations (or covariances) as matrix input for the FACTOR or RELIABILITY procedures has some preliminary work to do: They must transfer the EM correlations (or covariances) shown in the output viewer (and Figure 2) into an appropriately formatted matrix dataset (as in Figure 1). One way to do so is to send output from MVA to another dataset via the *Output Management System* (OMS), and then carry out several data management steps. (The OMS (Output Management System) command was introduced as part of the base module in release 13 of SPSS. See IBM, 2001b for details.) We have written a pair of macros to facilitate that process, and describe them next.

Description of the !PrepEMdata macros

We have written two macros called !PrepEMdata1 and !PrepEMdata2.[2] The common part of those macro names is short for *prepare EM data*. The macros are used on either side of an MVA command, as follows:

1. Call !PrepEMdata1.
2. Use DATASET ACTIVATE to activate the dataset containing the raw data and then issue an appropriate MVA command with an /EM sub-command.
3. Call !PrepEMdata2.

!PrepEMdata1 issues two OMS commands that cause EM output from MVA to be written to two new datasets. When the subsequent MVA command is issued, the results shown in the *Summary of Estimated Means* and *Summary of Estimated Standard Deviations* pivot tables (see Figure 2) are written to a new temporary dataset called @1; and the matrix of EM correlations is written to temporary dataset @2.

When !PrepEMdata2 is called, if first issues an OMSEND command that causes the @1 and @2 datasets to be finalized. It then goes through several data management steps that result in the creation of dataset EMcorr, a matrix dataset of EM correlations that is ready for use as input to the FACTOR and RELIABILITY procedures. (It also uses an MCONVERT command to create dataset EMcov, a matrix dataset of EM covariances. See IBM, 2011c for more information on MCONVERT.) For more detail concerning the data management steps, see the macro definitions in the Appendix.

Macro arguments

!PrepEMdata1 has no arguments: It is called simply by giving the macro name followed by a command terminator (i.e., a period). !PrepEMdata2 has two required arguments, *Vars* and *N*. *Vars* is a list of the variables appearing in the vectors of EM means and SDs and in the matrices of EM correlations and covariances. The keyword TO can be used in the variable list. The end of the variable list is indicated with a forward slash (/). *N* is a sample size (a *nominal* sample size, as Truxillo 2005 puts it) that will be plugged into the N-row of the final matrix files of EM correlations and covariances.

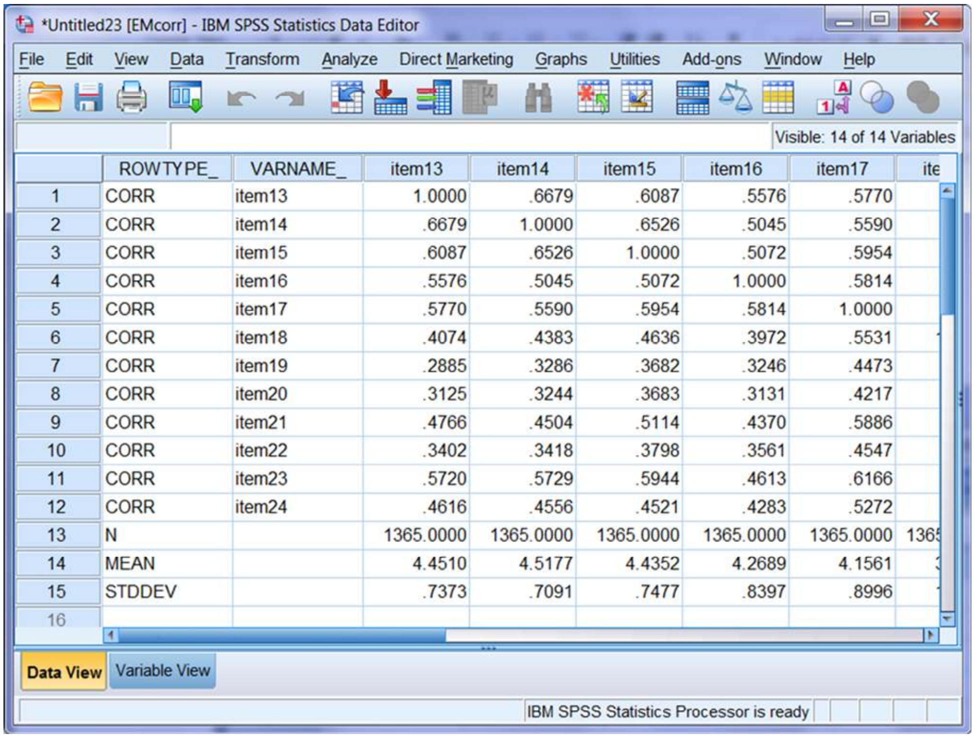
Examples

Exploratory Factor Analysis

To illustrate, we use data from the factor analysis example on this UCLA webpage: http://www.ats.ucla.edu/stat/spss/output/factor1.htm The data file can be downloaded from http://www.ats.ucla.edu/stat/spss/output/principal_components_files/M255.SAV. We start with syntax to do the following: 1) define two file handles, one for the data file, and one for the syntax file containing the two

[1] We generated the matrix dataset shown in Figure 1 via the CORRELATIONS procedure with a /MATRIX OUT sub-command, and with /MISSING=LISTWISE. Had we used /MISSING=PAIRWISE (the default setting), the matrix dataset would have contained a matrix of pairwise sample sizes on 4 rows with ROWTYPE_ equal to “N”.

[2] There is no requirement that macro names begin with an exclamation mark. We have chosen to start our macro names with "!" to be consistent with a practice recommended by Raynald Levesque on the well-known SPSS Tools website (www.spsstools.net). He recommends starting macro names with an exclamation mark because this makes it easier to distinguish between macros and regular SPSS commands when one is viewing a syntax file. Furthermore, searching for exclamation points makes it easy to find all macros that follow this naming convention.



	ROWTYPE_	VARNAME_	item13	item14	item15	item16	item17	ite
1	CORR	item13	1.0000	.6679	.6087	.5576	.5770	
2	CORR	item14	.6679	1.0000	.6526	.5045	.5590	
3	CORR	item15	.6087	.6526	1.0000	.5072	.5954	
4	CORR	item16	.5576	.5045	.5072	1.0000	.5814	
5	CORR	item17	.5770	.5590	.5954	.5814	1.0000	
6	CORR	item18	.4074	.4383	.4636	.3972	.5531	
7	CORR	item19	.2885	.3286	.3682	.3246	.4473	
8	CORR	item20	.3125	.3244	.3683	.3131	.4217	
9	CORR	item21	.4766	.4504	.5114	.4370	.5886	
10	CORR	item22	.3402	.3418	.3798	.3561	.4547	
11	CORR	item23	.5720	.5729	.5944	.4613	.6166	
12	CORR	item24	.4616	.4556	.4521	.4283	.5272	
13	N		1365.0000	1365.0000	1365.0000	1365.0000	1365.0000	1365
14	MEAN		4.4510	4.5177	4.4352	4.2689	4.1561	
15	STDDEV		.7373	.7091	.7477	.8397	.8996	
16								

Figure 3 ■ A partial view of the EMcorr dataset generated via three-step process of calling !PrepEMdata1, running MVA with an /EM sub-command, and calling !PrepEMdata2.

macro definitions; 2) run the syntax file containing the two macro definitions; 3) open the data file; 4) call the !PrepEMdata1 macro; 5) issue an MVA command with /EM sub-command; 6) call the !PrepEMdata2 macro; and 7) perform EFA via the FACTOR command.

```
* [1] Two FILE HANDLE commands.
FILE HANDLE TheDataFile
/NAME="C:\bw\SPSS\data\UCLA\M255.SAV".
FILE HANDLE MacroDefs
/NAME="C:\bw\SPSS\macros\PrepEMdata_macros.s
ps".

* [2] Run the macro definition file.
SET PRINTBACK = OFF. /* Suppress output.
INSERT FILE = "MacroDefs".
SET PRINTBACK = ON. /* Turn output back on.

* [3] Open the data file.
NEW FILE.
DATASET CLOSE all.
GET FILE = "TheDataFile".
DATASET NAME RawData.

* Now generate a matrix dataset of
* EM correlations.
```

```
* [4] Call first PrepEMdata macro to set up
* OMS commands.
!PrepEMdata1.

* [5] Use MVA with /EM to generate the
*      needed EM estimates.
*      Users are free to use whatever MVA
*      options they wish.
MVA VARIABLES= item13 to item24 /EM.

* [6] Call the 2nd PrepEMdata macro to do
* the required data management.
* I shall set the nominal N = 1365, which is
* the listwise N.
!PrepEMdata2 Vars = item13 to item24
  / N = 1365.

* At this point, datasets EMcorr and EMcov
* have been created.

* [7] Perform EFA with EM correlations as
* input.
* We shall extract 3 factors (via PAF) and
* use PROMAX rotation, as in the UCLA
* example.
DATASET ACTIVATE EMcorr.
FACTOR MATRIX IN(COR=*)
```

Table 1 ■ Rotated factor loadings and uniqueness values from SPSS and Stata.

	Factor 1		Factor 2		Factor 3		Uniqueness	
Variable	SPSS	Stata	SPSS	Stata	SPSS	Stata	SPSS	Stata
Item13	0.888	0.843					0.329	0.368
Item14	0.848	0.829					0.362	0.382
Item15	0.741	0.737					0.389	0.399
Item16	0.627	0.646					0.545	0.541
Item17	0.513	0.519					0.379	0.379
Item18			0.812	0.783			0.320	0.356
Item19			0.886	0.835			0.414	0.466
Item20			0.580	0.601			0.627	0.627
Item21			0.430	0.438			0.456	0.454
Item22			0.533	0.540			0.552	0.554
Item23					0.767	0.681	0.213	0.267
Item24					0.831	0.745	0.358	0.415

Note: Blanks represent factor loadings with absolute values < 0.3.

```
/ANALYSIS = item13 to item24
/PRINT INITIAL EXTRACTION ROTATION FSCORE
/FORMAT BLANK(.3)
/CRITERIA FACTORS(3) ITERATE(100) DELTA(0)
/EXTRACTION PAF
/ROTATION PROMAX
/METHOD=CORRELATION.
```

A partial view of the EMcorr dataset generated by that syntax listing is shown in Figure 3. Notice that it has the same basic structure as the matrix dataset shown in Figure 1A, the only differences being that the rows with ROWTYPE_ = N, MEAN and STDDEV appear at the bottom of the dataset rather than the top, and they are ordered differently (in Figure 1A, the order is MEAN, STDDEV, N). Neither of these differences affects how the matrix file works.

Full output for this example can be found in the online supplementary material. As a validity check, we implemented the approach shown on another UCLA website (http://www.ats.ucla.edu/stat/stata/faq/factor_missing.htm) to perform the same analysis using Stata (version 12.1). (The Stata example on that page used ML extraction and Varimax rotation. We modified the Stata commands to use principal factors (PF) for extraction and Promax(4) as the rotation method, thus matching the approach we used in SPSS.) The rotated factor loadings and uniqueness (or specificity) values from SPSS and Stata can be compared in Table 1. (The uniqueness values are not reported directly in the SPSS output. The values shown here are equal to 1 minus the communalities found in the Extraction column of the Communalities table.) Clearly, the two sets of rotated factor loadings are very similar.

Reliability Analysis

The following listing shows a RELIABILITY command that uses EMcorr (the matrix dataset of EM correlations) as input—notice the /MATRIX=IN(*) sub-command. We also ran two other RELIABILITY commands not shown here, one with /VARIABLES= item18 to item22 (for Factor 2), and one with /VARIABLES= item23 item24 (for Factor 3). The RELIABILITY results for Factor 1 are shown in Figure 4.

```
DATASET ACTIVATE EMcorr.
RELIABILITY
  /VARIABLES= item13 to item17
  /SCALE('Factor 1') ALL
  /MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE SCALE CORRELATIONS
  /SUMMARY=TOTAL MEANS VARIANCE
  /MATRIX=IN(*).
```

What is the correct nominal sample size?

One issue we have not yet addressed explicitly concerns the appropriate nominal sample size for the EMcorr and EMcov datasets. For those parts of the output from FACTOR and RELIABILITY that do not entail statistical inference (i.e., computation of standard errors, confidence intervals or *p*-values), the results are independent of the sample size one chooses. A demonstration of this is provided in the output file

Case Processing Summary

		N	%
Cases	Valid	1365	100.0
	Excluded[a]	0	.0
	Total	1365	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.870	.874	5

Item Statistics

	Mean	Std. Deviation	N
item13	4.451034	.7373051	1365
item14	4.517707	.7090572	1365
item15	4.435165	.7476744	1365
item16	4.268866	.8397183	1365
item17	4.156094	.8996052	1365

Inter-Item Correlation Matrix

	item13	item14	item15	item16	item17
item13	1.000	.668	.609	.558	.577
item14	.668	1.000	.653	.504	.559
item15	.609	.653	1.000	.507	.595
item16	.558	.504	.507	1.000	.581
item17	.577	.559	.595	.581	1.000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	4.366	4.156	4.518	.362	1.087	.022	5
Item Variances	.624	.503	.809	.307	1.610	.017	5

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
item13	17.377833	6.898	.729	.551	.835
item14	17.311159	7.060	.718	.556	.839
item15	17.393701	6.907	.713	.530	.839
item16	17.560001	6.756	.643	.425	.857
item17	17.672772	6.300	.699	.494	.844

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
21.828867	10.267	3.2042362	5

Figure 4 ▪ Results of RELIABILITY analysis of the five items loading uniquely on Factor 1, with EMcorr (the matrix dataset of EM correlations) as input.

included as part of the online supplement. As shown earlier, we used 1365 (the listwise N for our items) as the nominal sample size in our EMcorr and EMcov datasets. We then made a copy of EMcorr, called it EMcorrN10, and changed the nominal sample size to 10. When we re-ran our EFA and reliability analyses with EMcorrN10 as input, all parts of the output that did not entail inference were identical to the earlier analyses with the nominal $N = 1365$.

In light of that observation, it really doesn't matter what nominal sample size is handed to the !PrepEMdata2 macro when no inference is involved. But regardless of the nominal sample size one chooses, we recommend thorough reporting of how much data is missing, and suggest including the following items: The total number of cases in the data file, the number of valid cases for each variable, and the listwise (or complete case) sample size. A matrix of pairwise sample sizes might also be useful.

Conclusion

Missing data is a frequent problem for users of EFA and reliability analysis. The SPSS FACTOR procedure has three methods of handling missing data (listwise deletion, pairwise deletion and mean substitution), but all of them have well-known flaws. Graham (2009) suggests that EM covariances (and by extension EM correlations) provide a far superior basis for both of these types of analyses. The SPSS MVA procedure generates perfectly good vectors of EM means and SDs and matrices of EM covariances and correlations. However, MVA has no /MATRIX sub-command, and so the user who wishes to use a matrix dataset of EM correlations as input for FACTOR or RELIABILITY must first get the EM estimates from the output viewer into an appropriately formatted matrix dataset. One way to do so is to use OMS followed by a series of data management steps. The two !PrepEMdata macros described in this article facilitate that process. We hope that SPSS users who perform EFA or reliability analysis with missing data will find them useful.

References

Acock, A. C. (2005). Working with missing values. *Journal of Marriage and Family*, 67, 1012-1028.

Allison, P.D. (2012). Handling missing data by maximum likelihood. Keynote presentation at the SAS Global Forum, April 23, 2012, Orlando, Florida. Retrieved from http://www.statisticalhorizons .com/wp-content/uploads/MissingDataByML.pdf.

Donders, A. Rogier T., van der Heijden, Geert J.M.G., Stijnen, T., & Moons, K. G. M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, 59, 1087-1091.

IBM. (2011a). Help file for the MVA command. Retrieved from http://pic.dhe.ibm.com/infocenter/ spssstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Fsyn_mva.htm.

IBM. (2011b). Help file for the OMS command. Retrieved from http://pic.dhe.ibm.com/infocenter/ spssstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Fidh_idd_full_oms_gui.htm.

IBM. (2011c). Help file for the MCONVERT command. Retrieved from http://pic.dhe.ibm.com/infocenter/ spssstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Fsyn_mconvert_overview.htm.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annu. Rev. Psychol.*, 60, 549–576.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177.

Truxillo, C. (2005). Maximum likelihood parameter estimation with incomplete data. *Proceedings of the Thirtieth Annual SAS(r) Users Group International Conference.* Retrieved from http://www2.sas.com/ proceedings/sugi30/111-30.pdf.

UCLA Statistical Consulting Group. (n.d.) Annotated SPSS Output, Factor Analysis. Retrieved from http://www.ats.ucla.edu/stat/spss/output/factor1. htm on 5-Feb-2014.

von Hippel, Paul T. (2004). Biases in SPSS 12.0 missing value analysis. *The American Statistician*, 58(2), 160-164.

Appendix: Macro Definitions for !PrepEMdata1 and !PrepEMdata2

These macro definitions can also be found in syntax file *PrepEMdata_macros.SPS*. See syntax file *EFA_via_EM_correlations.SPS* for examples of how to use the macros. Both syntax files and output generated by the second syntax file can be downloaded from the journal website (http://www.tqmp.org/) or via the following links:
https://sites.google.com/a/lakeheadu.ca/bweaver/Home/statistics/files/PrepEMdata_macros.sps
https://sites.google.com/a/lakeheadu.ca/bweaver/Home/statistics/files/EFA_via_EM_correlations.sps
https://sites.google.com/a/lakeheadu.ca/bweaver/Home/statistics/files/EFA_via_EM_correlations_full_output.pdf

```
* First macro definition .
DEFINE !PrepEMdata1 ()

DATASET DECLARE  @1.
DATASET DECLARE  @2.
```

```
OMS
  /SELECT TABLES
  /IF COMMANDS=['MVA'] SUBTYPES=['CMPB_ COMPARE MEANS','CMPB_ COMPARE STANDARD DEVIATIONS']
  /DESTINATION FORMAT=SAV NUMBERED=TableNumber_ OUTFILE='@1'
   VIEWER=YES.

OMS
  /SELECT TABLES
  /IF COMMANDS=['MVA'] SUBTYPES=['EOUT_ EM CORRELATIONS']
  /DESTINATION FORMAT=SAV NUMBERED=TableNumber_ OUTFILE='@2'
   VIEWER=YES.

!ENDDEFINE.

* ------------------------------------------------------------------------- .

* Second macro definition .
* Macro arguments (both required):
* 1) Vars:  A list of variables to use (can include the keyword TO);
* 2) N:  The nominal sample size.

DEFINE !PrepEMdata2
 ( Vars = !CHAREND('/') /
   N = !CMDEND )

OMSEND.

DATASET ACTIVATE @1.    /* Means & SDs .
SELECT IF (TableNumber_ EQ 1) OR (Var1 EQ "EM").  /* Keep 3 rows.
COMPUTE @RowType = $casenum. /* 1=N, 2=Means, 3=SDs.

* Fill in the nominal sample size on row 1.
DO IF @RowType EQ 1.
DO REPEAT v = !Vars.
- COMPUTE v = !N.
END REPEAT.
END IF.
EXECUTE.

DELETE VARIABLES TableNumber_ to Var1.

* Fill in top half of the EM correlation matrix.
DATASET ACTIVATE @2. /* EM correlations */.

MATRIX.
GET M  / file = * / variables = !Vars / missing=0 .
COMPUTE M = M + T(M).

* At this point, the terms on the main diagonal are twice
* as large as they should be, so divide them by 2.

CALL SETDIAG(M,DIAG(M)/2).
MSAVE M /TYPE=CORR /OUTFILE=* /VARIABLES=!Vars.
END MATRIX.

DATASET NAME @EMcorr.
```

```
* Now stack datasets to get a matrix dataset of the
* EM correlations.  Start with the matrix of EM
* correlations, because it has the ROWTYPE_ and
* VARNAME_ variables in the correct position.

ADD FILES
 FILE = @EMcorr /
 FILE = @1

.

EXECUTE.

DATASET NAME EMcorr.
DATASET ACTIVATE EMcorr.

* Close datasets that are no longer needed.
DATASET CLOSE @1.
DATASET CLOSE @2.
DATASET CLOSE @EMcorr.

* Fill in the values of ROWTYPE_ where it is missing.
DATASET ACTIVATE EMcorr.
IF @RowType EQ 1 ROWTYPE_ = "N".
IF @RowType EQ 2 ROWTYPE_ = "MEAN".
IF @RowType EQ 3 ROWTYPE_ = "STDDEV".
EXECUTE.

DELETE VARIABLES @RowType.

* Use MCONVERT to create a matrix dataset
* containing the EM covariances.
DATASET DECLARE EMcov.
DATASET ACTIVATE EMcorr.
MCONVERT matrix=out(EMcov).
DATASET ACTIVATE EMcov.

!ENDDEFINE.
```

Citation

Weaver, B., & Maxwell, H. (2014). Exploratory factor analysis and reliability analysis with missing data: A simple method for SPSS users. *The Quantitative Methods for Psychology*, *10* (2), 143-152.

Received: 2/04/14 ~ Accepted: 2/05/14