

Inhaltsverzeichnis

Abbildungsverzeichnis	XI
Tabellenverzeichnis	XV
1 Einleitung	1
2 Unvollständige Datenmatrizen und ihre Beschreibung	5
2.1 Unvollständige Datenmatrizen: Ein reales Problem?	5
2.2 Grundlegende Definitionen und Annahmen	8
2.3 Ausfallmuster	9
2.4 Ausfallmechanismen	12
2.4.1 Missing Completely at Random (MCAR)	13
2.4.2 Missing at Random (MAR)	15
2.4.3 Missing Not at Random (MNAR)	18
3 Missing-Data-Verfahren	21
3.1 Eliminierungsverfahren	21
3.1.1 Objekteliminierung	22
3.1.2 Merkmalseliminierung	26
3.2 Imputationsverfahren	28
3.3 Parameterschätzverfahren	32
3.3.1 Full Information Maximum Likelihood	34
3.3.2 EM-Algorithmus	35
3.4 Anpassung von Analyseverfahren	39
3.5 Sensitivitätsbetrachtungen	42
4 Imputationsverfahren für unvollständige Datenmatrizen	45
4.1 Einfache Imputationsverfahren	46
4.1.1 Deduktive Imputation und Expertenschätzungen	46
4.1.2 Imputation eines vorgegebenen Werts	47
4.1.3 Lageparameterimputation	48
4.1.4 Zufallszahlenimputation	51
4.1.5 Imputation des Verhältnisschätzers	54
4.2 Deck-Verfahren	56
4.2.1 Hot-Deck-Verfahren	56
4.2.1.1 Berücksichtigung von Ähnlichkeiten	59
4.2.1.2 Hot-Deck bei multivariaten Ausfallmustern	64

4.2.1.3	Mehrfache Verwendung von Spendern	66
4.2.1.4	Weitere Aspekte	70
4.2.2	Cold-Deck-Verfahren	72
4.3	Multivariate Imputationsverfahren	73
4.3.1	Imputation mittels Regressionsanalyse	73
4.3.1.1	Die Methode von Buck und ihre Erweiterungen . . .	77
4.3.1.2	Iterative Ansätze	79
4.3.1.3	Adaptive Regressionsimputation	80
4.3.1.4	Lokale Regressionsimputation	84
4.3.1.5	Weitere Ansätze und Imputation qualitativer Daten .	86
4.3.2	Imputation mittels Hauptkomponentenanalyse und Singulärwertzerlegung	90
4.3.2.1	Verfahren ohne Regularisierung	90
4.3.2.2	Verfahren mit Regularisierung	96
4.3.2.3	Bayesscher Ansatz	99
4.3.3	EM-Imputation	102
4.4	Imputation mittels Verfahren des maschinellen Lernens	105
4.4.1	Imputation mittels k-Nächste-Nachbarn	105
4.4.2	Imputation mittels Entscheidungsbäumen	108
4.4.2.1	Imputation mittels einzelner Bäume	109
4.4.2.2	Imputation mittels Ensemble-Methoden	110
4.4.3	Imputation mittels Clustering	113
4.5	Genereller Aufbau von Imputationsverfahren	117
5	Analyse existierender Simulationsstudien	123
5.1	Literaturrecherche	124
5.2	Vorgehensweisen zum Vergleich von Imputationsverfahren	128
5.3	Simulationsdesign der untersuchten Studien	131
5.3.1	Datenmatrizen	131
5.3.2	Erzeugung fehlender Werte	134
5.3.3	MD-Verfahren	137
5.3.4	Gütekriterien	140
5.3.5	Auswirkungen der variierten Faktoren	145
5.4	Bewertung der Imputationsverfahren	147
5.4.1	Bewertung der Verfahrensgruppen	149
5.4.2	Einzelbetrachtung der Imputationsverfahren	155
5.4.3	Paarvergleich der Imputationsverfahren	163
5.5	Zusammenfassung und Forschungslücken	168
6	Simulationsstudie: Vergleich der besten Verfahren	171
6.1	Design der Simulationsstudie	171
6.1.1	Datenmatrizen	172
6.1.2	Erzeugung fehlender Werte	173
6.1.3	Imputationsverfahren	175
6.1.4	Gütekriterien	177

6.1.5	Ablaufplan	180
6.2	Datenaufbereitung und Verlässlichkeit der Ergebnisse	181
6.3	Ergebnisse der Simulationsstudie	186
6.3.1	Genauigkeit der Imputationswerte	187
6.3.2	Auswirkungen auf die Erwartungswertschätzung	192
6.3.3	Auswirkungen auf die Varianzschätzung	197
6.3.4	Auswirkungen auf die Kovarianzschätzung	201
6.3.5	Auswirkungen auf die Regressionskoeffizientenschätzung	205
6.3.6	Auswirkungen auf die Prognosewerte	209
6.4	Zusammenfassung und Interpretation	213
6.4.1	Ergebnisse der einzelnen Imputationsverfahren	214
6.4.2	Einfluss der Gütekriterien	220
6.4.3	Auswirkungen der variierten Faktoren	222
6.4.4	Vergleich mit existierenden Simulationsstudien	224
6.4.5	Kritische Würdigung und Limitationen	225
6.4.6	Praktische Implikationen	229
7	Zusammenfassung und Ausblick	233
Anhang		239
A	Alternative Definitionen der Ausfallmechanismen	239
B	Stichprobe aus ACS PUMS 2015	243
C	Lösbarkeit des Optimierungsproblems (4.15) - (4.19)	247
D	Details und Erläuterungen zum Kapitel 5	249
E	Details zur Simulationsstudie	257
E.1	Varianzzerlegung der abhängigen Variable	257
E.2	Verwendete Software für die Simulationsstudie	258
E.2.1	Datenerzeugung	258
E.2.2	Erzeugung fehlender Werte	258
E.2.3	Imputationsverfahren	259
E.2.4	Analyse der imputierten Datenmatrizen	260
E.2.5	Informationen zur R Session	260
E.3	Monte Carlo Standardfehler	261
E.4	Ablehnung beim Differenzentest	263
E.5	Tabellen zu den Simulationsergebnissen	264
E.5.1	Genauigkeit der Imputationswerte	265
E.5.2	Auswirkungen auf die Erwartungswertschätzung	271
E.5.3	Auswirkungen auf die Varianzschätzung	277
E.5.4	Auswirkungen auf die Kovarianzschätzung	283

E.5.5	Auswirkungen auf die Regressionskoeffizientenschätzung . . .	289
E.5.6	Auswirkungen auf die Prognosewerte	295
Symbolverzeichnis		301
Abkürzungsverzeichnis		307
Literaturverzeichnis		309