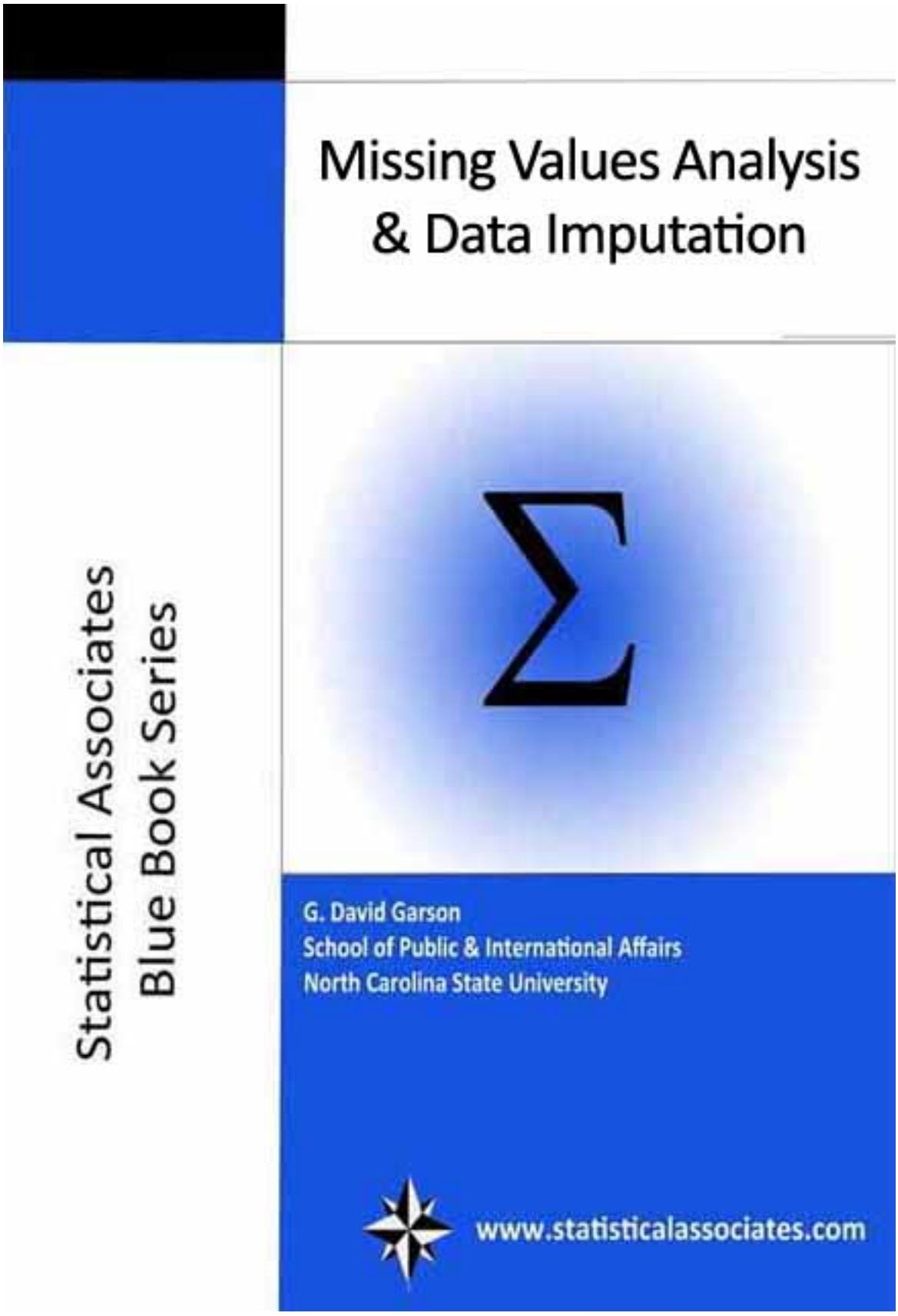


PREVIEW OF FIRST 21 PAGES

ISBN-10: 1626380333
ISBN-13: 978-1-62638-033-2

Contact:

G. David Garson, President
Statistical Publishing Associates
274 Glenn Drive
Asheboro, NC 27205 USA

Email: sa.publishers@gmail.com
Web: www.statisticalassociates.com

Table of Contents

Missing Values Analysis and Data Imputation

Overview

Proper handling of missing values is important in all statistical analyses. Improper handling of missing values will distort analysis because, until proven otherwise, the researcher must assume that missing cases differ in analytically important ways from cases where values are present. That is, the problem with missing values is not so much reduced sample size as it is the possibility that the remaining data set is biased. The imputation of values where data are missing is an area of statistics which has developed much since the 1980s.

Some authors disparage imputing values for a dependent variable on the reasoning that this reduces the variance of the dependent variable, biases estimates, and incorporates noise in the data into imputed dependent values. However, other statisticians (ex., Landerman, Land, & Pieper, 1997; Little & Rubin, 2002) argue that imputation of dependent variables "is essential for getting unbiased estimates of the regression coefficients" (Allison, 2001: 52).

Statistical objections can be raised about any of the methods which might be used for data imputation. Missing data are a form of measurement error. As such missing data may both bias the sample and attenuate effect sizes. Data imputation may reduce bias but also may introduce systematic regularities in the data arising from the prediction method.

Different statistical packages handle missing values analysis and data imputation in different ways. All contain options and variations which go beyond the introductory topics covered in this volume.

SPSS

The SPSS add-on module "Missing Value Analysis" (MVA) has long supported several imputation algorithms, the most popular being expectation maximization (EM). MVA is also useful for analyzing and understanding patterns of missingness in the data, as will be illustrated further below.

Since SPSS 17 a separate module, "Multiple Imputation," has supported the newer, preferred MI estimation method. Both are discussed below. Note that maximum likelihood data imputation, an EM method, can also be implemented in AMOS, the structural equation program supported by SPSS. As with SAS and Stata, the default MI method in SPSS is based on Markov Chain Monte Carlo ("fully conditional explanation") approaches developed by Rubin (1987, 1996; Little & Rubin, 2002).

SAS

In SAS, PROC MI performs multiple imputation and outputs multiple imputed datasets, using algorithms which depend on patterns of missingness. A large number of options are available. PROC MI became available with SAS 8.2. Its default method of imputation is the Markov Chain Monte Carlo (MCMC) method, based on the multivariate normal model and associated with the work of Rubin (1987, 1996; Little & Rubin, 2002).

A companion module, PROC MIANALYZE, is used after PROC MI and after usual SAS procedures (ex., PROC REGRESS). Taking the multiple imputed datasets from PROC MI and results of SAS procedures, PROC MIANALYZE can generate valid statistical inferences about the parameters associated with these procedures. In essence, PROC MIANALYZE combines results from analyses of each of the m imputed datasets to compute standard errors and significance tests in a more valid manner.

In SAS, handling missing values is a three-step process:

1. PROC MI generates m complete, imputed data sets.
2. The m data sets are analyzed using usual statistical analyses.
3. The results are combined by PROC MIANALYZE to produce more valid statistical inferences.

SAS also has the PROC HPIMPUTE procedure, where the first two letters stand for "high performance". It can perform large-scale imputations with precision when run in SAS High-Performance Server Distributed Mode. PROC HPIMPUTE is not covered here.

Stata

Stata does not have a separate missing value analysis module, but the suite of commands which accompanies its multiple imputation procedure is very extensive very extensive, based on its `mi impute` command. Its most common multivariate method of imputation is the "mvn" method based on the multivariate normal model and work by Rubin (1987, 1996; Little & Rubin, 2002). This is closely related to the default MCMC method used by SAS.

Data examples in this volume

The example dataset used in this, with versions for SPSS (.sav), SAS (.sas7bdat), and Stata (.dta), is a survey data set generously provided by survey researcher Michael D. Cobb of North Carolina State University, here carrying the root name "cobb_survey2". Data pertain to the 2008 U.S. presidential elections.

Variables used in examples below include those listed below. All are binary.

- debwatch Did R watch debate?
- rvpos R's position on vouchers
- rposvst R strength of attitude on vouchers
- gsupv Does R think Gore supports vouchers?
- bsupv Does R think Bush supports vouchers?
- rposuhc R position on universal health care
- rposuhcs R strength of attitude on health care
- rposdp R position on death penalty

These data are selected and, while suitable for pedagogical use, are not intended for substantive research.

- Click here to download cobb_survey2.sav for SPSS.
- Click here to download cobb_survey2.sas7bdat for SAS.
- Click here to download cobb_survey2.dta for Stata.

Key Concepts and Terms

Causes of non-response

Data may be missing for three reasons, as Kalton (1983) noted:

1. *Non-coverage:* the sample is not representative of the population to which the researcher wishes to generalize. Some portions of the intended population were not covered.
2. *Subject non-response:* Also called unit non-response, some subjects are included in the sample but do not provide any information, even to demographic items.
3. *Item non-response:* Some subjects only give information for some of the items.

Data imputation addresses the third type of missing data. For missing values, data imputation estimates what a subject's response would have been, based on the subject's other responses and responses of other subjects similar to the given subject. Data imputation is relevant to missingness due to item non-response but not due to non-coverage or subject non-response.

Item non-response

Item non-respondents are subjects who answer some but not all items on the instrument. Unit non-respondents, in contrast, do not answer any items and must be dropped from analysis.

Item non-response arises for many reasons, including those listed below.

- Fatigue with the instrument
- Sensitivity of the item
- Interruptions while taking the survey
- Information is unknown or not readily available
- Item is not applicable
- The item is ambiguous, encompasses two dimensions, or has other item validity issues
- In face to face surveys, the interviewer skips or fails to record an item
- In multi-stage instruments, absence of the subject at one or more stages

- Loss of data due to transmission problems (ex., power fails at the respondent's home in the middle of an online survey)
- In scientific and medical studies, problems with monitoring and recording equipment
- Loss of data during coding and storing
- Items are missing by design (ex., branching surveys skip some items for some respondents based on previous responses)

Item non-response forces the researcher to decide whether to leave cases with missing data out of analysis when data are missing for a variable being analyzed, or whether a value should be imputed for the case and the blank replaced by the imputed value. Similar issues arise with archival data, where the researcher may find no recorded data for certain values of certain records. Imputation is usual when non-response is not large (not over 20% is cited by some authors; Bagheri et al. (2014) say not over 50%) and other variables in the dataset have the capacity to predict missingness.

Listwise deletion of cases with missing values

There is no simple decision rule for whether to leave data as they are, to drop cases with missing values, or to impute values to replace missing values. Generally the first two choices amount to the same thing since most statistical packages by default drop cases listwise when a missing value is encountered. Listwise deletion (LD) is sometimes called the "complete case" (CC) method because only cases with no missing values are analyzed.

When data are missing completely at random (MCAR, discussed below), the complete cases are a random subset of all cases. While estimates may be unbiased, the smaller sample size of the CC data involve diminished power, larger standard errors, and increased chance of Type II error (false negatives). Moreover, MCAR is rare.

When the number of cases with missing data is small (typically but arbitrarily defined as less than 5% in larger samples, though there is no accepted consensus on the cutoff what a "larger sample" is), it is common simply to drop these cases from analysis, which, as noted above, most statistical programs do by default in the process of listwise deletion.

In an earlier period, it was sometimes suggested that the researcher might opt for dropping cases listwise rather than imputing values since imputation can distort significance and effect size coefficients (see Kalton and Kasprzyk, 1982). On the other hand, smaller sample size due to LD may mean lower power and increased chance of Type II error (false negatives) due to larger standard errors.

Today, imputation is almost universally recommended over LD. Thus Graham (2009: 559) wrote, "when the MAR assumption is violated, the violation affects the old procedures (e.g., listwise deletion) as well, and typically this violation has greater effect on the old procedures. In short, MI and ML methods are always at least as good as the old procedures". Imputation, by using as much information as is available, results in best-guess estimates of significance and effect size coefficients and is today the preferred approach. While imputation assumes data are missing at random (MAR), discussed below, imputation is as good as LD even when MAR assumptions are violated.

Dropping cases vs. data imputation is not a dichotomous choice. The researcher may run all analyses twice, once with data imputation and once with complete case analysis using LD. Often the same substantive results occur in each method. If they do not, the dual approach enables the researcher to determine what difference data imputation may make. Regardless, data imputation is preferred as it tends to reduce estimation bias.

Types of Missingness

Missing completely at random (MCAR)

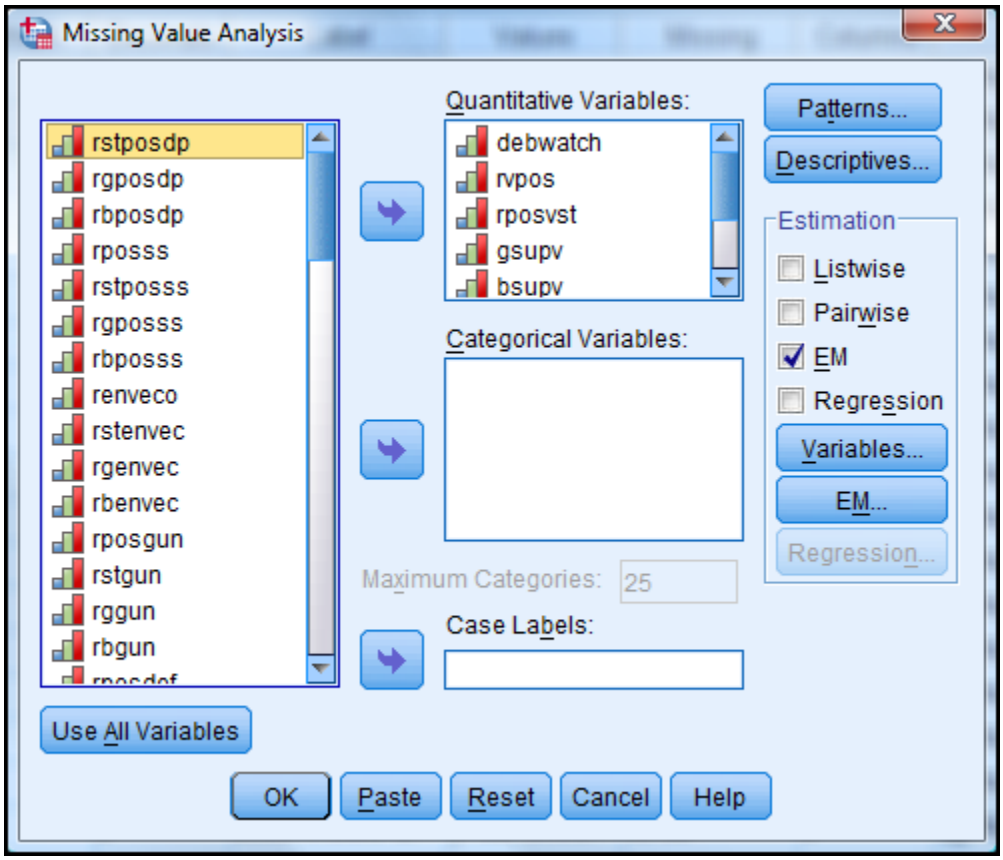
MCAR exists when missing values are randomly distributed across all observations. Missingness in given variable does not depend on any other variable, whether observed or unobserved. MCAR can be confirmed by dividing respondents into those with and without missing data, then using t-tests of mean differences on income, age, gender, and other key variables to establish that the two groups do not differ significantly on any variable in the model, including the dependent variable. If missing data are MCAR in a sufficiently large sample, cases with missing values may be dropped listwise from the analysis without biasing the estimates. If dropping MCAR cases appreciably reduces sample size, however, standard errors will be increased, increasing the chance of Type II error (false negative inferences; statistical power is diminished). MCAR, however, is unusual.

Little's MCAR test

Little's MCAR test is the most common test for missing cases being missing completely at random. If the p value for Little's MCAR test is not significant, then the data may be assumed to be MCAR and missingness is assumed not to matter for the analysis. Listwise deletion of observations with missing values is appropriate, provided the number of missing values is not very large. The seminal article on MCAR is Little (1988).

Little's MCAR test in SPSS

The SPSS Missing Values Analysis (MVA) option supports Little's MCAR test, which is a chi-square test for missing completely at random. In SPSS, select Analyze > Missing Value Analysis and check EM as the estimation method. Little's test will be printed below the EM Means, EM Covariances, and EM Correlations tables (and will have the same value in each), as illustrated below.



The data in the example below are not missing completely at random since Little's MCAR test is significant at the .006 level.

EM Means[a]

debwatch	rvpos	rposvst	gsupv	bsupv	rposuhc	rposuhcs	rposdp
1.4290	1.5878	1.4616	1.6587	1.2972	1.2390	1.3095	1.3111

a. Little's MCAR test: Chi-Square = 285.027, DF = 228, Sig. = .006

If data are MCAR, then the researcher may choose listwise or pairwise deletion of cases. If data are not MCAR, the usual recommendation is to impute missing values unless missing at random (MAR),discussed below, applies and meets the needs of the researcher.

Little's MCAR test in SAS

SAS does not support Little's MCAR test directly, but at least two SAS macro programs to do so may be found on the web. These are those by:

Yuming Ning, Yale Pepper Center (2008)
http://grasp.med.yale.edu/mediawiki/images/0/0b/MCARProgramNing.txt

Craig K. Enders, to accompany Enders (2010)
http://www.appliedmissingdata.com/littles-mcar-test.sas

Little's MCAR test in Stata

Though Stata does not directly support Little's MCAR test, a third-party add-on module for this purpose is available from Cheng Li (2013) as the `mcartest` program.

To install from Stata, type `findit mcartest`, then in the window which comes up, click on `st0318 from http://www.stata-journal.com/software/sj13-4`,

To run for the example data, use these commands:

`. use c:\\data\cobb_survey2.dta` (but substitute the directory where you have downloaded the sample dataset).

`. mcartest debwatch rvpos rposvst gsupv bsupv rposuhc rposuhcs rposdp, emoutput nolog` (this runs Little's test for the same eight variables as in the SPSS example above, using the `emoutput option`, which requests display of intermediate output from EM estimation. The nolog option simply suppresses display of the EM iteration log.)

Output appears as below. Little's MCAR test is significant at the .0059 level, indicating that the data are not missing completely at random. To be MCAR, Little's test should be non-significant.

```
. mcartest debwatch rvpos rposvst gsupv bsupv rposuhc rposuhcs rposdp, emoutput nolog

Expectation-maximization estimation      Number obs           =       338
                                         Number missing       =       561
                                         Number patterns      =        46
Prior: uniform                           Obs per pattern: min =         1
                                                          avg =  7.347826
                                                          max =       101

Observed log likelihood =  653.79412 at iteration 29
```

	debwatch	rvpos	rposvst	gsupv	bsupv	rposuhc	rposuhcs	rposdp
Coef								
_cons	1.428994	1.587747	1.461637	1.659418	1.297729	1.238989	1.309476	1.311127
Sigma								
debwatch	.2449582	-.0117876	.0291007	-.0333665	.018719	-.0110355	.008198	-.0124369
rvpos	-.0117876	.2412698	-.0289247	.0213159	-.0069713	-.0075369	-.0093574	.0164885
rposvst	.0291007	-.0289247	.2481825	-.027518	.0244348	-.0156531	.0127134	.0035276
gsupv	-.0333665	.0213159	-.027518	.2194688	-.1708497	.0109374	.0193596	.008678
bsupv	.018719	-.0069713	.0244348	-.1708497	.2024051	-.018014	-.0170322	-.0116049
rposuhc	-.0110355	-.0075369	-.0156531	.0109374	-.018014	.1817712	.0349505	-.0076662
rposuhcs	.008198	-.0093574	.0127134	.0193596	-.0170322	.0349505	.2135075	-.0030165
rposdp	-.0124369	.0164885	.0035276	.008678	-.0116049	-.0076662	-.0030165	.2146286

```
Little's MCAR test

Number of obs       = 338
Chi-square distance = 285.3742
Degrees of freedom  = 228
Prob > chi-square   = 0.0059
```

The `mcartest` command has additional options not discussed here, including:

noconstant suppresses constant term.
unequal specifies that unequal variances between missing-value patterns be allowed. By default, the test assumes equal variances between different missing-value patterns.
emoutput specifies that intermediate output from EM estimation be displayed.

As with other Stata commands, after installation of `mcartest, help mcartest` brings up a screen of additional information on the command.

Missing at random (MAR)

The phrase "missing at random" is misleading since MAR data reflect a systematic rather than random pattern of missingness. Data are missing at random (MAR) when (1) not MCAR, indicated by Little's MCAR test being significant; and (2) missingness may be predicted by other observed variables and does not depend on any unobserved variables. If missingness may be well predicted from observed variables, then multiple imputation (MI) is appropriate. In fact, MI assumes MAR as defined here. Listwise deletion will introduce bias if data are MAR. MAR is much more common than MCAR. Exploratory testing for MAR is discussed below.

To elaborate, for MAR data, missingness is not independent of the values of other variables in the model but is predictable by them. This implies that for MAR to be demonstrated, it must be assumed that missingness does not depend on unobserved variables. This assumption may be wrong (this is the model specification problem). If missingness is not predictable from observed variables, data are "missing not at random" (MNAR, discussed below).

MAR is a spectrum, depending on how much of missingness can be explained by other observed variables. A pure MAR example would be if there were test scores, test1 and test2, representing scores on two sequential tests. If students scoring 90 or greater on test1 were excused from test2, and if there were no other dropouts, missingness on test2 would be completely determined by the test1 variable. At the other end of the spectrum, in a large dataset it might happen that missingness on a given variable was significantly related to another observed variable (hence not MCAR) but the relation was so trivial in effect size that missingness could not be predicted from that variable. The point on this spectrum where prediction ceases to be useful is the point separating MAR from MNAR.

Missing not at random (MNAR)

Missing not at random (MNAR), also called non-ignorable missingness, is the most problematic form. It exists when missing values are neither MCAR nor MAR. This happens when missingness depends at least in part on unobserved variables (which is why observed variables fail to predict missingness, making data not MAR). Under MNAR conditions, variables in the dataset are inadequate predictors of missingness because the variable with missing cases is insufficiently correlated

with other variables in the dataset, undermining the effectiveness of the usual imputation methods, including multiple imputation (MI).

One approach to non-ignorable missingness is to impute values based on data otherwise external to the research design, as, for instance, estimating race based on Census block data associated with the address of the respondent, but while missingness cannot be ignored, there is no well-accepted method of dealing with non-ignorable missingness.

See Muthén, Asparouhov, Hunter, & Leuchter (2011) on analyzing MNAR data using MPlus statistical software.

Multiple imputation

Multiple imputation is the currently prevailing method of estimating missing values. Though it may be implemented by various methods, by default in SPSS, SAS, and Stata, it uses Markov Chain Monte Carlo (MCMC) simulation methods, which are probabilistic in nature. Multiple implementation involves three steps:

1. Creating multiple datasets in which missing values have been imputed.
2. Pooling the estimates from the multiple imputed datasets.
3. Running the pooled data on statistical procedures such as linear or logistic regression.

Van Buuren (2012: 27) stated, "Nowadays multiple imputation is almost universally accepted, and in fact acts as the benchmark against which newer methods are being compared."

When imputation should not be used

While data imputation is routinely used for missing data, there are circumstances under which it should not be used.

- If data are MCAR, imputation may not be not needed.
- If missingness is due to unmeasured variables related to the dependent variable, data are MNAR and should not be imputed.
- Imputation assumes data are MAR and should not be used with sparse data. Sparse data occur when missingness is non-random, such as a shopping cart survey of items purchased (coded 1) or not purchased (coded

0), because the null response (0) is non-random, due to unmeasured factors possibly not even known to the shopper.

- Imputation should not be used to impute all the data for a subject.
- Imputation should not be used for a missing value for a given observation if that observation is also missing values on predictively critical variables in the imputation model. While this is difficult to check for each value to be imputed, a table of missing value patterns will show how many cases missing on a given variable also have missing values on other variables. In some cases this may lead a researcher to reject imputation.
- Imputation should not be used if over 50% of data are missing (some authors use lower cutoffs, such as 20%).
- Imputation is used with cross-sectional or historical data and is not appropriate for imputing future data in a time series.
- Use of imputation is suspect if it generates values outside valid ranges.
- Imputation based on a single pass is not acceptable due to the probabilistic nature of imputation. While as few as 3 – 5 imputations may suffice for reliability, today 20 – 100 or more imputations are usual.

When data are distributed, the original as well as the imputed dataset should be provided, so that researchers may have the opportunity to experiment with alternative forms of data imputation. Accompany documentation should describe fully the imputation methods and options taken by the researcher.

Summary

- MCAR: Data are MCAR if missingness on any variable in the analytic model is unrelated to the values of any other variable in the model and there is no autocorrelation (missingness in Y is unrelated to values of Y itself). Little's MCAR test should be non-significant. Listwise deletion is appropriate provided the number of deleted cases is not large.
- MAR: Missingness on any variable in the analytic model may be explained solely using observed variables in the model. Unobserved variables do not explain missingness in any variable in the model. Multiple imputation (MI) is appropriate if the number of missing values is not high and if missingness may be predicted from observed variables. (There is no agreed cutoff for how high is too high but at some point the "best guess" reflected by MI ceases to be useful.)

- MNAR: Missingness is not MCAR but observed variables in the model cannot well explain missingness. There is no well-accepted remedy for MNAR.

Testing for missing at random (MAR)

Overview

Whether data are missing at random (MAR) cannot be determined with any simple test. Ultimately, proving conclusively that data are MAR would require showing the values which are missing are distributed randomly but that is impossible as missing values are, of course, unknown. For this reason, Schafer & Graham (2002: 153) state, "In general, there is no way to test whether MAR holds in a data set, except by obtaining follow-up data from non-respondents or by imposing an unverifiable model."

"Testing for MAR" instead refers to exploratory tests to see if data are consistent with what is implied by "missing at random" and with imputing MAR data. In checking the effects of missingness, some exploratory tests require creation of a set of dummy variables for missingness for each variable of interest, coded 0 = not missing on the given variable, 1 = missing. Note that the researcher may wish to explore whether auxiliary variables not in the original analytic model may also predict missingness, and if so, add them prior to imputation.

Some of the common exploratory procedures are outlined below.

1. *Little's MCAR test:* If data can be shown to be MCAR, they are not MAR.
2. *Scale independence*: If scales, indexes, or latent variables can be shown to be uncorrelated with the set of missingness dummy variables, this is consistent with missingness in the scales, indexes, or latent variables being MCAR, not MAR.
3. *Significance tests of missingness*: If missingness can be shown to be correlated significantly with the values of one or more other variables in the dataset, missingness may be predicted to some degree, consistent with the requirements for imputing MAR data. This may be accomplished using independent samples t-tests, described below in the SPSS section. Other tests, such as the significance of logistic coefficients when missingness dummy variables are predicted in logistic regression described below in the

SAS section, are available for the same purpose. Categorical variables may be cross-tabulated with missingness dummy variables and a chi-square test may be applied.

4. *Effect size in tests of missingness*: While significant relationship of other measured variables to missingness in a given variable is a prerequisite for multiple imputation of that variable, it is not sufficient. The stronger the relation of other measured variables to missingness in the given variable, the better the resulting predictions which result in imputed values for the given variable. Odds ratios from logistic regressions predicting the missingness dummy for a given variable are one measure of effect size. Others include measures of association for the crosstabulation of categorical variables with missingness dummies. Even the absolute t values in separate variance t-tests are reflect effect size (though conflated with sample size, sample size will be the same for all t-tests, allowing comparison). Unfortunately, there is no rule on how strong effect size should be to proceed with MI, which is why researchers often allow significance to suffice to establish MAR and thus to justify MI.
5. *Residual autocorrelation for longitudinal data*: If residuals in longitudinal data are autocorrelated, missing values in one period may be predicted to some degree by data from prior periods, consistent with the requirements for imputing MAR data. The Durbin-Watson test is the usual test of residual autocorrelation for time series data.
6. *Simulation*: Although tedious and therefore little used, it is possible to take the complete cases dataset resulting from listwise deletion, then randomly recode a proportion of values to be missing (the same proportion as in the original dataset), then use alternative methods (ex. listwise deletion, multiple imputation, expectation maximization) to see which method's imputations are closest to the true values prior to random recoding and to see how large the residuals are. More rigorously, this whole process may be repeated multiple times so that stable averages are obtained. The simulation method assumes that the dataset with values recoded as missing has the same covariance matrix as the hypothetical dataset with all cases and no missing values - not necessarily a valid assumption. On an exploratory basis, however, the simulation method may provide evidence on how to treat the missing data problem. If the MI and EM methods are superior to the LD method, this is circumstantial evidence that the data were MAR.

7. *Theory and literature*: In some cases, the assumption that data are MAR may be supported by theoretical arguments or by studies in the literature which have demonstrated MAR for similar models.

Little's MCAR test

Little's MCAR test is performed as described above. If data are MCAR, no further steps are necessary. If Little's MCAR test is significant (meaning data are not MCAR), other exploratory tests of missingness at random are warranted.

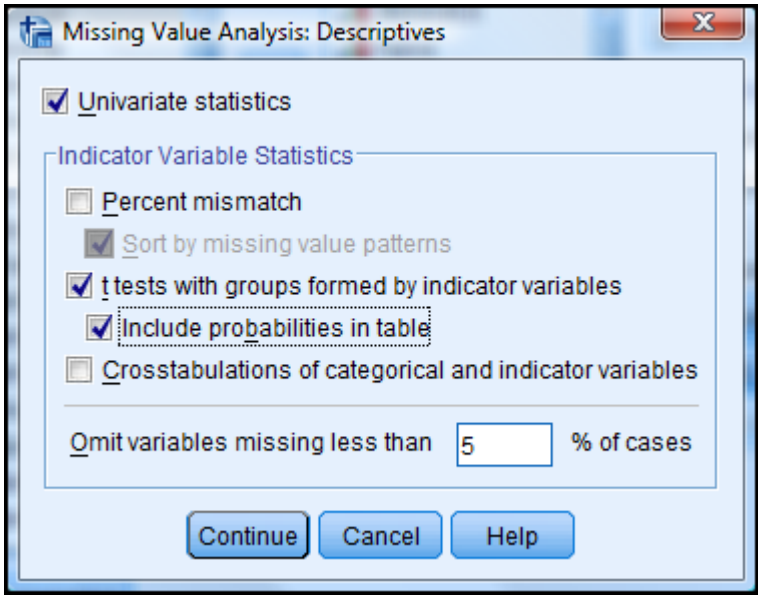
Testing missingness for independence

The usual procedure for testing missingness for independence may be described in terms of these conceptual steps:

1. A binary variable for 0 = not missing, 1 = missing, is created for each variable of in the model. This may or may not be done "behind the scenes" by statistics software. SPSS performs this step automatically. SAS and Stata do not. These binary variables are the missingness indicators.
2. A statistical test is performed to see if other modeled variables are significantly associated with the binary missingness indicator variable associated with each measured variable in the statistical model. Common tests are t-tests, chi-square tests, or tests of significance in binomial logistic regression. For continuous variables, t-tests and logistic coefficient tests are commonly used. For categorical variables, the researcher may cross-tabulate the missingness indicator with the categorical variable and use chi-square to determine significance. SPSS performs this step automatically. SAS and Stata do not.
3. Tests which return a finding of significance indicate that missingness in the variable of interest is significantly correlated with a measured variable in the model. This implies data are not MCAR but rather are consistent with MAR. If the level of association is sufficient to predict missingness, this also is consistent with imputation of MAR data. Even when the level of association is moderately imperfect, data are commonly assumed to be MAR and MI is commonly used. If the level of association is insufficient, data are effectively MNAR as MNAR implies that missingness cannot be well predicted using observed variables.

Testing for MAR in SPSS

The SPSS missing values analysis (MVA) module is accessed in SPSS by selecting Analyze > Missing Value Analysis. In the main MVA dialog illustrated above, click the "Descriptives" button, and then in the "Descriptives" dialog illustrated below, check t-tests and "Include probabilities" (t-test output is not the default).



The MCAR test

The SPSS MVA module will generate Little's MCAR test, shown below the "EM Means" table as described above. For the example data, the researcher may rule out the possibility that data are MCAR since Little's test is significant. The researcher moves on to testing for MAR.

Separate variance t-tests

SPSS will generate a table of "Separate Variance t Tests," illustrated in partial output below.

END OF PREVIEW OF FIRST 21 PAGES

To buy the Kindle version for $6.00, click here.

MISSING VALUES ANALYSIS & DATA IMPUTATION

Overview

An illustrated tutorial and introduction to missing values analysis and data imputtion using SPSS, SAS, and Stata. Suitable for introductory graduate-level study.

The 2015 edition is a major update to the 2012 edition. Among the new features are these:

- Was 40 pages with 25 figures, now 113 pages with 51 figures
- Now covers Stata and SAS as well as SPSS
- Totally revised throughout
- New coverage comparing MI and EM estimation methods
- New coverage on testing for MI convergence
- New coverage on exploratory testing for missing at random
- New section on when MI should not be used
- Twelve FAQ sections
- Links to all datasets used in the text.

The full content is now available from Statistical Associates Publishers. Click here.

Below is the unformatted table of contents.

Statistical Associates Publishing
Blue Book Series

Canonical Correlation
Case Studies
Cluster Analysis
Content Analysis
Correlation
Correlation, Partial
Correspondence Analysis
Cox Regression
Creating Simulated Datasets
Crosstabulation
Curve Estimation & Nonlinear Regression
Delphi Method in Quantitative Research
Discriminant Function Analysis
Ethnographic Research
Evaluation Research
Factor Analysis
Focus Group Research
Game Theory
Generalized Linear Models/Generalized Estimating Equations
GLM (Multivariate), MANOVA, and MANCOVA
GLM (Univariate), ANOVA, and ANCOVA
Grounded Theory
Life Tables & Kaplan-Meier Survival Analysis
Literature Review in Research and Dissertation Writing
Logistic Regression: Binary & Multinomial
Log-linear Models,
Longitudinal Analysis

Missing Values & Data Imputation
Multidimensional Scaling
Multiple Regression
Narrative Analysis
Network Analysis
Neural Network Models
Nonlinear Regression
Ordinal Regression
Parametric Survival Analysis
Partial Correlation
Partial Least Squares Regression
Participant Observation
Path Analysis
Power Analysis
Probability
Probit and Logit Response Models
Research Design
Scales and Measures
Significance Testing
Social Science Theory in Research and Dissertation Writing
Structural Equation Modeling
Survey Research & Sampling
Testing Statistical Assumptions
Two-Stage Least Squares Regression
Validity & Reliability
Variance Components Analysis
Weighted Least Squares Regression

Statistical Associates Publishing
http://www.statisticalassociates.com
sa.publishers@gmail.com