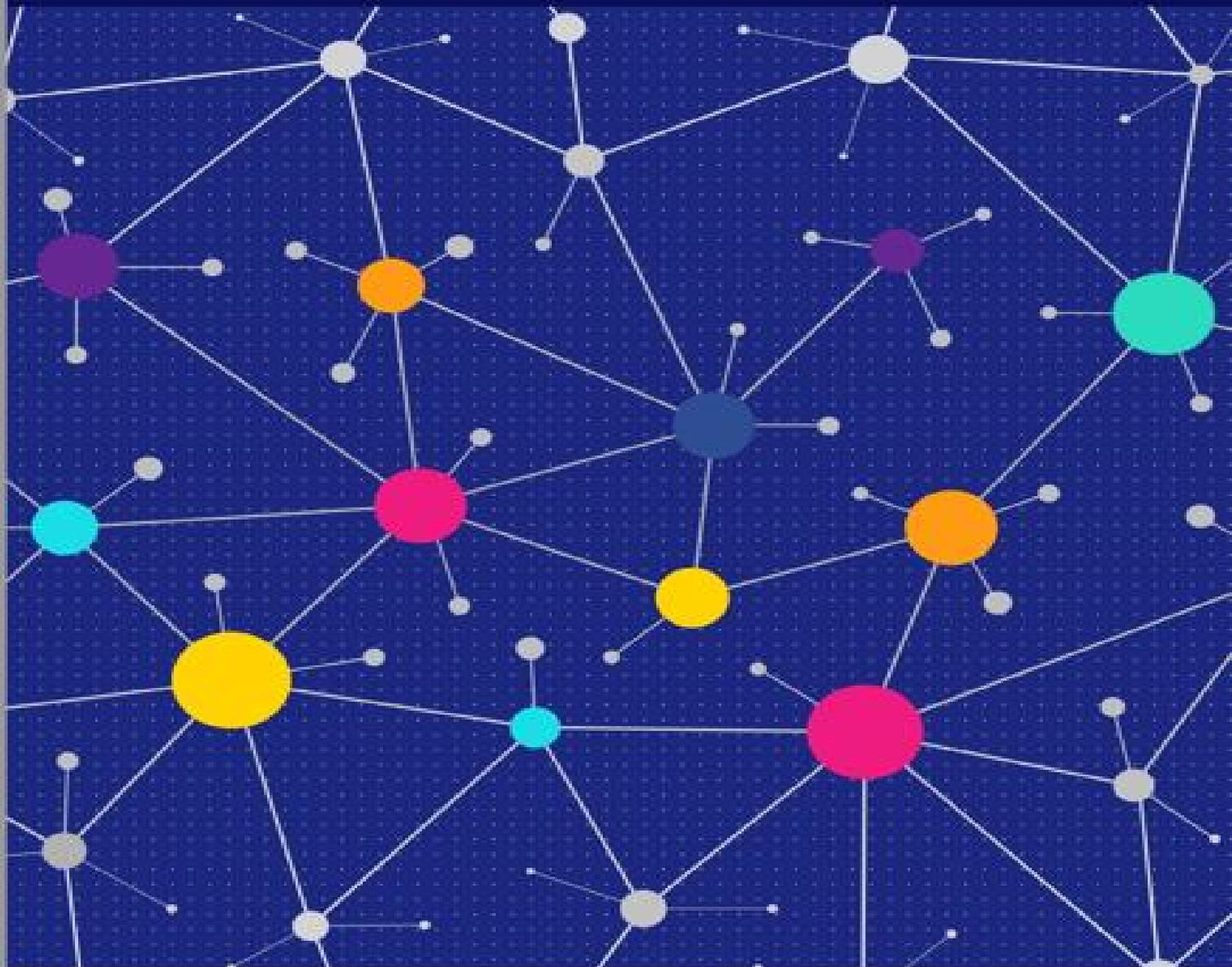




Multivariate Analysemethoden

Multivariate Statistik und
Datenanalyse mit SPSS einfach erklärt



M. Mittelstädt & J. Winke

Multivariate Analysemethoden

Multivariate Statistik und
Datenanalyse mit SPSS
einfach erklärt

Inhaltsverzeichnis

ABBILDUNGSVERZEICHNIS

DIE ANALYSE FEHLENDER WERTE

Eine Einführung

Was sind fehlende Werte?

Wodurch entstehen fehlende Werte?

Warum sind fehlende Werte ein Problem?

Die Ursache fehlender Werte

Folgen und Konsequenzen fehlender Werte

Beeinflussung der Validität

Klassifikation der fehlenden Werte

Vorgehen bei der Analyse fehlender Werte

Kann man fehlende Werte ignorieren?

Das Ausmaß der fehlenden Werte analysieren

Die Löschung von unvollständigen Daten

Tests zur Analyse der Zufälligkeit

Auswahl der Methode

1. Klassische Verfahren

2. Imputationsbasierte Verfahren

Allgemeine Handlungsempfehlung

DIE CLUSTERANALYSE

Einführung

Vorbereitung der Daten

Voraussetzung für Clusteranalyse

Das Vorgehen bei der Clusteranalyse

1) Erstellung der Proximitätsmatrix

Binäre/ nominale Variablenstruktur

Ordinale Variablenstruktur

Metrische Variablenstruktur

Gemischt skalierte

Variablenstruktur

2) Auswahl des Clusterverfahrens

Agglomerative Algorithmen

(hierarchisch)

Partitionierende Verfahren (K-

Means)

3) Festlegung Clusterzahl

4) Clusterdiagnose

EXPLORATIVE FAKTORENANALYSE

Einleitung

Voraussetzungen der Faktorenanalyse

Ablauf der Faktorenanalyse

Variableneignung

Kommunalität & Faktorenextraktion

Kommunalität

Extraktionsmethoden

Die Zahl der Faktoren bestimmen

Faktorinterpretation

Die Rotation

Benennung der Faktoren

Faktorwerte bestimmen

Ablauf in SPSS

DIE ENTSCHEIDUNGSBAUMANALYSE

Eine Einführung

Modellannahmen/ Voraussetzungen

Die Baumstruktur

Aufbaumethoden des Baums & Ablauf

Merging-Phase

Split-Phase

Vorgehen in SPSS

Interpretation der Ergebnisse

Der Baumeditor

Schritt-für-Schritt-Anleitung

MULTIPLE REGRESSION

Einsatzbereich

Vorgehensweise

1. Modellformulierung

**2. Schätzung der
Regressionsfunktion**

3. Prüfung der Regressionsfunktion

**4. Prüfung der
Regressionskoeffizienten**

5. Prüfung der Modellprämissen

Analyse einflussreicher Beobachtungen

Ausreißer

Hebelwerte (eng. leverage)

Cook-Distanzen

Ergänzungen

Dummy Variablen

Methoden der Multiplen Regression

KORRESPONDENZANALYSE

Allgemeines

Grundlagen der Korrespondenzanalyse

Einordnung

Fragestellung und Anspruch an die

Daten

Zielsetzung

Anwendungsbereiche

Vorgehensweise

Erstellung der Zeilen- und

Spaltenprofile

Festlegung der geometrischen

Schwerpunkte der

Das geometrische Modell - Marken

im Raum der Merkmale

Ermittlung der Streuungen

(Distanzen)

Standardisierung der Daten

Extraktion der Dimensionen

Ermittlung der Koordinaten

Interpretation

Symmetrische Normalisierung

Asymmetrische Normalisierung

Übersicht

Fazit

Vorgehensweise in SPSS

Abkürzungen

VARIANZANALYSE

Problemstellung

Einfaktorielle ANOVA

Modellgleichung

Hypothesen

Streuungszerlegung (sum of squares)

F-Statistik

Modellannahmen

SPSS Durchführung

Interpretation der SPSS-Ergebnisse

Zweifaktorielle ANOVA

Modellgleichung

Hypothesen

Streuungszerlegung

F-Statistik

Modellannahmen

Wechselwirkungen

SPSS Durchführung

SPSS Interpretation

Einfaktorielle ANOVA mit
Messwiederholung

Modellgleichung
Hypothesen
Streuungszerlegung
F-Statistik
Modellannahmen
SPSS Durchführung
SPSS Interpretation

Mehrfaktorielle ANOVA mit
Messwiederholung (Mixed ANOVA)

Hypothesen
Streuungszerlegung
F-Statistik
Modellgüte
Modellannahmen
SPSS Durchführung
SPSS Interpretation

QUELLENVERZEICHNIS

Abbildungsverzeichnis

ABBILDUNG 1: ÜBERSICHT VALIDITÄT

ABBILDUNG 2: MISSING AT COMPLETELY RANDOM

ABBILDUNG 3: MISSING AT RANDOM

ABBILDUNG 4: MISSING NOT AT RANDOM

ABBILDUNG 5: UNIVARIATE STATISTIK (VORGEHEN IN SPSS)

ABBILDUNG 6: KREISDIAGRAMM ÜBER FEHLENDE WERTE

ABBILDUNG 7: VARIABLENZUSAMMENFASSUNG (OUTPUT SPSS)

ABBILDUNG 8: EIN MUSTER FEHLENDER WERTE (SPSS OUTPUT)

ABBILDUNG 9: MUSTER FEHLENDER WERTE (SPSS OUTPUT)

ABBILDUNG 10: TEST DER FEHLENDEN WERTE (VORGEHEN IN SPSS)

ABBILDUNG 11: T- TESTS BEI UNTERSCHIEDLICHER VARIANZ

ABBILDUNG 12: KREUZTABELLE 1 (SPSS OUTPUT)

ABBILDUNG 13: KREUZTABELLE 2 (SPSS OUTPUT)

ABBILDUNG 14: BETRACHTUNG MUSTER (VORGEHEN IN SPSS)

ABBILDUNG 15: MUSTER IN TABELLEN (OUTPUT IN SPSS)

ABBILDUNG 16: MCAR-TEST NACH LITTLE

ABBILDUNG 17: LISTENWEISER FALLAUSSCHLUSS (VOR- UND NACHTEILE)

ABBILDUNG 18: PAARWEISER FALLAUSSCHLUSS (VOR- UND NACHTEILE)

ABBILDUNG 19: MITTELWERT-IMPUTATION (VOR- UND NACHTEILE)

ABBILDUNG 20: REGRESSIONS-IMPUTATION (VOR- UND NACHTEILE)

ABBILDUNG 21: HOT UND COLD DECK VERFAHREN (VOR- UND NACHTEILE)

ABBILDUNG 22: MULTIPLE IMPUTATION

ABBILDUNG 23: VORGEHEN DER MULTIPLLEN IMPUTATION IN SPSS

ABBILDUNG 24: INTERPRETATION SPSS OUTPUT MULTIPLE IMPUTATION

ABBILDUNG 25: DIE CLUSTERANALYSE

ABBILDUNG 26: ABWÄGUNG VARIABLENZAHL

ABBILDUNG 27: ROHDATENMATRIX IN ÄHNLICHKEITSMATRIX ÜBERFÜHREN

ABBILDUNG 28: ÄHNLICHKEITS- UND DISTANZMASS,

ABBILDUNG 29: PROXIMITÄTSMASSE IN SPSS (VORGEHEN)

ABBILDUNG 30: PROXIMITÄTSMASSE MIT SKALENNIVEAU

ABBILDUNG 31: BINÄRE OPTION IN SPSS

ABBILDUNG 32: KOMBINATIONSMÖGLICHKEITEN BINÄRE VARIABLENSTRUKTUR

ABBILDUNG 33: NÄHERUNGSMATRIX NACH JACCARD

ABBILDUNG 34: NÄHERUNGSMATRIX NACH RUSSELL UND RAO

ABBILDUNG 35: NÄHERUNGSMATRIX NACH M-KOEFFIZIENT

**ABBILDUNG 36: BEISPIELE AUSWAHL BINÄR
KOEFFIZIENTEN**

**ABBILDUNG 37: NÄHERUNGSMATRIX NACH CHI-
QUADRAT-MASS**

**ABBILDUNG 38: NÄHERUNGSMATRIX CITY-BLOCK-
METRIK (DISTANZMASS)**

**ABBILDUNG 39: NÄHERUNGSMATRIX (QUADRIERTE)
EUKLIDISCHE DISTANZ**

**ABBILDUNG 40: NÄHERUNGSMATRIX PEARSON-
KORRELATIONSKOEFFIZIENT**

**ABBILDUNG 41: ÜBERSICHT ÄHNLICHKEITS- UND
DISTANZMASSE**

ABBILDUNG 42: ÄHNLICHKEITS- UND DISTANZMASSE

**ABBILDUNG 43: GENAUES VORGEHEN IN SPSS
(DISTANZ- UND ÄHNLICHKEITSMASSE)**

ABBILDUNG 44: ÜBERSICHT CLUSTERVERFAHREN

ABBILDUNG 45: SINGLE LINKAGE VERFAHREN

ABBILDUNG 46: COMPLETE-LINKAGE-VERFAHREN

ABBILDUNG 47: AVERAGE LINKAGE-VERFAHREN

ABBILDUNG 48: VORGEHEN CLUSTERANALYSE IN SPSS

**ABBILDUNG 49: ZUORDNUNGSÜBERSICHT (SPSS-
OUTPUT)**

**ABBILDUNG 50: EISZAPFENDIAGRAMM (SPSS-
OUTPUT)**

ABBILDUNG 51: DENDROGRAMM (SPSS-OUTPUT)

ABBILDUNG 52: VORGEHEN K-MEANS IN SPSS

**ABBILDUNG 53: INTERPRETATION SPSS-OUTPUT K-
MEANS-VERFAHREN**

ABBILDUNG 54: ÜBERSICHT CLUSTERVERFAHREN

ABBILDUNG 55: ZIELKONFLIKT DER CLUSTERZAHL

ABBILDUNG 56: ZUORDNUNGSÜBERSICHT

ABBILDUNG 57: DAS ELBOW-KRITERIUM

ABBILDUNG 58: ERSTELLUNG MULTIVARIATE PROFILE

**ABBILDUNG 59: CLUSTERANALYSE MIT HILFE
MULTIVARIATER PROFILE**

**ABBILDUNG 60: EXPLORATIVE VS.
KONFIRMATORISCHE FAKTORENANALYSE**

ABBILDUNG 61: BILDUNG DER FAKTOREN

**ABBILDUNG 62: ZIELKONFLIKT DER
FAKTORENANALYSE**

ABBILDUNG 63: DAS FUNDAMENTALTHEOREM

ABBILDUNG 64: ABLAUF DER FAKTORENANALYSE

**ABBILDUNG 65: ES SOLLTEN HOHE UND NIEDRIGE
KORRELATIONEN VORLIEGEN**

ABBILDUNG 66: SIGNIFIKANZ DER KORRELATIONEN

ABBILDUNG 67: INVERSE KORRELATIONSMATRIX

ABBILDUNG 68: IMAGE VS. ANTI-IMAGE

ABBILDUNG 69: ANTI-IMAGE-MATRIX

ABBILDUNG 70: DER BARTLETT-TEST

ABBILDUNG 71: INTERPRETATION DER MSA WERTE

**ABBILDUNG 72: MSA-WERTE DER EINZELNEN
VARIABLEN**

**ABBILDUNG 73: KOMPONENTEN DER GESAMTVARIANZ
BEI EINER 1 FAKTOR**

**ABBILDUNG 74: GRAFISCHE DARSTELLUNG (VEKTOR-
DIAGRAMM)**

**ABBILDUNG 75: ÜBERSETZUNG VON WINKEL IN
KORRELATION**

ABBILDUNG 76: TOTALE VARIANZAUFKLÄRUNG

**ABBILDUNG 77: HAUPTKOMPONENTENANALYSE (SPSS
OUTPUT)**

ABBILDUNG 78: HAUPTACHSENANALYSE (SPSS OUTPUT)

ABBILDUNG 79: ITERATIONSMAXIMUM ERREICHT

ABBILDUNG 80: ITERATIONEN IN SPSS MANUELL EINSTELLEN

ABBILDUNG 81: VERGLEICH HAUPTACHSEN- & HAUPTKOMPONENTENANALYSE

ABBILDUNG 82: VERGLEICH DER KOMMUNALITÄTEN

ABBILDUNG 83: BERECHNUNG KAISER KRITERIUM

ABBILDUNG 84: BERECHNUNG DER KOMMUNALITÄT

ABBILDUNG 85: FAKTORAUSWAHL NACH KAISER KRITERIUM

ABBILDUNG 86: ZUSAMMENHANG EIGENWERT UND KOMMUNALITÄT

ABBILDUNG 87: BEISPIEL SCREE-PLOT

ABBILDUNG 88: FAKTORINTERPRETATION

ABBILDUNG 89: ROTIERTE LÖSUNG DER ORTHOGONALEN ROTATION

ABBILDUNG 90: UNTERSCHIEDE IN DER FAKTORENMATRIX NACH ROTATION

ABBILDUNG 91: ROTIERTE LÖSUNG DER OBLIQUE ROTATION

ABBILDUNG 92: SPSS OUTPUT OBLIQUE ROTATION

ABBILDUNG 93: VERGLEICH DER ORTHOGONALEN UND OBLIQUEN ROTATION

ABBILDUNG 94: BEISPIEL FAKTORENBEZEICHNUNG

ABBILDUNG 95: SCHÄTZVERFAHREN ZUR BESTIMMUNG DER FAKTORWERTE

ABBILDUNG 96: 3-DIMENSIONALES-LADUNGSDIAGRAMM

**ABBILDUNG 97: EIN-DIMENSIONALES-
LADUNGSDIAGRAMM**

**ABBILDUNG 98: VORGANG DER FAKTORENANALYSE IN
SPSS**

**ABBILDUNG 99: EINORDNUNG DER
ENTSCHEIDUNGSBAUMANALYSE**

**ABBILDUNG 100: ZUORDNUNG DES RICHTIGEN
DATENNIVEAUS**

ABBILDUNG 101: WERTELABELS VERGEBEN.

**ABBILDUNG 102: UNTERSCHIED CHAID- &
EXHAUSTIVE CHAID-VERFAHREN**

**ABBILDUNG 103: AUFBAUMETHODE EXHAUSTIVE
CHAID (SPSS)**

**ABBILDUNG 104: ÜBERSICHT TESTS FÜR
DATENNIVEAUS**

**ABBILDUNG 105: MEHRDIMENSIONALE
KONTIGENZTABELLE**

**ABBILDUNG 106: KREUZTABELLEN MERGEN SCHRITT 1
(BEISPIEL)**

**ABBILDUNG 107: KREUZTABELLEN MERGEN SCHRITT 2
(BEISPIEL)**

**ABBILDUNG 108: ÜBERSICHT P-WERT DER
UNTERTEILUNGSEBENEN**

**ABBILDUNG 109: MERGING DER KATEGORIE 1 UND 2
(BEISPIEL)**

**ABBILDUNG 110: „SIEGERVARIANTE“ DES MERGING-
PROZESSES**

ABBILDUNG 111: KREUZTABELLE SPLIT-PHASE

**ABBILDUNG 112: VORGEHEN IN SPSS
(ENTSCHEIDUNGSBAUMANALYSE)**

ABBILDUNG 113: INTERPRETATION DER ERGEBNISSE

ABBILDUNG 114: DER BAUMEDITOR
ABBILDUNG 115: DATENPUNKTE
ABBILDUNG 116: SPSS KOEFFIZIENTEN (EINFACH)
ABBILDUNG 117: SPSS KOEFFIZIENTEN
ABBILDUNG 118: STREUUNG
ABBILDUNG 119: SPSS ANOVA
ABBILDUNG 120: T-WERT
ABBILDUNG 121: T-WERT KONFIDENZINTERVALLE
**ABBILDUNG 122: KEIN LINEARER ZUSAMMENHANG
ERKENNBAR**
**ABBILDUNG 123: LINEARER ZUSAMMENHANG
ERKENNBAR**
**ABBILDUNG 124: STANDARDISIERTES
STREUDIAGRAMM**
**ABBILDUNG 125: DREIECKSMUSTER MIT GRÖßER
BZW. KLEINER WERDENDEN VARIANZ)**
**ABBILDUNG 126: RHOMBUS MUSTER (ANSTEIGENDE
UND KLEINER WERDENDE VARIANZ)**
**ABBILDUNG 127: RESIDUALWERTE VERÄNDERN SICH
SYSTEMATISCH**
ABBILDUNG 128: QQ-DIAGRAMM
ABBILDUNG 129: HISTOGRAMM
ABBILDUNG 130: POSITIVE AUTOKORRELATION
ABBILDUNG 131: NEGATIVE AUTOKORRELATION
**ABBILDUNG 132: AUTOKORRELATION
(DURBIN/WATSON)**
**ABBILDUNG 133: VENN-DIAGRAMM, GERINGE
MULTIKOLLINEARITÄT**
**ABBILDUNG 134: VENN-DIAGRAMM, KEINE
MULTIKOLLINEARITÄT**

**ABBILDUNG 135: VENN-DIAGRAMM, HOHE
MULTIKOLLINEARITÄT (KATASTROPHE)**

ABBILDUNG 136: KORRELATIONSMATRIX

**ABBILDUNG 137: TOLERANZ & VIF FÜR DAS
FALLBEISPIEL**

ABBILDUNG 138: MULTIVARIATE ANALYSEVERFAHREN

ABBILDUNG 139: FALLBEISPIEL DATENSATZ

ABBILDUNG 140: ZEILENPROFILE

ABBILDUNG 141: SPALTENPROFILE

ABBILDUNG 142: MASSEN DER ZEILEN

ABBILDUNG 143: MASSEN DER SPALTEN

ABBILDUNG 144: GEOMETRISCHES MODELL

ABBILDUNG 145: GRAFISCHE DARSTELLUNG

ABBILDUNG 146: KREUZTABELLE MARKE*MERKMAL

ABBILDUNG 147: CHI-QUADRAT-TESTS

ABBILDUNG 148: TOTALE INERTIA

ABBILDUNG 149: RELATIVE HÄUFIGKEITEN

ABBILDUNG 150: ZENTRIERTE WERTE

ABBILDUNG 151: STANDARDISIERTE DATEN

ABBILDUNG 152: SPSS OUTPUT

ABBILDUNG 153: ÜBERSICHT ZEILENPUNKTE

ABBILDUNG 154: ÜBERSICHT SPALTENPUNKTE

ABBILDUNG 155: SPSS ZEILENPUNKTE

ABBILDUNG 156: SPSS SPALTENPUNKTE

ABBILDUNG 157: ZEILENPUNKTE FÜR MARKE

ABBILDUNG 158: ZEILENPUNKTE FÜR MERKMAL

ABBILDUNG 159: ZEILEN- UND SPALTENPUNKTE

**ABBILDUNG 160: ZEILEN-PRINZIPAL-
NORMALISIERUNG**

ABBILDUNG 161: SPALTEN-PRINZIPAL-NORMALISIERUNG

ABBILDUNG 162: VORGEHENSWEISE SPSS

ABBILDUNG 163: STRUKTUR DER ANOVA

ABBILDUNG 164: STREUUNGSZERLEGUNG

ABBILDUNG 165: STREUUNGSZERLEGUNG

ABBILDUNG 166: EFFEKTSTÄRKE

ABBILDUNG 167: VORGEHEN IN SPSS

ABBILDUNG 168: SPSS AUSGABEN

ABBILDUNG 169: ODINAL

ABBILDUNG 170: DISORDINAL

ABBILDUNG 171: HYBRID

ABBILDUNG 172: VORGEHEN IN SPSS

ABBILDUNG 173: SPSS AUSGABEN

ABBILDUNG 174: VORGEHEN IN SPSS

ABBILDUNG 175: SPSS AUSGABEN

ABBILDUNG 176: VORGEHEN IN SPSS

ABBILDUNG 177: SPSS AUSGABEN

Die Analyse fehlender Werte

Eine Einführung

Was sind fehlende Werte?

„Missing data, where valid values on one or more variables are not available for analysis, are a fact of life in multivariate analysis.“ (Hair, Black, Babin, & Anderson, 2010, S. 42).

Missing Values sind fehlende Werte in den erhobenen Daten, die sich auf wenigen Variablen/ Fällen sammeln oder über den gesamten Datensatz verteilt vorhanden sein können. Sie bedrohen die Validitäten der durchgeführten Studie und können das Verständnis für Effekte beeinflussen. Besonders in der multivariaten Datenanalyse kommen fehlende Werte sehr häufig vor. Aus diesem Grund ist ihre Betrachtung von großer Bedeutung. (Bankhofer, 1995) & (Hair, 2010, S. 42-43)

Wodurch entstehen fehlende Werte?

Missing Values können aus drei folgenschweren Gründen fehlen. Ein erster Aspekt ist der Proband selbst, der das Vorkommen fehlender Werte verursacht. Das ist beispielsweise der Fall, wenn jemand eine Frage nicht beantworten möchte. Ein zweiter Grund ist das Studiendesign selbst, das z.B. durch seine Länge zu viel Zeit von den Probanden verlangt und aus diesem Grund zu fehlenden Werten führt. Ein letzter grundlegender Aspekt ist die Interaktion zwischen Probanden und dem Studiendesign/ Versuchsleiter. Probanden sind eventuell nicht fähig genug

eine Frage zu beantworten. Es könnte auch sein, dass die Fragen zu intim sind und der Proband nicht antworten möchte. (Ausführliche Darbietung der Gründe siehe: die Ursache fehlender Werte) (Bankhofer, 1995, S. 5-12)

Warum sind fehlende Werte ein Problem?

Fehlende Werte haben einen Einfluss auf die Generalisierbarkeit der Daten. Beispielsweise ist die Verallgemeinerung der Ergebnisse nicht möglich, wenn die Mehrheit der Studienteilnehmer, denen es während des Experiments schlecht erging, abbricht und deswegen die Ergebnisse vor allem auf den Probanden beruhen, die positiv auf die Bedingung reagiert haben. Ebenfalls können falsche Schlussfolgerungen gezogen werden, wenn die fehlenden Daten die Beziehungen zwischen den Variablen beeinflussen. Außerdem können Missing Values (fehlende Werte) zu nichtrepräsentativen Stichproben führen, wenn als Verfahren z.B. der listenweise Fallausschluss (siehe Kapitel: Vorgehen) genutzt wird. Ebenfalls haben fehlende Werte einen Einfluss auf die unterschiedlichen Validitäten, was mit dem Problem der Generalisierbarkeit einhergeht. (Hair, 2010, S. 42-43) & (Bankhofer, 1995, S. 25-27)

Die Ursache fehlender Werte

Die Ursachen fehlender Werte sind in drei großen Bereichen der empirischen Forschung zu finden. Beginnend mit dem Untersuchungsdesign, über die Datenerhebung selbst und

auch in der Datenauswertung können Gründe für das Entstehen fehlender Werte gefunden werden.

a) Untersuchungsdesign

Grundsätzlich gilt, dass das Untersuchungsdesign so umfangreich wie nötig und so knapp wie möglich sein sollte, um möglichst viele Informationen durch möglichst wenige Variablen und wenig (Zeit)Aufwand der Probanden zu generieren. Folgende Ursachen können Gründe für das Entstehen von fehlenden Werten sein: (Bankhofer, 1995, S. 8-12)

- Fehlerhaftes Untersuchungsdesign:

Das Untersuchungsdesign wird so bestimmt, dass es sehr wahrscheinlich abzusehen ist, dass es zu fehlenden Werten kommt. Beispiel: Wenn in einer Studie das Merkmal „Alter des Ehepartners“ abgefragt wird und sich jedoch auch unverheiratete Personen unter den Probanden befinden, kommt es, soweit keine „Non-Option“ gibt, zu fehlenden Werten. Ein weiteres Beispiel wäre: Es wird nach dem Alter der Kinder gefragt, ohne angeben zu können, keine zu haben. Bei kinderlosen Personen kommt es hier zu fehlenden Werten. (Bankhofer, 1995, S. 8-12)

- Mangelhaftes Untersuchungsdesign:

Das Design kann durch unübersichtliche Anordnungen, missverständliche Fragen, unbekanntem Wörtern oder Redewendungen in den Fragestellungen fehlende Werte verursachen. (Bankhofer, 1995, S. 8-12)

b) Die Datenerhebung

Auch die Datenerhebung selbst kann für fehlende Werte verantwortlich sein. Folgende Gründe könnte es geben:

Übersehen von Fragen:

Bei schriftlichen Befragungen kann es sein, dass der Proband eine Frage übersieht. Das kann dem Interviewer auch bei einer mündlichen Untersuchung passieren. Gründe dafür sind zumeist Ablenkung oder technische Mängel in der Untersuchung. (Bankhofer, 1995, S. 8-12)

Mangelndes Wissen:

Der Proband ist selbst mit Mühe nicht in der Lage eine Frage zu beantworten. Das ist besonders dann der Fall, wenn die Fragen auf Ereignisse abzielen, die längere Zeit zurückliegen. (Bankhofer, 1995, S. 8-12)

Dieser Art der fehlenden Werte klammern Angaben, wie z.B. „Ich weiß nicht“ aus, da diese vor allem auf eine Unentschlossenheit oder Unentschiedenheit der Probanden zurückgehen. Solche unentschlossenen Angaben sind keine fehlenden Werte, sondern Informationen.

Antwortverweigerung:

Ein Proband verweigert die Antwort auf eine Frage, unabhängig ob diese schriftlich oder mündlich gestellt wurde. Insbesondere tritt diese Ursache fehlender Werte dann auf, wenn die befragten Personen eine Verletzung der Privatsphäre vermuten. Beispiel: Probanden beantworten Fragen nach dem eigenen Trink-, Rauch- oder Sexualverhalten bewusst nicht. (Bankhofer, 1995, S. 8-12)

Motivationsprobleme:

Geringe Motivation von dem Interviewer oder der Testperson sorgen ebenfalls für Missing Values. Die fehlende Motivation geht auf eine Abneigung gegen die Befragungssituation, den Interviewer, die Erhebung allgemein oder auf die Länge der Befragung zurück. Ein Mangel an Motivation kann, z.B. auf geringe Vergütung, zurückgeführt werden. (Bankhofer, 1995, S. 8-12)

Verständnisprobleme:

Der Befragte kann die ihm gestellte Frage nicht richtig verstehen und hat deswegen Probleme zu antworten. (Bankhofer, 1995, S. 8-12)

Meinungslosigkeit:

Der Proband gibt keine Antwort auf eine Frage nach seiner Meinung, weil er entweder keine Meinung zu dieser Thematik hat oder sie nicht formulieren kann. (Bankhofer, 1995, S. 8-12)

Zeitprobleme:

Eine befragte Person hat zur Beantwortung eines Fragebogens nur limitiert Zeit und ist nicht in der Lage, alle ihr gestellten Fragen in dieser Zeit zu beantworten. In solchen Fällen sammeln sich die fehlenden Werte am Ende der Befragung. (Bankhofer, 1995, S. 8-12)

Filterführung des Interviewers:

Während der Befragung überspringt der Interviewer absichtlich bestimmte Fragen. Beispiel: Die Befragten, die angeben Raucher zu sein, werden weitere Fragen zu diesem Thema gestellt. Andere, die angeben Nichtraucher zu sein,

haben dadurch bei den Fragen über Rauchen fehlende Werte. (Bankhofer, 1995, S. 8-12)

Unaufmerksamkeit des Beobachters:

Werden in einer Studie auf Beobachtungen wert gelegt, erhält man fehlende Werte, wenn der Beobachter in der Zeit unaufmerksam ist und ihm deshalb der zu beobachtenden Sachverhalt entgeht. (Bankhofer, 1995, S. 8-12)

Technische Mängel:

Fehlende Werte können auf technische Defekte zurückgeführt werden. Beispiel: Bei Video- und Blickaufzeichnungen (z.B. Eye-Tracking-Studien) wird durch einen technischen Mangel der zu beobachtende Sachverhalt (die Blickrichtung) verpasst bzw. nicht erfasst. (Bankhofer, 1995, S. 8-12)

c) Die Datenauswertung

Auch in der Datenauswertung können einige Ursachen für fehlende Werte gefunden werden.

Codierungsfehler:

Fehlende Werte entstehen dann, wenn eigentlich vorhandene Werte irrtümlicherweise als fehlend codiert werden. (Bankhofer, 1995, S. 8-12)

Übertragungsfehler:

Wenn Daten beispielsweise von einem schriftlichen Fragebogen in SPSS überführt werden und dabei manuell

eingetippt werden müssen, kann es zu fehlenden Werten kommen. Diese können auch durch Fehler beim elektronischen Einlesen von Daten und deren Abspeicherung hervorgerufen werden. (Bankhofer, 1995, S. 8-12)

Löschung von unmöglichen Daten:

Ein Wert wird aus dem Datensatz entfernt, weil er in diesem Kontext unmöglich sein kann, was bei der Überprüfung der Daten sichtbar wird. Beispiel: Alphanumerische Zeichen werden gelöscht, wenn sie bei einer Variablen auftreten, die eigentlich nur numerische Ausprägungen besitzen darf. (Bankhofer, 1995, S. 8-12)

Löschung von fehlerhaften Daten:

Daten werden bei der Überprüfung als fehlerhaft erkannt und vom Forscher oder der Forscherin entfernt, obwohl sie grundsätzlich realisierbar wären und eigentlich korrekt sind. (Bankhofer, 1995, S. 8-12)

Folgen und Konsequenzen fehlender Werte

Aus dem Problem der fehlenden Werte können konkrete Folgen für die Praxis abgeleitet werden. Diese sind unter anderem Fehler in den statistischen Daten, Biases (=systematische Fehler) in den Parameterschätzungen, fehlerhafte Interpretationen der Ergebnisse sowie das Begrenzen der Repräsentativität und der Stärke der Stichprobe. So kann es durch Missing Values zu fehlerhaften statistischen Ergebnissen kommen, die bei Beachtung der

fehlenden Werte anders ausgefallen wären. Werden fehlende Werte in den Datensätzen von Forschern vernachlässigt, ist dieses Vorgehen nicht mit einer wissenschaftlichen Arbeitsweise vereinbar. Besonders die Validität einer Studie wird von Missing Values beeinflusst. (Bankhofer, 1995, S. 25-27) & (Hair, 2010, S. 42-43)

Beeinflussung der Validität

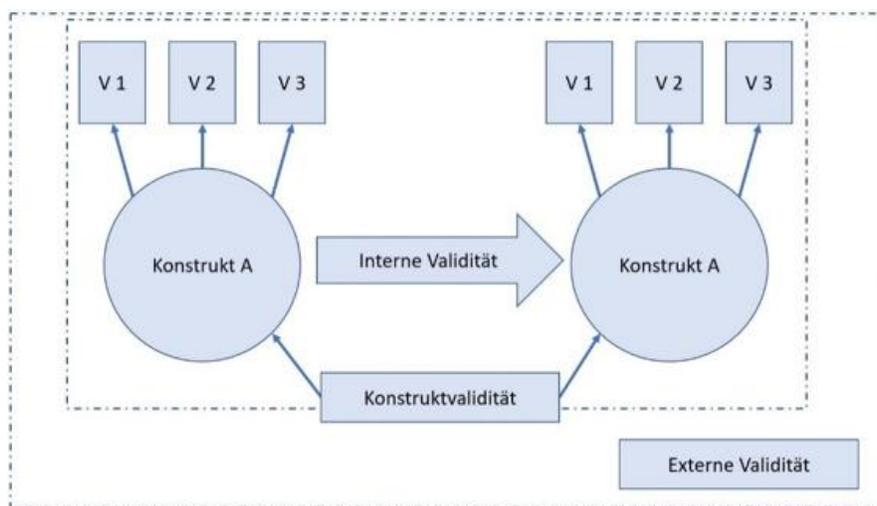


Abbildung 1: Übersicht Validität
Eigene Darstellung, Quelle: vgl. (Lüdtke & Trautwein, 2007)

a) Konstruktvalidität

Konstruktvalidität gibt an, wie gut die erfassten Variablen das vorgesehene Konstrukt messen und wie gut die Werte die relevanten Variablen voraussagen. Ist die Konstruktvalidität hoch, dann fallen die Testwerte so aus, wie es die Theorie und die Hypothese vorgibt. (Döring & Bortz, 2016)

Durch fehlende Werte in den Variablen ist es möglich, dass das zu erfassende Konstrukt (eine Fragestellung oder eine Hypothese) dahinter nicht vollständig oder fehlerhaft erschlossen wird. Besonders im Fall der mono-Operations sind fehlende Daten besonders nachteilig. Mono-Operations sind einzelne Messungen, die durch ein einzelnes Item oder multiple Items bestehen und zur Repräsentation des Konstrukts bzw. Konzepts dienen. Das ist z.B. dann der Fall, wenn das Rauchverhalten (als Konstrukt) durch die Anzahl konsumierter Zigaretten in den letzten 24 Stunden gemessen wird. Fehlt der Wert für die konsumierten Zigaretten, kann das Konstrukt Rauchen nicht erfasst werden. (Allison, 2002) & (McKnight, McKnight, Sidani, & Figueredo, 2007) & (Schafer & Graham, 2002)

Ein kleinerer Item-Pool führt ebenfalls mit fehlenden Werten zu einer größeren Fehlervarianz und diese dann wiederum zu geringerer Reliabilität in den Messungen. Der Informationsverlust, welcher mit Missing Values einhergeht, führt außerdem zu unvollständigen Repräsentationen des Konstrukts und damit automatisch zu einer kleineren Konstruktvalidität.

b) Interne Validität

Die interne Validität ist gegeben, wenn Unterschiede in der abhängigen Variablen mit Sicherheit auf Unterschiede in der unabhängigen Variablen zurückgeführt werden können.

Die interne und externe Validität beeinflussen sich gegenseitig. Durch die mit fehlenden Werten einhergehende Stichprobenselektion gibt es Unterschiede in den Gruppen, die wiederum zu einem systematischen Fehler führen. Dadurch entstehen nicht-repräsentative Stichproben, die

neben der internen auch die externe Validität verkleinern. Durch Randomisierung, d.h. Probanden werden zufällig auf unterschiedliche und identische Gruppen aufgeteilt, kann es auch zu fehlenden Werten kommen. Wenn es teilnehmerbedingt zu fehlenden Werten kommt, verursacht dieser Grund der fehlenden Werte ungleiche Stichprobengrößen, was wiederum die statistischen Voraussetzungen für viele Vorgänge in SPSS verletzt, wodurch es zu einer kleineren Validität in den Ergebnissen der statistischen Auswertung kommt. (Allison, 2002) & (McKnight, McKnight, Sidani, & Figueredo, 2007) & (Schafer & Graham, 2002)

c) Externe Validität

Die externe Validität gibt an, inwieweit sich die Testergebnisse über die Untersuchungssituation verallgemeinern lassen (Döring & Bortz, 2016). Bei homogenen Stichproben (viele Probanden ähneln sich) ist die Streuung der Werte/ Antworten gering. Der Test wird also schneller signifikant, aber die Repräsentativität der Ergebnisse für eine heterogene Grundgesamtheit sinkt. Wenn fehlende Werte in den Daten auftreten, dann kann es zu Schwierigkeiten mit den Schlussfolgerungen und Interpretationen der Ergebnisse kommen. Dies kann falsche Rückschlüsse hervorrufen, die dann möglicherweise falsche Richtlinien verursachen. (Allison, 2002) & (McKnight, McKnight, Sidani, & Figueredo, 2007) & (Schafer & Graham, 2002)

Klassifikation der fehlenden Werte

Um die richtigen Methoden zur Behandlung fehlender Werte auswählen zu können, ist es von Bedeutung zu wissen, ob hinter den fehlenden Werten eine Systematik liegt. Die Klassifikation der fehlenden Werte erfolgt nach der Arbeit von Rubin in drei Arten (Rubin, 1976):

a) MCAR – Missing completely at random

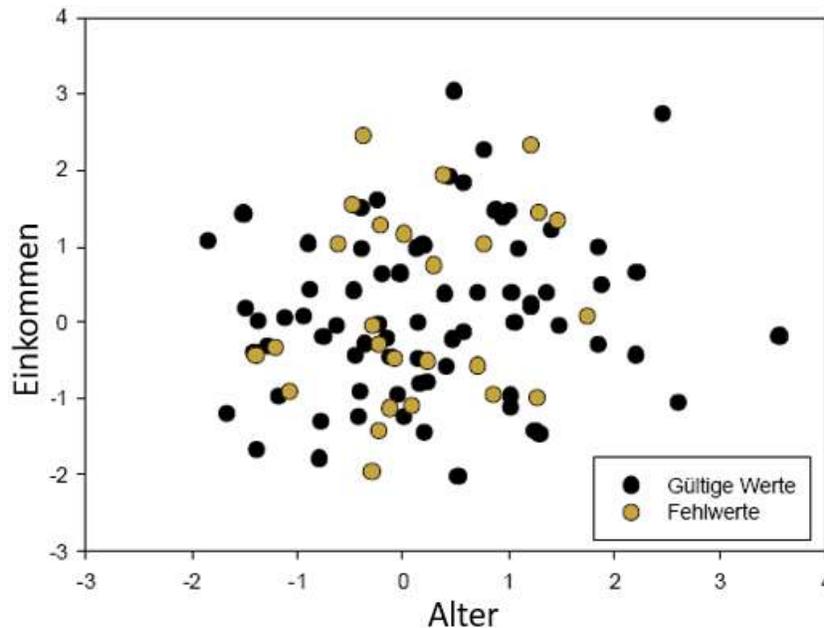


Abbildung 2: Missing at completely random
Eigene Darstellung, Quelle: (IBM®, 2020)

Die Werte in den Daten fehlen vollständig zufällig. Das Auftreten der fehlenden Werte (sowohl X als auch Y) lassen sich weder auf Ausprägungen der betreffenden Variablen