



Missings

Umgang mit und Ersetzen von fehlenden Werten
bei multivariaten Analysen

Warum überhaupt Gedanken machen? Was fehlt, ist doch weg, oder?

- Allgegenwärtiges Problem in psychologischer Forschung
- Bringt Fehlerquellen in die Studie
- Verzerrung der Ergebnisse möglich
- Verringerung der Effizienz von statistischen Verfahren
- Analyse kann Aufschluss über Verbesserungen der Studie geben
- Mögliche Verzerrungen können bei der Interpretation berücksichtigt werden

Missing Values: Ursachen (Schnell, 1986)

- Fehlerhaftes oder mangelhaftes Design (unpräzise Items, nicht kongruente Items, Darbietung, ...)
- Antwortverweigerung bei einer Untersuchung
- Wissensdefizite beim Befragten
- Mangelnde Antwortmotivation beim Befragten
- (Unaufmerksamkeit der Beobachter)
- (Unvollständigkeit von Sekundärdaten)
- Codierungs- und Übertragungsfehler

Missing Value-Mechanismen

- Systematisch (NMAR) vs. unsystematisch (MAR) fehlende Werte

MAR	Missing at random. Antwortrate ist unabhängig von der Ausprägung des gemessenen Merkmals
OAR	Observed at random. Antwortrate ist unabhängig von der Ausprägung anderer Merkmale
MCAR	Missing completely at random MAR und OAR treffen zu

Diagramm von MCAR-Fehlwerten

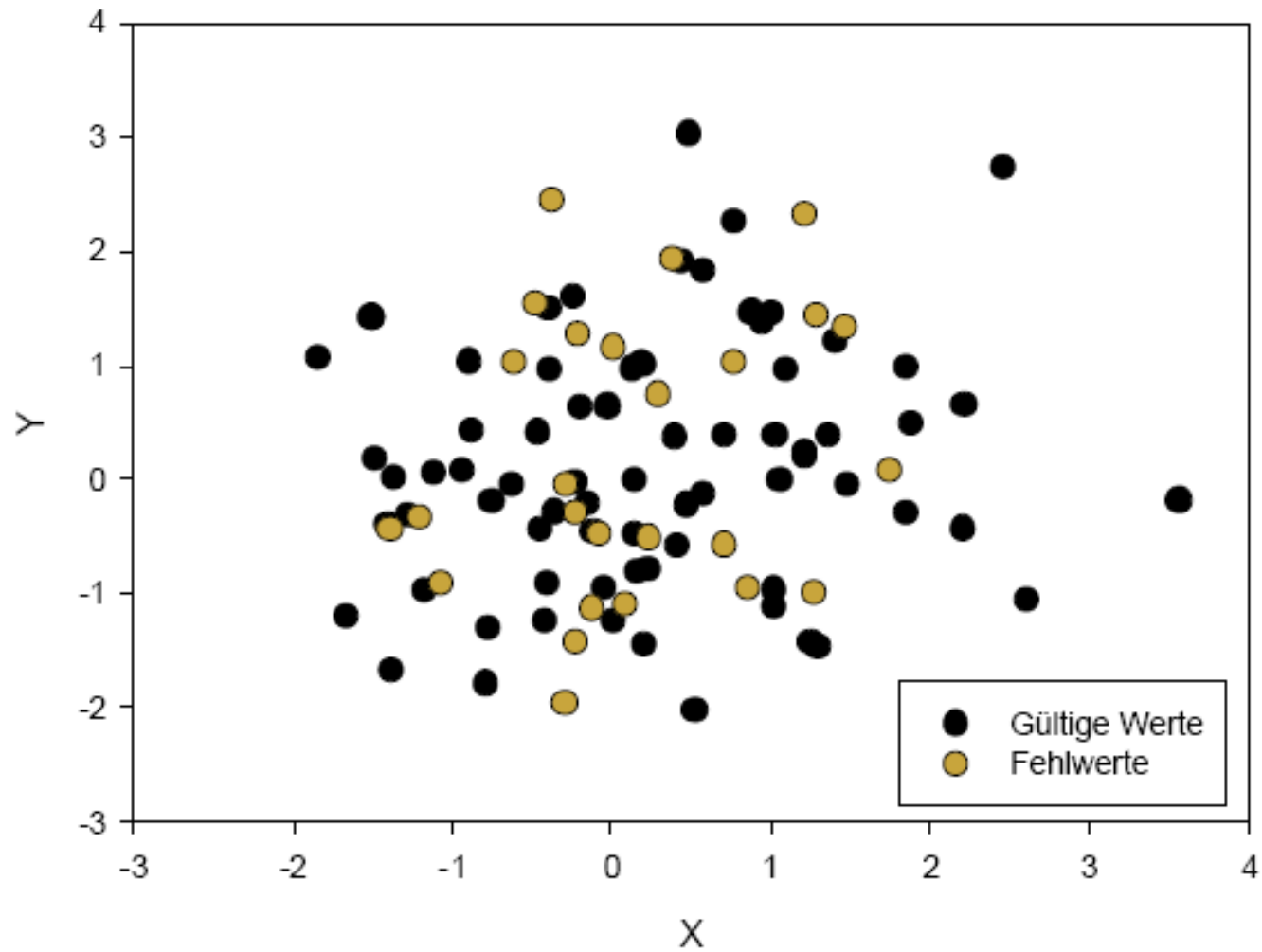


Diagramm von MAR-Fehlwerten

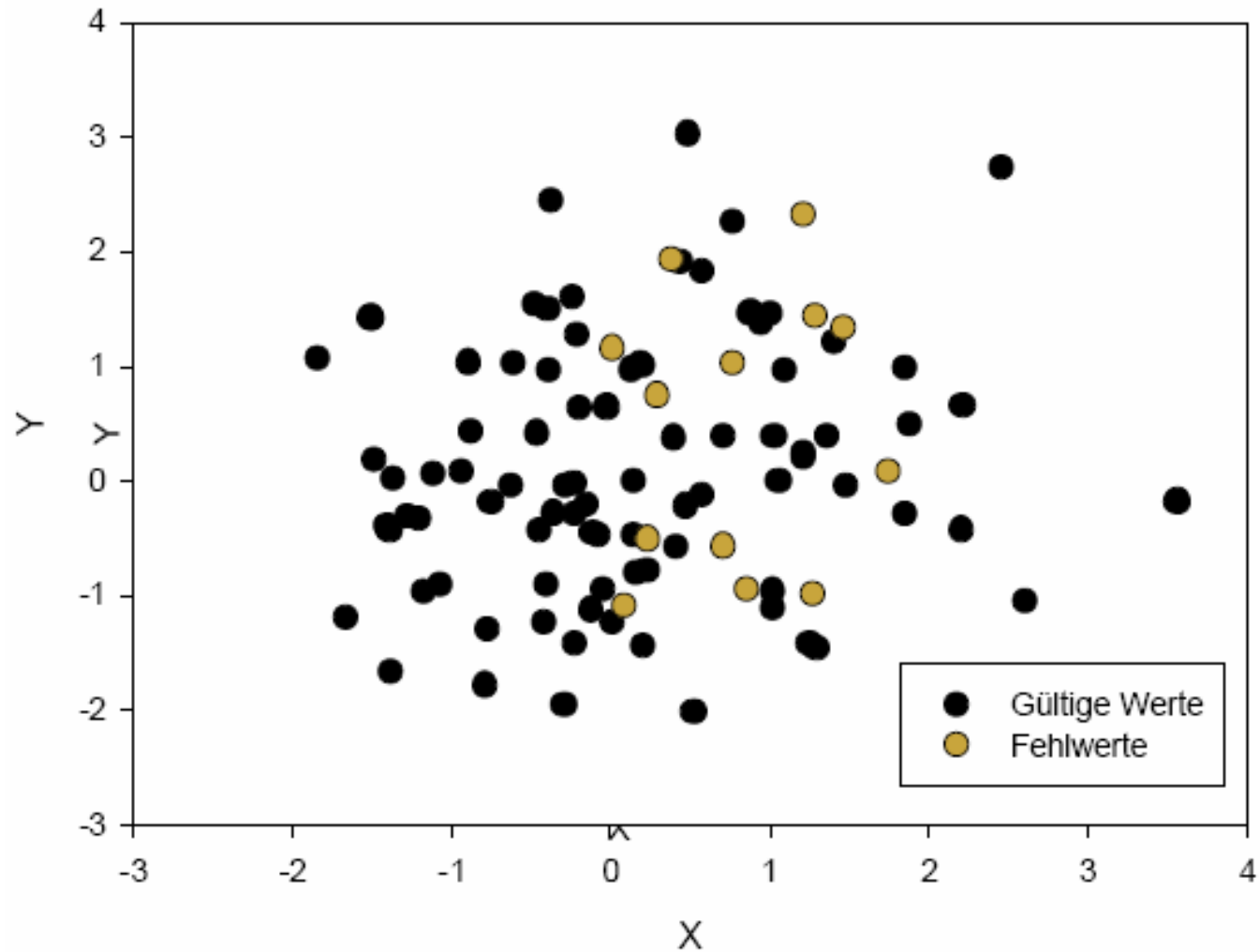
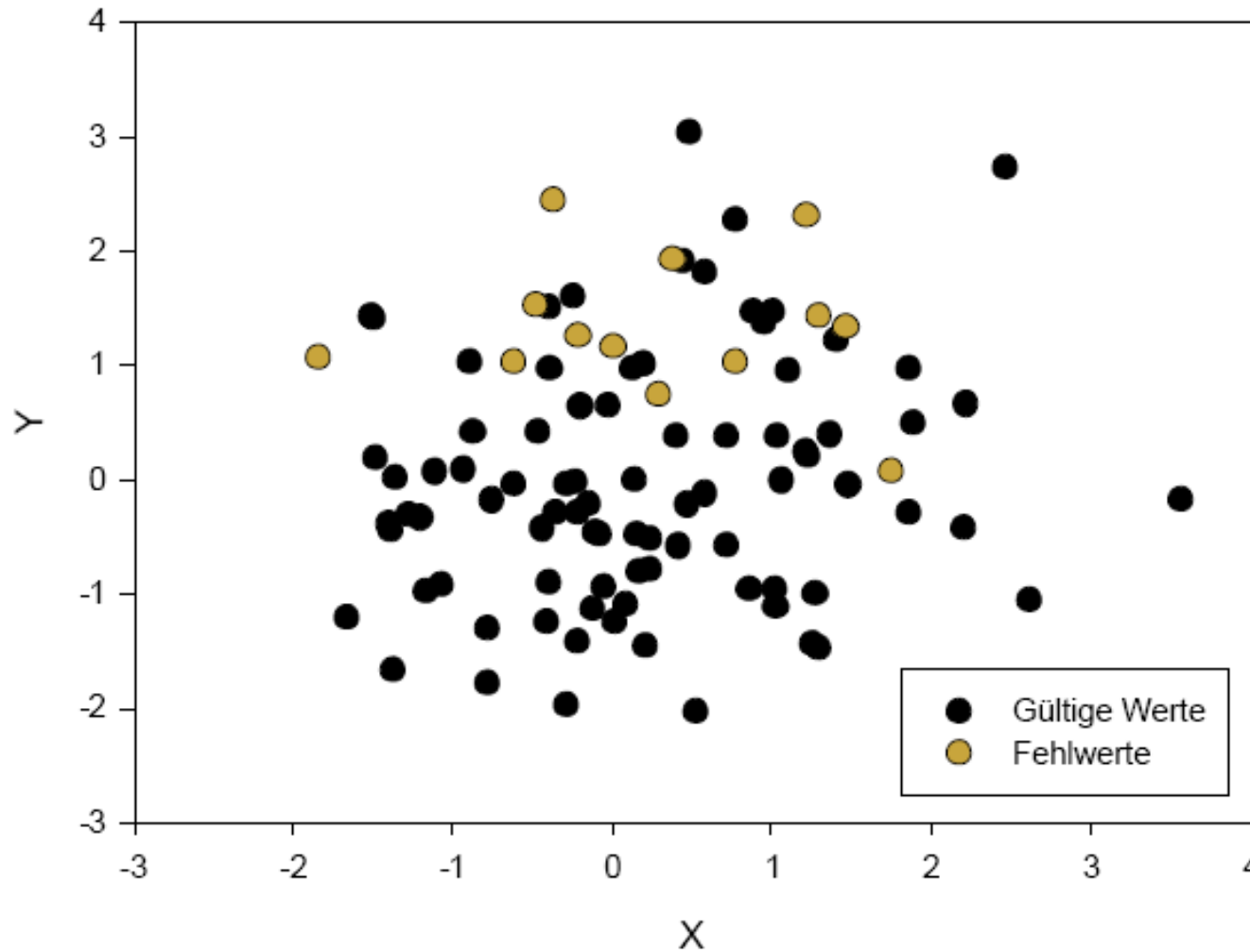


Diagramm von NMAR-Fehlwerten



Systematisch fehlende Werte (NMAR)

- Fehlende Werte sind mit gemessener und anderen Variablen assoziiert
- z.B.: Erfassen von Altern und Einkommen, fehlende Werte bei Einkommen
- NMAR: Fehlende Werte hängen vom Einkommen in bestimmten Altersgruppen ab
- MAR: Fehlende Werte unabhängig von Einkommen, aber assoziiert mit Alter
- MCAR: Fehlende Werte unabhängig von Einkommen und Alter

Kann man das überprüfen?

- Schwer bis überhaupt nicht (Teilnehmer kontaktieren, Antworten erzwingen?)
- Ausschluss der MAR und MCAR – Annahmen über Missing Value Analysis (MVA) möglich
- „As with other statistical assumptions,[..] the missing at random assumption may be a useful approximation even if it is believed to be false.“
(Allison, 1987)

Missing Value Analysis (MVA)

- **SPSS MVA**

- Descriptives: Univariate Statistiken, Percent Mismatch, t-Test, Crosstabs
- Patterns: Muster von fehlenden Werten werden angezeigt

Univariate Statistics

	N	Mean	Std. Deviation	Missing		No. of Extremes ^a	
				Count	Percent	Low	High
N_1	9	2,444444	1,130388	2	18,2	0	0
N_2	9	,777778	,833333	2	18,2	0	0
N_3	10	2,500000	1,178511	1	9,1	1	0
N_4	11	2,454545	1,213560	0	,0	1	0
AGE	11	75,7273	6,7393	0	,0	0	0
SEX	11			0	,0		

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

SPSS MVA

Tabulated Patterns

	Missing Patterns ^a						Complete if ... ^t
	N_4	AGE	SEX	N_3	N_2	N_1	
Number of Cases							
8							8
1				X			9
2					X	X	10

Patterns with less than 1% cases (0 or fewer) are not displayed.

- a. Variables are sorted on missing patterns.
- b. Number of complete cases if variables missing in that pattern (marked with X) are not used.

SPSS MVA

Missing Patterns (cases with missing values)

Case	# Missing	% Missing	Missing and Extreme Value Patterns ^a					
			N_4	AGE	SEX	N_3	N_2	N_1
12	1	16,7				S		
10	2	33,3					S	S
1	2	33,3					S	S

- indicates an extreme low value, while + indicates an extreme high value. The range used is (Q1 - 1.5*IQR, Q3 + 1.5*IQR).

a. Cases and variables are sorted on missing patterns.

Tukey's Robust
Boxplot Criterion
(Ansonsten $M \pm 2 SD$)

SPSS-MVA

Data Patterns (all cases)

Case ^a	# Missing	% Missing	Missing and Extreme Value Patterns						Variable Values				
			N_1	N_2	N_3	N_4	AGE	SEX	N_1	N_2	N_3		
			2	0	,0			-					
1	2	33,3	S	S							,	,	3,0000
8	0	,0									1,0000	,0000	3,0000
3	0	,0									3,0000	,0000	1,0000
11	0	,0									2,0000	2,0000	3,0000
10	2	33,3	S	S							,	,	3,0000
7	0	,0									1,0000	1,0000	3,0000
9	0	,0					-				2,0000	1,0000	4,0000
4	0	,0									4,0000	,0000	2,0000
5	0	,0									3,0000	2,0000	3,0000
12	1	16,7			S						2,0000	1,0000	,

- indicates an extreme low value, while + indicates an extreme high value. The range used is (Q1 - 1.5*IQR, Q3 + 1.5*IQR).

a. Cases are sorted by E_1.

SPSS-MVA

Percent Mismatch of Indicator Variables^{a,b}

	N ₃	N ₂	N ₁
N ₃	9,09		
N ₂	27,27	18,18	
N ₁	27,27	,00	18,18

The diagonal elements are the percentages missing, and the off-diagonal elements are the mismatch percentages of indicator variables.

- a. Variables are sorted on missing patterns.
- b. Indicator variables with less than 5% missing values are not displayed.

SPSS-MVA: t-Tests

Separate Variance t Tests^a

	N ₁	N ₂	N ₃	N ₄	AGE
t	,	,	-1,4	-,6	,6
df	,	,	7,0	1,4	2,1
P(2-tail)	,	,	,217	,624	,584
# Present	9	9	8	9	9
# Missing	0	0	2	2	2
N ₁ Mean(Present)	2,444444	,777778	2,375000	2,333333	76,2222
N ₁ Mean(Missing)	,	,	3,000000	3,000000	73,5000
t	,	,	-1,4	-,6	,6
df	,	,	7,0	1,4	2,1
P(2-tail)	,	,	,217	,624	,584
# Present	9	9	8	9	9
# Missing	0	0	2	2	2
N ₂ Mean(Present)	2,444444	,777778	2,375000	2,333333	76,2222
N ₂ Mean(Missing)	,	,	3,000000	3,000000	73,5000
t	,	,	,	,	,
df	,	,	,	,	,
P(2-tail)	,	,	,	,	,
# Present	8	8	10	10	10
# Missing	1	1	0	1	1
N ₃ Mean(Present)	2,500000	,750000	2,500000	2,500000	75,0000
N ₃ Mean(Missing)	,	,	,	,	,

For each quantitative variable, pairs of groups are formed by indicator variables (present, missing).

a. Indicator variables with less than 5% missing are not displayed.

SPSS-MVA: Crosstabs

SEX

			Total	male	female
N_1	Present	Count	9	4	5
		Percent	81,8	80,0	83,3
	Missing	% SysMis	18,2	20,0	16,7
N_2	Present	Count	9	4	5
		Percent	81,8	80,0	83,3
	Missing	% SysMis	18,2	20,0	16,7
N_3	Present	Count	10	5	5
		Percent	90,9	100,0	83,3
	Missing	% SysMis	9,1	,0	16,7

Indicator variables with less than 5% missing are not displayed.

Wie mit Missings umgehen?

Ignorieren

(Listenweiser/Pairweiser
Fallausschluss):

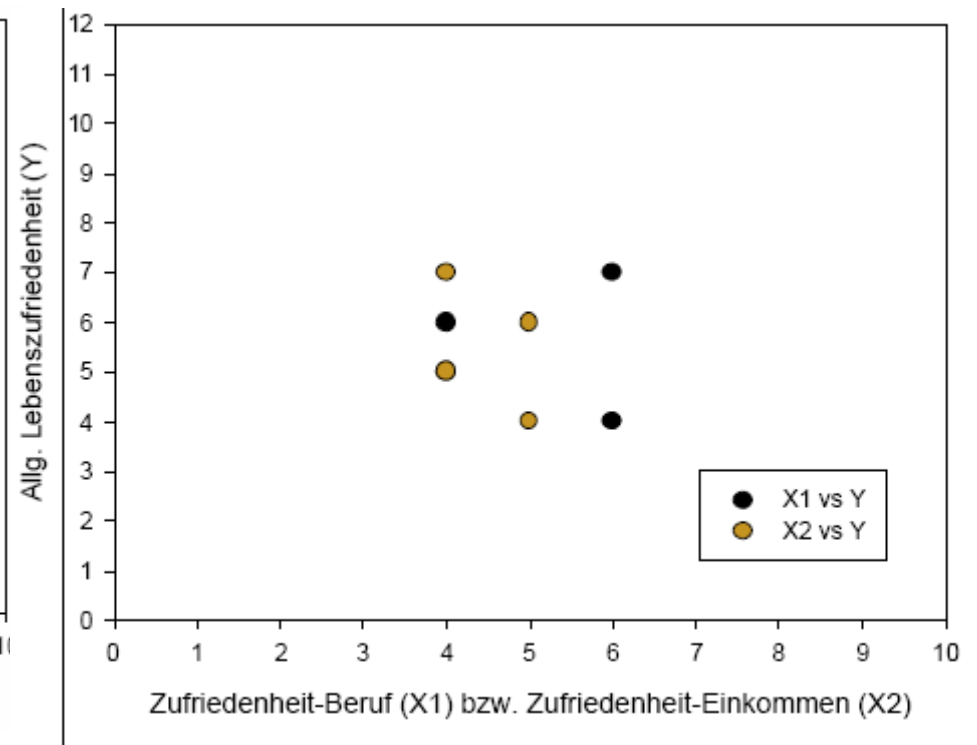
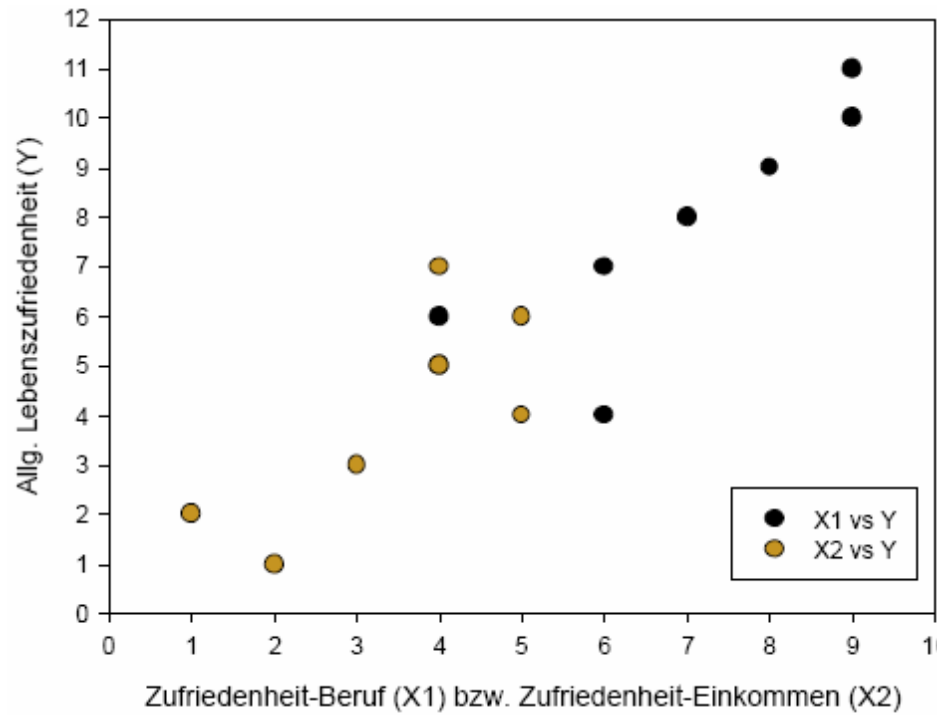
Schlecht.



Ignorieren (Fallausschluss)

- z.B. Listenweiser Fallausschluss:
- Analyse der kompletten Datensätze, Datensätze mit fehlenden Werten fliegen raus
- Bei sehr wenig fehlenden Werten (MCAR) kein Problem
- Probleme bei vielen fehlenden Werten
- Verzerrte Daten bei NMAR
- Reduktion der Stichprobe bis zur Unbrauchbarkeit

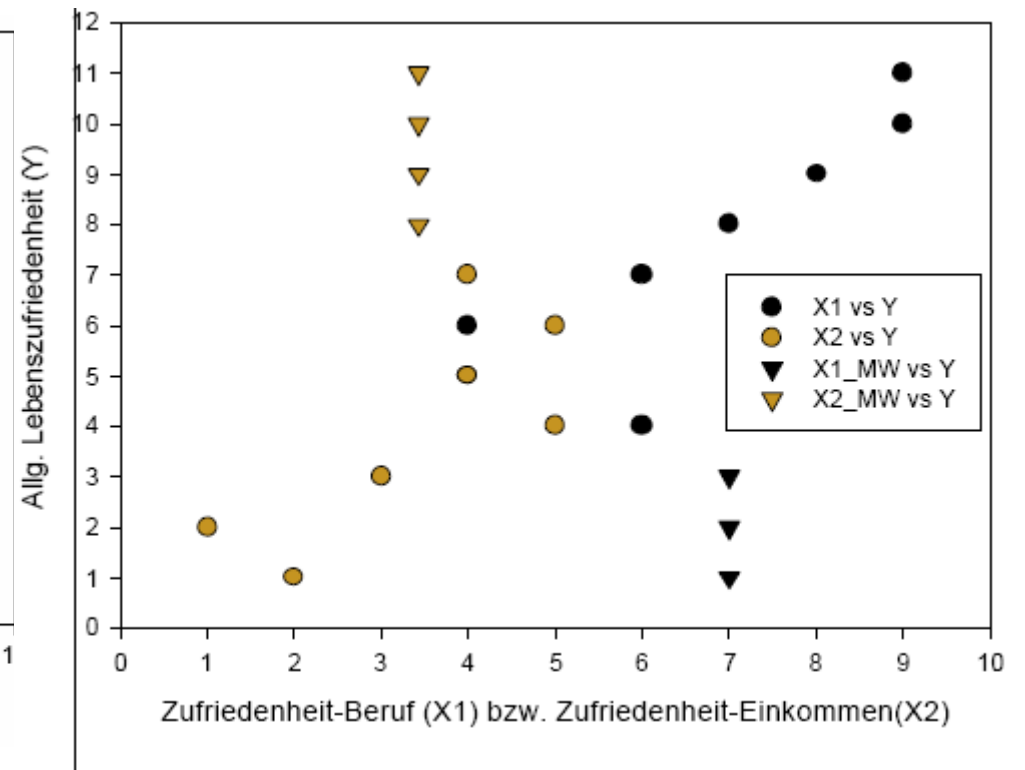
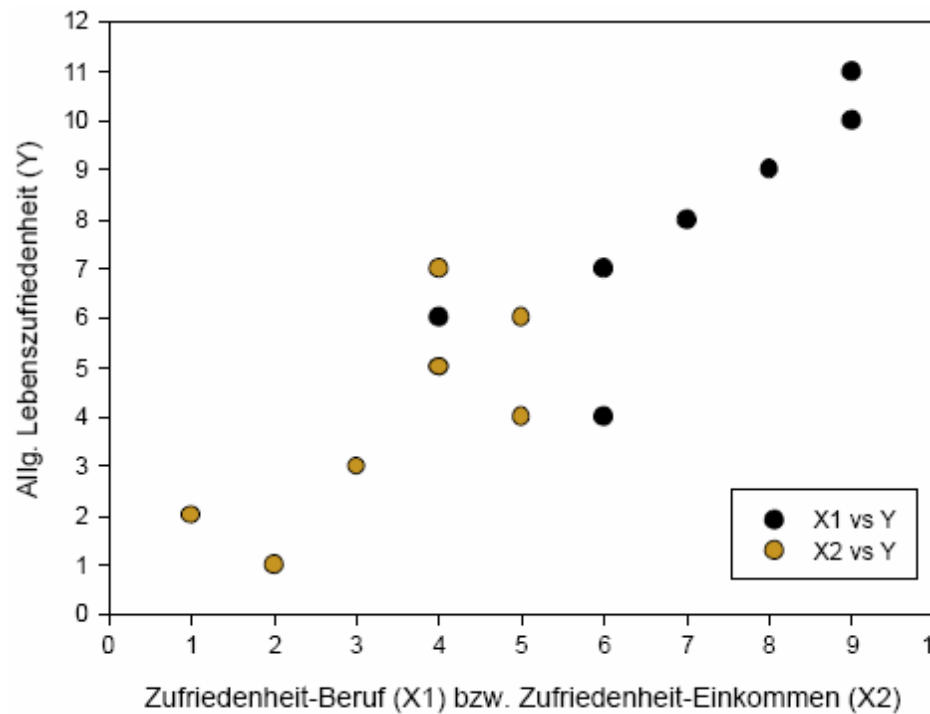
Auswirkungen



Ignorieren (Mittelwerte einsetzen)

- Ersetzen der fehlenden Werte durch den Mittelwert der Variable
- Nur bei der Berechnung von Summen- und Mittelwerten verzerrungsfrei
- Verzerrung der wahren Verteilung
- Unterschätzung der wahren Varianz
- Unterschätzung der wahren Zusammenhänge

Auswirkungen



Wie mit Missings umgehen?

Ignorieren

(Listenweiser/Pairweiser
Fallausschluss): Schlecht.

Besser: Imputieren

Faustregel:

Bei weniger als 5% fehlender
Werte auf einer Variablen gibt
es kaum Unterschiede
zwischen Imputationsverfahren



EM-Imputation (SPSS MVA)

- E-Schritt (Estimation): Finden der erwarteten Werte für die fehlenden Werte unter Gültigkeit der beobachteten (und momentan geschätzten) Parameter
- M-Schritt: Maximum Likelihood-Schätzung der fehlenden Werte gegeben die durch den E-Schritt aufgefüllte Verteilung
- Iteration, bis es passt
- Besser viele Prädiktoren (default: Alle quantitativen Variablen)

FIML (Full Information Maximum Likelihood)

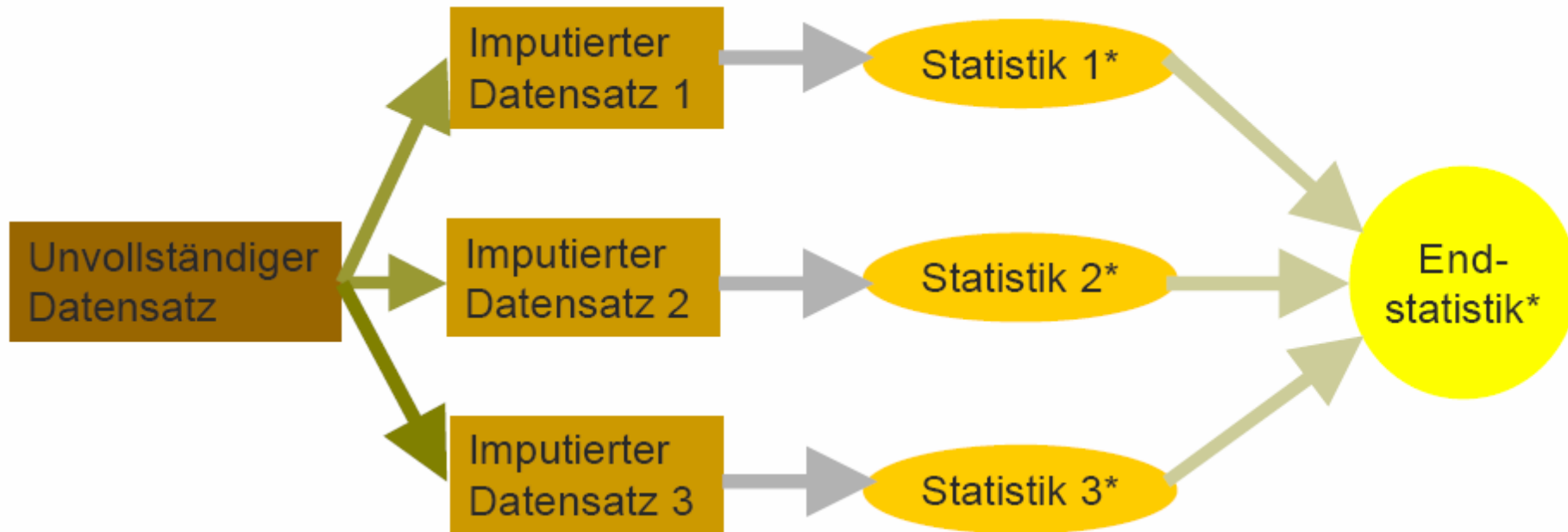
- Anwendung bei Strukturgleichungsmodellen
- Default-Einstellung in AMOS zur Schätzung fehlender Werte
- Maximum-Likelihood-Schätzung der fehlenden Werte aufgrund der Kovarianzmatrizen der beobachteten Werte
- ML-Schätzung: Suche nach einer Kombination von Parametern, die die Wahrscheinlichkeit der Kovarianzmatrix der beobachteten Werte maximiert

Multiple Imputation (MI; viel Arbeit)

1) IMPUTATION ⇨

2) ANALYSE ⇨

3) INTEGRATION



*Punktschätzer und Standardfehler

MI: Wie wird das gemacht?

- Mehrere Kopien des unvollständigen Datensatzes
- Mehr: Je mehr Missings, desto mehr Datensätze
- Faustregel: ca. 2-10 Datensätze
- Imputation der fehlenden Werte mit randomisierten Variationen (simuliert Zufallsfehler)
- Bestimmen des Fits der einzelnen Modelle
- Rekombination der verschiedenen imputierten Datensätze

Mehr?

- http://www.rehawissenschaft.uni-wuerzburg.de/methodenberatung/Igl_040604_Halle_Fehlende_Werte.pdf
- http://www.runte.de/matthias/publications/missing_values.pdf
- http://www.sc.uevora.pt/spss/pdf/manual_spss_12/SPSS_Missing_Value_Analysis_7.5