FORS

explore.understand.share.

UNIL | Université de Lausanne

FORS
explore.understand.share.

Multiple Imputation of Missing Values: Why, how and the Do's and Don'ts

Methods and Research Meetings

Michael Ochsner

UNIL | Université de Lausanne

Outline

- What are missing data?
- Treatments of missing data
- Multiple Imputation: Theory
- Reservations against MI in the social sciences
- Examples with data of the ESS 2012
- Do's and Don'ts

FORS
explore.understand.share.

Missing Data

There are always missing values (Rubin, 1976):

- Due to respondents:
  - People don‘t answer single questions
  - People stop answering questions
  - Panels: Some respondents drop out
- Due to design:
  - Filter questions
  - Experiments, differential treatments
- Due to the researcher analysing the data:
  - Outliers

Missingness (1/2)

- For statistical analysis, the *distribution* of missingness matters
- It was ignored until Rubin‘s 1976 article
- It is still ignored by a vast majority of researchers
- Rubin (and Little) introduced missingness as a *probabilistic* phenomenon

Missingness (2/2)

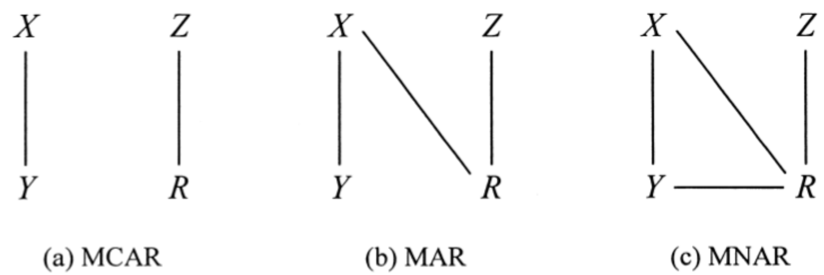
- Randomvariables *R* indicate missingness (for each variable with missing values, R is an indicator variable)
- *R* is also called *distribution of missingness* (Schafer & Graham, 2002)

Y	X	Z	R_Y	R_X	R_Z
3	.	4	0	1	0
.	4	5	1	0	0
4	5	8	0	0	0
4	7	1	0	0	0
.	.	7	1	1	0
6	.	7	0	1	0
10	7	10	0	0	0

- The relation between *R* and the data is crucial for the choice of treatment of missing values
- There is NO way of „no treatment“ of missing values if there are missings

FORS
explore.understand.share.

Three Types of Missing Values



Source:
Schafer & Graham 2002, S. 155-159

- Missing Completely at Random (MCAR)
  - Missingness does not depend on any variable related to the model
- Missing at Random (MAR)
  - Missingness may depend on a variable in the model but not on the variable with the missing value
- Missing not at Random (MNAR)
  - Missingness depends on the variable with the missing value

FORS
explore.understand.share.

Treatments of Missing Values: Older Methods

- Deletion methods
  - Listwise (Complete Case)
    - Assumption MCAR: Estimates unbiased, but inefficient
    - Small fraction of missing data: efficient, but check missing patterns
  - Pairwise Deletion
    - Assumption MCAR: Estimates unbiased, but inefficient; no SE’s
- Imputation methods
  - Mean substitution
    - decreases standard errors by a) reducing the variance and b) by artificially increasing N
    - Error rates are biased even under MCAR (also correlations)
  - Single imputation of conditional means or distributions
    - As complicated as full MI
    - Without correction of SE, Error rates are biased
    - Full MI has better properties with small additional effort

Multiple Imputation: Basics

- Assumption: MAR
- Basic idea:
  - Keep information of available data
  - Account for uncertainty
    - Add noise
    - Adjust the standard errors.
  - Independence on method of analysis
    - Two steps: Imputation and analysis
    - **However**: ≠ Independence of imputation model and **model** of analysis!!!
- Implementations
  - Chained equations (no sound statistical theory but good results under all conditions)
  - Multivariate normal regression (sound statistical theory but biased results if not multivariate normal)

FORS
explore.understand.share.

Multiple Imputation: Main steps

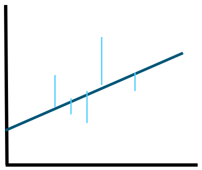
1. Each missing value is replaced by m>1 simulated values conditional on the other variables in the model
   - Result: m data sets where only the values vary that are missing in m=0

X_0	Y_0	X_1	Y_1	X_m	Y_m
2	.	2	3	2	6
.	.	4	2	6	3
6	5	6	5	6	5
.	2	5	2	6	2
8	.	8	2	8	4
2	4	2	4	2	4

2. Each of the m data sets is analysed by the same complete-data method
3. Combination of the estimates and calculation of the standard errors

Step One: Imputing missing values

- Imputing values conditional on all variables in model
  - Iterative process: Many imputations needed to get independent imputations → every 100th imputation is saved
  - Add random noise, i.e., add random residual



- Issues:
  - What is the right «m»?
    - Given 50% missings: m=2 is 90% efficient. m=10 is 95%
    - Some more sophisticated ways to choose m in White, Royston & Wood, 2011
  - What is the right model?
    - IMPORTANT: We **do not** want to find the correct value of the **person**. BUT to best predict missingness; the «true» joint distribution
    - Best way: add all variables.
    - Minimum: ALL variables used in the analysis, also the dependent variable

Step Two: Analyse m data sets

- Analyse all m>0 data sets using exactly the same method and model
  - All data sets have to be complete data sets!
- Advantage of MI:
  - Use the same data for more than one analysis
    - E.g. index as dependent variable:
      you can analyse each component using the same data separately
      and you can analyse the index/latent variable using the same data
    - Use different methods on the same data set

Step Three: Combine Estimates

- Point estimates: Just use the mean of the m point estimates (regression coefficients, means…)

$$\overline{Q} = m^{-1}\sum_{j=1}^{m}\hat{Q}^{(j)}$$

- Standard errors: Must be adjusted for uncertainty
  - Rubin's Rules (1987): Two parts of uncertainty
    - Within-imputation variance

$$\overline{U} = m^{-1}\sum_{j=1}^{m}U^{(j)}$$

    - Between-imputation variance

$$B = (m-1)^{-1}\sum_{j=1}^{m}[\hat{Q}^{(j)} - \overline{Q}]^2$$

      - Total variance:

$$T = \overline{U} + (1 + m^{-1})B$$

    - Inference: t with adjusted degrees of freedom

$$v = (m-1)\left[1 + \frac{\overline{U}}{(1+m^{-1})B}\right]^2$$

- R-squared needs also an adjustment (Harel, 2009)

Multiple Imputation: In Practice

- Statistically, MI and ML are proved to be superior to other treatments of missingness since 10-20 years.
  - Only in very esoteric circumstances they are inferior
  - Most older methods are almost always biased
- Modern computers and statistical software provide means at least since 10 years
- However: Almost nobody uses it in social sciences
  - EVEN WORSE: Most people discredit it informally
  - YOUR MAKING UP DATA!
  - WoS: 374 Papers with MI. Most methodological/clinical

Why so reluctant?

- Misunderstanding of the concept:
  - MI is **not** about finding the «true» answer for the missings
  - It's about finding the **right joint distribution** of the variables
    - Wrong applications:
      Researchers use models that explain the variable having missing values
      Income = job + education + gender + attitudes on taxes + age
      (e.g., Busemeyer et al., 2009)
    - This **introduces** bias!
    - It's never going to work as we have r-squares of 0.40 at best in the SocSci. We can't really explain the variables
- Reservations against simulation studies
  - «Well, the statisticians say… but in real life…»
    - Plenty of simple simulation studies (two artificial variables)
    - Very few studies using data situations related to practice (Eekhout et al., 2014)

Example with Data from the ESS 2012 for CH

- Real social sciences-question and variables
  - What influences the legitimacy of the state?
  - Legitimacy is an index
    indices are often used and are prone to issues with missings
    - Legitimacy: Formative index of 3 aspects
      - Legality: measured by two variables
      - Justification: measured by three variables
      - Consent: measured by three variables
  - 16 variables that (can) explain legitimacy
    - 2 binary, 1 ordinal, 13 quasi-continuous

Example Data: Introducing Missing Values

- ESS 2012 data for Switzerland
  - For the 24 variables used, 21% of the cases have a missing
  - Deleted them to get a complete data set
  - Population of our study: Swiss people who answer all questions in the ESS ☺
- Simulated missingness:
  - MCAR: randomly deleted values
  - MAR: defined which variables determine missing value
    - E.g., Vote: satisfaction with life is high, and satisfaction with government is high but not very high.
    - Plus 10% of the missings are random missings (reality check)
  - MNAR: missingness depends on variable
    - Plus 10% of the missings are random missings (reality check)

FORS
explore.understand.share.

Variables and missingness

Variable	Question	Variable	Question
Implvdm (5)	Important to live in Democracy	Dfprtalc	Parties offer alternatives
Dmcntov	How democratic is CH	Rghmgprc	Rights of minorities protected
Stflife (2)	Life Satisfaction	Votedirc (20)	Citizens have final say
Stfeco (2)	Satisfaction w/ economy	Dscrgrp (20)	Member of discriminatet Group
Stfgov (5)	Satisfaction w/ government	Vote (5)	Voted in last election
Stfdem	Satisfaction w/ democracy	Woman	Dummy for Women
Stfedu (2)	State of educ. System in CH	Gincdif	Gov. should reduce inequality
Stfhlth	State of health system in CH	Optfftr	Always optimistic about future
Rlgdgr (2)	How religious are you	Marital	Marital status
Fairelcc (5)	Elections are fair in CH	Hincfel	Feeling about HH income

FORS
explore.understand.share.

Missing Completely At Random

	Full (N=1174) Coef	(SE)	MI (N=1174) Coef	(SE)	CC (N=441) Coef	(SE)	Zeros (N=1174) Coef	(SE)
Implvdm (5)	0.03	(0.02)	0.04	(0.03)	0.08	(0.04)	0.02	(0.01)
Dmcntov	0.09**	(0.03)	0.09**	(0.03)	0.15**	(0.05)	0.12***	(0.03)
Stflife (2)	0.10***	(0.02)	0.10***	(0.02)	0.06	(0.04)	0.04**	(0.02)
Stfeco (2)	-0.01	(0.02)	-0.01	(0.02)	0.00	(0.03)	0.02	(0.02)
Stfgov (5)	0.24***	(0.02)	0.24***	(0.03)	0.26***	(0.04)	0.09***	(0.02)
Stfdem	0.11***	(0.03)	0.11***	(0.03)	0.08	(0.04)	0.19***	(0.03)
Stfedu (2)	0.02	(0.02)	0.02	(0.02)	0.04	(0.03)	0.01	(0.02)
Stfhlth	0.03	(0.02)	0.03	(0.02)	0.00	(0.03)	0.06**	(0.02)
Rlgdgr (2)	0.01	(0.01)	0.02	(0.01)	0.00	(0.02)	0.01	(0.01)
Fairelcc (5)	0.05*	(0.02)	0.04	(0.02)	0.04	(0.04)	0.04**	(0.01)
Dfprtalc	-0.04*	(0.02)	-0.04*	(0.02)	-0.06*	(0.03)	-0.02	(0.02)
Rghmgprc	0.04*	(0.02)	0.03	(0.02)	0.02	(0.03)	0.06**	(0.02)
Votedirc (20)	0.04*	(0.02)	0.05*	(0.02)	0.00	(0.03)	0.01	(0.01)
Dscrmiss (20)								
No							-0.09	(0.07)
Yes	0.12	(0.14)	0.02	(0.17)	0.00	(0.24)	0.02	(0.2)
Votemiss (5)								
Yes							0.34**	(0.12)
No	-0.36***	(0.07)	-0.35***	(0.07)	-0.30**	(0.11)	-0.06	(0.14)
Ineligible	0.00	(0.09)	0.01	(0.1)	0.05	(0.16)	0.33*	(0.15)
Woman	0.19**	(0.06)	0.17**	(0.06)	0.04	(0.1)	0.21**	(0.07)
Constant	0.80**	(0.28)	0.76**	(0.29)	1.09*	(0.46)	1.4***	(0.33)
R^2	0.43		0.43		0.40		0.34	
Adj. R^2	0.42		0.42		0.37		0.33	

* p<0.05; ** p<0.01; *** p<0.001

Black: unbiased | blue: estimate biased | pink: wrong inference | red: estimate biased and wrong inference

FORS
explore.understand.share.

Missing At Random

	Full (N=1174) Coef	(SE)	MI (N=1174) Coef	(SE)	CC (N=502) Coef	(SE)	Zeros (N=1174) Coef	(SE)
Implvdm (5)	0.03	(0.02)	0.02	(0.02)	0.07	(0.04)	0.00	(0.01)
Dmcntov	0.09**	(0.03)	0.09**	(0.03)	0.05	(0.04)	0.11***	(0.03)
Stflife (2)	0.10***	(0.02)	0.09***	(0.02)	0.05	(0.03)	0.07***	(0.02)
Stfeco (2)	-0.01	(0.02)	0.00	(0.02)	-0.03	(0.03)	0.05**	(0.02)
Stfgov (5)	0.24***	(0.02)	0.24***	(0.03)	0.25***	(0.04)	0.11***	(0.02)
Stfdem	0.11***	(0.03)	0.11***	(0.03)	0.16***	(0.04)	0.17***	(0.03)
Stfedu (2)	0.02	(0.02)	0.02	(0.02)	0.02	(0.03)	0.03	(0.02)
Stfhlth	0.03	(0.02)	0.03	(0.02)	0.04	(0.03)	0.03	(0.02)
Rlgdgr (2)	0.01	(0.01)	0.01	(0.01)	0.02	(0.02)	0.00	(0.01)
Fairelcc (5)	0.05*	(0.02)	0.06**	(0.02)	0.05	(0.03)	0.04**	(0.02)
Dfprtalc	-0.04*	(0.02)	-0.04*	(0.02)	-0.03	(0.02)	-0.02	(0.02)
Rghmgprc	0.04*	(0.02)	0.04*	(0.02)	0.06*	(0.03)	0.05*	(0.02)
Votedirc (20)	0.04*	(0.02)	0.03	(0.02)	0.02	(0.03)	0.03**	(0.01)
Dscrmiss (20)								
No							-0.07	(0.08)
Yes	0.12	(0.14)	0.08	(0.15)	0.25	(0.23)	0.04	(0.18)
Votemiss (5)								
Yes							-0.04	(0.13)
No	-0.36***	(0.07)	-0.38***	(0.07)	-0.36*	(0.14)	-0.41**	(0.14)
Ineligible	0.00	(0.09)	-0.01	(0.09)	-0.03	(0.13)	0.03	(0.16)
Woman	0.19**	(0.06)	0.21***	(0.06)	0.15	(0.09)	0.22**	(0.06)
Constant	0.80**	(0.28)	0.90**	(0.3)	0.85	(0.47)	1.46***	(0.35)
R^2	0.43		0.42		0.44		0.38	
Adj. R^2	0.42		0.42		0.42		0.37	

* p<0.05; ** p<0.01; *** p<0.001

Black: unbiased | blue: estimate biased | pink: wrong inference | red: estimate biased and wrong inference

FORS
explore.understand.share.

Missing Not At Random

	Full (N=1174) Coef	(SE)	MI (N=1174) Coef	(SE)	CC (N=463) Coef	(SE)
Implvdm (5)	0.03	(0.02)	0.01	(0.02)	0.04	(0.04)
Dmcntov	0.09**	(0.03)	0.10***	(0.03)	0.16**	(0.05)
Stflife (2)	0.10***	(0.02)	0.11***	(0.02)	0.06	(0.03)
Stfeco (2)	-0.01	(0.02)	-0.01	(0.02)	0.00	(0.03)
Stfgov (5)	0.24***	(0.02)	0.23***	(0.03)	0.25***	(0.04)
Stfdem	0.11***	(0.03)	0.11***	(0.03)	0.10*	(0.04)
Stfedu (2)	0.02	(0.02)	0.01	(0.02)	-0.01	(0.03)
Stfhlth	0.03	(0.02)	0.04	(0.02)	0.03	(0.03)
Rlgdgr (2)	0.01	(0.01)	0.01	(0.01)	0.01	(0.01)
Fairelcc (5)	0.05*	(0.02)	0.06*	(0.02)	0.02	(0.03)
Dfprtalc	-0.04*	(0.02)	-0.04*	(0.02)	0.00	(0.02)
Rghmgprc	0.04*	(0.02)	0.05**	(0.02)	0.05	(0.03)
Votedirc (20)	0.04*	(0.02)	0.02	(0.02)	0.02	(0.03)
Dscrmiss (20)						
No						
Yes	0.12	(0.14)	0.12	(0.14)	0.05	(0.19)
Votemiss (5)						
Yes						
No	-0.36***	(0.07)	-0.38***	(0.07)	-0.4***	(0.11)
Ineligible	0.00	(0.09)	-0.01	(0.10)	-0.15	(0.17)
Woman	0.19**	(0.06)	0.20**	(0.06)	0.12	(0.09)
Constant	0.80**	(0.28)	0.91**	(0.29)	1.02**	(0.47)
R^2	0.43		0.43		0.45	
Adj. R^2	0.42		0.42		0.43	

* p<0.05; ** p<0.01; *** p<0.001

Black: unbiased | blue: estimate biased | pink: wrong inference | red: estimate biased and wrong inference

FORS
explore.understand.share.

Missing At Random (Small)

	Full (N=1174) Coef	(SE)	MI (N=1174) Coef	(SE)	CC (N=797) Coef	(SE)
Implvdm (5)	0.03	(0.02)	0.03	(0.02)	0.04	(0.03)
Dmcntov	0.09**	(0.03)	0.09**	(0.03)	0.05	(0.03)
Stflife (2)	0.10***	(0.02)	0.10***	(0.02)	0.08**	(0.02)
Stfeco (2)	-0.01	(0.02)	-0.01	(0.02)	-0.02	(0.03)
Stfgov (5)	0.24***	(0.02)	0.24***	(0.02)	0.26***	(0.03)
Stfdem	0.11***	(0.03)	0.11***	(0.03)	0.13***	(0.03)
Stfedu (2)	0.02	(0.02)	0.01	(0.02)	0.01	(0.03)
Stfhlth	0.03	(0.02)	0.03	(0.02)	0.03	(0.03)
Rlgdgr (2)	0.01	(0.01)	0.01	(0.01)	0.02	(0.01)
Fairelcc (5)	0.05*	(0.02)	0.06**	(0.02)	0.06*	(0.03)
Dfprtalc	-0.04*	(0.02)	-0.04*	(0.02)	-0.03	(0.02)
Rghmgprc	0.04*	(0.02)	0.04*	(0.02)	0.05*	(0.02)
Votedirc (20)	0.04*	(0.02)	0.03	(0.02)	0.04	(0.02)
Dscrmiss (20)						
No						
Yes	0.12	(0.14)	0.12	(0.14)	0.20	(0.17)
Votemiss (5)						
Yes						
No	-0.36***	(0.07)	-0.37***	(0.07)	-0.21*	(0.10)
Ineligible	0.00	(0.09)	-0.02	(0.09)	0.05	(0.11)
Woman	0.19**	(0.06)	0.19**	(0.06)	0.16*	(0.07)
Constant	0.80**	(0.28)	0.80**	(0.29)	0.86*	(0.35)
R^2	0.43		0.43		0.43	
Adj. R^2	0.42		0.42		0.42	

* p<0.05; ** p<0.01; *** p<0.001

Black: unbiased | blue: estimate biased | pink: wrong inference | red: estimate biased and wrong inference

FORS
explore.understand.share.

Categorical Data: MAR

	Full (N=1163) Coef	(SE)	MI (N=1163) Coef	(SE)	CC (N=935) Coef	(SE)	Zeros (N=1163) Coef	(SE)	Zeros (1 as base, N=1163) Coef	(SE)
Gincdif (5)							0.66**	(0.20)	-0.66**	(0.20)
2	0.30**	(0.10)	0.31***	(0.10)	0.25*	(0.11)	0.95***	(0.19)	0.29**	(0.10)
3	0.20	(0.13)	0.24	(0.13)	0.19	(0.14)	0.88***	(0.20)	0.22	(0.13)
4	0.15	(0.14)	0.15	(0.15)	0.04	(0.15)	0.79***	(0.21)	0.13	(0.14)
5	-0.11	(0.29)	0.11	(0.29)	-0.31	(0.32)	0.79*	(0.34)	0.13	(0.30)
Optftr (2)							-0.13	(0.27)	0.13	(0.27)
2	-0.01	(0.08)	-0.01	(0.08)	-0.05	(0.09)	-0.15	(0.26)	-0.03	(0.08)
3	-0.11	(0.14)	-0.07	(0.14)	-0.10	(0.15)	-0.2	(0.28)	-0.08	(0.13)
4	-0.08	(0.19)	-0.08	(0.19)	-0.09	(0.22)	-0.21	(0.32)	-0.09	(0.19)
5	0.70**	(0.27)	0.69*	(0.32)	-0.13	(0.10)	0.65	(0.41)	0.77*	(0.32)
Marital (5)							0.17	(0.20)	-0.17	(0.20)
2	-0.17	(0.13)	-0.15	(0.13)	-0.24	(0.15)	0.14	(0.24)	-0.03	(0.14)
3	-0.22	(0.20)	-0.16	(0.19)	-0.34	(0.25)	0.14	(0.29)	-0.03	(0.21)
4	0.02	(0.09)	0.01	(0.09)	0.07	(0.09)	0.21	(0.20)	0.04	(0.09)
Hincfel (10)							0.72	(0.15)	-0.72***	(0.15)
2	-0.32***	(0.08)	-0.3***	(0.09)	-0.26***	(0.09)	0.47***	(0.16)	-0.25**	(0.08)
3	-0.41*	(0.18)	-0.43*	(0.21)	-0.4	(0.21)	0.37**	(0.25)	-0.36	(0.20)
4	-0.84*	(0.35)	-0.85*	(0.39)	-0.74*	(0.33)	-0.21	(0.42)	-0.93*	(0.41)
Not voted	-0.42***	(0.09)	-0.43***	(0.09)	-0.42***	(0.10)	-0.43***	(0.09)	-0.43***	(0.09)
Not eligible	0.21	(0.12)	0.23	(0.12)	0.31**	(0.13)	0.24*	(0.12)	0.24*	(0.12)
Dscrgrp	-0.23	(0.23)	-0.21	(0.23)	-0.34	(0.35)	-0.03	(0.25)	-0.03	(0.25)
Woman	0.09	(0.07)	0.09	(0.07)	0.06	(0.08)	0.08	(0.07)	0.08	(0.07)
Constant	6.19***	(0.13)	6.17***	(0.13)	6.33***	(0.14)	4.81***	(0.41)	6.24***	(0.13)
R^2	0.08		0.18		0.08		0.11		0.11	
Adj. R^2	0.07		0.11		0.06		0.09		0.09	

* p<0.05; ** p<0.01; *** p<0.001

Black: unbiased | blue: estimate biased | pink: wrong inference | red: estimate biased and wrong inference

FORS
explore.understand.share.

Conclusions: The Don'ts

- There is NO way of «no treatment» of the missings
- Do NOT substitute missings with zeros (or dummies) (Allison, 2002)
  - Except if you're interested in them (e.g., answer patterns; theory)
- Do NOT try to EXPLAIN missingness
  - i.e., model the values, income = age + job + education
  - Esp. in social sciences: R-squared often <0.40
  - This changes the joint distribution → this is **making up** data!!
- Do NOT analyse imputed data without Rubin's Rules
  - E.g., do NOT use multiply imputed values provided by someone else if only one data set → no info on missingness
- Do NOT plot imputed values in your graphs
  - We do not predict values for missings. We're not even interested in them.

FORS
explore.understand.share.

Conclusions: The Do's (I/II)

- Analyse the patterns of missings
- Actively decide on the treatment of the missings
- Use a probabilistic approach to missingness, esp. if more than 10% overall missings (Langkamp et al., 2010)
- If you feel save about your missings (MCAR), do at least a sensitivity analysis for evidence

Conclusions: The Do's (II/II)

- Imputation Model
  - Use ALL variables you use in your model of analysis, also the dependent! (Schafer & Graham, 2002)
  - Include higher order terms in imputations (von Hippel, 2009)
  - For multi-item constructs: Impute single items (Eekhout et al., 2014)
  - Add auxiliary variables (Collins et al., 2001)
  - Use a suitable method (reg, trunc, pmm, ologit, logit)
- Always use Rubin's Rules!
  - Standard errors
  - Significance
  - R-squared
- Plot only the observations with complete data (m=0)

References (1/2)

- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. Journal of School Psychology, 48(1), 5–37. doi:10.1016/j.jsp.2009.10.001
- Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures. Psychological Methods, 6(4), 330–351. doi:10.1037/1082-989X
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society Series B (Methodological), 39(1).
- Eekhout, I., de Vet, H. C. W., Twisk, J. W. R., Brand, J. P. L., de Boer, M. R., & Heymans, M. W. (2014). Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. Journal of Clinical Epidemiology, 67(3), 335–342. doi:10.1016/j.jclinepi.2013.09.009
- Enders, C. K. (2006). A Primer on the Use of Modern Missing-Data Methods in Psychosomatic Medicine Research. Psychosomatic Medicine, 68(3), 427–436. doi:10.1097/01.psy.0000221275.75056.d8
- Goldstein, H., Carpenter, J., Kenward, M. G., & Levin, K. A. (2009). Multilevel models with multivariate mixed response types. Statistical Modelling, 9(3), 173–197. doi:10.1177/1471082X0800900301
- Graham, J. W. (2009). Missing Data Analysis: Making It Work in the Real World. Annual Review of Psychology, 60, 549–576. doi:10.1146/annurev.psych.58.110405.085530
- Langkamp, D. L., Lehman, A., & Lemeshow, S. (2010). Techniques for Handling Missing Data in Secondary Analyses of Large Surveys. Academic Pediatrics, 10(3), 205–210. doi:10.1016/j.acap.2010.01.005
- Lee, K. J., & Carlin, J. B. (2010). Multiple Imputation for Missing Data: Fully Conditional Specification Versus Multivariate Normal Imputation. American Journal of Epidemiology, 171(5), 624–632. doi:10.1093/aje/kwp425
- Little, R. J. A. & Rubin, D. B. (1987). Statistical analysis with missing data. New York, NY: Wiley.

FORS explore.understand.share.

References (2/2)

- Royston, P. (2004). Multiple imputation of missing values. The Stata Journal, 4(3), 227–241.
- Royston, P. (2005). Multiple imputation of missing values: update. The Stata Journal, 5(2), 1–14.
- Rubin, D. B. (1976). Inference and Missing Data. Biometrika, 63(3), 581–592. doi:10.1093/biomet/63.3.581
- Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York, NY: Wiley.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. Psychological Methods, 7(2), 147–177. doi:10.1037/1082-989X.7.2.147
- van Buuren, S. (2011). Multiple Imputation of Multilevel Data. In J. J. Hox & J. K. Roberts (Eds.), Handbook of Advanced Multilevel Analysis (pp. 173–196). New York, NY: Routledge.
- von Hippel, P. T. (2009). How to impute squares, interactions, and other transformed variables. Sociological Methodology, 39(1), 265–291. doi:10.1111/j.1467-9531.2009.01215.x
- White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine, 30(4), 377–399. doi:10.1002/sim.4067

FORS

explore.understand.share.

MERCI!