

결측(缺測) 데이터와 그 대책

Chanseok Park (박찬석)

Applied Statistics Laboratory
Department of Industrial Engineering
Pusan National University

August 5, 2020

Hosted by SEC



부산대학교
PUSAN NATIONAL UNIVERSITY

1 Missing · Incomplete data

- Types of Missing and Incomplete
- Illustration of Missing Mechanism
- Ad-hoc methods

Missing · Incomplete data

Types of Missing and Incomplete

- Missing data: no value is observed. (Little and Rubin, 2002)
 - Missing Completely AT Random (MCAR)
if missingness does not depends on the data, $Y = (Y_{\text{obs}}, Y_{\text{mis}})$.
 $f_{\theta}(M|Y) = f_{\theta}(M)$, where M is a missing indicator.
 - Missing At Random (MAR)
if missingness depends only on the observed data Y_{obs} .
 $f_{\theta}(M|Y) = f_{\theta}(M|Y_{\text{obs}})$.
 - Missing Not At Random (MNAR)
if missingness depends on the data $Y = (Y_{\text{obs}}, Y_{\text{mis}})$.
 $f_{\theta}(M|Y) =$ as it is
- Incomplete data: value is partially observed.
 - Truncation
 - Censoring
 - Grouping
 - Masking

Missing · Incomplete data

Illustration of Missing Mechanism

Complete Data		MCAR		MAR		MNAR	
GPA	IQ	GPA	IQ	GPA	IQ	GPA	IQ
2.0	93	2.0	?	2.0	?	2.0	?
2.2	115	2.2	115	2.2	?	2.2	115
2.4	96	2.4	96	2.4	?	2.4	?
2.6	116	2.6	?	2.6	?	2.6	116
2.8	94	2.8	94	2.8	?	2.8	?
3.0	106	3.0	106	3.0	106	3.0	106
3.2	98	3.2	?	3.2	98	3.2	?
3.4	103	3.4	103	3.4	103	3.4	103
3.6	95	3.6	95	3.6	95	3.6	?
3.8	112	3.8	?	3.8	112	3.8	112
4.0	100	4.0	100	4.0	100	4.0	100
4.2	120	4.2	?	4.2	120	4.2	120

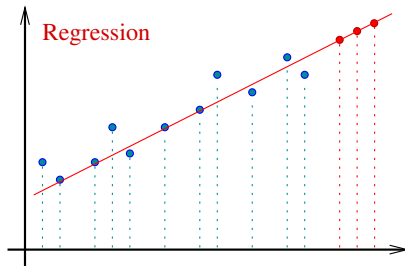
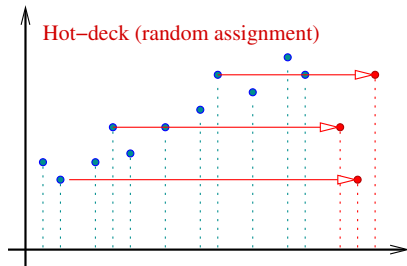
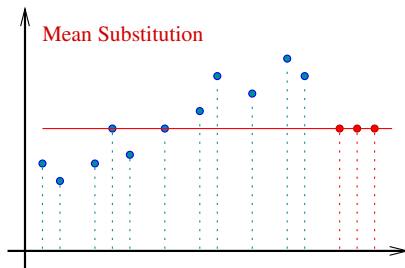
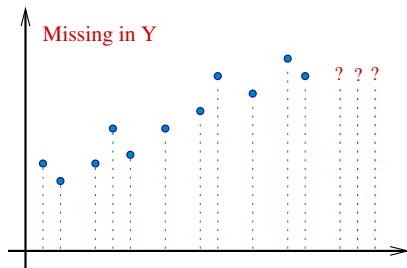
Deletion

- Complete-case analysis (listwise deletion, casewise deletion)
- Available-case analysis (pairwise deletion)

Single imputation

- Mean substitution
- Hot-deck, Cold-deck
- Regression imputation

Ad-hoc methods (single imputation)



- Little, R. J. A. and Rubin, D. B. (2002). Statistical Analysis with Missing Data. John Wiley & Sons, New York, 2nd edition.
- Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data. Chapman & Hall, Boca Raton, FL.