

# **Methoden zur umfassenden Darstellung des Lebensmittelverzehrs am Beispiel von Bio-Käufern und Nicht-Bio-Käufern – Auswertungen auf Basis der Nationalen Verzehrsstudie II**

Dissertation zur Erlangung des akademischen Grades Dr. oec. troph.

Fachbereich Agrarwissenschaften, Ökotrophologie und Umweltmanagement der  
Justus-Liebig-Universität Gießen

vorgelegt von

**Friederike Anna Wittig**

Karlsruhe 2015

**Begutachtet von:**

**Prof. Dr. Ingrid Hoffmann**

Max Rubner-Institut Karlsruhe,  
Institut für Ernährungsverhalten

**Prof. Dr. Michael Krawinkel**

Justus-Liebig-Universität Giessen,  
Institut für Ernährungswissenschaft

*Die umfassende Darstellung des Lebensmittelverzehrs:  
Der griechische Philosoph Aristoteles (384-322 v. Chr.) soll bereits gesagt haben,  
dass das Ganze mehr sei als die Summe seiner Teile (Klir 2001 S 37).*

## Danksagung

Die vorliegende Dissertation ist am Max Rubner-Institut, Bundesforschungsinstitut für Ernährung und Lebensmittel, am Institut für Ernährungsverhalten in Karlsruhe entstanden. Ich danke allen die mich bei der Arbeit unterstützt haben.

- Frau Prof. Dr. Ingrid Hoffmann danke ich für die Bereitstellung des spannenden Themas und die intensive Betreuung der Arbeit am Max Rubner-Institut.
- Bei Herrn Prof. Dr. Michael Krawinkel möchte ich mich für die Betreuung der Arbeit an der Justus-Liebig-Universität bedanken.
- Ein besonderer Dank gilt auch den Kolleginnen und Kollegen am Max Rubner-Institut, vor allem Marianne, Eva, Thorsten, Carina und Suse für die vielen kritischen Anmerkungen, konstruktiven Gespräche und motivierenden Worte.
- Herrn Prof. Joachim Selbig und seiner Arbeitsgruppe Bioinformatik in Potsdam möchte ich für die zielführenden Diskussionen vor allem über den Entscheidungsbaum danken.
- Alexander Roth, Björn Eggert, Franziska Lindner und Iris Zöllner danke ich für die statistische Beratung.
- Monika danke ich für das Korrekturlesen der Arbeit.
- Meiner Familie danke ich für die vielen ermutigenden Worte und die viele Zeit, die sie mir für das Anfertigen der Dissertation geschenkt hat.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis .....</b>	<b>IV</b>
<b>Tabellenverzeichnis .....</b>	<b>VI</b>
<b>Abkürzungsverzeichnis .....</b>	<b>X</b>
<b>1 Einführung und Fragestellung.....</b>	<b>1</b>
<b>2 Datenbasis für die vorliegenden Auswertungen .....</b>	<b>9</b>
2.1 Die Nationale Verzehrsstudie II .....	9
2.2 Statistische Grundlagen für die vorliegenden Auswertungen .....	10
2.3 Einkauf ökologisch erzeugter Lebensmittel.....	12
2.3.1 Einteilung der Studienteilnehmenden anhand des Einkaufs ökologisch erzeugter Lebensmittel .....	13
2.3.2 Lebensmittelverzehr der Bio-Käufer/innen und Nicht-Bio-Käufer/innen.....	15
<b>3 Hypothesenorientierter Ansatz zur umfassenden Darstellung des Lebensmittelverzehrs.....</b>	<b>19</b>
3.1 Indizes.....	19
3.1.1 Theoretischer Hintergrund .....	20
3.1.2 Anwendung von Indizes im Bereich Ernährung.....	23
3.1.3 Anwendung des HEI-NVS II auf den Lebensmittelverzehr der Teilnehmenden der NVS II .....	26
3.1.4 Inhaltliche Ergebnisse.....	30
3.1.5 Methodische Ergebnisse und Diskussion.....	48
<b>4 Explorative Ansätze zur umfassenden Darstellung des Lebensmittelverzehrs.....</b>	<b>52</b>
4.1 Faktorenanalyse .....	53
4.1.1 Wahl eines Verfahrens für die eigene Anwendung .....	56
4.1.2 Theoretischer Hintergrund der PCA.....	57
4.1.3 Anwendung der Faktorenanalyse im Bereich Ernährung .....	61
4.1.4 Anwendung der Hauptkomponentenanalyse auf den Lebensmittelverzehr der Teilnehmenden der NVS II .....	64
4.1.5 Inhaltliche Ergebnisse.....	65
4.1.6 Methodische Ergebnisse und Diskussion.....	79
4.2 Clusteranalyse.....	89
4.2.1 Wahl eines Verfahrens für die eigene Auswertung .....	91
4.2.2 Theoretischer Hintergrund .....	92
4.2.3 Anwendung der Clusteranalyse im Bereich Ernährung .....	94
4.2.4 Anwendung der Clusteranalyse auf den Lebensmittelverzehr der Teilnehmenden der NVS II .....	96

4.2.5	Inhaltliche Ergebnisse.....	97
4.2.6	Methodische Ergebnisse und Diskussion.....	105
4.3	Entscheidungsbaum .....	110
4.3.1	Wahl eines Verfahrens für die eigene Anwendung .....	112
4.3.2	Theoretischer Hintergrund .....	113
4.3.3	Anwendung von Entscheidungsbäumen .....	116
4.3.4	Anwendung des Entscheidungsbaums auf den Lebensmittelverzehr der Teilnehmenden der NVS II .....	118
4.3.5	Inhaltliche Ergebnisse.....	118
4.3.6	Methodische Ergebnisse und Diskussion.....	132
<b>5</b>	<b>Übergreifende Diskussion und Fazit.....</b>	<b>139</b>
5.1	Gegenüberstellung der Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs .....	141
5.2	Limitierungen der Arbeit.....	151
5.3	Fazit .....	152
<b>6</b>	<b>Zusammenfassung.....</b>	<b>154</b>
<b>7</b>	<b>Summary.....</b>	<b>157</b>
<b>8</b>	<b>Literatur.....</b>	<b>159</b>
<b>9</b>	<b>Anhang.....</b>	<b>171</b>
9.1	Lebensmittelgruppen .....	171
9.2	Faktorenanalyse .....	172
9.3	Clusteranalyse.....	184
9.4	Entscheidungsbaum .....	186

## Abbildungsverzeichnis

Abb. 1.1:	Mögliche Ebenen zur Untersuchung des Zusammenhangs zwischen Ernährungsweise/Ernährungsmuster und der Gesundheit des Menschen (modifiziert nach Hoffmann und Wittig 2011 S 113, Hoffmann 2003 S 514f) .....	1
Abb. 1.2:	Methoden zur Darstellung des Lebensmittelverzehrs (Schulze und Hoffmann 2006 S 861, mit eigener Ergänzung).....	4
Abb. 2.1:	Frage zur Kaufhäufigkeit einzelner Lebensmittel(-gruppen) als Bioprodukt (Ausschnitt aus dem Fragebogen der NVS II) .....	14
Abb. 3.1:	Anteil der Bio-Käuferinnen und Nicht-Bio-Käuferinnen, mit einem Indexwert von 10 in einzelnen Lebensmittelgruppen (Frauen).....	32
Abb. 3.2:	Anteil der Bio-Käufer und Nicht-Bio-Käufer, mit einem Indexwert von 10 in einzelnen Lebensmittelgruppen (Männer) .....	34
Abb. 3.3:	Häufigkeitsverteilung der HEI-NVS II-Indexsummen der Bio-Käuferinnen (Frauen) .....	35
Abb. 3.4:	Häufigkeitsverteilung der HEI-NVS II-Indexsummen der Bio-Käufer (Männer) .....	36
Abb. 3.5:	Gegenüberstellung von Bio-Kaufintensität, Gesundheitsverhalten und der HEI-NVS II-Indexsumme (Frauen).....	44
Abb. 3.6:	Gegenüberstellung von Bio-Kaufintensität, Gesundheitsverhalten und der HEI-NVS II-Indexsumme (Männer).....	45
Abb. 3.7:	Gegenüberstellung von Bio-Kaufintensität, Ernährungskennnissen und der HEI-NVS II-Indexsumme (Frauen).....	46
Abb. 3.8:	Gegenüberstellung von Bio-Kaufintensität, Ernährungskennnissen und der HEI-NVS II-Indexsumme (Männer).....	47
Abb. 4.1:	Ermittelte Eigenwerte der Hauptkomponenten der Frauen und Männer .....	66
Abb. 4.2:	Zuordnung der Frauen zu den ersten beiden Hauptkomponenten auf Basis der Faktorwerte .....	82
Abb. 4.3:	Differenzierung der Bio-Käuferinnen und Nicht-Bio-Käuferinnen über die ersten beiden Hauptkomponenten auf Basis der Faktorwerte .....	85
Abb. 4.4:	Zuordnung der Frauen zu den ersten beiden Hauptkomponenten auf Basis der Faktorwerte, getrennt nach der HEI-NVS II-Indexsumme.....	86
Abb. 4.5:	Schematische Darstellung einer Clusterbildung (eigene Darstellung).....	89
Abb. 4.6:	Beispiel für ein Dendrogramm zur Bestimmung der optimalen Clusteranzahl (eigene Darstellung) .....	93
Abb. 4.7:	Obst-, Gemüse- und Fleischverzehrsmenge (g/Tag) aller Frauen (links) und Männer (rechts) zur Prüfung der Eignung der Daten zur Berechnung einer Clusteranalyse .....	98

Abb. 4.8:	Schematische Darstellung eines Entscheidungsbaums (eigene Darstellung) ...	110
Abb. 4.9:	Entscheidungsbaum der Bio-Käuferinnen und Nicht-Bio-Käuferinnen auf Basis der Verzehrsmenge von 13 Lebensmittelgruppen (Frauen) .....	119
Abb. 4.10:	Entscheidungsbaum der Bio-Käufer und Nicht-Bio-Käufer auf Basis der Verzehrsmenge von 13 Lebensmittelgruppen (Männer) .....	126
Abb. 9.1	Zuordnung der Männer zu den ersten beiden Hauptkomponenten auf Basis der Faktorwerte .....	182
Abb. 9.2	Differenzierung der Bio-Käufer und Nicht-Bio-Käufer über die ersten beiden Hauptkomponenten auf Basis der Faktorwerte .....	183
Abb. 9.3:	Entscheidungsbaum der Bio-Käufer/innen, klassifiziert nach Frauen und Männern .....	189
Abb. 9.4:	Entscheidungsbaum der Bio-Käuferinnen, klassifiziert nach „Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“, gemessen an der HEI-NVS II-Indexsumme .....	190
Abb. 9.5:	Entscheidungsbaum der Bio-Käufer, klassifiziert nach „Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“, gemessen an der HEI-NVS II-Indexsumme .....	191

## Tabellenverzeichnis

Tab. 2.1:	Altersgruppen der Unterstichprobe der NVS II (absolute und relative Häufigkeiten) (modifiziert nach Eisinger-Watzl et al. 2010 S 7) .....	13
Tab. 2.2:	Anteile der Bio-Käufer/innen nach Kaufintensität (absoluter und relativer Anteil).....	15
Tab. 2.3:	Verzehnte Lebensmittelmengen (g/Tag) der Bio-Käuferinnen und Nicht-Bio-Käuferinnen (Frauen) (erweitert nach Eisinger-Watzl und Hoffmann 2010 S 34ff).....	16
Tab. 2.4:	Verzehnte Lebensmittelmengen (g/Tag) der Bio-Käufer und Nicht-Bio-Käufer (Männer) (erweitert nach Eisinger-Watzl und Hoffmann 2010 S 34ff) .....	17
Tab. 3.1:	Lebensmittelvariablen und Bewertungsbasis des HEI-NVS II.....	28
Tab. 3.2:	Indexwerte für einzelne Lebensmittelgruppen von Bio-Käuferinnen und Nicht-Bio-Käuferinnen (Frauen).....	31
Tab. 3.3:	Indexwerte für einzelne Lebensmittelgruppen von Bio-Käufern und Nicht-Bio-Käufern (Männer).....	33
Tab. 3.4:	HEI-NVS II-Indexsummen der Bio-Käuferinnen und Nicht-Bio-Käuferinnen (Frauen) .....	34
Tab. 3.5:	HEI-NVS II-Indexsummen der Bio-Käuferinnen unterschiedlicher Kaufintensität und Nicht-Bio-Käuferinnen (Frauen) .....	35
Tab. 3.6:	HEI-NVS II-Indexsummen der Bio-Käufer und Nicht-Bio-Käufer (Männer) .....	36
Tab. 3.7:	HEI-NVS II-Indexsummen der Bio-Käufer unterschiedlicher Kaufintensität und Nicht-Bio-Käufer (Männer).....	37
Tab. 3.8:	Indexgrenzwerte zur Einteilung der Bio-Käuferinnen und Bio-Käufer in die Terzilen: Empfehlungen wenig, mittelmäßig und gut erreicht.....	38
Tab. 3.9:	Charakterisierung der Gesamtgruppe sowie der Bio-Käuferinnen mit unterschiedlicher Lebensmittelauswahl (Frauen).....	39
Tab. 3.10:	Charakterisierung der Gesamtgruppe sowie der Bio-Käufer mit unterschiedlicher Lebensmittelauswahl (Männer).....	41
Tab. 3.11:	Variablen und Bewertungsbasis des Indexes zum Gesundheitsverhalten .....	43
Tab. 4.1:	Beispiel und Interpretation für eine Hauptkomponente (eigene Darstellung) .....	58
Tab. 4.2:	Erklärung der berechneten Faktorladung und des Faktorwertes einer Hauptkomponente am Beispiel von Person A (eigene Darstellung).....	60
Tab. 4.3:	Korrelationsmatrix der in der PCA einbezogenen Lebensmittelgruppen der Frauen und Männer.....	66
Tab. 4.4:	Hauptkomponenten der Frauen mit den jeweiligen Faktorladungen und dem erklärten Varianzanteil.....	67

Tab. 4.5:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der ersten Hauptkomponente .....	68
Tab. 4.6:	Charakterisierung der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der ersten Hauptkomponente .....	70
Tab. 4.7:	Lebensmittelverzehr (g/Tag) Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der zweiten Hauptkomponente .....	71
Tab. 4.8:	Charakterisierung der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der zweiten Hauptkomponente .....	72
Tab. 4.9:	Hauptkomponenten der Männer mit den jeweiligen Faktorladungen und dem erklärten Varianzanteil.....	74
Tab. 4.10:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der ersten Hauptkomponente .....	75
Tab. 4.11:	Charakterisierung der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der ersten Hauptkomponente .....	76
Tab. 4.12:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der zweiten Hauptkomponente .....	77
Tab. 4.13:	Charakterisierung der Gesamtgruppe sowie Männer der ersten und vierten Quartile der zweiten Hauptkomponente.....	78
Tab. 4.14:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen, die eine hohe Faktorladung für die erste oder zweite Hauptkomponente aufweisen .....	83
Tab. 4.15:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen des ersten und zweiten Clusters .....	99
Tab. 4.16:	Charakterisierung der Frauen der Gesamtgruppe sowie der Frauen des ersten und zweiten Clusters .....	100
Tab. 4.17:	Lebensmittelverzehr der Gesamtgruppe sowie der Männer des ersten, zweiten und dritten Clusters .....	102
Tab. 4.18:	Charakterisierung der Gesamtgruppe sowie der Männer des ersten, zweiten und dritten Clusters .....	103
Tab. 4.19:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen, die den einzelnen Blättern des Entscheidungsbaums zugeordnet wurden .....	120
Tab. 4.20:	Prozentuale Verteilung der Bio-Käuferinnen und Nicht-Bio-Käuferinnen auf die einzelnen Blätter des Entscheidungsbaums (Frauen) .....	121
Tab. 4.21:	Korrekt- zu Falschklassifikationsrate des Entscheidungsbaums der Frauen in die Gruppe der Bio-Käuferinnen und Nicht-Bio-Käuferinnen auf Basis der 13 Lebensmittelgruppen.....	123
Tab. 4.22:	Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Bio-Käuferinnen und Nicht-Bio-Käuferinnen mit und ohne 10fache Kreuzvalidierung des Entscheidungsbaums (Frauen).....	123

Tab. 4.23:	Charakterisierung der Gesamtgruppe sowie der Frauen, die dem ersten und vierten Blatt des Entscheidungsbaums zugeordnet wurden.....	124
Tab. 4.24:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer, die den einzelnen Blättern des Entscheidungsbaums zugeordnet wurden .....	127
Tab. 4.25:	Prozentuale Verteilung der Bio-Käufer und Nicht-Bio-Käufer auf die einzelnen Blätter des Entscheidungsbaums (Männer).....	127
Tab. 4.26:	Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Männer in die Gruppe der Bio-Käufer und Nicht-Bio-Käufer auf Basis der 13 Lebensmittelgruppen.....	128
Tab. 4.27:	Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Bio-Käufer und Nicht-Bio-Käufer mit und ohne 10fache Kreuzvalidierung des Entscheidungsbaums (Männer).....	129
Tab. 4.28:	Charakterisierung der Gesamtgruppe sowie der Männer, die dem dritten und fünften Blatt des Entscheidungsbaums zugeordnet wurden .....	130
Tab. 4.29:	Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Bio-Käufer/innen in die Gruppen der Frauen und Männer auf Basis von 13 Lebensmittelgruppen.....	135
Tab. 4.30:	Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Bio-Käuferinnen in die Gruppen „Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“, gemessen an der HEI-NVS II-Indexsumme, auf Basis von 13 Lebensmittelgruppen.....	136
Tab. 4.31:	Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Bio-Käufer, in die Gruppen „Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“, gemessen an der HEI-NVS II-Indexsumme, auf Basis von 13 Lebensmittelgruppen.....	136
Tab. 5.1:	Vergleich der angewendeten Methoden zur umfassenden Darstellung des Lebensmittelverzehrs .....	140
Tab. 5.2:	Übersicht über die Erfüllung der Kriterien, die zur Prüfung der Eignung der angewendeten Methoden zur umfassenden Darstellung des Lebensmittelverzehrs herangezogen wurden .....	143
Tab. 9.1:	Einteilung der Lebensmittel zu den einzelnen Lebensmittelgruppen.....	171
Tab. 9.2:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der dritten Hauptkomponente .....	172
Tab. 9.3:	Charakterisierung der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der dritten Hauptkomponente .....	173
Tab. 9.4:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der vierten Hauptkomponente.....	174
Tab. 9.5:	Charakterisierung der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der vierten Hauptkomponente .....	175

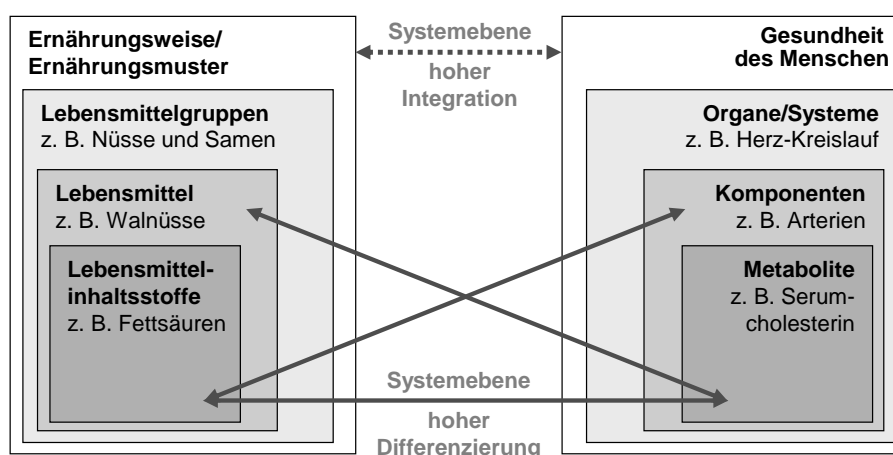
Tab. 9.6:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der dritten Hauptkomponente .....	176
Tab. 9.7:	Charakterisierung der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der dritten Hauptkomponente .....	177
Tab. 9.8:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der vierten Hauptkomponente.....	178
Tab. 9.9:	Charakterisierung der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der vierten Hauptkomponente .....	179
Tab. 9.10:	Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der fünften Hauptkomponente.....	180
Tab. 9.11:	Charakterisierung der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der fünften Hauptkomponente .....	181
Tab. 9.12:	Kriterien zur Festlegung der Clusteranzahl (Frauen) .....	184
Tab. 9.13:	Kriterien zur Festlegung der Clusteranzahl (Männer) .....	185
Tab. 9.14:	Charakterisierung der Gesamtgruppe sowie der Frauen der einzelnen Blätter des Entscheidungsbaums - Teil 1 .....	186
Tab. 9.15:	Charakterisierung der Gesamtgruppe sowie der Frauen der einzelnen Blätter des Entscheidungsbaums - Teil 2 .....	187
Tab. 9.16:	Charakterisierung der Gesamtgruppe sowie der Männer der einzelnen Blätter des Entscheidungsbaums .....	188

## Abkürzungsverzeichnis

AICR	American Institute for Cancer Research
AID	aid Infodienst Ernährung, Landwirtschaft, Verbraucherschutz e. V.
BLS	Bundeslebensmittelschlüssel
BMELV	Bundesministerium für Ernährung und Landwirtschaft, jetzt: BMEL Bundesministerium für Ernährung und Landwirtschaft
BMI	Body-Mass-Index (Gewicht durch Körpergröße in Meter zum Quadrat)
BÖL	Bundesprogramm ökologischer Landbau (jetzt: BÖLN Bundesprogramm Ökologischer Landbau und andere Formen nachhaltiger Landwirtschaft)
CAPI	Computer assisted personal interview
CART	Classification and regression tree
CHAID	Chi-square based automatic interaction detector
CLS	Concept learning system
DGE	Deutsche Gesellschaft für Ernährung e. V.
DISHES	Dietary Interview Software for Health Examination Studies
EFA	explorative/exploratorische Faktorenanalyse
EMSIG	Ernährungsmodellstudie in Gießen
FKE	Forschungsinstitut für Kinderernährung
HEI-EPIC	Healthy Eating Index of the European Prospective Investigation into Cancer and Nutrition study
HEI-NVS II	Healthy Eating Index der Nationalen Verzehrsstudie II
KI-MW	95. Konfidenzintervall des arithmetischen Mittelwertes
ML	maschinelles Lernen
MRI	Max Rubner-Institut, Bundesforschungsinstitut für Ernährung und Lebensmittel
MW	arithmetischer Mittelwert
n. s.	nicht signifikant
NVS II	Nationale Verzehrsstudie II
P5	5. Perzentil
P25	25. Perzentil
P75	75. Perzentil
P95	95. Perzentil
P50	50. Perzentil (Median)
PCA	Principal component analysis
SÖS	sozioökonomischer Status
Std dev	Standard deviation
TDIDT	Top-down induction of decision tree
WCRF	World Cancer Research Fund
WHO	World Health Organisation

## 1 Einführung und Fragestellung

Es gibt eine Vielzahl an Möglichkeiten, die Ernährung von Personengruppen darzustellen und zu analysieren. Im Wesentlichen lassen sich diese Möglichkeiten in vier Untersuchungsebenen beschreiben (van Dam 2005 S 573, Jacobs und Steffen 2003 S 508S). Zu diesen Untersuchungsebenen gehören: die unterste, differenzierteste Ebene der **Lebensmittel-inhaltsstoffe**, z. B. Fettsäuren, die Ebene der **Lebensmittel**, z. B. Walnüsse, die Ebene der **Lebensmittelgruppen**, z. B. Nüsse und Samen, sowie die oberste Ebene der **Ernährungsweisen/Ernährungsmuster**, z. B. eine mediterrane Ernährung (Abb. 1.1). Diese Ebenen werden verwendet, um auf ebenso unterschiedlichen Ebenen wie die der Metabolite, Komponenten, Organe/Systeme sowie der Gesundheit des Menschen Zusammenhänge zu untersuchen. Dabei wird nicht zwangsläufig der Zusammenhang zwischen der untersten Ebene auf Seiten der Ernährung, z. B. Fettsäuren, mit der untersten Ebene der Metabolite untersucht. Es können auch Untersuchungen zwischen verschiedenen Ebenen durchgeführt werden (Hoffmann und Wittig 2011 S 113, Hoffmann 2003 S 514f).



**Abb. 1.1: Mögliche Ebenen zur Untersuchung des Zusammenhangs zwischen Ernährungsweise/Ernährungsmuster und der Gesundheit des Menschen**  
(modifiziert nach Hoffmann und Wittig 2011 S 113, Hoffmann 2003 S 514f)

Für die Bearbeitung ernährungsassoziierter Fragestellungen ist es daher wichtig zu klären, „auf welchen Systemebenen Wissen generiert werden soll und wie die Forschung bzw. die Forschungsergebnisse der verschiedenen Ebenen in Verbindung gebracht werden können, um zu einer Synthese des Wissens zu kommen“ (Schneider und Hoffmann 2011 S 38).

Auf der untersten Ebene werden einzelne oder wenige **Nährstoffe oder andere Inhaltsstoffe** auf der gleichen Ebene mit spezifischen Metaboliten aus dem Blut oder Urin oder auf der obersten Ebene mit dem Risiko für das Auftreten verschiedener Krankheiten verknüpft. Der Ansatz, auf der untersten Ebene Untersuchungen durchzuführen, ist von Bedeutung, um

die Wirkung einzelner Lebensmittelinhaltsstoffe im Menschen zu verstehen (Messina et al. 2001 S 1416). Beispielsweise werden Zusammenhänge zwischen der Zufuhr von Fettsäuren über Lebensmittel (unterste Ebene) und dem Krebsrisiko (oberste Ebene) (Beispiel: Jakovljevic et al. 2002) oder zwischen der Zufuhr von Ballaststoffen über Lebensmittel (unterste Ebene) und weiteren Krankheiten (oberste Ebene) untersucht (Beispiel: Ganesan et al. 2012).

Durch den Ansatz, ausschließlich einzelne oder wenige Lebensmittelinhaltsstoffe zu betrachten, werden Beziehungen zwischen der Ernährung und Erkrankung vereinfacht dargestellt. Die Beziehungen zwischen den verzehrten Lebensmitteln und der menschlichen Physiologie sind jedoch sehr komplex (Jacobs und Steffen 2003 S 508S). Einzelne Inhaltsstoffe beeinflussen sich gegenseitig (Mursu et al. 2013 S 444, Schulze und Hoffmann 2006 S 860, 867, Jacques und Tucker 2001 S 1) in ihrer Bioverfügbarkeit und Absorption (Schulze und Hoffmann 2006 S 860). Daher werden bei der Untersuchung einzelner oder weniger Nährstoffe vielen Interaktionen zwischen der Ernährung und Erkrankungen nicht Rechnung getragen (Jacques und Tucker 2001 S 1). Beispielsweise werden durch eine ballaststoffreiche Ernährung in der Regel auch viel Vitamin C, Folsäure, Carotinoide, Magnesium und Kalium zugeführt. Wenn der Zusammenhang zwischen Ballaststoffen und einem Erkrankungsrisiko festgestellt wird, ist nicht sichergestellt, ob der Effekt möglicherweise auch durch einen oder mehrere andere mit der Nahrung aufgenommene Lebensmittelinhaltsstoffe wie Folsäure oder Carotinoide verursacht wird (Jacques und Tucker 2001 S 1).

Der Zusammenhang der Ernährung und dem Gesundheitszustand des Menschen lässt sich nicht durch das Wissen über einzelne oder wenige isolierte Untersuchungen zur Wirkung von Lebensmittelinhaltstoffen und dem Gesundheitszustand erklären, sondern nur durch die kombinatorische Wirkung aller Inhaltsstoffe. Dieses Zusammenspiel lässt sich auch mit dem Begriff Emergenz erklären und soll bereits von Aristoteles mit den Worten „Das Ganze ist mehr als die Summe seiner Teile“ ausgedrückt worden sein (Klir 2001 S 37). Wie von Mahner und Bunge (2000 S 32) beschrieben, wird die Eigenschaft eines Ganzen, die keines seiner Teile besitzt, als emergent bezeichnet. Die Eigenschaft kann nicht aus dem Wissen über die einzelnen Teile abgeleitet werden. Die alleinige Betrachtung isolierter Substanzen begrenzt das Verständnis komplexer Interaktionen.

Es wird empfohlen, neben Untersuchungen zur Wirkung einzelner Lebensmittelinhaltsstoffe auch Untersuchungen zu deren Interaktionen mit anderen Lebensmittelinhaltsstoffen durchzuführen (Messina et al. 2001 S 1416f). Hinzu kommt, dass durch die Grundlagenforschung noch nicht alle Funktionen der Lebensmittelinhaltsstoffe bekannt sind (Oltersdorf 1995 S 30) und das Wissen über die Zusammensetzung einzelner Lebensmittel nicht vollständig geklärt ist (Jacobs et al. 2009 S 1544S).

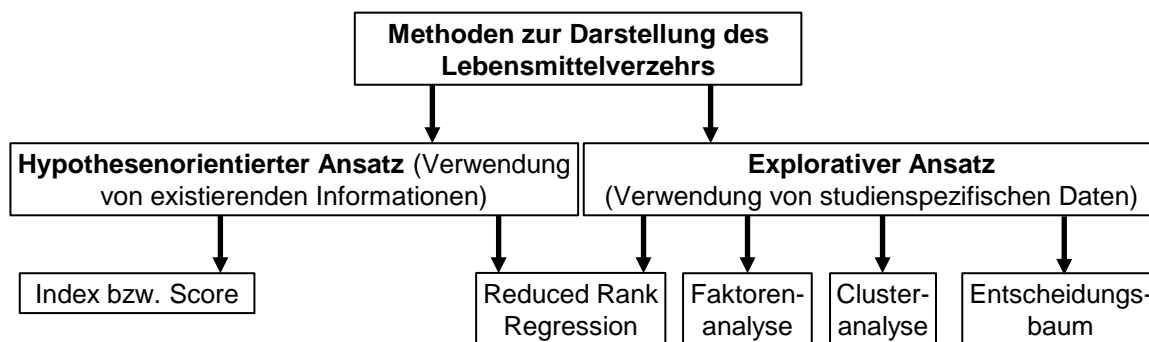
Auf Ebene der **Lebensmittel** werden einzelne oder wenige Lebensmittel beispielsweise im Zusammenhang mit Krankheiten untersucht. Die Analyse auf dieser Ebene kann dabei unterstützen, Erkenntnisse zu gewinnen, die durch nährstoffbasierte Analysen unerkannt bleiben (Jacques und Tucker 2001 S 1). Interaktionen, wie die gegenseitige Beeinflussung von Nährstoffen bezüglich der Bioverfügbarkeit und Absorption (Schulze und Hoffmann 2006 S 860) innerhalb eines Lebensmittels werden durch diesen Ansatz automatisch berücksichtigt. Zum Beispiel wird durch diesen Ansatz berücksichtigt, dass Vitamin C die Resorptionsquote von pflanzlichem Eisen (Biesalski und Grimm 2004 S 230) bei Spinat erhöht. Allerdings werden Lebensmittel in Kombination mit weiteren Lebensmitteln verzehrt und wirken in dieser Kombination (Hoffmann et al. 2004 S 935). So wird durch den gemeinsamen Verzehr von Kartoffeln und Ei die biologische Wertigkeit limitierender Aminosäuren verbessert (Kasper 2000 S 27, 343). Ebenso wirkt die Kombination verschiedener Lebensmittel z. B. auf die Entstehung verschiedener Krankheiten (Hoffmann et al. 2004 S 935, Jacobs und Steffen 2003 S 508S). Wie auf Ebene der Lebensmittelinhaltsstoffe ist es auch auf Ebene der Lebensmittel schwer, aufgrund vielfältiger Interaktionen bei dem Verzehr verschiedener Lebensmittel, spezifische Gesundheitseffekte zu isolieren (Jacques und Tucker 2001 S 1f).

Auf Ebene der **Lebensmittelgruppen** wird der Zusammenhang zwischen einzelnen oder wenigen Lebensmittelgruppen einer Personengruppe und beispielsweise einer Erkrankung untersucht. Hierdurch werden Interaktionen einzelner Lebensmittel einer Lebensmittelgruppe automatisch berücksichtigt. Je nach Untersuchungsziel kann auch diese Betrachtung mit Nachteilen behaftet sein. So weisen Jacques und Tucker (2001 S 1) darauf hin, dass Personen, die Vollkornprodukte verzehren, häufig weniger Fleisch und mehr Obst, Gemüse sowie Fisch verzehren. Damit ist bei Untersuchungen auf dieser Ebene nicht ersichtlich, ob z. B. ein positiver Effekt auf eine Erkrankung durch den Verzehr von Vollkornprodukten oder auf den gleichzeitigen Verzehr von Obst- und Gemüse zurückzuführen ist (Jacques und Tucker 2001 S 1). Multikausalitäten mehrerer Lebensmittelgruppen können bei Untersuchungen auf dieser Ebene nicht berücksichtigt werden. Es kann beispielsweise nicht erfasst werden, ob mehrere Lebensmittelgruppen gleichzeitig durch deren Zusammenspiel einen Effekt auf den untersuchten Zusammenhang ausüben (Hoffmann und Wittig 2011 S 114).

Auf der obersten Ebene, der Ebene der höchsten Integration, der **Ernährungsweisen/ Ernährungsmuster**, wird der Verzehr diverser Lebensmittelgruppen gemeinsam berücksichtigt (Hoffmann 2003 S 514f). Ziel dabei ist, die Ernährung möglichst umfassend abzubilden (Heidemann 2006 S 1131, Naska et al. 2006 S 182, Heidemann et al. 2005 S 1131). Dazu werden die verzehrten Lebensmittel diverser Lebensmittelgruppen gemeinsam erfasst und damit die komplexen Wechselwirkungen zwischen verschiedenen Lebensmittel-

gruppen berücksichtigt. Durch diesen Ansatz wird der Tatsache Rechnung getragen, dass Menschen weder isoliert einzelne Nährstoffe noch Lebensmittel oder Lebensmittel einer Gruppe verzehren, sondern die Ernährung aus einer Vielzahl an Lebensmitteln und den darin enthaltenen Substanzen besteht (Schulze und Hoffmann 2006 S 860).

Die Untersuchung von **Ernährungsweisen/Ernährungsmustern** hat sich neben der Untersuchung isolierter, einzelner Aspekte der Ernährung in der epidemiologischen Forschung als weiterer Forschungsansatz zur Untersuchung der Ernährung etabliert (Heidemann 2006 S 21, Schulze und Hoffmann 2006 S 860, Newby und Tucker 2004 S 177, Hu 2002 S 3). Mit diesem Ansatz wird der Tatsache Rechnung getragen, dass nicht einzelne Lebensmittelinhaltsstoffe, Lebensmittel oder Lebensmittelgruppen für die Untersuchung der Ernährung, sondern deren Kombination von Bedeutung sind. In der Literatur werden hierfür zwei Ansätze beschrieben: der **hypothesenorientierte** und der **explorative** Ansatz (Abb. 1.2).



**Abb. 1.2: Methoden zur Darstellung des Lebensmittelverzehrs**  
(Schulze und Hoffmann 2006 S 861, mit eigener Ergänzung)

Bei einem **hypothesenorientierten** Ansatz (a-priori) werden bereits existierende Informationen zu Grunde gelegt (Schulze und Hoffmann 2006 S 861, Hu 2002 S 4). Hierzu gehören beispielsweise Empfehlungen zur Nährstoffzufuhr, zum Lebensmittelverzehr sowie Informationen über den Zusammenhang zwischen einzelnen Nährstoffen und Krankheiten. Hierfür stehen Indizes bzw. Scores zur Verfügung.

Bei einem **explorativen** Ansatz (a-posteriori) werden dagegen studienbezogene Daten, die unabhängig von bisher existierenden Informationen wie Lebensmittelempfehlungen sind, herangezogen (Schulze und Hoffmann 2006 S 861). Hierfür stehen Faktorenanalysen, Clusteranalysen sowie Entscheidungsbäume zur Verfügung. Bei der Methode Reduced Rank Regression handelt es sich um eine Mischform aus dem hypothesenorientierten und explorativen Ansatz (Schulze und Hoffmann 2006 S 861).

In zahlreichen Studien werden Indizes, Clusteranalysen und Faktorenanalysen zur Untersuchung von Ernährungsweisen/Ernährungsmustern angewendet. Sowohl bei dem hypothesenorientierten als auch bei dem explorativen Ansatz wird üblicherweise das Ziel verfolgt, den Ursache-Wirkungs-Zusammenhang zwischen Nährstoffzufuhr bzw. Lebens-

mittelverzehr und Bioparametern wie Blutwerten oder dem Risiko für die Entwicklung einer bestimmten Erkrankung zu untersuchen (Beispiele: Devlin et al. 2012, Brennan et al. 2010, Edefonti et al. 2009, Waijers et al. 2007, Newby und Tucker 2004, Kant 1996).

Daneben wird mit Indizes, Faktoren- und Clusteranalysen das Ziel verfolgt, die Ernährung umfassend darzustellen und nicht den Ursache-Wirkungs-Zusammenhang mit weiteren Variablen wie Bioparametern oder Erkrankungen zu untersuchen (Beispiele: del Mar Bibiloni et al. 2012, McGowan und McAuliffe 2012, Richter et al. 2012, Torjusen et al. 2012, Hare-Bruun et al. 2011, Olinto et al. 2011). Die Anwendung und Analyse der genannten Methoden mit dem Ziel der umfassenden Darstellung der Ernährung wurde, im Vergleich zur Anwendung und Analyse von Ursache-Wirkungs-Zusammenhängen, bisher wenig untersucht. Für die vorliegende Arbeit stellt sich die Frage, welche Methode geeignet ist, um den Lebensmittelverzehr als solchen, unabhängig von der Untersuchung von Zusammenhängen mit Krankheiten oder Bioparametern, umfassend darzustellen.

### **Zielsetzung und Vorgehensweise**

In der vorliegenden Dissertation werden verschiedene Methoden in Bezug auf ihre Eignung für eine umfassende Darstellung des Lebensmittelverzehrs von Personengruppen am Beispiel von Bio-Käufer/innen und Nicht-Bio-Käufer/innen untersucht und verglichen. Die hierfür herangezogenen Daten zum Lebensmittelverzehr wurden im Rahmen der Nationalen Verzehrsstudie II (NVS II) erfasst.

Den Lebensmittelverzehr umfassend darzustellen erfordert, dass der Lebensmittelverzehr möglichst vollständig erfasst und dargestellt wird. Es sollen möglichst viele Lebensmittel mit den entsprechenden Verzehrsmengen berücksichtigt werden. Zudem ist es erforderlich, dass sowohl Personengruppen, die sich ähnlich ernähren identifiziert, als auch spezifische Personengruppen wie Bio-Käufer/innen und Nicht-Bio-Käufer/innen über den Lebensmittelverzehr differenziert werden können. Werden solche Personengruppen identifiziert und differenziert, ist es bedeutend, diese mit weiteren Variablen charakterisiert zu können. Um verschiedene Methoden entsprechend auf ihre Eignung hierfür zu analysieren, werden in der vorliegenden Arbeit fünf Bewertungskriterien herangezogen:

1. **Berücksichtigung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen bei Anwendung der Methode:** Die Methoden werden daraufhin überprüft, inwiefern sie erlauben, möglichst alle für das Untersuchungsziel relevanten Lebensmittelgruppen und die Mengen, in denen Lebensmittel verzehrt werden, bei der Anwendung der Methode, einzubeziehen. Je nach Untersuchungsziel können sich die relevanten Lebensmittelgruppen unterscheiden. Für das vorliegende Ziel, eine umfassende Darstellung des Lebensmittelverzehrs vorzunehmen, ist es von Bedeutung, die Lebensmittelgruppen Obst,

Gemüse, Milch, Getreide, Kartoffeln, Fisch, Fleisch, Eier, Fett, Alkohol, Getränke, Nüsse und Samen sowie Süßwaren zu berücksichtigen.

2. **Darstellung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse:** Neben der Frage, ob möglichst alle Lebensmittelgruppen bei der Anwendung der Methoden berücksichtigt wurden, stellt sich die Frage, ob auch durch die Ergebnisse möglichst alle für das Untersuchungsziel relevanten und einbezogenen Lebensmittelgruppen und die Mengen, in denen die Lebensmittel verzehrt werden, dargestellt werden können.
3. **Identifizierung von Personengruppen mit ähnlichem Lebensmittelverzehr:** Hierbei wird überprüft, ob die eingesetzten Methoden es erlauben, Personengruppen mit einem ähnlichen Lebensmittelverzehr zu identifizieren.
4. **Differenzierung von Bio-Käufer/innen und Nicht-Bio-Käufer/innen:** Dieses Kriterium dient der Untersuchung, ob die eingesetzten Methoden es erlauben, Personengruppen wie Bio-Käufer/innen und Nicht-Bio-Käufer/innen aufgrund ihres Lebensmittelverzehrs zu differenzieren, unabhängig davon, ob die Personengruppen auf Basis ihres Lebensmittelverzehrs direkt vor oder durch Anwendung einer Methode oder indirekt nach Anwendung einer Methode differenziert werden.
5. **Charakterisierung der identifizierten und differenzierten Personengruppen:** Es wird untersucht, inwiefern es die verwendeten Methoden erlauben, die über das dritte Kriterium identifizierten Personengruppen mit einem ähnlichen Lebensmittelverzehr sowie die über das vierte Kriterium differenzierten Personengruppen, mit weiteren ernährungs- und gesundheitsrelevanten Variablen zu charakterisieren.

Inwieweit diese fünf Kriterien erfüllt werden, wird als entscheidend dafür angesehen, ob eine Methode zur umfassenden Darstellung des Lebensmittelverzehrs von Personengruppen wie der Bio-Käufer/innen und Nicht-Bio-Käufer/innen der NVS II geeignet ist. Ziel dabei ist, dass alle fünf Kriterien gleichzeitig erfüllt werden.

Die Anwendung der Methoden mit dem Ziel, die Ernährung ohne Untersuchung von Ursache-Wirkungs-Zusammenhängen darzustellen, erfolgte bereits für Indizes (Beispiele: Bingham et al. 2010, Martinez et al. 2010, Previdelli et al. 2010, Bau et al. 2003, Sell et al. 2003), Faktorenanalysen (Beispiele: del Mar Bibiloni et al. 2012, Richter et al. 2012, Torjusen et al. 2012, Charreire et al. 2011, Cutler et al. 2011, Hare-Bruun et al. 2011, Olinto et al. 2011, Northstone und Emmett 2010, Rezazadeh et al. 2010, Arkkola et al. 2008) und Clusteranalysen (Beispiele: McGowan und McAuliffe 2012, Okubo et al. 2011, Ovaskainen et al. 2009, Knol et al. 2005). Um neben diesen Methoden weitere potentielle Methoden zur umfassenden Darstellung des Lebensmittelverzehrs zu identifizieren und untersuchen zu können, wurde eine Literaturrecherche und mehrere Expertengespräche in anderen Forschungsbereichen wie der Lebensstilforschung, dem Marketing und der Bioinformatik

durchgeführt. In den Bereichen Lebensstilforschung und Marketing werden zwar ebenfalls multivariate Reduktionsverfahren angewendet (Pepels 2007b S 95f), die allerdings aufgrund der Zielsetzung, Ursache-Wirkungs-Zusammenhänge zu untersuchen, nicht verwendet werden können. In diesen Bereichen, in denen Typologien gebildet werden, konnten keine potentiellen Methoden identifiziert werden. Aus dem Bereich Bioinformatik wurde die Methode des Entscheidungsbaums, die zur Darstellung des Lebensmittelverzehrs bisher nicht beschrieben wird, als potentielle Methode identifiziert. Da mit der Methode Reduced Rank Regression das Ziel verfolgt wird, direkt den Ursache-Wirkungs-Zusammenhang zwischen Prediktoren (Ursache) und einem Responsefaktor (Auswirkung), z. B. zwischen Lebensmittelverzehr und einer Erkrankung zu untersuchen (Hoffmann et al. 2004 S 935f), ist diese Methode für die vorliegende Auswertung nicht geeignet.

Für die Zielsetzung der vorliegenden Arbeit wird ein Index, eine Faktorenanalyse, eine Clusteranalyse und ein Entscheidungsbaum angewendet. Diese vier Methoden wurden ausgewählt, da sie zwar unterschiedliche Schwerpunkte haben, aber alle zumindest einen Teil der Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs erfüllen. Beispielsweise können über den Index verschiedene Personengruppen wie Bio-Käufer/innen und Nicht-Bio-Käufer/innen differenziert und bei der Clusteranalyse, Faktorenanalyse und dem Entscheidungsbaum alle Lebensmittelgruppen mit den entsprechenden Verzehrsmengen bei der Anwendung der Methode berücksichtigt werden.

Für die Vergleichbarkeit der Methoden für die genannte Forschungsfrage ist es bedeutend, die Verzehrsdaten der gleichen Studiengruppe heranzuziehen. Ein Vergleich der Eignung der vier Methoden zur umfassenden Darstellung des Lebensmittelverzehrs wird in der Literatur, weder mit gleichen noch mit unterschiedlichen Studiendaten, beschrieben. Diese Lücke wird mit der vorliegenden Arbeit gefüllt.

Vor diesem Hintergrund, die vier genannten Methoden zu vergleichen, werden die Daten einer gleichen Studiengruppe, die Daten der NVS II, herangezogen. Diese sind durch die große Studienpopulation, detaillierte Daten zum Lebensmittelverzehr, Daten zu weiteren Aspekten wie Lebensstil und der Möglichkeit nach verschiedenen Einflussfaktoren wie Geschlecht zu unterscheiden, besonders dafür geeignet. Die mit den vier Methoden ermittelten Ergebnisse wurden mit zahlreichen weiteren in der NVS II erhobenen Variablen verknüpft, um identifizierte Personengruppen mit einem ähnlichen Lebensmittelverzehr und differenzierte Personengruppen wie Bio-Käufer/innen und Nicht-Bio-Käufer/innen weiter zu charakterisieren. Zu diesen Variablen gehören Alter, sozioökonomischer Status (SÖS), Ernährungskennntnisse, Body-Mass-Index (BMI), Taillenumfang, Verhältnis Taillen- zu Hüftumfang (WHR), Selbstangaben der Teilnehmenden zur sportlichen Aktivität, zum Rauchverhalten, zum allgemeinen Gesundheitszustand und zur Schlafqualität.

In der vorliegenden Arbeit wird zunächst auf die NVS II eingegangen (Kapitel 2). Neben einer allgemeinen Beschreibung der NVS II werden die jeweiligen Erhebungsmethoden der NVS II, die für die vorliegende Arbeit relevant sind, beschrieben. Es wird die Datenbasis für die vorliegenden Auswertungen, die im Rahmen eines Projektes der NVS II über das Ernährungs- und Gesundheitsverhalten von Bio-Käufer/innen und Nicht-Bio-Käufer/innen festgelegt wurde, vorgestellt. Die mit diesem Datensatz ermittelten Ergebnisse der vier Methoden Index (Kapitel 3.1), Faktorenanalyse (Kapitel 4.1), Clusteranalyse (Kapitel 4.2) sowie des Entscheidungsbaums (Kapitel 4.3) werden vorgestellt.

In den jeweiligen Kapiteln werden zunächst die Methoden allgemein beschrieben und über die Literatur Anwendungsbeispiele der Methoden aufgezeigt. Aufbauend auf der Beschreibung der einzelnen Methoden wird das Vorgehen für die eigene Anwendung erläutert. Anschließend werden die inhaltlichen Ergebnisse der jeweiligen Methoden, d. h. der Lebensmittelverzehr, dargestellt und Personengruppen mit weiteren Variablen des Ernährungs- und Gesundheitsverhaltens charakterisiert. Darauf aufbauend wird analysiert, ob sich die angewendeten Methoden anhand der oben genannten Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs eignen.

In einer abschließenden übergreifenden Diskussion und einem Fazit (Kapitel 5) werden die angewendeten Methoden diskutiert und in Bezug auf die Forschungsfrage einander gegenübergestellt.

In der vorliegenden Arbeit steht die Analyse der Methoden zur umfassenden Darstellung des Lebensmittelverzehrs anhand der genannten Kriterien und nicht die Analyse oder Bewertung der Ernährung im Fokus. Zudem wird durch die vorliegende Auswertung nicht angestrebt, Ursachen für nicht plausible Ergebnisse zu finden. Inhaltliche Ergebnisse werden nur diskutiert, wenn sie für die Analyse der Methoden, bezogen auf die Kriterien, relevant sind.

## **2 Datenbasis für die vorliegenden Auswertungen**

Mit dem Ziel, die Eignung verschiedener Methoden zur umfassenden Darstellung des Lebensmittelverzehr zu prüfen, werden die Daten der NVS II verwendet. Im Folgenden werden die Methoden der NVS II sowie die verwendeten Daten der NVS II beschrieben. Anschließend wird ein Projekt über das Ernährungs-, Gesundheits- und Einkaufsverhalten von Bio-Käufer/innen und Nicht-Bio-Käufer/innen vorgestellt, in dem die vorliegende Dissertation begonnen und damit der verwendete Datensatz für die vorliegende Auswertung festgelegt wurde.

### **2.1 Die Nationale Verzehrsstudie II**

Die nationale Verzehrsstudie II, ist eine für die deutschsprachige Bevölkerung repräsentative Erhebung zum Ernährungsverhalten. Die Studie fand in den Jahren 2005 bis 2007 mit etwa 20.000 Personen im Alter von 14 bis 80 Jahren statt. Die Stichprobenziehung wurde über eine zweistufig geschichtete Zufallsauswahl durchgeführt. Neben einer Ziehung von etwa 500 Gemeinden nach Bundesland und Gemeindetyp, erfolgte eine zufällige Auswahl der Adressen möglicher Teilnehmenden über Einwohnermelderegister nach Alter und Geschlecht (MRI 2008a S XV).

Über ein persönliches Eingangsinterview CAPI (Computer Assisted Personal Interview) und einem ergänzenden Fragebogen wurden allgemeine Informationen u. a. aus den Bereichen Bildung, Erwerbstätigkeit, berufliche Stellung, Familienstand, Einkommen, Ernährungswissen, Gesundheitszustand, Rauchverhalten, sportliche Aktivität, Schlafqualität und Kriterien beim Lebensmitteleinkauf wie der Kauf von Bio-Produkten, erfasst. Ergänzend wurden anthropometrische Messungen (Körpergewicht, Körpergröße, Taillen- und Hüftumfang) durchgeführt (MRI 2008a S XVI, Brombach et al. 2006 S 7).

Der Lebensmittelverzehr wurde unter anderem über ein Diet-History-Interview mit der Software DISHES (Dietary Interview Software for Health Examination Studies) erhoben. Hiermit wurden die verzehrten Lebensmittel- und Getränkemengen der letzten vier Wochen erfasst. Die Verzehrsmengen wurden anhand eines Fotobuchs und anhand von Modellgeschirr abgeschätzt. Aus den ermittelten Verzehrdaten wurde mit Hilfe des Bundeslebensmittelschlüssels (BLS, Version 3.01 MRI 2010) die Nährstoffzufuhr ermittelt (Brombach et al. 2006 S 6f, Krems et al. 2006 S 44f).

## 2.2 Statistische Grundlagen für die vorliegenden Auswertungen

Der Kolmogorov-Smirnov-Test (bei Stichprobengrößen  $\geq 2000$ ) (Schneider 1997 S 86) und der Shapiro-Wilk-Test (bei Stichprobengrößen  $< 2000$ ) (Royston 1992 S 117), die Histogramme sowie die Ergebnisse der „Skewness“ (Schiefe der Verteilung) (Schneider 1997 S 28, 81) zeigen, dass die metrischen Daten, wie die Daten zum Lebensmittelverzehr, nicht normalverteilt sind. Signifikante Unterschiede metrischer Daten werden mit nicht-parametrischen Testverfahren berechnet.

Aufgrund der schiefen Verteilung müsste zur Darstellung des Lebensmittelverzehrs der Median herangezogen werden. Da dieser bei selten verzehrten Lebensmitteln wie Fisch, Nüssen und Samen bei null liegen kann und dann, wenn mehr als die Hälfte einer Gruppe die Lebensmittel nicht verzehrt, einen Nicht-Verzehr der Lebensmittel suggeriert, obwohl tatsächlich ein Verzehr vorliegt, ist der Median für die vorliegende Auswertung nicht geeignet. Angaben zum Lebensmittelverzehr erfolgen durch den arithmetischen Mittelwert (MW)  $\pm$  Standardabweichung (Std dev) und beziehen sich auf die verzehrten Lebensmittelmengen in Gramm pro Tag. Bei metrischen Daten wie dem Lebensmittelverzehr wird für den Vergleich von Unterschieden zwischen zwei Gruppen der Rangsummentest Mann-Whitney-U-Test verwendet. Der Kruskal-Wallis-Test kommt beim Vergleich von drei Gruppen zur Anwendung. Treten häufige Bindungen bei metrischen Daten auf (viele gleiche Werte), wie bei den Indexwerten in Kapitel 3.1.4, wird zusätzlich durch Histogramme sowie durch die Skewness geprüft, ob eine ähnliche Verteilung der Daten der zur vergleichenden Personengruppen vorliegt. Ist dies der Fall, kann der Mann-Whitney-U-Test angewendet werden. Ist dies nicht der Fall, ist der Mann-Whitney-U-Test nicht geeignet. Bei mehr als drei Gruppenvergleichen werden die Konfidenzintervalle für die arithmetischen Mittelwerte berechnet. Die Unterschiede zwischen Konfidenzintervallen gelten als signifikant, wenn sich die Konfidenzintervalle nicht überschneiden.

Der Trendtest nach Jonckheere-Terpstra dient der Untersuchung, ob ein Verlauf z. B. von Lebensmittelverzehrsmengen über mehrere Gruppen hinweg vorliegt.

Kategoriale Variablen wie die sportliche Aktivität (ja/nein), Ernährungskennntnisse (gut, mittel, schlecht) und Raucherstatus (ja/nein) sind als relative Häufigkeit (Prozent) angegeben. Um unterschiedliche Ausprägungen dieser Variablen zu untersuchen, wird der Chi-Quadrat-Test nach Pearson verwendet. Bei einem multiplen Signifikanzvergleich wird die Bonferroni-Korrektur mit einem alpha-Niveau von 5 % berechnet.

Die statistischen Tests werden mit der Software SAS (Statistical Analysis Software) Version 9.2 (SAS Institute Inc 2012) durchgeführt. Der Index wird über ein eigenes Skript in SAS berechnet. Die Berechnung der Hauptkomponentenanalyse erfolgt mit der Prozedur PRINCOMP. Die einbezogenen Daten werden z-standardisiert, um so gemäß Backhaus et al. (2008 S 332) eine gleiche Größenordnung der einbezogenen Variablen zu erzielen. Die

Berechnung der Clusteranalyse erfolgt mit der Prozedur CLUSTER und Methode=Ward, bei welcher die Variablen ebenfalls z-standardisiert werden (Schendera 2010 S 13f, Schendera 2004 S 543). Zur Berechnung der Entscheidungsbäume wird die Open Source Software R Version „R-2.14.2 for Windows“ (R Core Team 2014) und das Package RWeka Version 0.4-13 mit dem Verfahren J48 verwendet (Hornik et al. 2009 S 225ff).

Zur Bewertung der Ergebnisse wird das Signifikanzniveau  $p < 0,05$  (\*),  $p < 0,01$  (\*\*) und  $p < 0,001$  (\*\*\*) herangezogen.

Die Beschreibung der angewendeten Methoden Index, Hauptkomponentenanalyse, Clusteranalyse und Entscheidungsbaum erfolgt in den entsprechenden Kapiteln.

Aufgrund der großen geschlechtsspezifischen Unterschiede im Lebensmittelverzehr (MRI 2008b S 29ff) wird der Lebensmittelverzehr der Frauen und Männer getrennt analysiert und dargestellt.

## **2.3 Einkauf ökologisch erzeugter Lebensmittel**

Die Dissertation wurde im Projekt mit dem Titel „Auswertung der Daten der Nationalen Verzehrsstudie II (NVS II): eine integrierte verhaltens- und lebensstilbasierte Analyse des Bio-Konsums“ begonnen. Der für die Promotion verwendete Datensatz für die vorliegenden Auswertungen wurde damit über dieses Projekt festgelegt. Ziel des Projektes war, Erkenntnisse über das Ernährungs-, Gesundheits- und Einkaufsverhalten von Bio-Käufer/innen und Nicht-Bio-Käufer/innen zu gewinnen, um diese für die Politikberatung, Wirtschaftszweige und Verbraucherberatung zu nutzen. Mit diesen Erkenntnissen sollten Informationen zur Stärkung und Etablierung ökologisch erzeugter Lebensmittel zur Verfügung gestellt werden (Hoffmann und Spiller 2010b S 2). Auf Basis der Daten der NVS II wurde neben Fragestellungen über die Kaufmotivation und Einstellungen der tatsächliche Lebensmittelverzehr von Bio-Käufer/innen und Nicht-Bio-Käufer/innen im Vergleich untersucht.

Das Projekt war ein Verbundprojekt des Max Rubner-Institutes in Karlsruhe und der Georg-August-Universität Göttingen, das vom damaligen Bundesprogramms ökologischer Landbau (BÖL, heute BÖLN: Bundesprogramm ökologischer Landbau und andere Formen nachhaltiger Landwirtschaft) finanziert wurde.

### **Studiengruppe**

Die Gesamtgruppe der Bio-Käufer/innen und Nicht-Bio-Käufer/innen umfasst 13074 Personen im Alter von 18 bis 80 Jahren, davon 7113 Frauen (54,4 %) und 5961 Männer (45,6 %). Nicht aufgenommen werden konnten die 14- bis 17-Jährigen, da für diese Altersgruppe keine Fragen zum Kauf von Bio-Produkten vorlagen, sowie weitere 80 Personen, die keine verwertbaren Angaben zum Kauf von Bio-Produkten machten (Eisinger-Watzl et al. 2010 S 6ff).

Entsprechend der gegebenen Altersstruktur der Gesamtbevölkerung sind auch in dieser Unterstichprobe sowohl bei den Männern als auch bei den Frauen die 35- bis 50-Jährigen am stärksten vertreten (Tab. 2.1) (Eisinger-Watzl et al. 2010 S 6ff).

**Tab. 2.1: Altersgruppen der Unterstichprobe der NVS II (absolute und relative Häufigkeiten)**  
(modifiziert nach Eisinger-Watzl et al. 2010 S 7)

Altersgruppe	Gesamtgruppe		Frauen		Männer	
	Absolut	Relativer Anteil (%)	Absolut	Relativer Anteil (%)	Absolut	Relativer Anteil (%)
<b>Gesamt</b>	13074	100	7113	54,4	5961	45,6
<b>18–34</b>	2647	20,3	1472	11,3	1175	9,0
<b>35–50</b>	4420	33,8	2510	19,2	1910	14,6
<b>51–64</b>	3227	24,7	1721	13,2	1506	11,5
<b>65–80</b>	2780	21,3	1410	10,8	1370	10,5


### Vergleich der Unterstichprobe der NVS II mit dem Mikrozensus 2006

Die Qualität der Stichprobe der vorliegenden Auswertungen, wurde über einen Vergleich mit den Daten des Mikrozensus 2006 (Statistisches Bundesamt 2010) geprüft. Die Daten des Mikrozensus basieren auf einer Stichprobe von etwa einem Prozent der Gesamtbevölkerung. Im soziodemographischen Bereich wurden die Variablen Geschlecht, Altersgruppe, Familienstand, Haushaltsgröße und Bundesland, im sozioökonomischen Bereich die Variablen Schulabschluss und Erwerbstätigkeit und im Gesundheitsbereich das Rauchverhalten verglichen. Fast alle Vergleiche zwischen der vorliegenden Stichprobe der NVS II und der Allgemeinbevölkerung (Mikrozensus) zeigen nur geringe Abweichungen. Die größte Abweichung liegt mit ca. acht Prozent bei der Schulbildung vor. Bei beiden Geschlechtern liegt in der Unterstichprobe ein höheres Bildungsniveau vor als im Bundesdurchschnitt (Eisinger-Watzl et al. 2010 S 7f).

### 2.3.1 Einteilung der Studienteilnehmenden anhand des Einkaufs ökologisch erzeugter Lebensmittel

Die Studienteilnehmenden wurden gefragt: „Kaufen Sie Bioprodukte? (Produkte aus ökologischem Anbau bzw. aus ökologischer Erzeugung)“, was sie mit „ja“, „nein“ oder „weiß nicht“ beantworten konnten. Daneben wurden die Teilnehmenden gefragt, wie häufig sie ausgewählte Lebensmittel(-gruppen) als Bioprodukte kaufen. Die zweite Frage diente der Klassifizierung der Käufer in die Gruppen Intensiv-, Gelegenheits- und Selten-Bio-Käufer/innen (Abb. 2.1) (Eisinger-Watzl et al. 2010 S 8).

16. Wie häufig kaufen Sie folgende Lebensmittel(-gruppen) als Bioprodukte?

 Bitte in jeder Zeile ein Kästchen ankreuzen!

	(Fast) Immer	Häufig	Selten	Nie	Esse / trinke ich nicht
Brot und Backwaren .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Getreide und Getreideprodukte (z.B. Müsli) .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Obst .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gemüse .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kartoffeln .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Milch und Milchprodukte (inkl. Käse, Joghurt) .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fleisch und Wurstwaren .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fisch .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Eier .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Fertigprodukte (z.B. Tiefkühlgemüse, Pizza, Soßen) .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Obst- und Gemüsesäfte .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wein, Bier .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Abb. 2.1: Frage zur Kaufhäufigkeit einzelner Lebensmittel(-gruppen) als Bioprodukt**  
(Ausschnitt aus dem Fragebogen der NVS II)

Teilnehmende, welche die Frage nach dem Kauf von Bioprodukten verneinten, wurden der Gruppe der Nicht-Bio-Käufer/innen zugeordnet. Ebenso wurden Personen dieser Gruppe zugeordnet, die bei der Frage nach dem Bio-Kauf: „weiß nicht“ ankreuzten und bei der nachfolgenden Frage nach der Kaufhäufigkeit verschiedener Bio-Produkte keine Angaben machten oder immer „esse/trinke ich nicht“ angaben (Eisinger-Watzl et al. 2010 S 8).

Der Gruppe der Bio-Käufer/innen wurden Personen zugeordnet, welche die erste Frage bejahten, sowie Personen, die in der ersten Frage „weiß nicht“ wählten und bei der zweiten Frage Angaben zur Kaufhäufigkeit machten (Eisinger-Watzl et al. 2010 S 9).

Insgesamt wurden 5875 Personen (44,9 %) der Gruppe der Bio-Käufer/innen und 7199 Personen (55,1 %) der Gruppe der Nicht-Bio-Käufer/innen zugeordnet. Während der Anteil der Bio-Käuferinnen und Nicht-Bio-Käuferinnen in der Gruppe der Frauen 50,3 % zu 49,7 % beträgt, liegt bei den Männern der Anteil bei 61,4 % Bio-Käufern und 38,6 % Nicht-Bio-Käufern (Eisinger-Watzl et al. 2010 S 9).

Bei der Frage nach dem Kauf von Bioprodukten wurde die höchste Kaufintensität „(fast) immer“ mit einem Punkt, gefolgt von zwei Punkten für die Angabe „häufig“, drei Punkten für „selten“ bis zur geringsten Kaufintensität „nie“, mit vier Punkten bewertet. Da jedoch viele Teilnehmende nicht für alle 12 Lebensmittelgruppen die Kaufhäufigkeit angaben, wurde die Summe der Einzelangaben durch die Anzahl der entsprechenden Nennungen dividiert. Auch die Angabe „esse/trinke ich nicht“ wurde als „nicht beantwortet“ behandelt. Der Quotient aus

der Summe der Häufigkeiten und Anzahl der Nennungen betrug mindestens einen Punkt und höchstens vier Punkte, wobei Personen mit 1 bis <2 Punkten als Intensivkäufer mit 2 bis <3 als Gelegenheitskäufer und mit 3 bis  $\leq 4$  als Seltenkäufer gruppiert wurden (Eisinger-Watzl et al. 2010 S 10).

Die Gesamtgruppe der 44,9 % Bio-Käufer/innen lässt sich in 13,7 % Selten-, 26,2 % Gelegenheits- und 5,1 % Intensiv-Bio-Käufer/innen aufteilen (Tab. 2.2).

**Tab. 2.2: Anteile der Bio-Käufer/innen nach Kaufintensität (absoluter und relativer Anteil)**

	Gesamtgruppe		Frauen		Männer	
	Absolut	Relativer Anteil (%)	Absolut	Relativer Anteil (%)	Absolut	Relativer Anteil (%)
<b>Nicht-Bio-Käufer/innen</b>	7199	55,1	3536	49,7	3663	61,4
<b>Bio-Käufer/innen</b>	5875	44,9	3577	50,3	2298	38,6
<b>Intensiv</b>	661	5,1	442	6,2	219	3,7
<b>Gelegentlich</b>	3428	26,2	2089	29,4	1339	22,5
<b>Selten</b>	1786	13,7	1046	14,7	740	12,4

### 2.3.2 Lebensmittelverzehr der Bio-Käufer/innen und Nicht-Bio-Käufer/innen

Zur Gewinnung erster Erkenntnisse über den Lebensmittelverzehr wurden die Verzehrsmengen einzelner Lebensmittelgruppen der Bio-Käufer/innen und Nicht-Bio-Käufer/innen getrennt nach Geschlecht den Empfehlungen gegenübergestellt (Tab. 2.3 und Tab. 2.4). Diese Ergebnisse dienen als Vergleich der Ergebnisse der Methoden, die zur Prüfung ihrer Eignung zur umfassenden Darstellung des Lebensmittelverzehrs angewendet werden. Die Lebensmittel, die den einzelnen Lebensmittelgruppen zugeordnet wurden, sind in Tab. 9.1 im Anhang aufgelistet.

**Tab. 2.3: Verzehrte Lebensmittelmengen (g/Tag) der Bio-Käuferinnen und Nicht-Bio-Käuferinnen (Frauen)**  
(erweitert nach Eisinger-Watzl und Hoffmann 2010 S 34ff)

Lebensmittel-		MW±	Perzentile					Signifikanzen <sup>1</sup>
gruppen		Std dev	Min.	Max.	P25	P50	P75	
Bio-Käuferinnen	Obst	310±233	0	3188	159	266	398	<0,001
Nicht-Bio-Käuferinnen		273±231	0	3589	121	228	364	
Bio-Käuferinnen	Gemüse	274±158	0	1610	165	242	347	<0,001
Nicht-Bio-Käuferinnen		228±138	0	1345	134	200	289	
Bio-Käuferinnen	Milch	245±200	0	2372	107	197	324	<0,001
Nicht-Bio-Käuferinnen		233±222	0	4490	95	178	299	
Bio-Käuferinnen	davon:	196±193	0	2354	64	147	271	<0,01
Nicht-Bio-Käuferinnen	Milch/Joghurt	190±215	0	4451	56	132	250	
Bio-Käuferinnen	davon:	48±43	0	587	21	38	63	<0,001
Nicht-Bio-Käuferinnen	Käse/Quark	44±43	0	600	17	33	56	
Bio-Käuferinnen	Getreide/ Kartoffeln	319±116	0	1176	241	307	384	<0,001
Nicht-Bio-Käuferinnen		303±104	6	1064	235	293	361	
Bio-Käuferinnen	davon: Getreide	249±104	0	1030	178	235	305	<0,001
Nicht-Bio-Käuferinnen		229±93	1	973	163	219	282	
Bio-Käuferinnen	davon: Kartoffeln	70±41	0	313	41	65	94	<0,001
Nicht-Bio-Käuferinnen		74±45	0	475	43	70	100	
Bio-Käuferinnen	Fisch	24±23	0	266	10	20	32	<0,001
Nicht-Bio-Käuferinnen		21±21	0	240	7	17	29	
Bio-Käuferinnen	Fleisch	68±46	0	416	36	62	92	<0,001
Nicht-Bio-Käuferinnen		82±50	0	592	49	74	103	
Bio-Käuferinnen	Eier	12±13	0	135	4	9	17	n.s.
Nicht-Bio-Käuferinnen		12±13	0	131	4	10	17	
Bio-Käuferinnen	Streichfett	20±17	0	143	9	16	26	<0,05
Nicht-Bio-Käuferinnen		21±18	0	150	9	17	28	
Bio-Käuferinnen	Alkohol <sup>2</sup>	6±8	0	91	1	3	8	<0,001
Nicht-Bio-Käuferinnen		5±9	0	172	0	2	7	
Bio-Käuferinnen	Getränke	2414±888	275	7493	1818	2333	2899	<0,001
Nicht-Bio-Käuferinnen		2256±883	170	7586	1642	2179	2753	
Bio-Käuferinnen	Nüsse und Samen	4±11	0	150	0	0	4	<0,001
Nicht-Bio-Käuferinnen		3±11	0	244	0	0	2	
Bio-Käuferinnen	Süßwaren	58±55	0	1020	23	45	79	<0,001
Nicht-Bio-Käuferinnen		63±57	0	1187	26	50	83	

<sup>1</sup>Signifikanzniveau p, getestet über den Rangsummentest Mann-Whitney-U-Test

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

**Tab. 2.4: Verzehrte Lebensmittelmengen (g/Tag) der Bio-Käufer und Nicht-Bio-Käufer (Männer)**  
(erweitert nach Eisinger-Watzl und Hoffmann 2010 S 34ff)

	Lebensmittel- gruppen	MW± Std dev	Min.	Max.	Perzentile			Signifikanzen <sup>1</sup>
					P25	P50	P75	
Bio-Käufer	Obst	275±228	0	1850	120	216	362	<0,001
Nicht-Bio-Käufer		233±217	0	2498	82	185	311	
Bio-Käufer	Gemüse	262±151	4	1337	157	227	338	<0,001
Nicht-Bio-Käufer		224±136	0	1178	128	199	292	
Bio-Käufer	Milch	254±243	0	2603	93	184	329	<0,001
Nicht-Bio-Käufer		246±267	0	3647	79	166	316	
Bio-Käufer	davon:	203±235	0	2601	47	132	270	<0,01
Nicht-Bio-Käufer	Milch/Joghurt	201±261	0	3538	37	116	258	
Bio-Käufer	davon: Käse/	51±48	0	598	21	38	65	<0,001
Nicht-Bio-Käufer	Quark	46±49	0	679	17	34	59	
Bio-Käufer	Getreide/	403±142	49	1221	308	386	472	<0,001
Nicht-Bio-Käufer	Kartoffeln	390±143	0	1372	296	373	465	
Bio-Käufer	davon: Getreide	315±130	17	1052	225	296	381	<0,001
Nicht-Bio-Käufer		295±125	0	1146	209	278	358	
Bio-Käufer	davon: Kartoffeln	88±53	0	367	52	80	114	<0,001
Nicht-Bio-Käufer		95±59	0	651	55	87	126	
Bio-Käufer	Fisch	32±31	0	314	13	25	43	<0,001
Nicht-Bio-Käufer		27±27	0	294	9	21	37	
Bio-Käufer	Fleisch	115±72	0	711	69	105	148	<0,001
Nicht-Bio-Käufer		145±88	0	963	88	128	179	
Bio-Käufer	Eier	15±16	0	180	4	11	20	<0,001
Nicht-Bio-Käufer		17±19	0	309	5	12	22	
Bio-Käufer	Streichfett	28±24	0	281	12	23	37	<0,05
Nicht-Bio-Käufer		31±28	0	223	12	24	42	
Bio-Käufer	Alkohol <sup>2</sup>	15±17	0	198	3	10	21	n.s.
Nicht-Bio-Käufer		16±19	0	228	2	10	23	
Bio-Käufer	Getränke	2379±949	274	7086	1714	2292	2926	<0,001
Nicht-Bio-Käufer		2302±1000	0	9091	1620	2189	2850	
Bio-Käufer	Nüsse und Samen	6±15	0	214	0	0	4	<0,001
Nicht-Bio-Käufer		4±12	0	190	0	0	3	
Bio-Käufer	Süßwaren	65±56	0	471	26	52	89	n.s.
Nicht-Bio-Käufer		70±66	0	924	25	54	95	

<sup>1</sup>Signifikanzniveau p, getestet über den Rangsummentest Mann-Whitney-U-Test

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

Die deskriptive Statistik zeigt, dass sich sowohl bei den Frauen als auch bei den Männern der Lebensmittelverzehr fast aller Lebensmittelgruppen zwischen den Bio-Käufer/innen und Nicht-Bio-Käufer/innen signifikant unterscheidet. Die Bio-Käufer/innen verzehren durchschnittlich mehr Obst, Gemüse, Milch, Getreide, Fisch, alkoholfreie Getränke, Nüsse und Samen sowie weniger Kartoffeln, Streichfett und Fleisch als die Nicht-Bio-Käufer/innen. Daneben verzehren in der Gruppe der Männer die Bio-Käufer weniger Eier, während bei der Zufuhr von Alkohol sowie beim Verzehr von Süßwaren zwischen den Bio-Käufern und Nicht-Bio-Käufern keine Unterschiede festgestellt werden konnten. In der Gruppe der Frauen konnten keine Unterschiede in der Verzehrsmenge von Eiern, aber in der Zufuhr von Alkohol und in der Verzehrsmenge von Süßwaren aufgezeigt werden. Bio-Käuferinnen nehmen mehr Alkohol auf und verzehren weniger Süßwaren als Nicht-Bio-Käuferinnen.

In der Gruppe der Frauen zeigt eine Gegenüberstellung der durchschnittlich verzehrten Mengen mit den Empfehlungen der Deutschen Gesellschaft für Ernährung e. V. (DGE) (DGE und AID 2013 S 7, DGE et al. 2013 S 66f), dass sowohl Bio-Käuferinnen als auch Nicht-Bio-Käuferinnen im Durchschnitt die Empfehlungen für Obst (250 g/Tag), Fisch (80–150 g/Woche), Fleisch (300–600 g/Woche), Eier (bis zu drei Eier pro Woche), Streichfett (15–30 g/Tag), Alkohol (bezogen auf Ethanol als Inhaltsstoff maximal: 10 g/Tag) sowie alkoholfreien Getränken (mindestens 1500 ml/Tag) erreichen bzw. im Bereich der Empfehlungen liegen (Tab. 2.3). Die Empfehlung täglich 200–250 g Milch/Joghurt und 50–60 g Käse/Quark zu verzehren, wurde im Durchschnitt fast erreicht, ebenso wie die Empfehlung insgesamt 350–560 g Getreide und Kartoffeln zu verzehren. Die Empfehlung für Gemüse (400 g/Tag) wurde bei den Frauen nicht erreicht.

In der Gruppe der Männer zeigt die Gegenüberstellung der durchschnittlich verzehrten Mengen mit den Empfehlungen der DGE (DGE und AID 2013 S 7, DGE et al. 2013 S 66f), dass sowohl Bio-Käufer als auch Nicht-Bio-Käufer im Durchschnitt lediglich die Empfehlungen für Getreide und Kartoffel (350–560 g/Tag), Eier (bis zu drei Eier pro Woche), Alkohol (bezogen auf Ethanol als Inhaltsstoff maximal: 20 g/Tag) sowie den alkoholfreien Getränken (mindestens 1500ml/Tag) erreichen. Daneben erreichen die Bio-Käufer im Gegensatz zu den Nicht-Bio-Käufern die Empfehlung für Obst (250 g/Tag) und liegen im Bereich der Empfehlung für Fisch (80–150 g/Woche) und Streichfett (15–30 g/Tag), während die Nicht-Bio-Käufer einen zu geringen Obstverzehr und einen zu hohen Verzehr an Fisch und Streichfett aufweisen. Die Empfehlung, täglich 200–250 g Milch/Joghurt zu verzehren wurde von beiden Gruppen und die Empfehlung täglich 50–60 g Käse/Quark zu verzehren, nur von den Bio-Käufern erreicht. Die Empfehlung für Gemüse (400 g/Tag) wurde wie bei den Frauen von keiner Gruppe erreicht und die Empfehlung für Fleisch (300–600 g/Woche) sowohl von den Bio-Käufern als auch Nicht-Bio-Käufern überschritten (Tab. 2.4).

### 3 Hypothesenorientierter Ansatz zur umfassenden Darstellung des Lebensmittelverzehrs

Zur Prüfung der Eignung verschiedener Methoden zur umfassenden Darstellung des Lebensmittelverzehrs, wird ein hypothesenorientierter Ansatz angewendet. Es handelt sich dabei um einen a-priori-Ansatz, bei dem auf Basis vorhandener wissenschaftlicher Daten, z. B. Empfehlungen, der Verzehr von Lebensmitteln und die Zufuhr von Nährstoffen einer ausgewählten Personengruppe untersucht werden kann. Zu dem hypothesenorientierten Ansatz wird der Index bzw. Score gezählt (Schulze und Hoffmann 2006 S 861, Hu 2002 S 4). Hierbei werden ausgewählte Nährstoffe, Lebensmittel oder Lebensmittelgruppen mit Empfehlungen verglichen und über ein Punktesystem bewertet.

Im Folgenden werden verschiedene Indexformen erläutert. Zu jedem Verfahren werden neben Hintergrund und Zielen auch Anwendungsbeispiele genannt. Auf dieser Basis wird ein Index für die Darstellung des Lebensmittelverzehrs der Bio-Käufer/innen und Nicht-Bio-Käufer/innen entwickelt und angewendet. Zunächst werden die inhaltlichen Ergebnisse präsentiert, um darauf aufbauend anhand der hierfür definierten Kriterien zu analysieren, ob sich der Index zur umfassenden Darstellung des Lebensmittelverzehrs der vorliegenden Studienteilnehmenden eignet.

#### 3.1 Indizes

Mit der Anwendung von Indizes bzw. Scores wird das Ziel verfolgt, die Ernährung (Nährstoffzufuhr und/oder Lebensmittelverzehr) von bestimmten Personengruppen zu untersuchen und zu bewerten. Indizes sind aus mehreren Variablen zusammengesetzt und als Messwert zu verstehen (Bortz und Döring 2009 S 143).

Mehrere für das Untersuchungsziel relevante Variablen auf den Ebenen der Nährstoffe, Lebensmittel und Lebensmittelgruppen können gleichzeitig berücksichtigt (Kant 1996 S 785), mit Empfehlungen verglichen und über ein Punktesystem bewertet werden (Beispiel: Kennedy et al. 1995 S 1104ff). Entspricht die verzehrte Menge der Empfehlung, wird für die ausgewählte Variable die Maximalpunktzahl vergeben. Wird die Empfehlung nicht erreicht, können Punkte abgezogen werden. Der sogenannte **Indexwert** jeder einzelnen Variable wird auf diesem Weg gebildet.

Anschließend kann die **Gesamtpunktzahl** über das Zusammenfassen der einzelnen Indexwerte gebildet werden. Die zu erreichende Gesamtpunktzahl (i. d. R. ein Maximalwert) entspricht, gemessen an den zugrunde gelegten Empfehlungen, einer idealen Ernährung (Kennedy et al. 1995 S 1104). Je mehr Auswahl und Menge der verzehrten Lebensmittel den

Empfehlungen entsprechen, umso höher ist die erreichte Gesamtpunktzahl (Beispiel: USDA 1995 S 7).

Bei der Entwicklung eines Indexes wird in Abhängigkeit der Fragestellung festgelegt, welche und wie viele Variablen berücksichtigt werden, auf Basis welcher Empfehlungen diese bewertet und wie die Bewertung vorgenommen werden soll. Auf die methodischen Möglichkeiten, die hierfür zur Verfügung stehen, wird im Folgenden eingegangen.

### 3.1.1 Theoretischer Hintergrund

Indizes können auf Ebene der Nährstoffe, Lebensmittel und Lebensmittelgruppen (Kant 1996 S 785) gebildet werden. Neben den genannten Variablen können weitere Variablen, wie die Lebensmittelvielfalt als Verzehr einer bestimmten Anzahl an Lebensmitteln, Lebensmittel aus verschiedenen Gruppen in einem bestimmten Zeitraum (Beispiele: Angelopoulos et al. 2009, Shatenstein et al. 2005, Drewnowski et al. 1996, Kennedy et al. 1995) sowie Variablen, beispielsweise aus dem Lebensstilbereich (Beispiel: Krems und Brombach 2007) in einem Index berücksichtigt werden.

Zur Bildung eines Indexes muss zunächst die Ebene, die für das Untersuchungsziel relevant ist, festgelegt werden. Die anschließende Bewertung der ausgewählten Variablen erfolgt anhand von Empfehlungen von Organisationen. Beispielsweise können nationale Empfehlungen wie die der DGE (DGE und AID 2013, DGE et al. 2013) sowie internationale Empfehlungen wie die des World Cancer Research Fund (WCRF und AICR 2007) oder der WHO (2003) herangezogen werden. Daneben können auch sonstige Werte, z. B. abgeleitet aus einer Vergleichsgruppe, als Bewertungsbasis zugrunde gelegt werden (Beispiele: Lockheart et al. 2007, Bosetti et al. 2003, Trichopoulou et al. 2003, 1995).

Als Ergebnis stehen die Indexwerte der einzelnen Variablen sowie die Gesamtpunktzahl zur Verfügung.

#### Indexwerte

Neben der Auswahl aller für die Fragestellung relevanten Variablen muss eine passende **Gewichtung** der Variablen festgelegt werden (Bortz und Döring 2009 S 144). Daneben muss festgelegt werden, welche **Punktzahl** für welche Variable vergeben werden soll und wie die **Indexwerte berechnet** werden.

Bei der **Gewichtung** werden zwischen *gleichgewichteten* und *ungleichgewichteten* Indizes unterscheiden. Bei einem *gleichgewichteten* Index erhalten alle Variablen den gleichen Wertebereich z. B. 15 Punkte (Schnell et al. 2008 S 171). Damit wird allen Variablen die gleiche Bedeutung zugesprochen. Die meisten Indizes sind gleichgewichtete Indizes (Beispiele: Mursu et al. 2013, Angelopoulos et al. 2009, Manios et al. 2009, Waijers et al. 2007, Trichopoulou et al. 2003, Kennedy et al. 1995, Patterson et al. 1994a). Bei einem

*ungleichgewichteten* Index unterscheidet sich der Wertebereich der einbezogenen Variablen (Schnell et al. 2008 S 171). Beispielsweise haben von Rüsten et al. (2009 S 452f) einen Obst- und Gemüseverzehr über der empfohlenen Verzehrshäufigkeit durch eine höhere Punktzahl höher gewichtet als die anderen berücksichtigten Lebensmittelgruppen. Begründet wurde dies damit, dass es sich bei Lebensmitteln dieser Gruppen um besonders nährstoffreiche Lebensmittel mit geringer Energiedichte handelt. McCullough et al. (2002 S 1262f) haben die ungleichgewichtete Indexform gewählt, um zu verhindern, dass die Variable „Multivitamineinnahme“ über Supplemente eine zu hohe Bedeutung im Index erhält. Durch diese Variable konnten weniger Punkte erzielt werden als durch die restlichen Variablen.

Nach Festlegung dieser Gewichtung wird definiert, wie die **Punktzahl** bzw. der **Indexwerte** einzelner Variablen berechnet werden soll. Die Punktzahl kann über ein *stufiges Verfahren*, eine *Verhältnisrechnung* oder über eine *Abhängigkeitsrechnung* berechnet werden. Beispielsweise können bei einem *stufigen Verfahren* null, ein oder zwei Punkte erreicht werden. Entspricht der Verzehr oder die Zufuhr einer Variablen dem Zielwert, wird ein Indexwert von zwei vergeben. Wird z. B. nur die Hälfte der Empfehlung erreicht, wird ein Indexwert von eins erreicht. Wird weniger verzehrt oder aufgenommen, wird für diese Kategorie ein Indexwert von null vergeben (Beispiele: Bau et al. 2003, Sell et al. 2003, Winkler und Döring 1995, Patterson et al. 1994a). Ein weiteres Beispiel für ein stufiges Verfahren ist das Einteilen der Studienteilnehmenden anhand ihrer Zufuhr von Nährstoffen in Quartilen. Beispielsweise erhielten in einer Studie von Hu et al. (2001) Studienteilnehmende, die dem ersten Quartil der Nährstoffzufuhr zugeordnet wurden, einen Indexwert von fünf und Personen des vierten Quartils einen Indexwert von eins für den jeweiligen Nährstoff.

Zum anderen kann die Berechnung der Indexwerte über eine *Verhältnisrechnung* erfolgen. Der tatsächliche Verzehr (Ist) kann mit dem empfohlenen Verzehr (Soll) verglichen werden. Werden z. B. 87 % der Empfehlung erreicht, werden 87 % der für diese Variable erreichbaren Punkte vergeben. Wird die empfohlene Menge überschritten, kann die Bewertung invers erfolgen (Beispiele: Kleiser et al. 2009, von Rüsten et al. 2009, Lockheart et al. 2007, McCullough et al. 2002). Beispielsweise haben von Rüsten et al. (2009) den Indexwert für eine Lebensmittelgruppe berechnet, indem sie für eine Verzehrsmenge unterhalb der empfohlenen Menge die Verzehrsmenge durch die empfohlene Menge dividierten. Für eine Verzehrsmenge oberhalb der empfohlenen Menge erfolgte die Berechnung invers, indem die empfohlene Menge durch die tatsächliche Verzehrsmenge dividiert wurde. Je mehr die Empfehlung überschritten wurde, desto geringer wurde auch der Indexwert für die Lebensmittelgruppe (von Rüsten et al. 2009 S 453). Durch die Anwendung der Verhältnisrechnung können im Gegensatz zu mehrstufigen Bewertungsverfahren größere Punktspannen genutzt und damit die Punkte differenzierter vergeben werden.

Daneben können Indexwerte in *Abhängigkeit der Energie- bzw. Nährstoffdichte* berechnet werden. So wurden beispielsweise im HEI-2005 die Indexwerte von Obst und Gemüse bezogen auf 1000 kcal und der Indexwert von gesättigten Fettsäuren bezogen auf 7 % der Gesamtenergiezufuhr berechnet (Guenther et al. 2007 S 3).

### **Gesamtpunktzahl (Indexsumme oder Indexprodukt)**

Die Gesamtpunktzahl kann **additiv** oder **multiplikativ** gebildet werden (Bortz und Döring 2009 S 145ff, Schnell et al. 2008 S 166ff), woraus die Indexsumme oder das Indexprodukt entsteht. Der *additive* Index wird am häufigsten verwendet. Hierfür wird die Indexsumme durch Addition der Einzelwerte berechnet (Schnell et al. 2008 S 171). Das Indexprodukt wird durch Multiplikation der Indexwerte gebildet. Bei einem *multiplikativen* Index wird eine bestimmte Mindestausprägung aller Variablen vorausgesetzt. Eine wechselseitige Kompensierung der einzelnen Indexwerte wie bei additiven Indizes ist nicht möglich. Erhält eine Variable den Wert Null, wird das Indexprodukt Null (Bortz und Döring 2009 S 145, Schnell et al. 2008 S 172). Wenn z. B. der Verzehr eines Lebensmittels als entscheidend angesehen wird, könnte über den Index definiert werden, dass ein Nicht-Verzehr dieses Lebensmittels grundsätzlich mit einer Gesamtpunktzahl von Null bewertet wird. Ein Beispiel für die Anwendung ist die Berücksichtigung eines toxischen Inhaltsstoffes, dessen Zufuhr auch durch alle weiteren Variablen nicht kompensieren werden kann (Leitzmann 2012 S 383).

Zusätzlich können für die Bildung der Indexsumme die gebildeten Indexwerte gewichtet werden. Damit werden ausgewählte Variablen stärker als andere in der Indexsumme berücksichtigt (Schnell et al. 2008 S 173). Drescher et al. (2007 S 649) haben sich bei der Bewertung einzelner Variablen an dem von der DGE (2005) beschriebenen Verhältnis 73 % pflanzliche Lebensmittel, 25 % tierische Lebensmittel und 2 % Fette und Öle orientiert. Damit werden Lebensmittelgruppen, die mengenmäßig den größten Anteil ausmachen sollen (hier: die pflanzlichen Lebensmittel), bei der Bildung der Indexsumme höher gewichtet.

Die gebildete Gesamtpunktzahl spiegelt die Qualität der Ernährung wider. Je höher die Gesamtpunktzahl, desto mehr entspricht die Ernährung der zur Bewertung herangezogenen nährstoff- oder lebensmittelbasierten Empfehlung.

Auch bei der Bewertung der Gesamtpunktzahl können verschiedene Einteilungsmöglichkeiten angewendet werden. Bei der Einteilung anhand fester Grenzen können die Studienteilnehmenden über zuvor definierte Grenzen gruppiert und anschließend verglichen werden. Beispielsweise haben Angelopoulos et al. (2009) sowie Manios et al. (2009) die Studienteilnehmenden anhand ihrer erreichten Punktzahl in die Gruppe 1 mit 80 bis 100 Prozent, in Gruppe 2 mit 50 bis 80 Prozent und in Gruppe 3 mit weniger als 50 Prozent der möglichen Punkte zugeordnet. Daneben können die Studienteilnehmenden anhand von

Perzentilen, Terzilen, Quartilen oder Quintilen verglichen werden. Die Studienteilnehmenden werden durch diesen Ansatz anhand der erreichten Gesamtpunktzahl in gleichgroße Gruppen aufgeteilt (Beispiele: Bazelmans et al. 2006, Kant et al. 2000, Winkler und Döring 1995).

### **3.1.2 Anwendung von Indizes im Bereich Ernährung**

Als Ergebnis von Indizes stehen sowohl die einzelnen Indexwerte als auch die Gesamtpunktzahl zur Verfügung. Die für jede einzelne Person berechneten Indexwerte, z. B. für einzelne Lebensmittelgruppen, werden für Einzelvergleiche verwendet, um die Übereinstimmung der Verzehrsmenge ausgewählter Lebensmittelgruppen mit den Empfehlungen zu prüfen. Die Gesamtpunktzahl und damit die Zusammenfassung der einzelnen Lebensmittelgruppen kann zur Beurteilung der Ernährung verschiedener Personengruppen herangezogen werden (Beispiele: Angelopoulos et al. 2009, Manios et al. 2009, Schulze und Hoffmann 2006). Sowohl die Indexwerte als auch die Gesamtpunktzahl können ferner eingesetzt werden, um eine Studiengruppen mit weiteren Variablen zu charakterisieren. So wird häufig der Zusammenhang zwischen den Indexwerten oder der Gesamtpunktzahl zu allgemeinen Variablen wie Alter, Geschlecht, SÖS oder Bildung beschrieben (Beispiele: Martinez et al. 2010, Previdelli et al. 2010, Shah et al. 2010).

Indizes werden seit Jahrzehnten eingesetzt, um die Ernährung unterschiedlicher Personengruppen zu beschreiben und zu bewerten. Beispielsweise haben Burke und Stuart (1938) bereits im Jahr 1938 über eine Fünf-Punkte-Skala die Zufuhr von Nährstoffen von Kindern bewertet und Biomarkern wie Hämoglobin gegenübergestellt. Hatte ein Kind die empfohlene Menge von beispielsweise 1 g Calcium am Tag über verschiedene Lebensmittel wie Milchprodukte aufgenommen, wurden die maximal zu erreichenden fünf Punkte vergeben. Wurde nur 3/4 der empfohlenen Menge erreicht, wurden vier Punkte vergeben. Mit Hilfe des Indexes wurde die Zufuhr einzelner Nährstoffe wie Calcium und Eisen bewertet.

Der bekannteste Index ist der Healthy Eating Index (HEI), der vom amerikanischen Landwirtschaftsministerium (U. S. Department of Agriculture) entwickelt wurde. Er dient der Überprüfung, wie gut die Ernährung der Amerikaner mit den „Dietary Guidelines for Americans“ und der „Food Guide Pyramid“ übereinstimmt. Der HEI besteht aus zehn Variablen: Obst, Gemüse, Getreide, Milch und Fleisch sowie Gesamtfett, gesättigte Fettsäuren, Cholesterin, Natrium und Lebensmittelvielfalt. Für jede Variable werden je nach Erreichungsgrad der entsprechenden Empfehlung zwischen null und zehn Punkten vergeben und diese additiv zu einem Index mit einer Maximalsumme von 100 zusammengefasst (USDA 1995 S 14).

Es wurde eine Vielzahl unterschiedlicher Indizes entwickelt. Im Wesentlichen lassen sich diese Indizes in Anlehnung an Kant (1996 S 785) in vier Kategorien einteilen. Indizes können aus (1) **Nährstoffen**, (2) **Lebensmitteln und Lebensmittelgruppen**, (3) **Nährstoffen, Lebensmitteln und Lebensmittelgruppen** sowie (4) **Nährstoffen, Lebensmitteln und Lebensmittelgruppen kombiniert mit weiteren nährstoff- und lebensmittel-unabhängigen Variablen** beispielsweise in Bezug auf den Lebensstil aufgebaut sein.

Nach Berechnung eines Indexes können Ursache-Wirkungs-Zusammenhänge zwischen der Ernährung und Bioparametern und Krankheiten untersucht werden oder eine Charakterisierung von Personengruppen über die unterschiedlichen Indexwerte und unterschiedliche Gesamtpunktzahl vorgenommen werden. Im Folgenden werden die vier Kategorien der Indizes beschrieben und Beispiele für die Untersuchung von Zusammenhängen und zur Charakterisierung der Studiengruppe mit weiteren Variablen genannt.

- (1) **Nährstoffbasierte Indizes** sind eine nicht sehr häufig verwendete Indexform. Sie enthalten ausschließlich Nährstoffe und werden verwendet, um die Nährstoffzufuhr über einen Vergleich mit Empfehlungen für die Nährstoffzufuhr zu beurteilen (Beispiele: Hu et al. 2001, Drewnowski et al. 1997, 1996, Burke et al. 1943, Burke und Stuart 1938).

Die Ergebnisse dieser Indizes, die Indexwerte und Gesamtpunktzahl, können anschließend mit weiteren Variablen verknüpft werden, um die Qualität eines Indexes zu testen oder zu validieren. Beispielsweise haben Drewnowski et al. (1997, 1996) untersucht, ob ein Zusammenhang zwischen der Gesamtpunktzahl und der Energiezufuhr, der Zufuhr von Fett (Energie%) und Kohlenhydrate (Energie%) besteht.

- (2) **Lebensmittel- und lebensmittelgruppenbasierte Indizes** werden verwendet, um zum einen die *Lebensmittelauswahl* und zum anderen um die *Lebensmittelvielfalt* zu untersuchen.

Mit Indizes, welche die *Lebensmittelauswahl* erfassen, wird das Ziel verfolgt, die verzehrten Lebensmittelmengen mit empfohlenen Verzehrsmengen zu vergleichen und zu beurteilen (Beispiele: Mursu et al. 2013, von Rüsten et al. 2010, Kleiser et al. 2009, von Rüsten et al. 2009, Bau et al. 2003, Sell et al. 2003, Kant et al. 2000).

Das Ergebnis dieser Indizes, die Gesamtpunktzahl, kann anschließend, wie bei nährstoffbasierten Indizes, im Zusammenhang mit zahlreichen weiteren Variablen untersucht werden. So haben von Rüsten et al. (2010) den Zusammenhang zwischen der Gesamtpunktzahl und der Inzidenz kardiovaskulärer Krankheiten, Diabetes mellitus Typ2 und Krebs, Kant et al. (2000) den Zusammenhang zur Mortalitätsrate, Bau et al. (2003) den Zusammenhang zum Migrationshintergrund und Sell et al. (2003) den Zusammenhang zu Trends in Bezug auf Veränderungen im Lebensmittelverzehr untersucht.

Indizes mit dem Fokus auf der *Lebensmittelvielfalt* verfolgen das Ziel, die Anzahl verschiedener verzehrter Lebensmittel oder Lebensmittel ausgewählter Lebensmittelgruppen zu erfassen (Beispiele: Moursi et al. 2008, Drescher et al. 2007, Mirmiran et al. 2006, 2004, Drewnowski et al. 1996, Kant et al. 1993, Miller et al. 1992).

Das Ergebnis dieser Indizes kann ebenfalls mit weiteren Variablen verknüpft werden. Beispielsweise haben Moursi et al. (2008), Mirmiran et al. (2006), Mirmiran et al. (2004) und Drewnowski et al. (1996) neben der Lebensmittelvielfalt den Zusammenhang zwischen der Gesamtpunktzahl und der Nährstoffzufuhr, Kant et al. (1993) den Zusammenhang mit dem Mortalitätsrisiko und Miller et al. (1992) den Zusammenhang mit Bluthochdruck untersucht.

- (3) **Kombinierte Indizes** können aus Nährstoffen, Lebensmitteln, Lebensmittelgruppen sowie Aspekten der Lebensmittelvielfalt bestehen. Häufig dienen diese Indizes der Prüfung der Übereinstimmung der Ernährung mit lebensmittel- und nährstoffbasierten Empfehlungen (Beispiele: Rifas-Shiman et al. 2009, Boynton et al. 2008, Guenther et al. 2008, Shatenstein et al. 2005, Kim et al. 2003, Stookey et al. 2000, Kennedy et al. 1995).

Die Ergebnisse dieser Indizes werden auch verwendet, um beispielsweise Personengruppen mit weiteren Variablen aus dem Lebensstilbereich (Beispiele: Drewnowski et al. 2009, von Normann 2007) zu charakterisieren, den Zusammenhang zum BMI (Beispiele: Tande et al. 2010, Guo et al. 2004), zu Biomarkern (Beispiele: Turner-McGrievy et al. 2008, Ford et al. 2005, Drewnowski et al. 1997) sowie der Prävalenz oder Inzidenz verschiedener Krankheiten (Beispiele: Ma et al. 2008, Pan und Pratt 2008, Bazelmans et al. 2006, Kruzich et al. 2004, Seymour et al. 2003, McCullough et al. 2002, Hu et al. 2001, Kant et al. 2000, Patterson et al. 1994a) zu untersuchen. Daneben haben Boynton et al. (2008) den Zusammenhang zwischen der Gesamtpunktzahl und BMI, Einkommen, Bildungsstand und Raucherstatus; Drewnowski et al. (2009), Ma et al. (2008), Guo et al. (2004) und McCullough et al. (2002) den Zusammenhang der Gesamtpunktzahl und Alter, Geschlecht, Bildung, BMI sowie verschiedenen Variablen aus dem Lebensstilbereich wie Rauchen und sportliche Aktivität untersucht.

- (4) Es gibt auch **Indizes**, in denen neben Aspekten aus dem Bereich Ernährung **weitere nährstoff- und lebensmittelunabhängige Variablen**, beispielsweise aus dem Lebensstilbereich, direkt einbezogen werden. So haben Krems und Brombach (2007) den sogenannten „Healthy Nutrition Lifestyle Index“ entwickelt, um mehrere Variablen des Ernährungsverhaltens über eine einfache Messgröße zusammenzufassen. Dieser Index besteht aus den Variablen Getreide, Obst, Gemüse, gesättigte Fettsäuren, Fett gesamt, Ballaststoffe, Alkohol, Lebensmittelvielfalt sowie Über-

gewicht/Adipositas und der körperlichen Aktivität. Ford et al. (2009) entwickelten einen „Healthy Lifestyle Factor“ bestehend aus BMI, sportlicher Aktivität, Raucherstatus und Ernährung (Obst, Gemüse, Vollkornbrot und rotes Fleisch), um den Zusammenhang zu chronischen Krankheiten zu untersuchen.

Durch die mit den bisher entwickelten Indizes verfolgten Ziele wird deutlich, dass diese sowohl zur Untersuchung von Ursache-Wirkungs-Zusammenhängen der Ernährung mit Bioparametern und Krankheiten als auch zur Darstellung der Ernährung Anwendung finden. Die Vielzahl der bisher entwickelten Indizes zur Darstellung der Nährstoffzufuhr und des Lebensmittelverzehrs und die zahlreichen Anwendungsbereiche werden auch durch die Übersichtsarbeiten von Kant (1996) mit etwa 60 Veröffentlichungen zu Indizes aus den Jahren 1943 bis 1994 sowie Waijers et al. (2007) mit etwa 40 Veröffentlichungen aus den Jahren 1972 bis 2005 deutlich. Kant (1996) schlussfolgert, dass Indizes besonders vielversprechend für epidemiologische Studien sind, da eine Vielzahl an Variablen gleichzeitig berücksichtigt werden können (Kant 1996 S 790). Waijers et al. (2007) bestätigen zwar zehn Jahre später ebenfalls den potentiellen Nutzen, weisen jedoch darauf hin, dass der Aufbau und die damit verbundenen Limitationen eines Indexes, wie die Abhängigkeit von vorhandenen Empfehlungen, bei der Interpretation der Ergebnisse unbedingt berücksichtigt werden müssen (Waijers et al. 2007 S 228).

### **3.1.3 Anwendung des HEI-NVS II auf den Lebensmittelverzehr der Teilnehmenden der NVS II**

Zur Prüfung der Anwendbarkeit eines Indexes zur umfassend Darstellung des Lebensmittelverzehrs wurde ein eigener Index, der Healthy Eating Index für die Daten der NVS II (HEI-NVS II) entwickelt. Dafür wurden die für die deutsche Bevölkerung geltenden lebensmittelbasierten Empfehlungen der DGE (DGE und AID 2013, DGE et al. 2013) herangezogen. Zu den Lebensmittelgruppen, für die für den Index verwendbare, also quantifizierbare Empfehlungen vorliegen, gehören: Obst, Gemüse, Getreide und Kartoffeln, Milch, Fisch, Fleisch, Eier, Fett sowie die Gruppe der alkoholfreien Getränke. Daneben wurde Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke berücksichtigt.

Auf Basis der Empfehlungen der DGE und weiterer Institutionen (AID, FKE) wurden bereits verschiedene Indizes entwickelt (Beispiele: von Rüsten et al. 2010, Kleiser et al. 2009, Drescher et al. 2007, Krems und Brombach 2007, von Normann 2007, Bau et al. 2003, Sell et al. 2003, Winkler und Döring 1995), die allerdings nicht alle oben genannten Lebensmittelgruppen, für die es quantitative Empfehlungen gibt, berücksichtigen. Deshalb wurde in der vorliegenden Arbeit ein eigener Index entwickelt.

Die für die Auswertungen zugrunde gelegten Empfehlungen der DGE (DGE und AID 2013, DGE et al. 2013) beziehen sich auf durchschnittliche gesunde erwachsene Personen, bei

denen weder zwischen Frauen und Männern, noch zwischen unterschiedlich körperlich aktiven Personen, Personen mit einer Erkrankung, mit besonderen Ernährungsweisen beispielsweise aufgrund von Allergien sowie Personen, die eine Diät einhalten, oder auch Schwangeren und Stillenden unterschieden wird. Zudem ist es nicht möglich, bei der Berechnung der Indexwerte zu unterscheiden, ob eine Empfehlung am Tag oder pro Woche erreicht werden soll. Die zur Verfügung stehenden Verzehrdaten der NVS II beziehen sich auf die durchschnittlich pro Tag verzehrten Lebensmittelmengen der letzten vier Wochen. Bei der Berechnung der Indexwerte kann beispielsweise nicht unterschieden werden, ob eine Person die akzeptable Menge von Alkohol der ganzen Wochen an einem Tag getrunken hat.

Bei den Auswertungen wurden deshalb allen Studienteilnehmenden diesen Empfehlungen undifferenziert gegenübergestellt. Mit dem HEI-NVS II sind nur Aussagen auf Personengruppenebene und nicht auf individueller Ebene möglich.

### **Indexwerte des HEI-NVS II**

Für jeden Teilnehmenden werden die verzehrten Lebensmittelmengen aller Lebensmittelgruppen (Ist) mit der empfohlenen Verzehrsmenge aller Lebensmittelgruppen (Soll) verglichen und mit Punkten bewertet. Für die Bewertung wurde eine Verhältnisrechnung gewählt (Tab. 3.1).

**Tab. 3.1: Lebensmittelvariablen und Bewertungsbasis des HEI-NVS II**

Lebensmittelgruppen	Zufuhrempfehlung/ Richtwert	Berechnung des Indexwertes <sup>1</sup>
Obst <sup>2,3</sup>	250 g/Tag	Ist * 10 / Soll = maximal 15 Punkte
Gemüse <sup>2,3</sup>	400 g/Tag	
Getreide und Kartoffeln <sup>3</sup>	350–560 g/Tag	Verzehrer = Empfehlung: = 10 Punkte  Verzehrer < Empfehlung: Ist * 10 / Soll = < 10 Punkte  Verzehrer > Empfehlung: Soll * 10 / Ist = < 10 Punkte
Milch <sup>3</sup>	2 Portionen pro Tag: 1 Portion = 200–250 g Milch/Joghurt oder 50–60 g Käse/Quark	
Fisch <sup>3</sup>	150–220 g/Woche	
Fleisch <sup>3</sup>	<300–600 g/Woche	
Eier <sup>3</sup>	≤3 Eier d. h. ≤180 g/Woche	Verzehrer > Empfehlung: Soll * 10 / Ist = < 10 Punkte
Alkohol <sup>4</sup>	Frauen: ≤10 g Ethanol/Tag Männer: ≤20 g Ethanol/Tag	
Streichfett <sup>3</sup>	≤15–30 g/Tag	
Getränke <sup>3</sup>	≥1,5 l/Tag	Verzehrer ≥ Empfehlung: Ist * 10 / Soll = maximal 10 Punkte  Verzehrer < Empfehlung: Ist * 10 / Soll (<10 Punkte)

Die Farben in der Tabelle spiegeln unterschiedliche Vorgehensweisen für die Berechnung wider.

Ist = verzehrte Menge in Gramm pro Tag

Soll = empfohlene Menge in Gramm pro Tag

<sup>1</sup>In Anlehnung an von Rüsten et al. (2010 S 3, 2009 S 452f).

<sup>2</sup>Bei der Berechnung der Indexsumme, nicht aber bei den Indexwerten, wurde bei Obst und Gemüse, gemäß der Empfehlung, insgesamt ein Glas Saft (maximal 200 ml) als Ersatz einer Portion Obst oder Gemüse berücksichtigt (DGE und AID 2013 S 7).

<sup>3</sup>Mengenvorschläge zur Orientierung pro Tag (DGE und AID 2013 S 7)

<sup>4</sup>Alkohol (Ethanol) als Inhaltsstoff: D-A-CH Referenzwerte für die Nährstoffzufuhr (DGE et al. 2013 S 66f)

Die Indexwerte für die zehn Lebensmittelgruppen wurden in Anlehnung an von Rüsten et al. (2010 S 3, 2009 S 452f) folgendermaßen berechnet:

- Obst und Gemüse** zeichnen sich durch eine besonders hohe Nährstoffdichte aus. Sie liefern viele Nährstoffe bei gleichzeitig niedrigem Energiegehalt und weisen, wie in einer Stellungnahme der DGE (2012) beschrieben, präventives Potential auf. Die Empfehlungen für Obst (250 g/Tag) und Gemüse (400 g/Tag) stellen die Mengen dar, die mindestens verzehrt werden sollten. Ein höherer Verzehr ist wünschenswert (DGE und AID 2013 S 12). Deshalb wurden bei einem Verzehr oberhalb der Empfehlungen jeweils bis zu fünf Zusatzpunkte vergeben. Wird die Empfehlung nicht erreicht, werden anteilmäßig nicht alle Punkte vergeben.
- Für **Getreide und Kartoffeln, Milch sowie Fisch** gibt die DGE als Empfehlung eine Spanne an (DGE und AID 2013 S 7). Die Empfehlungen gelten als erreicht, wenn die verzehrte Menge innerhalb der empfohlenen Spanne liegt. In diesem Fall wurden zehn Punkte vergeben.

Das Unterschreiten der Empfehlungen führt zu einer geringeren Punktzahl als zehn, da ein zu geringer Verzehr dieser Lebensmittel mit einer zu geringen Zufuhr derjenigen Nährstoffe einhergeht, für die Getreide und Kartoffeln, Milch oder Fisch wesentliche Lieferanten sind.

Bei Überschreitung der Empfehlungen wurden ebenfalls weniger als zehn Punkte vergeben, da ein zu hoher Verzehr dieser Lebensmittel u. a. zu einer zu hohen Energiezufuhr führt. In diesem Fall erfolgte die Berechnung invers.

Bei der Gruppe Milch wurde zusätzlich zwischen der Verzehrsmenge von Milch/erzeugnissen und Käse/Quark unterschieden, da Käse/Quark aufgrund des geringeren Wassergehaltes eine höhere Konzentration der entsprechenden Nährstoffe aufweist. Wurden insgesamt zwei Portionen Milch/-erzeugnissen oder Käse/Quark verzehrt, wurden 10 Punkte vergeben.

3. **Fleisch, Eier und Alkohol** sollen nur in Maßen verzehrt werden (DGE und AID 2013 S 14, 22), da ein zu hoher Verzehr die Gesundheit beeinträchtigen kann. Die Empfehlungen werden als akzeptable Verzehrsmenge interpretiert. Für diese Gruppen werden jeweils zehn Punkte vergeben, wenn der Verzehr dieser Lebensmittel maximal dem oberen Wert der Empfehlung und bei Alkohol maximal dem Richtwert entsprach, darunter lag oder diese Lebensmittel gar nicht verzehrt wurden.

Bei Überschreitung der maximal empfohlenen Menge wird wie bei Milch, Getreide und Kartoffeln sowie Fisch eine inverse Rechnung vorgenommen.

4. Fett (**Streichfett** und Öl) ist ein wichtiger Energielieferant. Der menschliche Organismus ist auf die Zufuhr der überwiegend in Ölen enthaltenen Linolsäure und Alpha-Linolensäure angewiesen (DGE et al. 2013 S 45f). Die Verzehrsmengen von Ölen z. B. in Salatdressing liegen für die Studienteilnehmenden allerdings nicht vor und konnten damit nicht in die Berechnungen einbezogen werden. Aufgrund dessen bezieht sich die Berechnung des Indexwertes für den Verzehr von Fetten ausschließlich auf die Streichfette und wird analog zu den Berechnungen für Fleisch, Eier und Alkohol durchgeführt. Bei Unterschreitung der Empfehlung werden ebenfalls 10 Punkte vergeben. Bei Überschreitung der Empfehlung erfolgt analog eine inverse Berechnung.

5. Eine regelmäßige Flüssigkeitszufuhr in Form von **alkoholfreien Getränken** ist lebensnotwendig. Die DGE und AID (2013 S 22) empfiehlt mindestens 1,5 Liter alkoholfreie Getränke pro Tag zu trinken. Wird die Empfehlung erreicht oder überschritten, werden 10 Punkte vergeben. Wird die Empfehlung unterschritten, erfolgte wie bei Milch, Getreide und Kartoffeln sowie Fisch ein entsprechender Punktabzug nach einer inversen Berechnung.

Durch die beschriebene Berechnung werden bei den Indexwerten im Gegensatz zur Gegenüberstellung der Verzehrsmengen in Form von Mittelwerten (siehe Kapitel 2.3.2) auch die

Überschreitung der Verzehrsmenge z. B. von Fleisch, Eier und Milchprodukten in Form der Verhältnisrechnung berücksichtigt.

Pro Lebensmittelgruppe können damit meist maximal 10 Punkte erreicht werden, bei Obst und Gemüse je maximal 15 Punkte. Da nicht alle Lebensmittelgruppen den gleichen Wertebereich besitzen, handelt es sich um ungleichgewichtete Indexwerte.

### **HEI-NVS II-Indexsumme**

Da in den Empfehlungen der DGE nicht von gesunden oder ungesunden Lebensmitteln gesprochen wird, sondern von einer „angemessenen Menge und Kombination nährstoffreicher und energiearmer Lebensmittel“ (DGE und AID 2013 S 6) wird kein multiplikativer Index, sondern ein additiver Index verwendet. Durch Addition der einzelnen Indexwerte wird jedem Indexwert das gleiche Gewicht und die gleiche Bedeutung in dem HEI-NVS II geben. Je mehr die verzehrte Menge den Empfehlungen entspricht, desto höher ist die erreichte Indexsumme und desto besser ist die Lebensmittelauswahl zu beurteilen. Bei einer optimalen Indexsumme von 100 Punkten bzw. maximalen Indexsumme von 110 Punkten entspricht der Lebensmittelverzehr in allen berücksichtigten Lebensmittelgruppen den Empfehlungen der DGE.

### **3.1.4 Inhaltliche Ergebnisse**

Zur Darstellung des Lebensmittelverzehrs werden im Folgenden, getrennt für Frauen und Männer, die Indexwerte und Indexsummen präsentiert. Diese inhaltliche Auswertung dient als Basis, um zu prüfen, inwiefern der Index die Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs erfüllt, was im nächsten Kapitel vorgenommen wird.

Im folgenden Kapitel werden zunächst die **Indexwerte** der einzelnen Lebensmittelgruppen dargestellt, um aufzuzeigen, welche Teilnehmenden näher an den Empfehlungen der DGE bei einzelnen Lebensmittelgruppen liegen. Im nächsten Schritt wird die **Indexsumme** zur Darstellung des Lebensmittelverzehrs herangezogen. Zur Beschreibung des Lebensmittelverzehrs wird der arithmetische Mittelwert der einzelnen Lebensmittelgruppen verwendet. Anschließend wird die Indexsumme gekoppelt mit anderen Variablen zur weiteren Charakterisierung der Teilnehmenden herangezogen.

#### **Indexwerte einzelner Lebensmittelgruppen bei den Frauen**

Die berechneten Indexwerte der einzelnen Lebensmittelgruppen (Tab. 3.2) werden verwendet, um die Verzehrsmenge ausgewählter Lebensmittel mit den Empfehlungen zu vergleichen. Je höher der Indexwert, desto eher ernährt sich eine Person nach den Empfehlungen der DGE.

Die Darstellung des Lebensmittelverzehr der Frauen anhand der berechneten Indexwerte zeigt, dass Bio-Käuferinnen im Vergleich zu Nicht-Bio-Käuferinnen für Obst, Gemüse, Getreide und Kartoffeln, Milch, Fisch, Fleisch, Streichfett sowie für Getränke im Durchschnitt höhere Indexwerte erreichen und sich damit eher nach den Empfehlungen der DGE ernähren als Nicht-Bio-Käuferinnen. Bei Alkohol erreichen dagegen die Nicht-Bio-Käuferinnen höhere Indexwerte (getestet über Mann-Whitney-U-Test,  $p < 0,01$ ). Bei der Lebensmittelgruppe Eier können keine Unterschiede festgestellt werden.

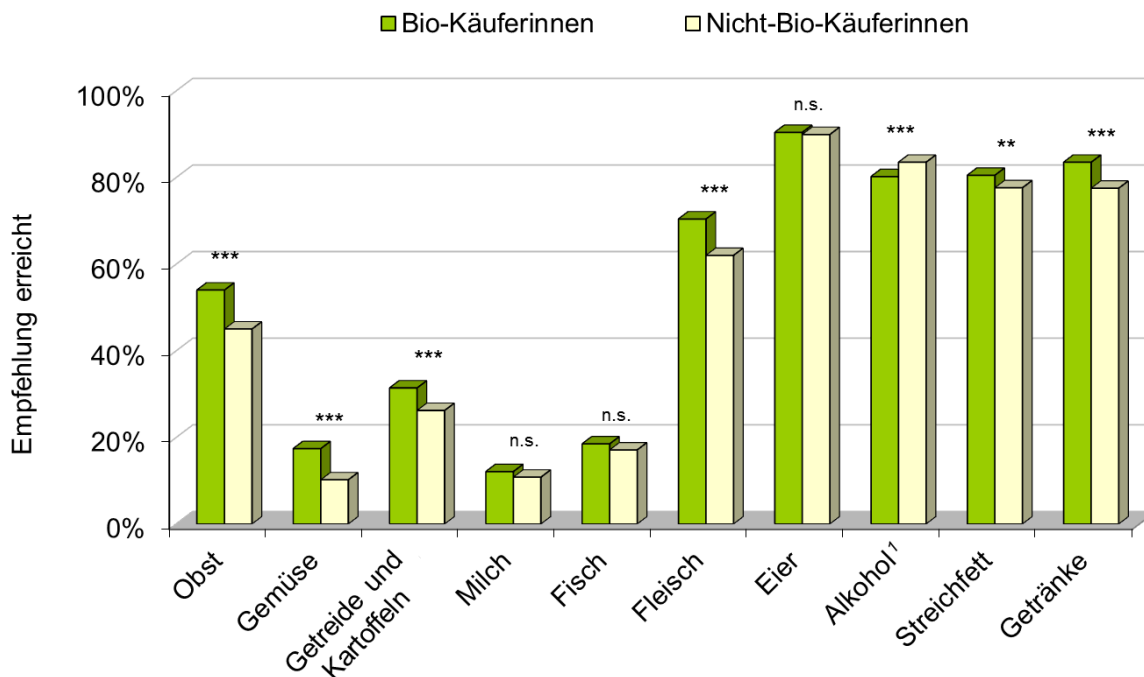
**Tab. 3.2: Indexwerte für einzelne Lebensmittelgruppen von Bio-Käuferinnen und Nicht-Bio-Käuferinnen (Frauen)**

	Lebensmittelgruppen	MW±Std dev	Min.	Max.	Perzentilen					Skewness <sup>2</sup>
					5.	25.	50.	75.	95.	
Bio-Käuferinnen n = 3577	<b>Obst</b>	10,1±4,5	0,0	15,0	1,9	6,4	10,7	15,0	15,0	-0,5
	<b>Gemüse</b>	6,7±3,5	0,0	15,0	2,0	4,1	6,0	8,7	14,2	0,7
	<b>Getreide und Kartoffeln</b>	8,2±2	0,0	10,0	4,4	6,9	8,7	10,0	10,0	-1,0
	<b>Milch</b>	6,7±2,6	0,0	10,0	1,9	4,8	6,9	9,0	10,0	-0,5
	<b>Fisch</b>	6,3±3,4	0,0	10,0	0,0	3,9	6,9	9,5	10,0	-0,6
	<b>Fleisch</b>	9,3±1,5	2,1	10,0	5,8	9,4	10,0	10,0	10,0	-2,1
	<b>Eier</b>	9,7±1	1,9	10,0	7,6	10,0	10,0	10,0	10,0	-4,4
	<b>Alkohol<sup>1</sup></b>	9,2±1,8	1,1	10,0	4,5	10,0	10,0	10,0	10,0	-2,4
	<b>Streichfett</b>	9,4±1,4	2,1	10,0	5,9	10,0	10,0	10,0	10,0	-2,8
	<b>Getränke</b>	9,6±1,2	0,9	10,0	6,6	10,0	10,0	10,0	10,0	-3,5
Nicht-Bio-Käuferinnen n = 3536	<b>Obst</b>	9±4,9	0,0	15,0	0,8	4,8	9,1	14,6	15,0	-0,2
	<b>Gemüse</b>	5,6±3,2	0,0	15,0	1,5	3,3	5,0	7,2	12,2	1,0
	<b>Getreide und Kartoffeln</b>	8±1,9	0,2	10,0	4,3	6,7	8,3	10,0	10,0	-0,8
	<b>Milch</b>	6,4±2,7	0,0	10,0	1,7	4,3	6,6	8,7	10,0	-0,3
	<b>Fisch</b>	5,9±3,5	0,0	10,0	0,0	3,1	6,4	9,2	10,0	3,5
	<b>Fleisch</b>	9±1,7	1,5	10,0	5,2	8,3	10,0	10,0	10,0	-1,7
	<b>Eier</b>	9,7±1	2,0	10,0	7,8	10,0	10,0	10,0	10,0	-4,6
	<b>Alkohol<sup>1</sup></b>	9,4±1,7	0,6	10,0	4,7	10,0	10,0	10,0	10,0	-2,8
	<b>Streichfett</b>	9,3±1,5	2,0	10,0	5,6	10,0	10,0	10,0	10,0	-2,4
	<b>Getränke</b>	9,4±1,4	0,5	10,0	5,9	10,0	10,0	10,0	10,0	-2,8

<sup>1</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

<sup>2</sup>Eine symmetrische Verteilung bzw. Normalverteilung weist eine Skewness von Null auf. Eine linksschiefe Verteilung besitzt eine negative und eine rechtsschiefe Verteilung eine positive Skewness.

Bio-Käuferinnen erreichen für die Lebensmittelgruppen Obst, Gemüse, Getreide und Kartoffel, Fleisch, Streichfett sowie den Getränken häufiger zehn Punkte und damit häufiger die Empfehlungen der DGE als die Nicht-Bio-Käuferinnen (Abb. 3.1).



\* =  $p < 0,05$  \*\* =  $p < 0,01$  \*\*\* =  $p < 0,001$  n.s. = nicht signifikant (getestet über: Chi-Quadrat)

<sup>1</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

**Abb. 3.1: Anteil der Bio-Käuferinnen und Nicht-Bio-Käuferinnen, mit einem Indexwert von 10 in einzelnen Lebensmittelgruppen (Frauen)**

Bei Alkohol erreichen die Nicht-Bio-Käuferinnen häufiger die maximal erreichbaren zehn Punkte. Bei den Lebensmittelgruppen Milch und Fisch konnten nur Unterschiede in den Indexwerten festgestellt werden, nicht aber im Anteil der Personen, die die Empfehlungen der DGE erreichen. Bei der Lebensmittelgruppe Eier konnten keine Unterschiede zwischen den Bio-Käuferinnen und Nicht-Bio-Käuferinnen festgestellt werden.

Die Empfehlungen für Eier, Streichfett, Alkohol und Getränke werden von etwa 80 % der Gesamtgruppe der Frauen erreicht. Deutlich weniger als ein Drittel der Frauen erreicht die empfohlene Menge für Gemüse, Milch und Fisch (Abb. 3.1). Erst nach der 75. Perzentile werden zehn Punkte erreicht. Die durchschnittliche Punktzahl für diese Lebensmittelgruppen liegen bei den Bio-Käuferinnen und Nicht-Bio-Käuferinnen zwischen 5,6 und 6,7 Punkten (Tab. 3.2).

### Indexwerte einzelner Lebensmittelgruppen bei den Männern

In der Gruppe der Männer zeigt sich ein ähnliches Bild wie bei den Frauen. Bio-Käufer erreichen im Vergleich zu Nicht-Bio-Käufern außer bei den Eiern höhere Indexwerte und ernähren sich damit eher nach den Empfehlungen der DGE. Lediglich bei Eiern erreichen Nicht-Bio-Käufer höhere Indexwerte (getestet über Mann-Whitney-U-Test,  $p < 0,05$ ) (Tab. 3.3).

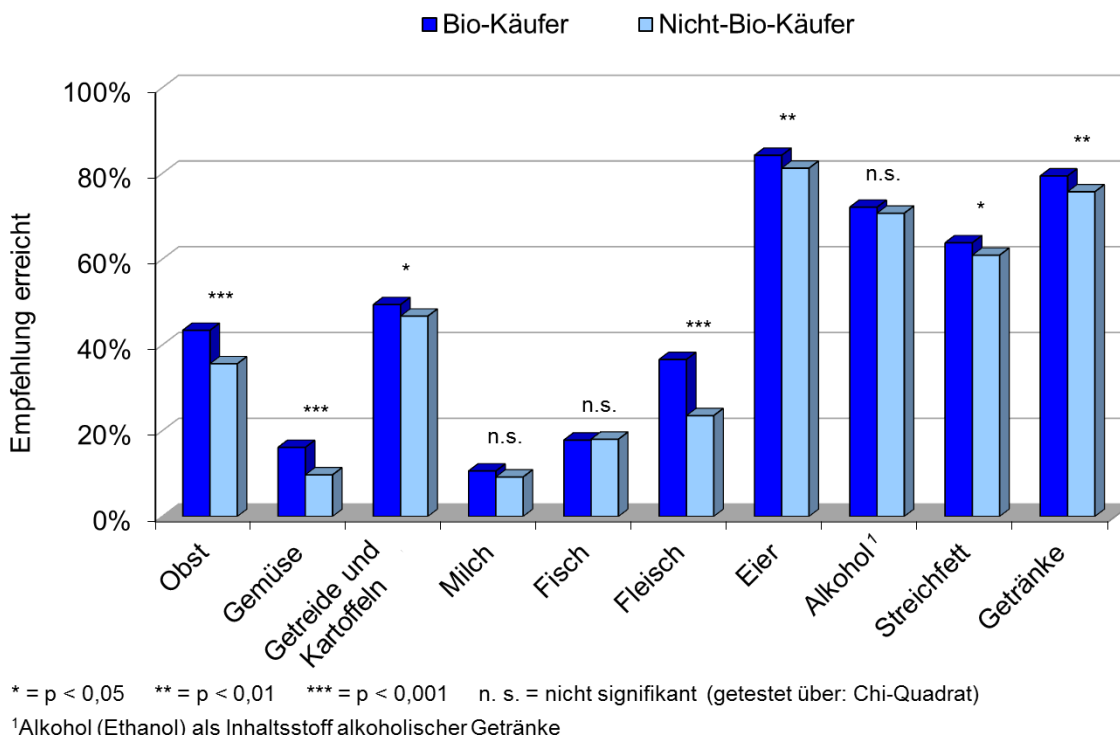
**Tab. 3.3 Indexwerte für einzelne Lebensmittelgruppen von Bio-Käufern und Nicht-Bio-Käufern (Männer)**

	Lebensmittelgruppen	MW± Std dev	Min.	Max.	Perzentilen					Skewness <sup>2</sup>
					5.	25.	50.	75.	95.	
Bio-Käufer n = 2298	<b>Obst</b>	8,8±4,9	0,0	15,0	0,8	4,8	8,6	14,5	15,0	-0,1
	<b>Gemüse</b>	6,4±3,4	0,1	15,0	2,0	3,9	5,7	8,4	13,6	0,8
	<b>Getreide und Kartoffeln</b>	9,0±1,5	1,4	10,0	5,7	8,3	10,0	10,0	10,0	-1,7
	<b>Milch</b>	6,4±2,7	0,0	10,0	1,7	4,4	6,6	8,7	10,0	-0,4
	<b>Fisch</b>	6,3±3,2	0,0	10,0	0,0	4,2	6,8	9,4	10,0	-0,6
	<b>Fleisch</b>	7,7±2,3	1,2	10,0	3,6	5,8	8,2	10,0	10,0	-0,6
	<b>Eier</b>	9,5±1,4	1,4	10,0	6,2	10,0	10,0	10,0	10,0	-3,2
	<b>Alkohol<sup>1</sup></b>	9,0±2,0	1,0	10,0	4,3	9,4	10,0	10,0	10,0	-1,9
	<b>Streichfett</b>	8,8±2,1	1,1	10,0	4,0	8,1	10,0	10,0	10,0	-1,6
	<b>Getränke</b>	9,5±1,4	0,9	10,0	6,0	10,0	10,0	10,0	10,0	-3,0
Nicht-Bio-Käufer n = 3663	<b>Obst</b>	7,8±5,1	0,0	15,0	0,0	3,3	7,4	12,4	15,0	0,1
	<b>Gemüse</b>	5,6±3,2	0,0	15,0	1,4	3,2	5,0	7,3	12,1	0,9
	<b>Getreide und Kartoffeln</b>	8,9±1,6	0,0	10,0	5,4	8,1	9,8	10,0	10,0	-1,7
	<b>Milch</b>	6,0±2,8	0,0	10,0	1,3	3,9	6,2	8,4	10,0	-0,2
	<b>Fisch</b>	6,0±3,5	0,0	10,0	0,0	3,4	6,7	9,4	10,0	-0,5
	<b>Fleisch</b>	6,8±2,5	0,9	10,0	2,8	4,8	6,7	9,8	10,0	-0,1
	<b>Eier</b>	9,4±1,6	0,8	10,0	5,2	10,0	10,0	10,0	10,0	-2,8
	<b>Alkohol<sup>1</sup></b>	8,9±2,1	0,9	10,0	3,6	8,8	10,0	10,0	10,0	-1,8
	<b>Streichfett</b>	8,5±2,3	1,3	10,0	3,6	7,2	10,0	10,0	10,0	-1,3
	<b>Getränke</b>	9,3±1,6	0,0	10,0	5,3	10,0	10,0	10,0	10,0	-2,7

<sup>1</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

<sup>2</sup>Eine symmetrische Verteilung bzw. Normalverteilung weist eine Skewness von Null auf. Eine linksschiefe Verteilung besitzt eine negative und eine rechtsschiefe Verteilung eine positive Skewness.

Die empfohlenen Verzehrsmengen und damit einen Indexwert von zehn erreichen männliche Bio-Käufer häufiger als Nicht-Bio-Käufer für Obst, Gemüse, Getreide und Kartoffel, Fleisch, Eier, Streichfett und Getränke (Abb. 3.2).



**Abb. 3.2: Anteil der Bio-Käufer und Nicht-Bio-Käufer, mit einem Indexwert von 10 in einzelnen Lebensmittelgruppen (Männer)**

Bei den Männern werden Eier, Alkohol und Getränke am häufigsten gemäß den Empfehlungen verzehrt. Etwa 70 % der Gesamtgruppe der Männer erreichen die Empfehlungen dieser Gruppen. Allerdings ist auch bei den Männern ersichtlich, dass sowohl bei den Bio-Käufern als auch bei den Nicht-Bio-Käufern jeweils deutlich weniger als ein Drittel der Männer die Empfehlungen für Gemüse, Milch und Fisch erreicht. Erst nach der 75. Perzentile werden zehn Punkte erzielt. Die durchschnittliche Punktzahl liegt bei den Bio-Käufern und Nicht-Bio-Käufern für diese Lebensmittelgruppen zwischen 5,6 und 6,4 Punkten (Tab. 3.2).

**HEI-NVS II-Indexsumme der Frauen**

Der Lebensmittelverzehr der Frauen wird im Folgenden mit Hilfe der Indexsumme, berechnet durch Addition der zehn Indexwerte, dargestellt (Tab. 3.4). Bio-Käuferinnen weisen mit durchschnittlich 86,6 von maximal 110 Punkten eine signifikant höhere Indexsumme auf als Nicht-Bio-Käuferinnen mit 83,0 Punkten (getestet über Mann-Whitney-U-Test, p <0,001).

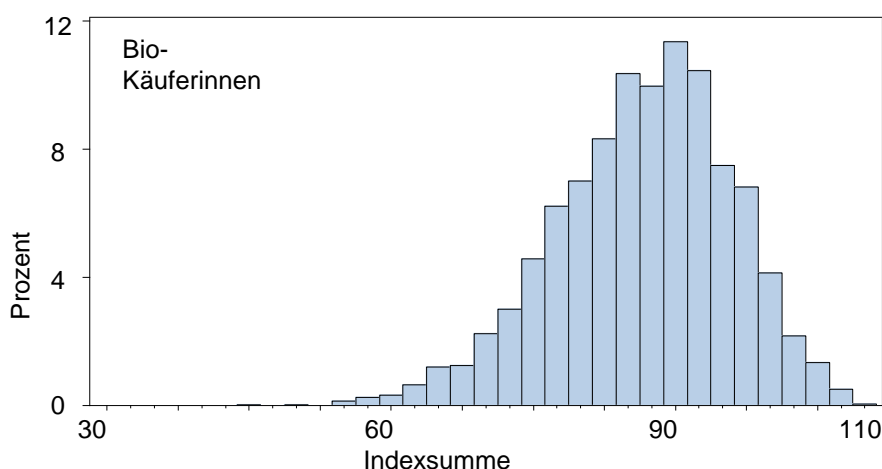
**Tab. 3.4: HEI-NVS II-Indexsummen der Bio-Käuferinnen und Nicht-Bio-Käuferinnen (Frauen)**

	MW±Std dev	Min.	Max.	Perzentilen					Skewness <sup>1</sup>
				5.	25.	50.	75.	95.	
<b>Bio-Käuferinnen</b>	86,6±9,3	45,5	110,0	70,1	80,6	87,4	93,1	100,4	-0,4
<b>Nicht-Bio-Käuferinnen</b>	83,0±10,0	42,6	108,3	65,5	76,4	83,7	90,1	98,4	-0,3

<sup>1</sup>Eine symmetrische Verteilung bzw. Normalverteilung weist eine Skewness von Null auf. Eine linksschiefe Verteilung besitzt eine negative und eine rechtsschiefe Verteilung eine positive Skewness.

Bio-Käuferinnen ernähren sich insgesamt günstiger als Nicht-Bio-Käuferinnen. Allerdings ist, wie in Abb. 3.3 ersichtlich, die Streuung der erreichten Indexsummen sehr groß. Bio-Käuferinnen erreichen Indexsummen zwischen 45,5 und 110,0 Punkten. Eine Bio-Käuferin erreicht die Maximalpunktzahl. Das bedeutet, dass es auch Bio-Käuferinnen gibt (<5 %), die nicht einmal die Hälfte der Maximalpunktzahl erreichen und deren Ernährung damit stark von den Empfehlungen abweicht.

Bei den Nicht-Bio-Käuferinnen ist ein ähnlicher Verlauf erkennbar (grafisch nicht dargestellt), nur auf einem niedrigeren Niveau, da diese Gruppe eine insgesamt geringere Indexsumme erreicht. Die niedrigste Indexsumme liegt bei 42,6 und die höchste Indexsumme bei 108,3.



**Abb. 3.3: Häufigkeitsverteilung der HEI-NVS II-Indexsummen der Bio-Käuferinnen (Frauen)**

Zudem zeigt sich ein positiver Zusammenhang zwischen der Bio-Kaufintensität und der erreichten Indexsumme. Die höchste Indexsumme erreicht die Gruppe der Intensiv-Bio-Käuferinnen mit durchschnittlich 88,5 von 110 möglichen Punkten, die geringste die Nicht-Bio-Käuferinnen mit 83,0 Punkten (Tab. 3.5). Es zeigt sich, dass die Spanne der geringsten und höchsten Indexsumme in allen vier Bio-Käufergruppen sehr weit ist.

**Tab. 3.5: HEI-NVS II-Indexsummen der Bio-Käuferinnen unterschiedlicher Kaufintensität und Nicht-Bio-Käuferinnen (Frauen)**

	MW±Std dev	Min.	Max.	Perzentilen					Skewness <sup>1</sup>
				5.	25.	50.	75.	95.	
<b>Intensiv-Bio-Käuferinnen</b>	88,5±9,1	58,6	106,9	72,4	82,8	89,4	95,0	102,3	-0,5
<b>Gelegenheits-Bio-Käuferinnen</b>	86,9±9,2	45,5	110,0	70,6	81,2	87,9	93,2	100,4	-0,5
<b>Selten-Bio-Käuferinnen</b>	85,1±9,4	54,4	107,7	68,6	78,5	85,6	91,8	99,6	-0,3
<b>Nicht-Bio-Käuferinnen</b>	83,0±10,0	42,6	108,3	65,5	76,4	83,7	90,1	98,4	-0,3

<sup>1</sup>Eine symmetrische Verteilung bzw. Normalverteilung weist eine Skewness von Null auf. Eine linksschiefe Verteilung besitzt eine negative und eine rechtsschiefe Verteilung eine positive Skewness.

Bio-Käuferinnen treffen einerseits eine Lebensmittelauswahl näher an den Empfehlungen als Nicht-Bio-Käuferinnen, andererseits gibt es einen steigenden Trend von einer Lebensmittelauswahl näher an den Empfehlungen von den Nicht-Bio-Käuferinnen über die Selten-Bio-Käuferinnen, die Gelegenheits-Bio-Käuferinnen zu den Intensiv-Bio-Käuferinnen (getestet über Jonckheere-Terpstra-Test,  $p < 0,001$ ).

**HEI-NVS II-Indexsumme der Männer**

Die HEI-NVS II-Indexsumme der Männer zeigt, dass Bio-Käufer mit 82,9 von maximal 110 Punkten eine signifikant höhere Indexsumme als Nicht-Bio-Käufer mit 78,6 Punkten aufweisen (getestet über Mann-Whitney-U-Test,  $p < 0,001$ ) (Tab. 3.6).

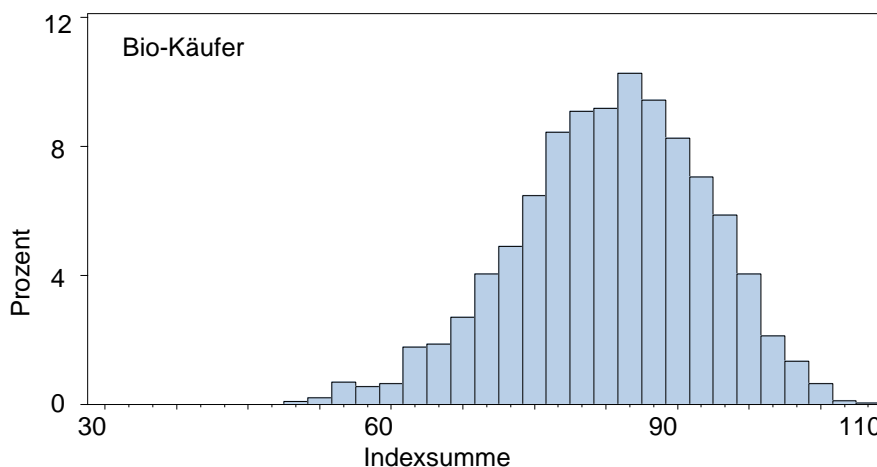
**Tab. 3.6: HEI-NVS II-Indexsummen der Bio-Käufer und Nicht-Bio-Käufer (Männer)**

	MW±Std dev	Min.	Max.	Perzentilen					Skewness <sup>1</sup>
				5.	25.	50.	75.	95.	
<b>Bio-Käufer</b>	82,9±10,0	51,0	108,8	65,5	76,5	83,6	90,1	98,4	-0,3
<b>Nicht-Bio-Käufer</b>	78,6±10,8	34,6	109,4	60,0	71,3	79,0	86,3	95,5	-0,2

<sup>1</sup>Eine symmetrische Verteilung bzw. Normalverteilung weist eine Skewness von Null auf. Eine linksschiefe Verteilung besitzt eine negative und eine rechtsschiefe Verteilung eine positive Skewness.

In Abb. 3.4 ist ersichtlich, dass auch bei den Männern die Streuung der erreichten Indexsummen sehr groß ist. Bio-Käufer erreichen Indexsummen zwischen 51,0 und 108,8 Punkten. Das bedeutet, dass es ebenso wie bei den Bio-Käuferinnen auch Bio-Käufer gibt, die nicht einmal die Hälfte der Maximalpunktzahl erreichen und deren Ernährung damit stark von den Empfehlungen abweicht. Die Maximalpunktzahl wird von keinem Mann erreicht.

Bei den Nicht-Bio-Käufern ist ein ähnlicher Verlauf wie bei den Bio-Käufern erkennbar (grafisch nicht dargestellt), nur auf einem niedrigeren Niveau, da diese Gruppe eine insgesamt geringere Indexsumme erreicht. Die niedrigste Indexsumme liegt bei 34,6 und die höchste Indexsumme bei 109,4.



**Abb. 3.4: Häufigkeitsverteilung der HEI-NVS II-Indexsummen der Bio-Käufer (Männer)**

Zudem zeigt sich auch in der Gruppe der Männer ein positiver Zusammenhang zwischen der Bio-Kaufintensität und der erreichten Indexsumme. Die höchste Indexsumme erreicht die Gruppe der Intensiv-Bio-Käufer mit durchschnittlich 84,2 von 110 möglichen Punkten, die geringste die Nicht-Bio-Käufer mit 78,6 Punkten (Tab. 3.7). Es zeigt sich, dass die Spanne der geringsten und höchsten Indexsumme in allen vier Bio-Käufergruppen sehr weit ist.

**Tab. 3.7: HEI-NVS II-Indexsummen der Bio-Käufer unterschiedlicher Kaufintensität und Nicht-Bio-Käufer (Männer)**

	MW±Std dev	Min.	Max.	Perzentilen					Skewness <sup>1</sup>
				5.	25.	50.	75.	95.	
<b>Intensiv-Bio-Käufer</b>	84,2±9,9	53,7	104,6	66,2	77,9	84,2	90,8	99,2	-0,5
<b>Selten-Bio-Käufer</b>	81,4±10,0	52,1	104,9	63,5	75,4	81,3	88,4	97,6	-0,3
<b>Gelegenheits-Bio-Käufer</b>	83,6±9,9	51,0	108,8	66,0	76,7	84,6	90,7	98,5	-0,3
<b>Nicht-Bio-Käufer</b>	78,6±10,8	34,6	109,4	60,0	71,3	79,0	86,3	95,5	-0,2

<sup>1</sup>Eine symmetrische Verteilung bzw. Normalverteilung weist eine Skewness von Null auf. Eine linksschiefe Verteilung besitzt eine negative und eine rechtsschiefe Verteilung eine positive Skewness.

Bio-Käufer/innen treffen einerseits eine Lebensmittelauswahl näher an den Empfehlungen als Nicht-Bio-Käufer/innen, andererseits gibt es bei beiden Geschlechtern einen steigenden Trend einer Lebensmittelauswahl näher an den Empfehlungen von den Nicht-Bio-Käufer/innen über die Selten-Bio-Käufer/innen, die Gelegenheits-Bio-Käufer/innen zu den Intensiv-Bio-Käufer/innen (getestet über Jonckheere-Terpstra-Test,  $p < 0,001$ ).

### **Charakterisierung von Personengruppen mit unterschiedlicher Indexsumme**

Durch die Indexsumme werden mehrere für das Untersuchungsziel relevante Lebensmittelgruppen gleichzeitig herangezogen und damit eine möglichst umfassende Darstellung des Lebensmittelverzehr angestrebt. Im Folgenden wird untersucht, inwiefern sich Personengruppen, die eine hohe Indexsumme erreicht haben und sich damit eher nach den Empfehlungen der DGE ernähren, von Personengruppen unterscheiden, die eine geringere Indexsumme erreicht haben. Hierfür werden Bio-Käufer/innen, die sich gemessen an den Empfehlungen besonders günstig ernähren, mit jenen Bio-Käufer/innen, die die Empfehlungen in einem geringeren Ausmaß erreichen, verglichen. Für diesen Vergleich wurden die Bio-Käuferinnen und Bio-Käufer anhand ihrer erreichten Indexsumme in Terzilen, berechnet auf drei Nachkommastellen, aufgeteilt. Die Gruppen werden entsprechend der Lebensmittelauswahl im Folgenden „Empfehlungen wenig erreicht“ (niedrige Indexsumme), „Empfehlungen mittelmäßig erreicht“ (mittlere Indexsumme) und „Empfehlungen gut erreicht“ (hohe Indexsumme) genannt (Tab. 3.8).

**Tab. 3.8: Indexgrenzwerte zur Einteilung der Bio-Käuferinnen und Bio-Käufer in die Terzilen: Empfehlungen wenig, mittelmäßig und gut erreicht**

	Lebensmittelauswahl (Terzilen der HEI-NVS II-Indexsumme)		
	Empfehlungen wenig erreicht	Empfehlungen mittel erreicht	Empfehlungen gut erreicht
<b>Bio-Käuferinnen</b>	<83,170	≥83,170 und <91,186	≥91,186
<b>Bio-Käufer</b>	<78,999	≥78,999 und <87,687	≥87,687

Bio-Käuferinnen der ersten Terzile weisen eine Indexsumme unter 83,170 und Bio-Käuferinnen der dritten Terzile eine Indexsumme von mindestens 91,186 Punkten auf. Die erreichte Punktzahl der Bio-Käufer liegt deutlich unter denen der Bio-Käuferinnen. Bio-Käufer der ersten Terzile erreichen weniger als 78,999 Punkte und die der dritten Terzile mindestens 87,687 Punkte.

Die Charakterisierung der Bio-Käuferinnen und Bio-Käufer anhand der Zuordnung in die Terzilen zeigt, dass sowohl bei den Bio-Käuferinnen als auch bei den Bio-Käufern Zusammenhänge zwischen der Lebensmittelauswahl einerseits und soziodemographischen sowie gesundheitsrelevanten Variablen andererseits bestehen (Tab. 3.9 und Tab. 3.10).

**Tab. 3.9: Charakterisierung der Gesamtgruppe sowie der Bio-Käuferinnen mit unterschiedlicher Lebensmittelauswahl (Frauen)**

		Gesamt- gruppe (%)  n=3577	Empfehlungen:			Signifikanzen <sup>1</sup>
			wenig erreicht (%)  n=1192	mittelmäßig erreicht (%)  n=1192	gut erreicht (%)  n=1193	
<b>Altersgruppe</b>	18–34 Jahre	18,4	24,1	17,6	13,5	<0,001
	35–50 Jahre	37,2	38,8	37,8	35,0	n.s.
	51–64 Jahre	25,5	21,9	25,2	29,5	<0,001
	65–80 Jahre	18,8	15,2	19,4	22,0	<0,001
<b>Soziale Schicht<sup>3</sup></b>	untere Schicht	15,7	17,8	14,4	14,9	<0,05
	mittlere Schicht	27,3	28,1	28,0	25,8	n.s.
	obere Schicht	57,0	54,1	57,6	59,3	<0,01
<b>BMI</b>	Untergewicht	1,8	2,4	1,5	1,5	<0,05
	Normalgewicht	49,6	51,0	49,3	48,4	n.s.
	Übergewicht	26,1	24,8	24,2	29,3	<0,01
	Adipositas	16,7	15,3	18,2	16,6	n.s.
	keine Angabe	5,8	6,5	6,8	4,2	... <sup>2</sup>
<b>Taillenumfang</b>	bis 88 cm	69,1	70,8	65,6	70,8	n.s.
	über 88 cm	26,1	23,5	29,0	25,7	n.s.
	keine Angabe	4,8	5,7	5,4	3,4	... <sup>2</sup>
<b>WHR</b>	bis 0,85	76,5	77,4	74,2	78,0	n.s.
	über 0,85	18,7	17,0	20,4	18,6	n.s.
	keine Angabe	4,8	5,7	5,4	3,4	... <sup>2</sup>
<b>Energiezufuhr</b>	kcal/Tag (MW)	1992,6	1846,4	1982,4	2149,0	<0,001
<b>Ernährungs- kenntnisse</b>	gut	49,2	42,4	49,8	55,5	<0,001
	mittel	31,7	34,7	31,6	28,7	<0,001
	schlecht	18,3	21,6	18,0	15,2	<0,001
	keine Angabe	0,9	1,3	0,6	0,7	... <sup>2</sup>
<b>Sportlich aktiv</b>	ja	65,9	60,0	64,5	73,3	<0,001
	nein	33,4	39,7	34,5	26,1	<0,001
	keine Angabe	0,7	0,3	1,0	0,7	... <sup>2</sup>
<b>Raucher</b>	ja	17,9	24,4	17,4	12,1	<0,001
	nein	81,7	75,4	82,1	87,5	<0,001
	keine Angabe	0,4	0,3	0,5	0,4	... <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	79,1	76,6	81,0	79,6	<0,05
	mittel	17,5	19,7	16,0	16,8	n.s.
	schlecht	3,2	3,5	2,6	3,4	n.s.
	keine Angabe	0,3	0,2	0,3	0,3	... <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	79,8	77,9	82,4	79,1	n.s.
	schlecht	19,6	21,1	17,5	20,4	n.s.
	keine Angabe	0,6	1,1	0,2	0,5	... <sup>2</sup>

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistisch signifikanter Trend von einer niedrigen bis zu einer hohen Indexsumme vorliegt (getestet über: Jonckheere-Terpstra-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Trendtest wurde hier nicht berechnet.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

In der jüngsten Altersgruppe, der 18- bis 34-Jährigen, sinkt der Anteil der Bio-Käuferinnen mit steigender Indexsumme. Ein Unterschied innerhalb der 35- bis 50-Jährigen konnte in Abhängigkeit der Indexsumme nicht festgestellt werden. In der Gruppe der 51- bis 64-Jährigen und 65- bis 80-Jährigen Bio-Käuferinnen ist ein Anstieg der Bio-Käuferinnen mit steigender Indexsumme zu beobachten. Daneben steigt in den oberen sozialen Schichten

mit steigender Indexsumme der Anteil an Bio-Käuferinnen. In der unteren sozialen Schicht sinkt der Anteil an Bio-Käuferinnen mit steigender Indexsumme.

Es konnte auch ein Zusammenhang zwischen der Indexsumme und dem BMI der Bio-Käuferinnen festgestellt werden. Der Anteil übergewichtiger (nicht aber adipöser) Bio-Käuferinnen steigt mit steigender Indexsumme. Im Anteil normalgewichtiger Bio-Käuferinnen konnten keine Unterschiede bei steigender Indexsumme festgestellt werden. Im Taillenumfang und im WHR, die neben dem BMI als Marker für das Fettverteilungsmuster herangezogen werden können (Schauder und Ollenschläger 2003 S 538, WHO 2000 S 7ff), konnte kein Zusammenhang mit der Indexsumme festgestellt werden. Allerdings ist mit steigender Indexsumme ein Anstieg der Energiezufuhr zu beobachten.

Bei den Ernährungskennnissen zeigt sich, dass mit der Indexsumme der Anteil der Bio-Käuferinnen steigt, die sich durch gute Ernährungskennnisse auszeichnen. Daneben steigt der Anteil an Nichtrauchern sowie sportlich Aktiven Bio-Käuferinnen mit steigender Indexsumme.

Als weiterer Marker eines gesundheitsbewussten Verhaltens wurden Selbstangaben zum allgemeinen Gesundheitszustand und zur Schlafqualität der Teilnehmenden herangezogen. Bei den Einschätzungen des allgemeinen Gesundheitszustands konnte bei den Bio-Käuferinnen gezeigt werden, dass mit steigender Indexsumme der Anteil an Bio-Käuferinnen steigt, die angaben, einen guten Gesundheitszustand zu besitzen. Ein Zusammenhang zwischen der Indexsumme und der Schlafqualität konnte bei den Bio-Käuferinnen nicht festgestellt werden.

**Tab. 3.10: Charakterisierung der Gesamtgruppe sowie der Bio-Käufer mit unterschiedlicher Lebensmittelauswahl (Männer)**

		Gesamt- gruppe (%)  n=2298	Empfehlungen:			Signifikanzen <sup>2</sup>
			wenig erreicht (%) n =766	mittelmäßig erreicht (%) n =766	gut erreicht (%) n =766	
<b>Altersgruppe</b>	18–34 Jahre	15,6	20,6	15,8	10,4	<0,001
	35–50 Jahre	35,1	36,4	38,4	30,4	n.s.
	51–64 Jahre	26,7	24,0	26,1	30,0	<0,01
	65–80 Jahre	22,6	18,9	19,7	29,1	<0,001
<b>Soziale Schicht<sup>3</sup></b>	untere Schicht	11,9	15,4	11,5	8,9	--- <sup>2</sup>
	mittlere Schicht	23,5	24,9	21,5	23,9	n.s.
	obere Schicht	64,6	59,7	67,0	67,2	<0,001
<b>BMI</b>	Untergewicht	0,6	1,0	0,4	0,4	<0,05
	Normalgewicht	31,8	31,3	35,1	29,0	n.s.
	Übergewicht	45,2	43,1	42,4	50,0	<0,01
	Adipositas	17,3	18,0	17,5	16,3	n.s.
	keine Angaben	5,1	6,5	4,6	4,3	--- <sup>2</sup>
<b>Tailenumfang</b>	bis 102 cm	71,6	69,1	73,1	72,7	n.s.
	über 102 cm	24,0	25,3	23,6	24,3	n.s.
	keine Angabe	41,0	5,6	3,3	3,0	--- <sup>2</sup>
<b>WHR</b>	bis 1,0	83,0	79,2	86,0	83,8	n.s.
	über 1,0	13,0	15,1	10,7	13,2	n.s.
	keine Angabe	4,0	5,6	3,3	3,0	--- <sup>2</sup>
<b>Energiezufuhr</b>	kcal/Tag (MW)	2528,6	2529,0	2488,3	2568,4	<0,05
<b>Ernährungs- kenntnisse</b>	gut	23,8	18,3	23,0	30,2	<0,001
	mittel	41,2	41,0	44,9	37,6	n.s.
	schlecht	34,1	39,6	31,1	31,7	<0,001
	keine Angabe	0,9	1,2	1,0	0,5	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	64,0	58,2	64,5	69,3	<0,001
	nein	35,2	41,1	34,5	30,0	<0,001
	keine Angabe	0,8	0,7	1,0	0,7	--- <sup>2</sup>
<b>Raucher</b>	ja	22,5	31,2	20,6	15,7	<0,001
	nein	77,2	68,5	79,0	83,9	<0,001
	keine Angabe	0,4	0,3	0,4	0,4	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	76,3	74,3	76,6	78,1	n.s.
	mittel	19,9	22,1	19,5	18,2	n.s.
	schlecht	3,7	3,7	3,9	3,5	n.s.
	keine Angabe	0,1	0,0	0,0	0,3	--- <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	85,3	83,4	85,4	87,1	<0,05
	schlecht	13,2	16,5	14,2	12,0	<0,05
	keine Angabe	0,5	0,1	0,4	0,9	--- <sup>2</sup>

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistisch signifikanter Trend von einer niedrigen bis zu einer hohen Indexsumme vorliegt (getestet über: Jonckheere-Terpstra-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Trendtest wurde hier nicht berechnet.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

Ebenso wie in der Gruppe der Frauen sinkt der Anteil der Bio-Käufer, in der Gruppe der 18- bis 34-Jährigen mit steigender Indexsumme. In allen drei Indexgruppen machen Bio-Käufer im Alter von 35- bis 50 Jahren den größten Prozentsatz aus. Ein Unterschied in Abhängigkeit der Indexsumme in der Altersgruppe der 35- bis 51-Jährigen konnte auch hier nicht festgestellt werden. In der Gruppe der 51- bis 64-Jährigen und 65- bis 80-Jährigen ist ein Anstieg des Prozentsatzes mit steigender Indexsumme zu beobachten. Daneben ist ein

Anstieg der Bio-Käufer aus den oberen sozialen Schichten mit steigender Indexsumme zu erkennen.

Zudem gibt es auch bei den Bio-Käufern einen Zusammenhang zwischen der Indexsumme und dem BMI. Der Anteil übergewichtiger (nicht aber adipöser) Bio-Käufer steigt mit steigender Indexsumme. Im Anteil Normalgewichtiger sowie bei den Variablen Taillenumfang und WHR konnte bei den Bio-Käufern kein Zusammenhang mit der Indexsumme festgestellt werden. Allerdings ist mit steigender Indexsumme ein Anstieg der Energiezufuhr zu beobachten.

Bei der Variable Ernährungskennnisse zeigt sich, dass der Anteil der Bio-Käufer, die sich durch gute Ernährungskennnisse auszeichnen, mit der Indexsumme steigt. Der Anteil, der sich durch schlechte Ernährungskennnisse auszeichnet, sinkt. Daneben geht mit steigender Indexsumme ein höherer Anteil an Nichtrauchern sowie sportlich Aktiven einher.

Bei den Einschätzungen des allgemeinen Gesundheitszustands konnten keine Unterschiede festgestellt werden. Der Anteil der Bio-Käufer, die angaben, eine gute Schlafqualität zu besitzen, steigt mit steigender Indexsumme.

### **Gleichzeitigkeit mehrerer Variablen**

Um zu prüfen, ob der Lebensmittelverzehr bei unterschiedlicher Kaufintensität von Bio-Lebensmitteln gleichzeitig mit einem gesundheitsbewussteren Verhalten oder besseren Ernährungskennnissen im Zusammenhang steht, wurde ebenfalls die Indexsumme herangezogen. Zum einen wurde der Zusammenhang zwischen der HEI-NVS II-Indexsumme, der Bio-Kaufintensität und dem Gesundheitsverhalten und zum anderen der Zusammenhang zwischen der HEI-NVS II-Indexsumme, der Bio-Kaufintensität und den Ernährungskennnissen untersucht.

Zur Beschreibung des gesundheitsrelevanten Verhaltens wurde der BMI in Kombination mit dem Taillenumfang (Janssen et al. 2005 S 683ff, Schauder und Ollenschläger 2003 S 538, WHO 2000 S 7ff), dem Raucherstatus sowie die Angabe zur sportlichen Aktivität, als Indikatoren für das Gesundheitsverhalten der Teilnehmenden, zu einem Index zusammengefasst (Tab. 3.11).

**Tab. 3.11: Variablen und Bewertungsbasis des Indexes zum Gesundheitsverhalten**

Variablen des Gesundheitsverhaltens		Berechnung des Indexwertes zum Gesundheitsverhalten	
<b>BMI und Taillenumfang</b>	BMI	BMI $\geq 18,5$ bis $< 25$	1
	Taillenumfang	Männer: $< 94$ cm Frauen: $< 80$ cm	1
	BMI	BMI $\geq 17$ bis $< 18,5$ BMI $\geq 25$ bis $< 30$	0,5
	Taillenumfang	Männer: $94$ cm $< 102$ cm Frauen: $80$ cm $< 88$ cm	0,5
	BMI	BMI $< 17$ oder $\geq 30$	0
	Taillenumfang	Männer: $\geq 102$ cm Frauen: $\geq 88$ cm	0
<b>Rauchverhalten</b>	Nichtraucher/innen		2
	Ehemalige Raucher/innen		1
	Raucher/innen		0
<b>Sportlich aktiv</b>	Ja		2
	Nein		0
<b>Index Gesundheitsverhalten<sup>1</sup></b>			
<b>(Summe / Anzahl der vorhandenen Angaben)</b>		Gut	=2
		Mittel	$\geq 1$ und $< 2$
		Schlecht	$< 1$

<sup>1</sup>In den drei berücksichtigten Gruppen konnten jeweils null bis zwei Punkte erreicht werden. Die Summe der drei Gruppen wird durch die Anzahl der Nennungen dividiert. Wenn z. B. keine Angaben zum Rauchverhalten vorliegen, wird die Gesamtsumme nur durch zwei Angaben dividiert, wodurch insgesamt null bis vier Punkte erreichbar sind.

Es zeigt sich, dass mit steigender Bio-Kaufintensität und einem **gleichzeitig** gesundheitsbewussteren Verhalten (d. h. geringerer Anteil an Raucherinnen, höherer Anteil sportlich Aktiver sowie höherer Anteil an Normalgewichtigen) eine höhere HEI-NVS II-Indexsumme erreicht wird (Abb. 3.5). Intensiv-Bio-Käuferinnen erreichen 89,9 Punkte, Gelegentlich-Bio-Käuferinnen 88,5, Selten-Bio-Käuferinnen 85,9 und Nicht-Bio-Käuferinnen 84,8 Punkte. Frauen, die keine Bio-Lebensmittel kaufen und ein eher ungünstiges Gesundheitsverhalten aufweisen, erreichen lediglich 80,9 Punkte.

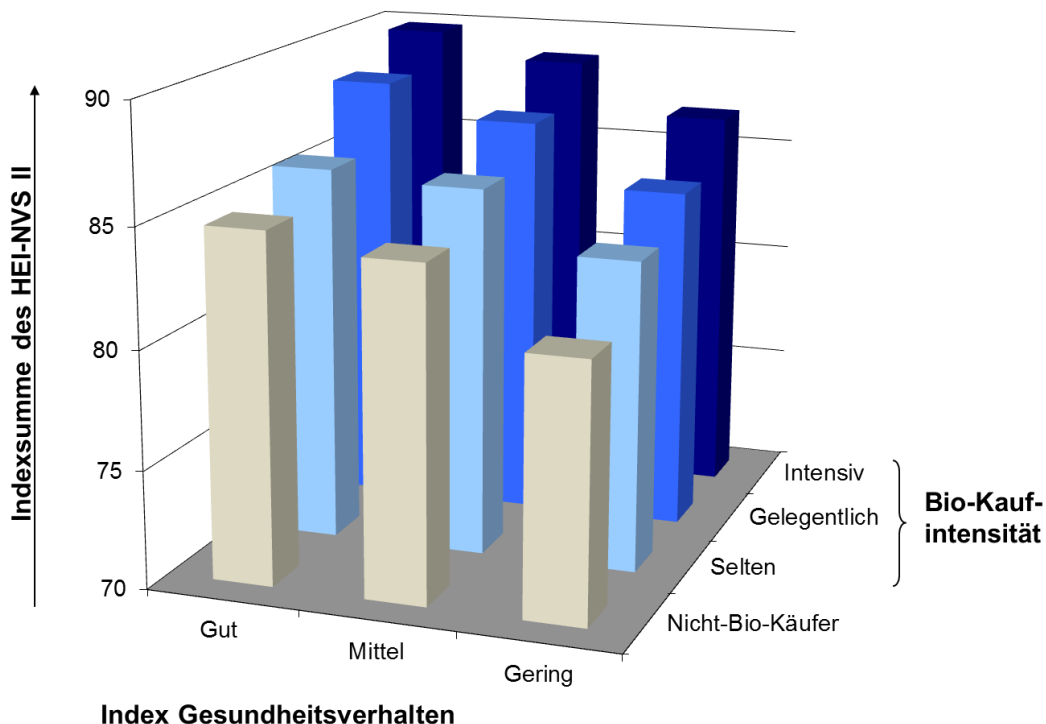


Abb. 3.5: Gegenüberstellung von Bio-Kaufintensität, Gesundheitsverhalten und der HEI-NVS II-Indexsumme (Frauen)

In der Gruppe der Männer zeigt sich ein anderes Bild. In dieser Gruppe erreichen die Gelegenheits-Bio-Käufer mit einem guten Gesundheitsverhalten 86,2 Punkte und die Intensiv-Bio-Käufer 83,4 Punkte. Männer, die keine Bio-Lebensmittel kaufen und ein eher ungünstiges Gesundheitsverhalten aufweisen, erreichen 77,1 von 110 möglichen Punkten (Abb. 3.6).

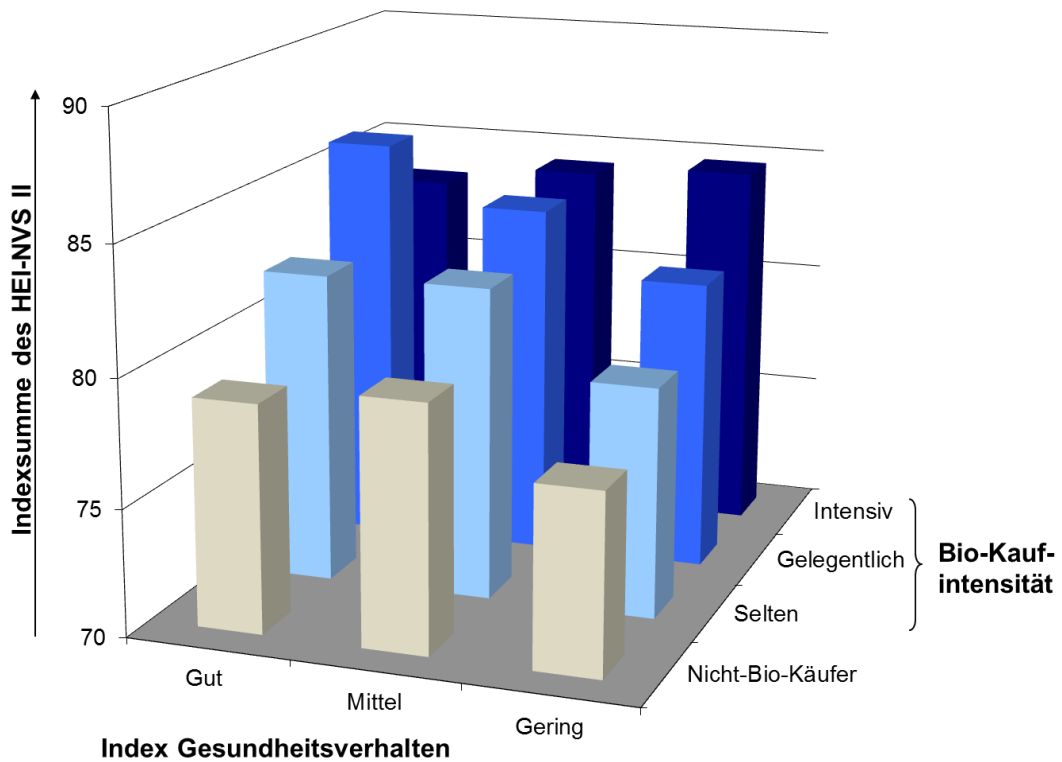


Abb. 3.6: Gegenüberstellung von Bio-Kaufintensität, Gesundheitsverhalten und der HEI-NVS II-Indexsumme (Männer)

Die gleichzeitige Untersuchung der HEI-NVS II-Indexsumme, der Bio-Kaufintensität und den Ernährungskennnissen zeigt, dass in der Gruppe der Frauen mit steigender Bio-Kaufintensität und gleichzeitig besseren Ernährungskennnissen eine höhere Indexsumme erreicht wird (Abb. 3.7). Intensiv-Bio-Käuferinnen mit guten Ernährungskennnissen erreichen 90,2 Punkte. Frauen, die keine Bio-Lebensmittel kaufen und nur geringe Ernährungskennnisse aufweisen, erreichen lediglich 81,5 Punkte.

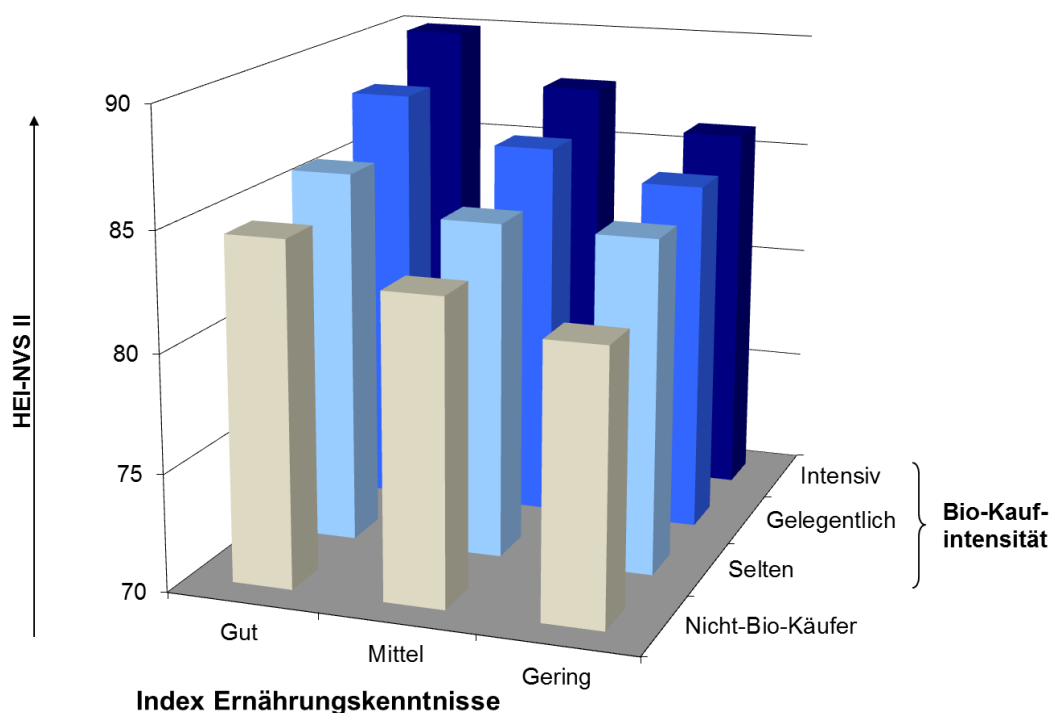


Abb. 3.7: Gegenüberstellung von Bio-Kaufintensität, Ernährungskennnissen und der HEI-NVS II-Indexsumme (Frauen)

In der Gruppe der Männer zeigt sich wiederum ein anderes Bild. Intensiv-Bio-Käufer mit guten Ernährungskennntnissen erreichen 83,9 und Gelegenheits-Bio-Käufer 85,8 Punkte. Männer, die keine Bio-Lebensmittel kaufen und nur geringe Ernährungskennntnisse aufweisen, erreichen lediglich 77,4 von 110 möglichen Punkten (Abb. 3.8).

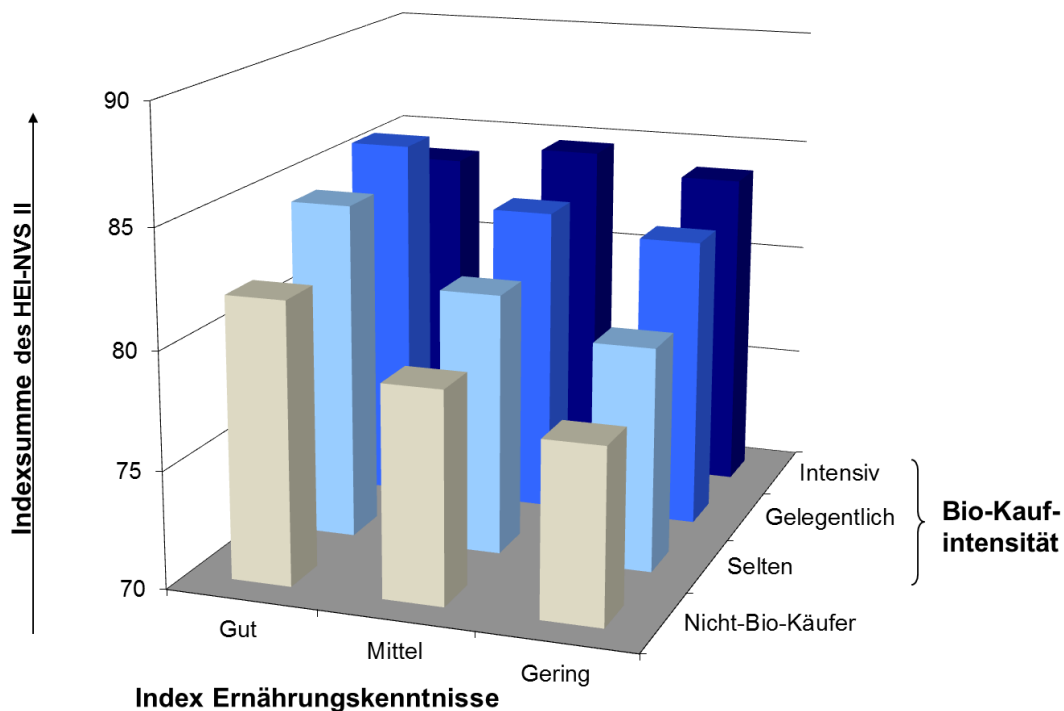


Abb. 3.8: Gegenüberstellung von Bio-Kaufintensität, Ernährungskennntnissen und der HEI-NVS II-Indexsumme (Männer)

### 3.1.5 Methodische Ergebnisse und Diskussion

Für die Zielsetzung der vorliegenden Arbeit, verschiedene Methoden auf ihre Eignung zur umfassenden Darstellung des Lebensmittelverzehr zu prüfen, wurde der HEI-NVS II angewendet. Hierbei wurden alle neun Lebensmittelgruppen, für die es quantifizierbare Empfehlungen der DGE gibt, sowie Alkohol als Inhaltsstoff alkoholischer Getränke als zehnte Variable berücksichtigt (DGE und AID 2013 S 7). Zur Berechnung der Indexwerte wurde ein ungleichgewichtetes Verfahren gewählt. Für Obst und Gemüse konnten maximal 15 Punkte, für alle anderen Lebensmittelgruppen maximal 10 Punkte erreicht werden. Die Berechnung der Indexwerte erfolgte über eine Verhältnisrechnung der verzehrten Lebensmittelmengen im Vergleich zu den empfohlenen Verzehrsmengen. Die Indexsumme wurde additiv auf Basis der Indexwerte gebildet.

#### **Kriterium: Berücksichtigung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen bei Anwendung der Methode**

Ein Index ist von vorhandenen wissenschaftlichen Daten abhängig (Schulze und Hoffmann 2006 S 861, Hu 2002 S 4). Damit können nur Lebensmittelgruppen einbezogen werden, für die quantifizierbare Empfehlungen vorliegen. Auch in dem für die vorliegende Arbeit entwickelten HEI-NVS II konnten Lebensmittelgruppen wie Süßwaren, die aufgrund ihres Energiegehaltes nur gelegentlich verzehrt werden sollen (DGE et al. 2013 S 6, 20) und für die es keine für den Index quantifizierbaren Mengenangaben als Empfehlung gibt, nicht berücksichtigt werden. Dennoch lässt sich festhalten, dass es für 10 der 13 für die Arbeit relevanten Lebensmittelgruppen quantifizierbare Empfehlungen gibt.

Über den HEI-NVS II wird das Kriterium, möglichst alle Lebensmittelgruppen mit den verzehrten Mengen in der Methode zu berücksichtigen, weitgehend erreicht. Es werden fast alle, aber nicht alle Lebensmittelgruppen bei Anwendung des HEI-NVS II einbezogen.

#### **Kriterium: Darstellung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse**

Zur Auswertung des HEI-NVS II stehen die sich ergänzenden Indexwerte und die Indexsumme zur Verfügung. Durch die Indexwerte erfolgt eine direkte und durch die Indexsumme eine indirekte Darstellung der Lebensmittelgruppen. Durch die einzelnen Indexwerte werden die 10 einbezogenen Lebensmittelgruppen mit den entsprechenden Verzehrsmengen im Vergleich zu den Empfehlungen dargestellt. Die Indexsumme entsteht additiv und liefert eine zusammengefasste Bewertung über die 10 einbezogenen Lebensmittelgruppen im Vergleich zu den Empfehlungen. Wie bei dem bekanntesten Index, dem HEI (Kennedy et al. 1995 S 1104, USDA 1995 S 7), gilt für den HEI-NVS II, dass je mehr Auswahl und Menge der

verzehrt Lebensmittel den Empfehlungen der Lebensmittelgruppen entsprechen, umso höher sind die erreichten Indexwerte und umso höher ist die erreichte Indexsumme. Ein Index spiegelt eine zuvor definierte ideale Ernährung wider (Miller et al. 2010 S 1272).

Das Kriterium möglichst alle Lebensmittelgruppen mit den dazugehörigen Verzehrsmengen über die Ergebnisse darzustellen, wird weitgehend erfüllt. Durch die Indexwerte und die Indexsumme werden die 10 Lebensmittelgruppen dargestellt, die bei der Bildung des HEI-NVS II berücksichtigt werden.

**Kriterium: Identifizieren von Personengruppen mit ähnlichem Lebensmittelverzehr**

Personengruppen, die sich stärker oder weniger stark nach den Empfehlungen der DGE ernähren, können durch die Aufteilung auf Basis einer hohen und niedrigen Indexsumme gruppiert werden. Die Indexsumme ist allerdings ein zusammengefasster Wert, der keine Auskunft über die charakteristische Kombination der verzehrten Lebensmittel gibt. Es ist nicht möglich, Aussagen darüber zu treffen, ob Personen mit der gleichen Indexsumme einen ähnlichen Lebensmittelverzehr aufweisen. Auch die Indexwerte einzelner Lebensmittelgruppen tragen nicht zur Erfüllung des Kriteriums bei, da in der vorliegenden Auswertung Personengruppen mit einem ähnlichen Lebensmittelverzehr nicht auf Basis einzelner, sondern auf Basis möglichst vieler Lebensmittelgruppen, identifiziert werden sollen.

Das Kriterium, Personengruppen mit ähnlichem Lebensmittelverzehr zu identifizieren, wird durch die Ergebnisse des HEI-NVS II nicht erfüllt.

**Kriterium: Differenzierung von Bio-Käufer/innen und Nicht-Bio-Käufer/innen**

Ein wesentlicher Vorteil von Indizes ist, dass die Indexwerte und Indexsummen für jede einzelne Person individuell berechnet werden. Durch die Ergebnisse des HEI-NVS II konnten die rein deskriptiven Ergebnisse (Kapitel 2.3.2) weitgehend bestätigt werden: Bio-Käufer/innen ernähren sich eher nach den Empfehlungen der DGE als Nicht-Bio-Käufer/innen. Abweichungen zwischen den Ergebnissen der deskriptiven Auswertung und den **Indexwerten** wurden bei den Lebensmittelgruppen Fisch und Milch festgestellt. Während die deskriptiven Ergebnisse der Mittelwerte (arithmetischer Mittelwert und Median) dieser Lebensmittelgruppen suggerieren, dass sich Frauen und Männer im Durchschnitt nach den Empfehlungen der DGE ernähren, weisen die Mittelwerte der Indexwerte auf einen suboptimalen Verzehr hin. Diese Ergebnisse beruhen darauf, dass durch die Berechnung der Indexwerte, im Gegensatz zur Bildung von arithmetischen Mittelwerten und den Medianwerten, eine Verhältnisrechnung in Anlehnung an einen Index von Rüsten et al. (2009 S 452f) angewendet wurde. Die Indexwerte wurden über das Verhältnis der Verzehrsmenge

einer Person zur empfohlenen Verzehrsmenge berechnet. Wurde mehr oder weniger als empfohlen verzehrt, konnte durch diese Berechnungsform die volle Punktzahl nicht erreicht werden (siehe Kapitel 3.1.3). Damit wird verhindert, dass Personen, die beispielsweise mehr Fisch als empfohlen verzehren, für diese Lebensmittelgruppe die maximale Punktzahl erreichen. Würde die Überschreitung nicht berücksichtigt werden, würde sich der Indexwert fälschlicherweise erhöhen. Das bedeutet, dass der Indexwert bei den vorliegenden Studienteilnehmenden, bezogen auf die Frage, wie gut die Empfehlungen der DGE erreicht werden, genauere Aussagen erlaubt, als dies über den Mittelwert oder Median möglich ist.

Aus den einzelnen Indexwerten entsteht additiv die **Indexsumme**. Diese zeigt, dass sich Bio-Käufer/innen insgesamt durch eine Lebensmittelauswahl näher an den Empfehlungen auszeichnen als Nicht-Bio-Käufer/innen. Es zeigt sich auch, dass Personen mit einer höheren Bio-Kaufintensität höhere Indexsummen erreichen und sich damit eher nach den Empfehlungen der DGE ernähren.

Die Anwendung des Indexes ist geeignet, Personengruppen wie Bio-Käufer/innen und Nicht-Bio-Käufer/innen anhand ihres Lebensmittelverzehr der zehn einbezogenen Lebensmittelgruppen zu beschreiben und voneinander zu differenzieren. Das Kriterium der Differenzierung der Bio-Käufer/innen und Nicht-Bio-Käufer/innen wird erfüllt.

### **Kriterium: Charakterisierung der identifizierten und differenzierten Personengruppen**

In der vorliegenden Arbeit wurde untersucht, inwiefern sich Bio-Käufer/innen, die eine hohe Indexsumme aufweisen, in Bezug auf weitere ernährungs- und gesundheitsrelevante Variablen von anderen Bio-Käufer/innen unterscheiden. Durch diese Charakterisierung konnte beispielsweise gezeigt werden, dass Personen, die sich insgesamt günstiger ernähren, gleichzeitig ein gesundheitsbewussteres Verhalten aufweisen.

Der HEI-NVS II wurde auch genutzt, um die Gleichzeitigkeit mehrerer Variablen (Indexsumme, Bio-Kaufintensität und Gesundheitsverhalten bzw. Indexsumme, Bio-Kaufintensität und Ernährungskennntnisse) zu untersuchen. Auch hierbei konnten positive Zusammenhänge zwischen den Variablen festgestellt werden. Beispielsweise zeigt sich bei den Frauen, mit steigender Bio-Kaufintensität, gleichzeitig besserer Ernährungskennntnisse und ein gesundheitsbewussteres Verhalten eine höhere Indexsumme. Diese Ergebnisse unterstreichen die vielfältigen Einsatzmöglichkeiten der Indexsumme zur Charakterisierung spezifischer Personengruppen.

Auch in anderen Studien wurden zahlreiche Zusammenhänge zwischen einer Indexsumme und weiteren Variablen untersucht. So haben Mirmiran et al. (2006), Drewnowski et al. (1997) und USDA (1995) den Zusammenhang zwischen der Indexsumme mit der Nährstoffzufuhr, Drewnowski et al. (2009), Boynton et al. (2008), Ma et al. (2008), Guo et al. (2004) und McCullough et al. (2002) den Zusammenhang mit Alter und mit Bildung, Mullie et al.

(2010), Manios et al. (2009), Boynton et al. (2008), Guo et al. (2004) und Stookey et al. (2000) mit sozioökonomischen Variablen wie Einkommen oder berufliche Stellung und Vitolins et al. (2007) mit der Anzahl im Haushalt lebender Personen untersucht. Daneben gibt es eine Vielzahl an Studien, die den Zusammenhang mit unterschiedlichen gesundheitsrelevanten Variablen wie sportliche Aktivität, Rauchverhalten und BMI untersucht haben (Beispiele: Previdelli et al. 2010, Tande et al. 2010, Angelopoulos et al. 2009, Drewnowski et al. 2009, Hurley et al. 2009, Manios et al. 2009, Boynton et al. 2008, Ma et al. 2008, Guo et al. 2004, McCullough et al. 2002, Stookey et al. 2000).

Eine Identifizierung von Personengruppen mit ähnlichem Lebensmittelverzehr ist nicht möglich. Daher konnten eine Charakterisierung nur für die differenzierten Personengruppen vorgenommen werden.

### **Fazit**

Es lässt sich festhalten, dass durch den HEI-NVS II die Kriterien der umfassenden Darstellung des Lebensmittelverzehrs nur zum Teil erfüllt werden. Durch die Einschränkung, dass nur Lebensmittelgruppen, für die quantifizierbare Empfehlungen vorliegen, berücksichtigt werden können, werden die Kriterien, möglichst alle Lebensmittelgruppen mit den entsprechenden Verzehrsmengen bei der Anwendung der Methode und Darstellung der Ergebnisse zu berücksichtigen, weitgehend erfüllt. Es konnten nur 10 der 13 relevanten Lebensmittelgruppen einbezogen und dargestellt werden. Auch das Identifizieren von Personengruppen mit ähnlichem Lebensmittelverzehr ist nicht möglich. Weder durch die Indexwerte einzelner Lebensmittelgruppen noch durch die Indexsumme, als zusammengefasste Größe, können Personengruppen mit einem ähnlichen Lebensmittelverzehr, bezogen auf möglichst viele Lebensmittelgruppen, identifiziert werden. Die Differenzierung von Personengruppen wie die der Bio-Käufer/innen und Nicht-Bio-Käufer/innen kann vorgenommen werden. Diese differenzierten Personengruppen konnten anschließend mit weiteren ernährungs- und gesundheitsrelevanten Variablen charakterisiert werden.

Damit wird durch die vorliegenden Analysen gezeigt, dass zwar nicht alle Kriterien zur umfassenden Darstellung erfüllt werden, sich der HEI-NVS II dennoch eignet, um den Lebensmittelverzehr spezieller Personengruppen wie Bio-Käufer/innen und Nicht-Bio-Käufer/innen darzustellen. Durch die Ergebnisse des HEI-NVS II kann der Lebensmittelverzehr damit umfassender dargestellt werden, als dies über rein deskriptive Vergleiche einzelner Lebensmittelgruppen möglich ist.

## 4 Explorative Ansätze zur umfassenden Darstellung des Lebensmittelverzehr

Zur Analyse von Methoden zur umfassenden Darstellung des Lebensmittelverzehr werden verschiedene explorative Ansätze angewendet. Bei einem explorativen Ansatz handelt es sich um einen a-posteriori-Ansatz, der im Gegensatz zum hypothesenorientierten Ansatz rein auf studienspezifischen Daten und nicht auf vorhandenen wissenschaftlichen Daten basiert (Schulze und Hoffmann 2006 S 861, Hu 2002 S 4). Mit explorativen Ansätzen können umfangreiche Datensätze strukturiert und vereinfacht werden. Zu diesen Ansätzen gehören u. a. die Faktoren- und die Clusteranalyse (Schulze und Hoffmann 2006 S 861, Newby und Tucker 2004 S 177).

Bei der **Faktorenanalyse** werden die **Variablen** eines Datensatzes auf Basis ihrer Korrelationen auf eine geringere Anzahl möglichst aussagekräftiger Faktoren oder Hauptkomponenten reduziert (Bortz und Schuster 2010 S 386) bzw. gebündelt (Backhaus et al. 2008 S 17). Bei der **Clusteranalyse** werden Cluster auf Basis von Ähnlichkeiten oder Distanzen zwischen den untersuchten **Objekten oder Fällen** ermittelt. Sie ist eine Methode zur Klassenbildung (Christof 2007 S 114). Objekte oder Fälle werden so zusammengefasst, dass sie innerhalb einer Gruppe möglichst ähnlich und die Gruppen untereinander möglichst unähnlich sind (Backhaus et al. 2008 S 17). Bei beiden Methoden handelt es sich um nicht-parametrische Methoden, was bedeutet, dass sie keine bestimmte Verteilungsform der Daten voraussetzen.

Neben den beiden genannten Methoden, gibt es die Methode des Entscheidungsbaums, die in anderen Forschungsgebieten angewendet wird, um ebenfalls Datensätze zu strukturieren. Über einen **Entscheidungsbaum** können **Objekte oder Fälle** in Abhängigkeit einer Klassifizierungsvariablen voneinander getrennt bzw. differenziert werden. Es entstehen sogenannte Blätter, die Teilgruppen enthalten, die in Bezug auf die Klassifizierungsvariable möglichst homogen sein sollten (Christof 2007 S 127f). Im Gegensatz zu den beiden erst genannten Methoden wird der Entscheidungsbaum zur Darstellung des Lebensmittelverzehr in der Literatur bisher nicht beschrieben. Ebenso wie bei der Faktoren- und Clusteranalyse handelt es sich bei dem Entscheidungsbaum um eine nicht-parametrische Methode. Auch hier wird keine bestimmte Verteilungsform vorausgesetzt (Rokach und Maimon 2005 S 484).

Im Folgenden werden zu jeder Methode neben Hintergrund und Zielen Anwendungsbeispiele genannt. Am Beispiel der vorliegenden Studiendaten werden zunächst die inhaltlichen Ergebnisse der einzelnen Methoden präsentiert, um darauf aufbauend zu analysieren, ob sich die Methoden zur umfassenden Darstellung des Lebensmittelverzehr eignen.

## 4.1 Faktorenanalyse

Der Begriff Faktorenanalyse ist ein Sammelbegriff für verschiedene Berechnungsverfahren, die der Vereinfachung und Strukturierung eines Datensatzes dienen (Bortz und Schuster 2010 S 385, Kessler 2007 S 22, Moosbrugger und Kelava 2007 S 308, Flury und Riedwyl 1983 S 135). Neben verschiedenen Arten der Faktorenanalyse wird hierzu auch die Hauptkomponentenanalyse gezählt (Kessler 2007 S 22).

Ihr gemeinsames Ziel ist, eine Vielzahl an Variablen auf wenige grundlegende, sogenannte latente Faktoren zu reduzieren (Schendera 2010 S 180, Bortz und Döring 2009 S 378, Kessler 2007 S 22). In der Faktorenanalyse werden diese Faktoren, in der Hauptkomponentenanalyse Hauptkomponenten genannt (Kessler 2007 S 22). Die Reduktion der Variablen erfolgt auf Basis ihrer Korrelationen zueinander. Unter der Korrelation wird dabei nicht nur der Zusammenhang zwischen zwei Variablen verstanden, sondern auch eine hinter diesen Variablen stehende, dritte (latente) Variable, die als eigentliche Ursache für die Korrelation gelten könnte (Schendera 2010 S 180).

Auf unterschiedliche Verfahren der Faktorenanalyse, die in der Literatur speziell zur Darstellung der Ernährung (Nährstoffzufuhr und Lebensmittelverzehr) eingesetzt werden, wird im Folgenden kurz eingegangen. Dazu gehört neben der explorativen Faktorenanalyse, die Hauptkomponentenanalyse sowie die konfirmatorische Faktorenanalyse.

### Explorative Faktorenanalyse

Die explorative Faktorenanalyse (EFA) ist ein datenstrukturierendes, reduzierendes und hypothesengenerierendes Verfahren (Bortz und Schuster 2010 S 387). Sie kommt zur Anwendung, wenn keine konkreten Hypothesen oder Vorannahmen (Bortz und Döring 2009 S 517) über die Anzahl der zugrundeliegenden Faktoren und über die Zuordnung der beobachteten Variablen zu den Faktoren vorliegen (Moosbrugger und Kelava 2007 S 308f).

Die Datenstrukturierung und -reduktion wird vorgenommen, indem Beziehungszusammenhänge in den Daten gesucht werden. Gruppen von Variablen, die hoch miteinander korrelieren, werden identifiziert (Bortz und Döring 2009 S 147, Backhaus et al. 2008 S 324) und von weiteren, weniger hoch korrelierten Variablen getrennt (Backhaus et al. 2008 S 324). Über sogenannte Faktoren werden diese hochkorrelierenden Variablen gebündelt, um das Gemeinsame zu erfassen (Bortz und Döring 2009 S 147).

Die allgemeine Formel der gebildeten Faktoren lautet (Park et al. 2002 S 564):

$$X_1 = \vartheta_{1(1)}CF_{(1)} + \vartheta_{1(2)}CF_{(2)} + \dots + \vartheta_{1(m)}CF_{(m)} + e_1$$

$$X_2 = \vartheta_{2(1)}CF_{(1)} + \vartheta_{2(2)}CF_{(2)} + \dots + \vartheta_{2(m)}CF_{(m)} + e_2$$

$$X_p = \vartheta_{p(1)}CF_{(1)} + \vartheta_{p(2)}CF_{(2)} + \dots + \vartheta_{p(m)}CF_{(m)} + e_p$$

$CF$  = *gemeinsame Varianz (common Faktor)*

$\vartheta_{j(i)}$  = *Gewicht der gemeinsamen Varianz*

$i = 1, 2, \dots, m$

$j = 1, 2, \dots, p$

$e_2$  = *spezifische Varianz*

Die gebildeten Faktoren sind künstliche Konstrukte, die nicht direkt gemessen werden können. Der Schwerpunkt der EFA liegt in der Varianzaufklärung, die durch das Trennen der gemeinsamen von der spezifischen Varianz berechnet wird. Unter der *gemeinsamen Varianz* wird die Varianz, die eine Variable mit einer anderen teilt, und unter der *spezifischen Varianz* wird der Varianzanteil, der spezifisch für jede einzelne Variable, die unabhängig von der Faktorstruktur ist, verstanden (Park et al. 2002 S 563f). Im Gegensatz zur Hauptkomponentenanalyse wird bei der EFA zwischen unterschiedlichen Varianzarten unterschieden (Schendera 2004 S 600).

Die Datenreduktion wird vorgenommen, indem Ausprägungen für die Faktoren ermittelt werden (Backhaus et al. 2008 S 328). Für jede Person wird ein Faktorwert für jeden Faktor berechnet. Dieser Wert gibt Auskunft darüber, wie gut die in einem Faktor zusammengefassten Variablen bei einer Person wiederspiegelt werden (Bortz und Schuster 2010 S 392f). Dieser Faktorwert wird anschließend anstelle der Originalwerte für weitere Berechnungen verwendet (Backhaus et al. 2008 S 324).

Die Informationen der Ausgangsvariablen werden durch wenige voneinander unabhängige Faktoren wiedergegeben (Bortz und Schuster 2010 S 387). Anschließend können Hypothesen aus den faktoranalytischen Ergebnissen gebildet werden (Bortz und Schuster 2010 S 387).

### **Hauptkomponentenanalyse**

Die Hauptkomponentenanalyse (engl. principal component analysis, PCA) ist wie die explorative Faktorenanalyse ein datenreduzierendes (Bortz und Schuster 2010 S 396, Kessler 2006 S 22, 24) und hypothesengenerierendes Verfahren (Bortz und Schuster 2010 S 396). Dieses Verfahren wird eingesetzt, wenn komplexe Daten auf einfache Strukturen reduziert werden sollen.

Die Datenstrukturierung und -reduktion wird vorgenommen, indem aus den Ausgangsvariablen auf Basis der Korrelationen eine geringere Anzahl an Hauptkomponenten

extrahiert werden (Kessler 2006 S 22, Schendera 2004 S 600), die so viele Informationen wie möglich enthalten (Park et al. 2002 S 563).

Eine Hauptkomponente ist die lineare Summe der unterschiedlich gewichteten Originaldaten und stellt mathematisch betrachtet eine Linearkombination dar (Kessler 2006 S 22). Dabei gibt die Gewichtung der Originaldaten, die sogenannte Faktorladung, Auskunft über die Korrelation einer Variablen mit einer Hauptkomponente (Bortz und Schuster 2010 S 393). Die allgemeine Funktion der extrahierten Hauptkomponenten lautet (Schulze et al. 2003a, James und Mcculloch 1990 S 134, Flury und Riedwyl 1983):

$$PC_1 = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_m X_m$$

$$PC_2 = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_m X_m$$

$$PC_3 = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_m X_m$$

$$PC_p = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_m X_m$$

$PC$  = Hauptkomponenten (*principle component*)

$p$  = Anzahl der Hauptkomponenten

$\beta_i$  = gewichtete Faktorladung der Variable,  $i = 1, 2, \dots, m$

$X_i$  = standardisierte Variable,  $i = 1, 2, \dots, m$

Es werden mehrere voneinander unabhängige Hauptkomponenten extrahiert. Durch die erste Hauptkomponente wird der größte und durch die letzte Hauptkomponente der geringste Varianzanteil der einbezogenen Variablen erklärt. Nach und nach wird durch die einzelnen Hauptkomponenten die maximale Varianz der Daten erreicht. Je höher die Variablen miteinander korrelieren, desto weniger Hauptkomponenten werden zur Erklärung der Gesamtvarianz benötigt (Bortz und Schuster 2010 S 391, Schendera 2004 S 600).

Zwischen unterschiedlichen Varianzarten wird im Gegensatz zur EFA nicht unterschieden (Schendera 2004 S 600). Es wird davon ausgegangen, dass die Varianz vollständig erklärt wird und keine Restvarianz existiert (Schendera 2010 S 183, 2004 S 600).

Anschließend können wie bei der explorativen Faktorenanalyse Hypothesen über die Ergebnisse generiert werden (Bortz und Schuster 2010 S 396). Hierfür stehen die *Faktorladungen* sowie der *Faktorwert* zur Verfügung. Die Faktorladung gibt Auskunft über die Korrelation einer Variablen mit einer Hauptkomponente (Bortz und Schuster 2010 S 393), also wie die einzelnen Variablen zusammenhängen. Daneben wird, wie bei der explorativen Faktorenanalyse, für jede Person ein Faktorwert für jede Hauptkomponente berechnet. Der Faktorwert gibt Auskunft, wie gut die in einer extrahierten Hauptkomponente zusammengefassten Variablen bei einer Person widerspiegelt werden (Bortz und Schuster 2010 S 392f).

## **Konfirmatorische Faktorenanalyse**

Die konfirmatorische Faktorenanalyse, auch konfirmative Faktorenanalyse genannt (engl. Übersetzung: bekräftigend, bestätigend), ist der explorativen Faktorenanalyse ähnlich. Allerdings handelt es sich um ein hypothesenprüfendes Verfahren (Moosbrugger und Kelava 2007 S 308f) und nicht um eine datenstrukturierendes und -reduzierendes Verfahren (Bortz und Schuster 2010 S 387). Zudem müssen im Gegensatz zur explorativen Faktorenanalyse Vorannahmen bzw. Hypothesen z. B. über die Anzahl der Faktoren vorgegeben werden (Bortz und Döring 2009 S 517). Dieses Verfahren wird genutzt, um Hypothesen über die Faktorenstruktur zu überprüfen (Bortz und Schuster 2010 S 430). Da dies in der vorliegenden Untersuchung keine Anwendung findet, wird auf dieses Verfahren nicht weiter eingegangen.

### **4.1.1 Wahl eines Verfahrens für die eigene Anwendung**

In der vorliegenden Auswertung liegen keine Vorannahmen oder Hypothesen über die Anzahl der zugrundeliegenden Faktoren bzw. Hauptkomponenten vor. Es ist nicht bekannt, ob und wie viele Faktoren oder Hauptkomponenten zur Darstellung des Lebensmittelverzehr der Frauen und Männer extrahiert werden können. Damit können die beiden hypothesengenerierenden Verfahren, die explorative Faktorenanalyse als auch die Hauptkomponentenanalyse angewendet werden.

Die explorative Faktorenanalyse und Hauptkomponentenanalyse stehen „im engeren Sinne in einem Konkurrenzverhältnis“ (Bortz und Schuster 2010 S 427). Neben der Gemeinsamkeit, Daten zu reduzieren, gibt es wesentliche Unterschiede (Park et al. 2002 S 563ff). Insbesondere unterscheiden sich die beiden Verfahren in der Berechnung der Faktoren bzw. der extrahierten Hauptkomponenten. Bei der explorativen Faktorenanalyse wird ein Faktor gebildet, der ein künstliches Konstrukt darstellt. Durch diesen wird die Varianz der Variablen durch Trennung der gemeinsamen und spezifischen Varianz fokussiert (Park et al. 2002 S 563f). Bei der Hauptkomponentenanalyse wird eine Hauptkomponente gebildet, die eine Linearkombination der Ursprungsdaten darstellt (Schendera 2004 S 600). Zwischen Varianzarten wird bei der Hauptkomponentenanalyse nicht unterschieden, sondern das sukzessive Aufklären der Gesamtvarianz der Variablen angestrebt (Park et al. 2002 S 563f). Mit dem Ziel, den Lebensmittelverzehr umfassend darzustellen, wird die Hauptkomponentenanalyse verwendet, da durch diese das Ziel verfolgt wird, soviel Gesamtvarianz (hier: Varianz des Lebensmittelverzehr) wie möglich zu erklären. Die in der Hauptkomponentenanalyse gebildete Linearkombination beschreibt eine nachvollziehbare, spezifische Kombination verzehrter Lebensmittelgruppen, was ebenso für die Anwendung der Hauptkomponentenanalyse spricht.

### 4.1.2 Theoretischer Hintergrund der PCA

Die Verzehrsmengen oder Nährstoffzufuhr der ausgewählten Variablen werden aufgrund ihrer Korrelationen zusammengefasst (Newby und Tucker 2004 S 177) und durch wenige möglichst aussagekräftige Hauptkomponenten dargestellt (Schulze und Hoffmann 2006 S 861). Ziel ist, mit wenigen Hauptkomponenten möglichst viel Varianz der einbezogenen Variablen zu erklären (Hoffmann et al. 2004 S 936).

Voraussetzung für die Durchführung einer Hauptkomponentenanalyse ist, dass die Ausgangsvariablen miteinander korrelieren. Eine Korrelationsmatrix (eingesetzt bei z-standardisierten Daten) oder Kovarianzmatrix (eingesetzt bei nicht standardisierten Daten) gibt erste Hinweise darauf, ob und wo Zusammenhänge zwischen den einzelnen Variablen bestehen (Schendera 2010 S 183, 2004 S 596). Hohe Korrelationen zwischen Variablen lassen vermuten, dass diese etwas Gemeinsames messen und so eine Hauptkomponente extrahierbar ist. Sind die Korrelationen niedrig, könnten mehrere Hauptkomponenten extrahiert werden. Werden allerdings ausschließlich niedrige Korrelationen festgestellt, kann vermutlich keine Hauptkomponente extrahiert werden (Schendera 2010 S 183).

#### Faktorladung

In jeder Hauptkomponente wird für jede einbezogene Variable eine Faktorladung berechnet. Die Faktorladung in einer Hauptkomponente entspricht der Korrelation einer Variablen mit der entsprechenden Hauptkomponente (Bortz und Schuster 2010 S 393). Eine positive Ladung einer Variablen lässt auf eine positive, eine negative Ladung auf eine negative bzw. inverse Korrelation mit der Hauptkomponente schließen (Bamia et al. 2005 S 103). Bezogen auf den Lebensmittelverzehr bedeutet dies, dass unter einer positiven Ladung ein hoher und unter einer negativen Ladung ein niedriger Lebensmittelverzehr im Vergleich zur Gesamtgruppe verstanden wird. Eine Ladung um Null bedeutet, dass die entsprechende Variable nicht charakteristisch für die Hauptkomponente ist. Der Lebensmittelverzehr einer solchen Variablen entspricht dem Durchschnitt der Gesamtgruppe. Der Grenzwert, ab dem eine Faktorladung als hoch und damit als relevant bezeichnet wird, variiert in veröffentlichten Studien von  $\pm 0,2$  (Beispiel: Schulze et al. 2003a S 410) bis  $\pm 0,4$  (Beispiel: Jordan et al. 2013 S 907) und ist willkürlich festgelegt.

Jede Hauptkomponente lässt sich anhand der Faktorladungen der Variablen durch eine charakteristische Kombination an Lebensmittelgruppen beschreiben (Tab. 4.1).

**Tab. 4.1: Beispiel und Interpretation für eine Hauptkomponente (eigene Darstellung)**

Lebensmittelgruppen	Hauptkomponente	
	Ladung der Variablen	Interpretation der Hauptkomponente
Obst	-0,05	
Gemüse	0,14	
<b>Getreide</b>	<b>0,31</b>	<b>hoher Getreideverzehr</b>
<b>Fleisch</b>	<b>0,57</b>	<b>hoher Fleischverzehr</b>

Die für die Hauptkomponente charakteristischen Lebensmittelgruppen sind fett gedruckt.

Für das vorliegende Beispiel wurde eine hohe Faktorladung bei einem Wert von  $\pm 0,2$  angenommen. Demnach zeichnet sich die in Tab. 4.1 dargestellte Hauptkomponente durch einen hohen Verzehr an Getreide und Fleisch aus. Die Lebensmittelgruppen Obst und Gemüse weisen eine niedrige Faktorladung auf, was bedeutet, dass der Lebensmittelverzehr dieser Gruppen in etwa dem der Gesamtgruppe entspricht.

Bei der Anwendung der Hauptkomponentenanalyse im Ernährungsbereich werden die extrahierten Hauptkomponenten als Ernährungsmuster bezeichnet (Schulze und Hoffmann 2006 S 861).

### Rotationsverfahren

Für die Anwendung der Hauptkomponentenanalyse stehen verschiedene Rotationsverfahren der Variablen (hier: Lebensmittelgruppen) in einer extrahierten Hauptkomponente zur Verfügung. Durch die Rotation wird das Ziel verfolgt, schwer zu interpretierende Ladungen durch leichter zu interpretierende zu ersetzen und damit subjektive Einflüsse bei der Interpretation einer Hauptkomponente auszuschließen (Schendera 2010 S 184, 2004 S 601f). Durch eine Rotation werden die Ladungen einer Variablen in einer Hauptkomponente auf/um Null oder Eins gesetzt. Damit wird erreicht, dass eine Variable nur in wenigen Hauptkomponenten eine hohe Ladung aufweist und charakteristisch für nur wenige Hauptkomponenten ist. Die Wahl des Rotationsverfahrens hat damit zwar einen Einfluss auf die Ladung der Variablen einer Hauptkomponente, nicht aber auf die statistische Erklärungskraft (Schendera 2010 S 184, 2004 S 601f).

Bei den Rotationsverfahren kann zwischen der orthogonalen und obliquen Rotation unterschieden werden (Moosbrugger und Kelava 2007 S 313f). Bei der *orthogonalen (rechtwinkligen) Rotation* bleibt die Rechtwinkligkeit der Achsen erhalten (Bortz und Schuster 2010). Das bedeutet, dass die extrahierten Hauptkomponenten nach wie vor *nicht* miteinander korrelieren. Diese Rotation wird verwendet, wenn noch keine Annahmen über die Hauptkomponenten vorhanden sind. Zum einen kann das Quartimax-Verfahren, das eine Vereinfachung der Variablen anstrebt, verwendet werden. Hierbei wird die Anzahl der Hauptkomponenten zur Interpretation einer Variablen minimiert. Die Erklärung der Gesamtvarianz soll möglichst mit einer Hauptkomponente erklärt werden (Schendera 2004 S 601). Zum

anderen kann das Varimax-Verfahren verwendet werden. Es zielt darauf ab, eine Varianzmaximierung zu erreichen, indem die Anzahl der Variablen mit hoher Ladung in einer Hauptkomponente minimiert werden (Schendera 2004 S 601).

Die *oblique (schiefe) Rotation* führt dazu, dass die bisher unkorrelierten extrahierten Hauptkomponenten nach der Rotation miteinander korrelieren, was die Unterscheidung der Hauptkomponenten erschwert. Sie wird durchgeführt, wenn beispielsweise aus der Theorie bekannt ist, dass die Hauptkomponenten miteinander korrelieren (Schendera 2004 S 602).

### **Simple Structure**

Daneben können die extrahierten Hauptkomponenten durch die Anwendung des Verfahrens Simple Structure (Simplifizierung) optimiert werden. Variablen, die keinen oder nur einen sehr geringen Beitrag in allen ausgewählten Hauptkomponente leisten, werden als irrelevant angesehen (Beispiele: Schulze et al. 2003a, Schulze 2002). Die Hauptkomponentenanalyse wird unter Ausschluss dieser irrelevanten Variablen zur Spezifizierung der Hauptkomponenten erneut berechnet (Schulze 2002 S 53). Ab wann von hoch korrelierten Variablen gesprochen werden kann, ist allerdings nicht fest definiert und variiert von  $\pm 0,2$  (Beispiel: Schulze et al. 2003a S 410) bis  $\pm 0,4$  (Beispiel: Jordan et al. 2013 S 907).

### **Faktorwert**

In der Hauptkomponentenanalyse wird für jedes Objekt (hier: Personen) ein Faktorwert für jede Hauptkomponente berechnet, der darüber Auskunft gibt, wie gut die in einer extrahierten Hauptkomponente zusammengefassten Variablen (hier: Lebensmittelgruppen) den tatsächlichen Lebensmittelverzehr einer Person widerspiegeln (Bortz und Schuster 2010 S 392f). Weist eine Person einen hohen Faktorwert für eine Hauptkomponente auf, bedeutet dies, dass der Lebensmittelverzehr dieser Person durch die in der Hauptkomponente zusammengefassten Merkmale dargestellt werden kann. Der Lebensmittelverzehr lässt sich in diesem Fall über eine Hauptkomponente beschreiben.

Der Faktorwert wird berechnet, indem die Faktorladungen (z-standardisierte Verzehrsmenge einer Variable multipliziert mit der Ladung der Variable in einer Hauptkomponente) der einzelnen Variablen summiert werden (Bamia et al. 2005 S 103, Schulze et al. 2001 S 364).

In Tab. 4.2 ist eine Hauptkomponente und deren Interpretation für eine Beispielperson dargestellt. Bei Person A, die mit 9,69 den höchsten Faktorwert der Gesamtgruppe besitzt, spiegelt sich das Bild der Hauptkomponente im tatsächlichen Lebensmittelverzehr für die meisten Lebensmittelgruppen wider. Person A verzehrt die für die Hauptkomponente charakteristischen Lebensmittel Getreide, Kartoffeln, Eier, Fleisch, Streichfett und Süßwaren in höheren Mengen als die Gesamtgruppe. Bei der Lebensmittelgruppe Eier konnte keine Übereinstimmung zwischen der Hauptkomponente und dem tatsächlichen Verzehr von

Person A festgestellt werden. Während die Hauptkomponente einen hohen Eierverzehr beschreibt, weist Person A einen Verzehr von 0 g Eier pro Tag auf.

**Tab. 4.2: Erklärung der berechneten Faktorladung und des Faktorwertes einer Hauptkomponente am Beispiel von Person A (eigene Darstellung)**

Lebensmittelgruppen	Hauptkomponente		Gesamtgruppe Frauen (MW±Std dev in g/d)	Person A		
	Ladung der Variablen	Interpretation der Hauptkomponente		Verzehrs- menge (g/d)	z-standard- isierter Verzehrs- menge	Faktorladungen der Haupt- komponente
Obst	-0,03		292±233	86	-0,88	0,02
Gemüse	0,13		251±150	282	0,21	0,03
<b>Getreide</b>	<b>0,41</b>	<b>hoher Getreideverzehr</b>	<b>239±99</b>	<b>582</b>	<b>3,45</b>	<b>1,41</b>
<b>Kartoffeln</b>	<b>0,37</b>	<b>hoher Kartoffelverzehr</b>	<b>72±43</b>	<b>213</b>	<b>3,28</b>	<b>1,22</b>
Nüsse und Samen	0,06		4±11	0	-0,34	-0,02
Milch	0,11		239±211	1062	3,91	0,42
<b>Eier</b>	<b>0,34</b>	<b>hoher Eierverzehr</b>	<b>12±13</b>	<b>0</b>	<b>-0,98</b>	<b>-0,34</b>
Fisch	0,15		23±22	38	0,68	0,10
<b>Fleisch</b>	<b>0,46</b>	<b>hoher Fleischverzehr</b>	<b>75±48</b>	<b>488</b>	<b>8,56</b>	<b>3,92</b>
<b>Streichfett</b>	<b>0,44</b>	<b>hoher Streichfettverzehr</b>	<b>21±17</b>	<b>55</b>	<b>1,99</b>	<b>0,88</b>
Getränke	0,01		2335±889	1900	-0,49	-0,01
Alkohol <sup>1</sup>	0,04		6±9	0	-0,66	-0,03
<b>Süßwaren</b>	<b>0,34</b>	<b>hoher Süßwarenverzehr</b>	<b>61±56</b>	<b>400</b>	<b>6,09</b>	<b>2,07</b>
					<b>Faktorwert:</b>	<b>9,69</b>

<sup>1</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

Die für die Hauptkomponente charakteristischen Lebensmittelgruppen sind fett gedruckt.

### Bestimmung der Anzahl der Hauptkomponenten

Durch die Hauptkomponentenanalyse werden mehrere Hauptkomponenten extrahiert. Durch die erste Hauptkomponente wird der größte und durch die letzte Hauptkomponente der geringste Anteil der Varianz des Lebensmittelverzehrs erklärt (Schulze et al. 2003a S 410). Meist werden nur wenige Hauptkomponenten für die Auswertung herangezogen, die einen großen Anteil der Varianz erklären (Schulze 2002 S 51). Durch die Auswahl einer bestimmten Anzahl an Hauptkomponenten, wird automatisch nur ein Teil der Gesamtvarianz erklärt. Der Varianzanteil der nicht berücksichtigten Hauptkomponenten geht bei der weiteren Auswertung verloren (Moosbrugger und Kelava 2007 S 309f). Zur Bestimmung der Anzahl der Hauptkomponenten gibt es verschiedene Auswahlkriterien (Backhaus et al. 2008 S 369):

- **Extraktion bis X % der Varianz erklärt ist** (z. B. 95 %) (Backhaus et al. 2008 S 369, Schendera 2004 S 602ff).
- **Extraktion aller Hauptkomponenten mit einem Eigenwert  $\geq 1$**  (Kaiser-(Guttman)-Kriterium) (Backhaus et al. 2008 S 369, Schendera 2004 S 602ff).

Dabei ist unter einem Eigenwert die Summe der quadrierten Faktorladung aller Variablen einer Hauptkomponente zu verstehen. Der Eigenwert gibt den Varianzanteil der Gesamtvarianz an, die eine Hauptkomponente erklärt (Bortz und Schuster 2010 S 393). Je größer der Eigenwert, desto mehr Varianzanteil wird aufgeklärt (Schendera 2010 S 184). Nur wenn die Hauptkomponente einen Eigenwert über eins

besitzt, wird das Ziel der Datenreduktion erfüllt. In diesem Fall wird mehr Varianz erklärt als durch eine einzelne Variable (Moosbrugger und Kelava 2007 S 311f).

- **Extraktion einer bestimmten Anzahl** (z. B.  $n = 3$  Hauptkomponenten) (Backhaus et al. 2008 S 369).
- **Extraktion aller Hauptkomponenten, die über einen Scree-Test als relevant identifiziert werden.**

Über den Scree-Test wird der Verlauf der Eigenwerte der einzelnen Hauptkomponenten anhand einer Grafik untersucht (Moosbrugger und Kelava 2007 S 311f). Die Hauptkomponenten werden nach Eigenwerten sortiert abgebildet. In der Grafik ist ein Knick im Verlauf der abgebildeten Eigenwerte zu erkennen. Die Hauptkomponenten nach einem Knick werden als nicht mehr relevant angesehen (Backhaus et al. 2008 S 369, Schendera 2004 S 602ff).

- **Anzahl der Hauptkomponenten ist kleiner als die Hälfte der Anzahl an Variablen** (Backhaus et al. 2008 S 369).
- **Extraktion aller Hauptkomponenten, die nach der Rotation interpretierbar sind** (Backhaus et al. 2008 S 369).
- **Extraktion aller Hauptkomponenten, die interpretierbar sind (Plausibilitätstest)** (Schulze 2002 S 52).

#### **4.1.3 Anwendung der Faktorenanalyse im Bereich Ernährung**

Die extrahierten Faktoren oder Hauptkomponenten werden in der Literatur auch als Ernährungsmuster bezeichnet (Schulze und Hoffmann 2006 S 861). Auf Basis der in der Faktorenanalyse berechneten Faktorwerte können unterschiedliche Auswertungen durchgeführt werden. Es kann untersucht werden, welche Personen sich nach welcher extrahierten Hauptkomponente überwiegend ernähren. Hierfür werden die Studienteilnehmenden anhand ihres erreichten Faktorwertes in gleichgroße Gruppen, z. B. Tertzile (Beispiele: Guo et al. 2012, Kim et al. 2012), Quartile (Beispiele: Hong et al. 2012, Okubo et al. 2012, Richter et al. 2012) oder Quintile (Beispiele: Azadbakht und Esmailzadeh 2012, Malik et al. 2012, Maruyama et al. 2012, Boggs et al. 2011) eingeteilt und miteinander verglichen. Die Personengruppe mit den höchsten Faktorwerten, deren Nährstoffaufnahmen oder Lebensmittelverzehr am ehesten einer extrahierten Hauptkomponente entspricht, wird in der Regel mit der Personengruppe, mit den niedrigsten Faktorwerten verglichen.

Daneben ist es möglich, den Zusammenhang zwischen dem Faktorwert und weiteren Variablen wie dem Erkrankungsrisiko (Beispiele: Bradshaw et al. 2012, Hong et al. 2012, Malik et al. 2012, Lee et al. 2011, Noethlings et al. 2008) zu untersuchen.

Bereits Anfang der 1980er Jahre wurden mit Hilfe einer Hauptkomponentenanalyse Untersuchungen zur Analyse des Lebensmittelverzehr veröffentlicht (Schwerin et al. 1982, 1981) und deren Bedeutung zur Erklärung des komplexen Zusammenhangs zwischen Ernährung und Gesundheit aufgezeigt. Ziel der Analyse von Schwerin et al. (1982, 1981) war, den Nutzen des Konzeptes von Ernährungsmustern durch den Zusammenhang zwischen dem Lebensmittelverzehr und Gesundheitsstatus der Studienteilnehmenden aufzuzeigen. Hierfür wurden Daten aus der Ten-State- und dem HANES I-Survey verwendet. Die verzehrten Lebensmittel wurden in 15 Lebensmittelgruppen eingeteilt. Basierend darauf wurde eine Hauptkomponentenanalyse durchgeführt. Es wurden sieben sogenannte Ernährungsmuster extrahiert (Schwerin et al. 1982, 1981).

Ebenso wie bei den Indizes (siehe Kapitel 3.1.1) kann die Faktorenanalyse auf Basis der einbezogenen Variablen in Kategorien eingeteilt werden. In die Faktorenanalyse können (1) **Nährstoffe**, (2) **Lebensmittel und Lebensmittelgruppen** sowie (3) **Nährstoffe, Lebensmittel und Lebensmittelgruppen kombiniert mit weiteren nährstoff- und lebensmittel-unabhängigen Variablen**, beispielsweise Variablen zum Lebensstil einbezogen werden. Nach Durchführung der Faktorenanalyse können Ursache-Wirkungs-Zusammenhänge zwischen den extrahierten Faktoren oder Hauptkomponenten und Biomarkern oder Krankheiten oder eine Charakterisierung von Personengruppen, die den einzelnen Faktoren oder Hauptkomponenten zugeordnet wurden, vorgenommen werden. Im Folgenden werden diese drei Kategorien der Faktorenanalyse beschrieben und Beispiele für die Untersuchung von Zusammenhängen und Charakterisierung von Personengruppen mit weiteren Variablen genannt.

- (1) **Faktorenanalysen auf Basis von Nährstoffen** werden insbesondere verwendet, um Ursache-Wirkungs-Zusammenhänge zwischen der Nährstoffzufuhr und dem Risiko, an verschiedenen Krankheiten wie Krebs zu erkranken, zu untersuchen und Risikogruppen anhand weiterer Variablen zu charakterisieren.

Beispielsweise haben de Stefani et al. (2008a) und Edefonti et al. (2010) neben dem Zusammenhang zwischen der Nährstoffzufuhr und dem Risiko, an Krebs zu erkranken, den Zusammenhang zwischen den extrahierten Faktoren oder Hauptkomponenten und dem Verzehr verschiedener Lebensmittel analysiert, um weitere Erkenntnisse über mögliche Zusammenhänge zwischen der Ernährung und dem Krebsrisiko zu erhalten. Zusätzlich haben de Stefani et al. (2008a) den Zusammenhang zwischen extrahierten Hauptkomponenten und Alter, Geschlecht sowie verschiedener gesundheitsrelevanter Variablen untersucht, um die Risikogruppen weiter zu charakterisieren.

- (2) **Faktorenanalysen auf Basis von Lebensmitteln oder Lebensmittelgruppen** sind die von den drei beschriebenen Kategorien am häufigsten durchgeführten Faktorenanalysen. Sie werden eingesetzt, um z. B. Zusammenhänge zwischen der Ernährung und Biomarkern wie Adiponectin (Beispiel: Guo et al. 2012) und Homocystein (Beispiel: Yakub et al. 2010) und Krankheiten (Beispiele: Jordan et al. 2013, Guallar-Castillon et al. 2012, Daniel et al. 2011, Butler et al. 2010, Erber et al. 2010, Hamer und Mishra 2010, Shaheen et al. 2010, Fung et al. 2001) zu untersuchen. Die Faktorenanalyse findet aber auch Anwendung, um unabhängig von der Untersuchung von Ursache-Wirkungs-Zusammenhängen die Ernährung verschiedener Personengruppen darzustellen. So wird die Faktorenanalyse auch verwendet, um die Ernährung von Schwangeren (Beispiele: Torjusen et al. 2012, Arkkola et al. 2008), den Zusammenhang zwischen den Hauptkomponenten von Männern und ihren Frauen, soziodemographischen Variablen und der Nährstoffzufuhr (Beispiele: Northstone und Emmett 2010, Schulze et al. 2001), um den Zusammenhang zur sportlichen Aktivität (Beispiel: Charreire et al. 2011) sowie zu sozioökonomischen Variablen und Lebensstilvariablen und der Nährstoffzufuhr (Beispiele: del Mar Bibiloni et al. 2012, Richter et al. 2012, Cutler et al. 2011, Hare-Bruun et al. 2011, Olinto et al. 2011, Rezazadeh et al. 2010) zu untersuchen.
- (3) Daneben gibt es **Faktorenanalysen**, bei denen neben Variablen der Ernährung **weitere nährstoff- und lebensmittelunabhängige Variablen** einbezogen wurden. Beispielsweise haben Slattery et al. (1999) neben Nährstoffen wie Calcium und Folsäure sowie Lebensmitteln wie Obst und Gemüse weitere lebensstilrelevante Variablen wie Bildung, BMI, körperliche Aktivität, Rauchverhalten, Mahlzeitenhäufigkeit und Medikamenteneinnahme direkt in die Faktorenanalyse einbezogen. Ziel dabei war, nicht wie bei den anderen Analysen den Zusammenhang zwischen der Ernährung und einer Erkrankung zu untersuchen, sondern den Zusammenhang zwischen dem *Lebensstil* und dem Risiko, an Darmkrebs zu erkranken, zu untersuchen.

Einen Überblick über die Anwendung von Faktorenanalysen geben verschiedene Übersichtsarbeiten (Beispiele: Brennan et al. 2010, Edefonti et al. 2009, Newby und Tucker 2004). Brennan et al. (2010) und Edefonti et al. (2009) haben Veröffentlichungen, in denen der Zusammenhang zwischen Ernährungsmustern und Brustkrebs beschrieben wird, zusammengetragen. Brennan et al. (2010) konnten 16 Veröffentlichungen aus den Jahren 2009 bis 2010 und Edefonti et al. (2009) 10 Veröffentlichungen aus den Jahren 1995 bis 2008 identifizieren. Beide Autorengruppen berichten von verschiedenen Zusammenhängen der identifizierten Faktoren oder Hauptkomponenten und Brustkrebs. Newby und Tucker (2004) konnten von 1980 bis 2004 58 Veröffentlichungen, in denen die Anwendung einer

Faktorenanalyse zur Untersuchung des Zusammenhangs der Ernährung und weiteren Variablen beschrieben wird, identifizieren.

Wie von Newby und Tucker (2004) gezeigt, wird bei einer faktoranalytischen Bestimmung der Ernährungsmuster überwiegend die Hauptkomponentenanalyse verwendet. Hierfür wird meistens die Varimax Rotation durchgeführt, um die Interpretierbarkeit der extrahierten Ernährungsmuster zu erhöhen. In der Regel werden Hauptkomponenten mit einem Eigenwert  $>1$  einbezogen (Newby und Tucker 2004 S 192).

Bei allen in der genannten Übersichtsarbeit beschriebenen Faktorenanalysen wurde der Zusammenhang zwischen der Nährstoffaufnahme oder dem Lebensmittelverzehr mit weiteren Variablen wie Bioparametern oder Krankheiten untersucht. Der Einsatz der Faktorenanalyse mit dem direkten Ziel, den Lebensmittelverzehr umfassend darzustellen, ohne den Zusammenhang zwischen der Ernährung mit Bioparametern oder Krankheiten zu untersuchen, findet seltener Anwendung (Beispiele: del Mar Bibiloni et al. 2012, Richter et al. 2012, Torjusen et al. 2012, Charreire et al. 2011, Cutler et al. 2011, Hare-Bruun et al. 2011, Olinto et al. 2011, Northstone und Emmett 2010, Rezazadeh et al. 2010, Arkkola et al. 2008).

#### **4.1.4 Anwendung der Hauptkomponentenanalyse auf den Lebensmittelverzehr der Teilnehmenden der NVS II**

Für die Zielsetzung, die Eignung verschiedener Methoden zur umfassenden Darstellung des Lebensmittelverzehrs zu prüfen, wurde die Hauptkomponentenanalyse angewendet. Der Vorteil explorativer Methoden ist, dass im Gegensatz zu hypothesenorientierten Ansätzen unabhängig von Empfehlungen für das Untersuchungsziel bedeutende Nährstoffe und Lebensmittel(-gruppen) einbezogen werden können. Zudem wird durch das Verfahren der Hauptkomponentenanalyse die Anzahl der Variablen, die zur Berechnung einbezogen werden sollen, nicht eingeschränkt. Es werden die 13 Lebensmittelgruppen Obst, Gemüse, Getreide, Kartoffeln, Milch, Fisch, Fleisch, Eier, Alkohol, alkoholfreie Getränke, Nüsse und Samen sowie Süßwaren einbezogen (Gruppierung der Lebensmittelgruppen siehe Anhang Tab. 9.1).

Für die vorliegenden Analysen wird das Kaisers Eigenwert-Kriterium zur Bestimmung der Anzahl der Hauptkomponenten verwendet. Demnach sind Hauptkomponenten mit einem Eigenwert  $\geq 1$  bedeutsam. Durch einen Eigenwert von  $<1$  würden nicht mehr als der durchschnittliche Eigenwert der Variablen erklärt werden (Schendera 2004 S 680).

Daneben wurde eine orthogonale Rotation über das Varimax-Verfahren durchgeführt, um die Interpretierbarkeit der Hauptkomponenten mit einer Varianzmaximierung zu erzielen, indem die Anzahl der Variablen mit hoher Ladung auf eine Hauptkomponente minimiert werden (Schendera 2004 S 601).

Über die Anwendung des Verfahrens Simple Structure wird geprüft, ob Lebensmittelgruppen, die keinen wesentlichen Beitrag in den extrahierten Hauptkomponenten leisten, zur Vereinfachung der Auswertung ausgeschlossen werden können.

Da Untersuchungsgegenstand der vorliegenden Arbeit eine umfassende Darstellung des Lebensmittelverzehrs ist, werden als Auswahlkriterium nur Variablen mit einer Ladung unter  $\pm 0,2$  ausgeschlossen.

Um die extrahierten Hauptkomponenten miteinander zu vergleichen, werden die Teilnehmenden anhand des Faktorwertes in vier gleich große Gruppen (Quartilen) eingeteilt. Es werden Personen der ersten Quartile (weniger gute Übereinstimmung mit einer Hauptkomponente) und vierten Quartile (gute Übereinstimmung mit einer Hauptkomponente), verglichen. Hierfür werden zum einen die durchschnittlich verzehrten Lebensmittelmengen verglichen und zum anderen die Personen der einzelnen Quartilen anhand weiterer Variablen charakterisiert. Unter anderem wird untersucht, in welcher Quartile sich die Bio-Käufer/innen und Nicht-Bio-Käufer/innen überwiegend wiederfinden und inwiefern ein Zusammenhang mit weiteren Variablen wie Alter, sozialer Schicht sowie gesundheitsrelevanten Variablen wie BMI, sportliche Aktivität sowie Raucherstatus vorliegt.

#### **4.1.5 Inhaltliche Ergebnisse**

Zur Darstellung des Lebensmittelverzehrs werden im Folgenden, getrennt für Frauen und Männer, die inhaltlichen Ergebnisse der Hauptkomponentenanalyse präsentiert. Diese inhaltliche Auswertung dient als Basis, um die Methode daraufhin zu prüfen, in wie fern sie die Kriterien für eine umfassende Darstellung des Lebensmittelverzehrs erfüllen, was im nächsten Kapitel vorgenommen wird.

Im folgenden Kapitel werden nach der Korrelationsmatrix die extrahierten Hauptkomponenten der Frauen und Männer dargestellt. Anschließend werden die Studienteilnehmenden über die Faktorwerte den Hauptkomponenten zugeordnet und anhand weiterer Variablen beschrieben.

#### **Korrelationsmatrix**

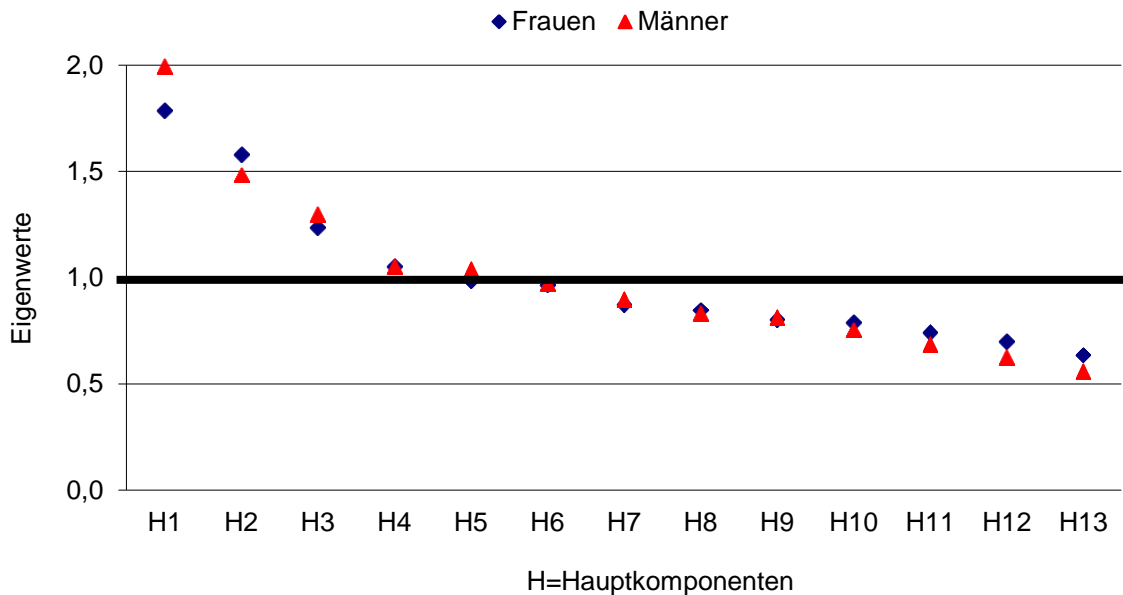
Zur Eignungsprüfung der Variablen wurde zunächst eine Korrelationsmatrix getrennt nach Geschlecht berechnet (Tab. 4.3). Sie liefert Hinweise, ob und wo Zusammenhänge zwischen den einzelnen Variablen bestehen. Nur wenn hohe Korrelationen vorliegen, ist die Durchführung einer Hauptkomponentenanalyse geeignet (Schendera 2010 S 183, 2004 S 596). Es zeigt sich, dass, sowohl hohe Korrelationen über  $\pm 0,2$ , wie für Obst und Gemüse, als auch niedrige Korrelation, wie für Milch und Fisch vorliegen. Demnach ist davon auszugehen, dass mehrere Hauptkomponenten extrahiert werden können und der Lebensmittelverzehr über mehrere Lebensmittelgruppen gleichzeitig dargestellt werden kann.

**Tab. 4.3: Korrelationsmatrix der in der PCA einbezogenen Lebensmittelgruppen der Frauen und Männer**

Lebensmittelgruppen	Obst	Gemüse	Getreide	Kartoffeln	Nüsse/Samen	Milch	Eier	Fisch	Fleisch	Streichfett	Getränke	Alkohol <sup>1</sup>	Süßwaren
<b>Frauen</b>													
Obst	1,00												
Gemüse	0,25	1,00											
Getreide	-0,03	0,09	1,00										
Kartoffeln	0,01	0,07	0,06	1,00									
Nüsse und Samen	0,08	0,08	0,04	-0,01	1,00								
Milch	0,09	0,06	0,06	0,02	0,03	1,00							
Eier	-0,02	0,06	0,07	0,11	0,01	0,04	1,00						
Fisch	0,14	0,20	-0,01	0,09	0,05	0,05	0,10	1,00					
Fleisch	-0,10	0,02	0,15	0,24	-0,02	-0,01	0,20	0,08	1,00				
Streichfett	-0,04	-0,02	0,27	0,15	0,02	0,01	0,12	0,01	0,18	1,00			
Getränke	0,14	0,17	0,06	-0,03	0,03	0,07	0,00	0,06	-0,01	-0,08	1,00		
Alkohol <sup>1</sup>	-0,07	0,04	-0,03	-0,01	0,06	-0,03	0,08	0,06	0,09	-0,02	0,02	1,00	
Süßwaren	-0,02	-0,05	0,21	0,09	0,02	0,08	0,07	-0,05	0,13	0,17	-0,01	-0,07	1,00
<b>Männer</b>													
Obst	1,00												
Gemüse	0,27	1,00											
Getreide	0,04	0,12	1,00										
Kartoffeln	0,02	0,11	0,07	1,00									
Nüsse und Samen	0,09	0,08	0,05	0,02	1,00								
Milch	0,06	0,03	0,10	0,03	0,04	1,00							
Eier	-0,02	0,05	0,10	0,14	0,07	0,12	1,00						
Fisch	0,13	0,21	0,04	0,08	0,05	0,02	0,07	1,00					
Fleisch	-0,11	0,07	0,23	0,28	0,00	0,00	0,24	0,04	1,00				
Streichfett	-0,01	-0,02	0,28	0,17	-0,01	0,01	0,13	0,04	0,18	1,00			
Getränke	0,10	0,13	0,15	0,00	0,02	0,03	0,07	0,03	0,14	-0,05	1,00		
Alkohol <sup>1</sup>	-0,08	0,04	-0,07	0,04	0,04	-0,08	0,06	0,06	0,16	0,02	-0,08	1,00	
Süßwaren	-0,02	-0,04	0,24	0,14	0,03	0,10	0,08	0,00	0,13	0,22	0,08	-0,09	1,00

<sup>1</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke  
Die Werte über ±0,2 sind gelb hinterlegt

Bei den Frauen und Männern konnten jeweils 13 Hauptkomponenten extrahiert werden. Der Verlauf der Eigenwerte ist in Abb. 4.1 präsentiert. Zur Bestimmung der Anzahl der Hauptkomponenten werden Hauptkomponenten mit einem Eigenwert  $\geq 1$  verwendet.



**Abb. 4.1: Ermittelte Eigenwerte der Hauptkomponenten der Frauen und Männer**

### Hauptkomponenten der Frauen

Über das Kaiser-Kriterium (Hauptkomponenten mit einem Eigenwert  $\geq 1$ ), konnten in der Gruppe der Frauen vier Hauptkomponenten extrahiert werden, die eine Gesamtvarianz von 43,5 % des Lebensmittelverzehr erklären. Mit der ersten Hauptkomponente wird mit 14 % der größte Anteil der Gesamtvarianz erklärt.

Die Anwendung des Verfahrens Simple Structure ergab, dass keine Lebensmittelgruppe ausgeschlossen werden kann, die in allen vier extrahierten Hauptkomponenten eine Faktorladung unter  $\pm 0,2$  aufweist. Es fand keine weitere Optimierung der extrahierten Hauptkomponenten statt.

Die Darstellung des Lebensmittelverzehr in einer Hauptkomponentenanalyse erfolgt durch die extrahierten Hauptkomponenten. Die vier Hauptkomponenten der Frauen enthalten eine charakteristische Kombination an Lebensmittelgruppen und -mengen (Tab. 4.4), wobei die Verzehrsmengen durch die Faktorladungen in einer Hauptkomponente ausgedrückt werden. Eine negative Faktorladung einer Lebensmittelgruppe in einer Hauptkomponente deutet auf eine geringere, eine positive Ladung auf eine höhere Verzehrsmenge im Vergleich zur Gesamtgruppe hin.

**Tab. 4.4: Hauptkomponenten der Frauen mit den jeweiligen Faktorladungen und dem erklärten Varianzanteil**

Lebensmittelgruppen	Faktorladungen der Hauptkomponenten			
	1	2	3	4
Obst		0,51	0,20	
Gemüse		0,54		
Getreide	0,41		0,36	0,27
Kartoffeln	0,37			-0,42
Nüsse und Samen		0,20		0,58
Milch			0,30	
Eier	0,34		-0,30	
Fisch		0,40	-0,29	
Fleisch	0,46		-0,28	
Streichfett	0,44			
Getränke		0,38		
Alkohol <sup>1</sup>			-0,51	0,55
Süßwaren	0,34		0,38	
<b>erklärte Varianz (%)</b>	<b>14</b>	<b>12</b>	<b>10</b>	<b>8</b>

<sup>1</sup>Alkohol (Ethanol) als Nahrungsinhaltsstoff alkoholischer Getränke

Die **erste Hauptkomponente** lässt sich durch einen gleichzeitigen hohen Verzehr der sechs Lebensmittelgruppen Getreide, Kartoffeln, Eier, Fleisch, Streichfett und Süßwaren beschreiben. Die Lebensmittelgruppen Obst, Gemüse, Nüsse und Samen, Milch sowie Fisch, Getränke und Alkohol spielen bei dieser Hauptkomponente keine Rolle. Die Verzehrsmengen

mengen der weiteren Lebensmittelgruppen sind im Vergleich zur Gesamtgruppe nicht auffällig. Die **zweite Hauptkomponente** zeichnet sich durch einen hohen Verzehr an Obst, Gemüse, Fisch, Getränke, Nüsse und Samen, die **dritte Hauptkomponente** durch einen hohen Verzehr an Obst, Getreide, Milch, Süßwaren und gleichzeitig einen geringen Verzehr an Eier, Fisch, Fleisch und Alkohol aus. Die **vierte Hauptkomponente** der Frauen ist durch einen hohen Verzehr an Getreide, Nüssen und Samen, Alkohol sowie durch eine gleichzeitig geringe Menge an Kartoffeln gekennzeichnet.

Die vier Hauptkomponenten enthalten die oben beschriebene charakteristische Kombination an Lebensmittelgruppen und -mengen. Für die Zuordnung der Studienteilnehmenden zu den einzelnen Hauptkomponenten wird der Faktorwert zu den vier Hauptkomponenten über die Faktorladungen der einzelnen Lebensmittelgruppen berechnet (Beispielrechnung siehe Tab. 4.2). Je höher der Faktorwert einer Person zu einer Hauptkomponente, desto besser stimmt der Lebensmittelverzehr mit der in der entsprechenden Hauptkomponente beschriebenen Lebensmittelkombination überein. Die Frauen werden anhand ihres Faktorwertes zu jeder einzelnen Hauptkomponente in jeweils vier gleichgroße Gruppen eingeteilt. Frauen aus der ersten Quartile (schlechte Übereinstimmung mit der Hauptkomponente) werden mit denen der vierten Quartile (gute Übereinstimmung mit der Hauptkomponente) verglichen.

Frauen, die der vierten Quartile der ersten Hauptkomponente zugeordnet wurden, verzehren die für diese Hauptkomponente charakteristischen Lebensmittel in höheren Mengen als Frauen, die der ersten Quartile dieser Hauptkomponente zugeordnet wurden (Tab. 4.5).

**Tab. 4.5: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der ersten Hauptkomponente**

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev)	1. Hauptkomponente (MW±Std dev)		Signifikanzen <sup>1</sup>	Faktorladung der 1. Hauptkomponente	Interpretation der 1. Hauptkomponente
		1. Quartile	4. Quartile			
Obst	292±233	309±281	284±236	n.s.		
Gemüse	251±150	219±147	283±163	<0,001		
Getreide	239±99	171±69	308±111	<0,001	<b>0,41</b>	hoher Getreideverzehr
Kartoffeln	72±43	45±30	99±48	<0,001	<b>0,37</b>	hoher Kartoffelverzehr
Nüsse und Samen	4±11	3±10	5±13	<0,001		
Milch	239±211	207±178	277±261	<0,001		
Eier	12±13	6±7	20±17	<0,001	<b>0,34</b>	hoher Eierverzehr
Fisch	23±22	17±18	28±26	<0,001		
Fleisch	75±48	43±27	115±59	<0,001	<b>0,46</b>	hoher Fleischverzehr
Streichfett	21±17	10±8	35±23	<0,001	<b>0,44</b>	hoher Streichfettverzehr
Getränke	2335±889	2337±930	2370±894	n.s.		
Alkohol <sup>2</sup>	6±9	5±8	6±9	<0,001		
Süßwaren	61±56	34±29	93±78	<0,001	<b>0,34</b>	hoher Süßwarenverzehr

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

Auf Basis der Einteilung in Quartilen können die Frauen mit weiteren Variablen wie dem Bio-Kauf sowie ernährungs- und gesundheitsrelevanten Variablen charakterisiert und verglichen

werden. Im Gegensatz zur Beschreibung von Zusammenhängen zwischen dem Verzehr einzelner ausgewählter Lebensmittel und weiteren Variablen, können über den Ansatz der Hauptkomponentenanalyse Personen, die sich durch eine charakteristische Kombination an verzehrten Lebensmitteln auszeichnen, umfangreicher mit weiteren Variablen charakterisiert werden. Hierfür werden Frauen, die eine hohe Übereinstimmung zur ersten Hauptkomponente aufweisen, d. h. viel Getreide, Kartoffeln, Eier, Fleisch, Streichfett und gleichzeitig viel Süßwaren verzehren, mit Personen, die keine gute Übereinstimmung zu dieser Hauptkomponente aufweisen, verglichen. Hierdurch wird auch geprüft, wie gut sich die Bio-Käuferinnen und Nicht-Bio-Käuferinnen den einzelnen Hauptkomponenten zuordnen lassen (Tab. 4.6).

**Tab. 4.6: Charakterisierung der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der ersten Hauptkomponente**

		Gesamtgruppe (%)	1. Hauptkomponente 1. Quartile (%)	4. Quartile (%)	Signifikanzen <sup>1</sup>
<b>Bio-Kauf</b>	ja	50,3	51,8	49,4	n.s.
	nein	49,7	48,2	50,6	n.s.
<b>Altersgruppe</b>	18–34 Jahre	20,7	23,7	21,2	n.s.
	35–50 Jahre	35,3	33,0	39,8	<0,001
	51–64 Jahre	24,2	24,5	22,5	n.s.
	65–80 Jahre	19,8	18,8	16,5	n.s.
<b>Soziale Schicht<sup>3</sup></b>	untere Schicht	21,0	20,3	22,0	n.s.
	mittlere Schicht	29,9	30,9	29,7	n.s.
	obere Schicht	49,1	48,8	48,3	n.s.
<b>BMI</b>	Untergewicht	1,6	1,9	2,0	n.s.
	Normalgewicht	46,0	47,1	46,0	n.s.
	Übergewicht	27,9	26,9	25,4	n.s.
	Adipositas	18,9	17,3	21,3	<0,01
	keine Angabe	5,6	6,8	5,3	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 88 cm	65,8	67,4	64,4	n.s.
	über 88 cm	29,4	27,5	30,9	<0,05
	keine Angabe	4,8	5,2	4,7	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüft- umfang)</b>	bis 0,85	74,2	74,8	76,2	n.s.
	über 0,85	21,0	20,0	19,1	n.s.
	keine Angabe	4,8	5,2	4,7	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	1968±589	1463,5	2571,6	<0,001
<b>Ernährungskennnisse</b>	gut	40,9	40,2	41,7	n.s.
	mittel	33,9	34,1	34,0	n.s.
	schlecht	24,2	24,4	23,5	n.s.
	keine Angabe	1,0	1,4	0,8	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	58,9	62,3	53,1	<0,001
	nein	40,6	37,1	46,2	<0,001
	keine Angabe	0,6	0,6	0,7	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	21,7	23,8	24,1	n.s.
	nein	78,0	75,9	75,7	n.s.
	keine Angabe	0,3	0,3	0,3	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	76,2	75,8	75,7	n.s.
	mittel	20,2	20,1	20,6	n.s.
	schlecht	3,5	3,9	3,7	n.s.
	keine Angabe	0,1	0,2	0,0	--- <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	78,8	77,8	78,6	n.s.
	schlecht	20,7	21,4	21,0	n.s.
	keine Angabe	0,5	0,7	0,5	--- <sup>2</sup>
<b>HEI-NVS II</b>	arithm. MW±SD	84,8±9,8	82,2	84,5	<0,001

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und dem HEI-NVS II über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

Es konnte kein Unterschied im Anteil der Bio-Käuferinnen und Nicht-Bio-Käuferinnen zwischen der ersten und vierten Quartile festgestellt werden. Frauen der vierten Quartile sind zu einem höheren Anteil in der Altersgruppe der 35- bis 50-Jährigen vertreten, sind zu einem

höheren Anteil adipös, zu einem höheren Anteil in der Gruppe mit einem erhöhten Taillenumfang, weisen eine höhere Energiezufuhr auf und wurden zu einem geringeren Anteil der Gruppe der sportlich Aktiven zugeordnet. Zudem erreichen Frauen der vierten Quartile eine höhere HEI-NVS II-Indexsumme, die dem Durchschnitt der Gesamtgruppe der Frauen entspricht, während Frauen der ersten Quartile darunter liegen. In den weiteren Variablen konnten keine Unterschiede festgestellt werden.

Durch die zweite Hauptkomponente wird ein gleichzeitig hoher Verzehr der sechs Lebensmittelgruppen Obst, Gemüse, Milch, Fisch, Getränke, Nüsse und Samen beschrieben. Wie bei der ersten Hauptkomponente unterscheidet sich der Lebensmittelverzehr der Frauen, die der ersten und vierten Quartile zugeordnet wurden (Tab. 4.7).

**Tab. 4.7: Lebensmittelverzehr (g/Tag) Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der zweiten Hauptkomponente**

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev)	2. Hauptkomponente (MW±Std dev)		Signifikanzen <sup>1</sup>	Faktorladung der 2. Hauptkomponente	Interpretation der 2. Hauptkomponente
		1. Quartile	4. Quartile			
Obst	292±233	144±103	484±315	<0,001	<b>0,51</b>	hoher Obstverzehr
Gemüse	251±150	144±70	391±182	<0,001	<b>0,54</b>	hoher Gemüseverzehr
Getreide	239±99	239±99	238±105	n.s.		
Kartoffeln	72±43	71±43	73±46	n.s.		
Nüsse und Samen	4±11	1±4	7±18	<0,001	<b>0,20</b>	hoher Nuss- und Samenverzehr
Milch	239±211	177±149	310±278	<0,001	<b>0,20</b>	hoher Milchverzehr
Eier	12±13	12±12	13±14	<0,001		
Fisch	23±22	12±11	38±32	<0,001	<b>0,40</b>	hoher Fischverzehr
Fleisch	75±48	84±53	68±49	<0,001		
Streichfett	21±17	26±21	16±13	<0,001		
Getränke	2335±889	1794±671	2914±974	<0,001	<b>0,38</b>	hoher Getränkeverzehr
Alkohol <sup>2</sup>	6±9	5±7	6±10	<0,001		
Süßwaren	61±56	78±75	50±48	<0,001		

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

Im Gegensatz zur ersten Hauptkomponente konnten bei der zweiten Hauptkomponente mehr Unterschiede in den beschreibenden Variablen festgestellt werden (Tab. 4.8).

**Tab. 4.8: Charakterisierung der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der zweiten Hauptkomponente**

		Gesamtgruppe (%)	2. Hauptkomponente 1. Quartile (%)	4. Quartile (%)	Signifikanzen <sup>1</sup>
<b>Bio-Kauf</b>	ja	50,3	36,4	62,3	<0,001
	nein	49,7	63,6	37,7	<0,001
<b>Altersgruppe</b>	18–34 Jahre	20,7	32,7	14,1	<0,001
	35–50 Jahre	35,3	34,2	33,9	n.s.
	51–64 Jahre	24,2	15,9	32,5	<0,001
	65–80 Jahre	19,8	17,1	19,6	n.s.
<b>Soziale Schicht<sup>2</sup></b>	untere Schicht	21,0	26,0	18,2	<0,001
	mittlere Schicht	29,9	33,5	26,2	<0,001
	obere Schicht	49,1	40,5	55,6	<0,001
<b>BMI</b>	Untergewicht	1,6	2,4	1,2	<0,01
	Normalgewicht	46,0	47,6	43,3	<0,01
	Übergewicht	27,9	25,1	31,3	<0,001
	Adipositas	18,9	18,2	19,1	n.s.
	keine Angabe	5,6	6,8	5,1	... <sup>2</sup>
<b>Taillenumfang</b>	bis 88 cm	65,8	66,2	67,2	n.s.
	über 88 cm	29,4	27,8	28,6	n.s.
	keine Angabe	4,8	6,0	4,3	... <sup>2</sup>
<b>WHR (Taillen- zu Hüft- umfang)</b>	bis 0,85	74,2	74,6	75,4	n.s.
	über 0,85	21,0	19,5	20,3	n.s.
	keine Angabe	4,8	6,0	4,3	... <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	1968±589	1901,3	2132,2	<0,001
<b>Ernährungskennnisse</b>	gut	40,9	32,4	48,5	<0,001
	mittel	33,9	35,0	31,9	<0,05
	schlecht	24,2	31,8	19,0	<0,001
	keine Angabe	1,0	0,7	0,6	... <sup>2</sup>
<b>Sportlich aktiv</b>	ja	58,9	46,7	70,4	<0,001
	nein	40,6	53,1	28,9	<0,001
	keine Angabe	0,6	0,2	0,7	... <sup>2</sup>
<b>Raucherstatus</b>	ja	21,7	28,2	19,0	<0,001
	nein	78,0	71,6	80,7	<0,001
	keine Angabe	0,3	0,2	0,3	... <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	76,2	72,6	79,2	<0,001
	mittel	20,2	23,2	17,1	<0,001
	schlecht	3,5	4,2	3,5	n.s.
	keine Angabe	0,1	0,1	0,2	... <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	78,8	77,6	80,9	<0,05
	schlecht	20,7	21,6	18,8	<0,05
	keine Angabe	0,5	0,7	0,3	... <sup>2</sup>
<b>HEI-NVS II</b>	arithm. MW±SD	84,8±9,8	75,3	92,5	<0,001

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und dem HEI-NVS II über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

Frauen, die der vierten Quartile zugeordnet wurden, sind zu einem höheren Anteil Bio-Käuferinnen, zu einem höheren Anteil in der Altersgruppe der 51- bis 64-Jährigen, zu einem geringeren Anteil in der Altersgruppe der 18- bis 34-Jährigen, zu einem höheren Anteil in der

oberen sozialen Schicht und in der Gruppe der Frauen mit guten Ernährungskennntnisse vertreten. Daneben sind sie zu einem höheren Anteil sportlich aktiv und Nichtraucherinnen. Zudem ist der Anteil derjenigen, die ihren Gesundheitsstatus sowie die Schlafqualität als gut einschätzen, höher als in der Personengruppe, die der ersten Quartile zugeordnet wurde. Der Anteil Normalgewichtiger ist in der vierten Quartile niedriger und der Anteil Übergewichtiger höher. Der Anteil Adipöser sowie derjenigen, die einen zu hohen Taillenumfang oder ein zu hohes WHR besitzen, unterscheidet sich nicht. Zudem erreichen Frauen der vierten Quartile eine höhere HEI-NVS II-Indexsumme als Frauen, die der ersten Quartile zugeordnet wurden, sowie im Vergleich zur Gesamtgruppe.

Die Ergebnisse der dritten und vierten Hauptkomponente der Frauen sind im Anhang dargestellt (Kapitel 0).

### **Hauptkomponenten der Männer**

Über das Kaiser-Kriterium konnten bei den Männern fünf Hauptkomponenten, die eine Gesamtvarianz von 52,8 % erklären, extrahiert werden. Mit der ersten Hauptkomponente wird mit 15 % der größte Anteil der Gesamtvarianz erklärt.

Die Anwendung des Verfahrens Simple Structure ergab, dass keine Lebensmittelgruppe ausgeschlossen werden kann.

Durch die fünf Hauptkomponenten, die in der Gruppe der Männer extrahiert wurden, wird eine charakteristische Kombination an Lebensmittelgruppen und -mengen abgebildet (Tab. 4.9).

**Tab. 4.9: Hauptkomponenten der Männer mit den jeweiligen Faktorladungen und dem erklärten Varianzanteil**

Lebensmittelgruppen	Männer				
	Faktorladungen der Hauptkomponenten				
	1	2	3	4	5
Obst		0,57			
Gemüse	0,22	0,53			
Getreide	0,42		-0,28		
Kartoffeln	0,35		0,22		
Nüsse und Samen		0,21		0,27	0,43
Milch			-0,29	0,20	0,59
Eier	0,33				0,48
Fisch		0,35	0,27	0,22	
Fleisch	0,43	-0,21	0,27	-0,32	
Streichfett	0,37	-0,25		0,36	-0,33
Getränke	0,20	0,23		-0,73	
Alkohol <sup>1</sup>			0,62		
Süßwaren	0,34		-0,37		
erklärte Varianz (%)	15	11	10	8	8

<sup>1</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

Männer, die der **ersten Hauptkomponente** zugeordnet werden, weisen einen gleichzeitigen hohen Verzehr an Gemüse, Getreide, Kartoffeln, Eier, Fleisch, Streichfett, Süßwaren und Getränken auf. Die Lebensmittelgruppen Obst, Nüsse und Samen, Milch, Fisch und Alkohol spielen bei dieser Hauptkomponente keine Rolle. Die **zweite Hauptkomponente** zeichnet sich durch einen gleichzeitig hohen Verzehr an Obst, Gemüse, Nüsse und Samen, Fisch, Getränke und einen niedrigen Verzehr an Fleisch und Streichfett aus. Die **dritte Hauptkomponente** weist einen niedrigen Getreide-, Milch- und Süßwarenverzehr sowie gleichzeitig einen hohen Verzehr an Kartoffeln, Fisch, Fleisch und Alkohol auf. Die **vierte Hauptkomponente** zeichnet sich durch einen hohen Verzehr an Nüssen und Samen, Milch, Fisch und Streichfett sowie gleichzeitig einen geringen Fleisch- und Getränkeverzehr aus. Die **fünfte Hauptkomponente** ist durch einen hohen Verzehr an Nüssen und Samen, Milch, Eier und gleichzeitig einen geringen Verzehr an Streichfett gekennzeichnet.

Männer, die der vierten Quartile der ersten Hauptkomponente zugeordnet wurden, verzehren die für diese Hauptkomponente charakteristischen Lebensmittel in höheren Mengen als Männer, die der ersten Quartile zugeordnet wurden. Zwar wurden auch in den Verzehrsmengen der weiteren Lebensmittelgruppen Unterschiede zwischen den Männern der ersten und vierten Quartile festgestellt, diese sind aber nicht charakteristisch für diese Hauptkomponente (Tab. 4.10).

**Tab. 4.10: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der ersten Hauptkomponente**

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev)	1. Hauptkomponente (MW±Std dev)		Signifikanzen <sup>1</sup>	Faktorladung der 1. Hauptkomponente	Interpretation der 1. Hauptkomponente
		1. Quartile	4. Quartile			
Obst	249±223	214±206	282±251	<0,001		
Gemüse	239±143	183±109	298±173	<0,001	<b>0,22</b>	hoher Gemüseverzehr
Getreide	303±127	212±80	401±143	<0,001	<b>0,42</b>	hoher Getreideverzehr
Kartoffeln	92±57	60±36	129±71	<0,001	<b>0,35</b>	hoher Kartoffelverzehr
Nüsse und Samen	5±13	2±8	7±18	<0,001		
Milch	249±258	188±203	328±325	<0,001		
Eier	16±18	8±9	27±27	<0,001	<b>0,33</b>	hoher Eierverzehr
Fisch	29±29	21±20	37±37	<0,001		
Fleisch	133±83	84±44	197±110	<0,001	<b>0,43</b>	hoher Fleischverzehr
Streichfett	30±27	16±13	49±37	<0,001	<b>0,37</b>	hoher Streichfettverzehr
Getränke	2332±981	1987±865	2704±1107	<0,001	<b>0,20</b>	hoher Getränkeverzehr
Alkohol <sup>2</sup>	16±18	15±18	17±21	<0,001		
Süßwaren	68±63	37±33	107±85	<0,001	<b>0,34</b>	hoher Süßwarenverzehr

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

Die für die Hauptkomponente charakteristische Kombination an verzehrten Lebensmitteln wird im Zusammenhang mit weiteren Variablen beschrieben. Hierfür werden die Männer der vierten Quartile mit denen der ersten Quartile verglichen (Tab. 4.11).

**Tab. 4.11: Charakterisierung der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der ersten Hauptkomponente**

		Gesamtgruppe (%)	1. Hauptkomponente 1. Quartile (%)	4. Quartile (%)	Signifikanzen <sup>1</sup>
<b>Bio-Kauf</b>	ja	38,6	37,2	35,6	n.s.
	nein	61,5	62,8	64,4	n.s.
<b>Altersgruppe</b>	18–34 Jahre	19,7	17,7	24,0	<0,001
	35–50 Jahre	32,0	28,4	36,9	<0,001
	51–64 Jahre	25,3	26,4	22,2	< 0,01
	65–80 Jahre	23,0	27,6	16,9	<0,001
<b>Soziale Schicht<sup>3</sup></b>	untere Schicht	17,6	18,9	18,7	n.s.
	mittlere Schicht	28,4	27,7	31,2	<0,05
	obere Schicht	54,1	53,4	50,1	n.s.
<b>BMI</b>	Untergewicht	0,7	0,7	0,9	n.s.
	Normalgewicht	30,0	25,5	35,8	<0,001
	Übergewicht	45,0	45,5	42,4	n.s.
	Adipositas	20,1	21,8	17,9	< 0,01
	keine Angabe	4,3	6,5	3,0	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 102 cm	68,6	65,1	72,6	<0,001
	über 102 cm	28,0	30,2	24,8	<0,001
	keine Angabe	3,4	4,7	2,6	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüft- umfang)</b>	bis 1,0	81,2	77,2	84,0	<0,001
	über 1,0	15,4	18,1	13,4	<0,001
	keine Angabe	3,4	4,7	2,6	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	2569±850	1790±411	3526±857	<0,001
<b>Ernährungskennnisse</b>	gut	17,1	14,5	18,7	< 0,05
	mittel	38,2	36,2	37,4	n.s.
	schlecht	43,8	48,0	43,4	< 0,05
	keine Angabe	0,9	1,3	0,5	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	56,2	50,7	58,0	<0,001
	nein	43,4	48,8	41,6	<0,001
	keine Angabe	0,5	0,5	0,3	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	28,1	27,6	33,8	<0,001
	nein	71,6	72,2	65,8	<0,001
	keine Angabe	0,2	0,1	0,4	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	75,2	73,3	76,6	<0,05
	mittel	20,7	21,3	20,2	n.s.
	schlecht	3,9	5,4	3,1	<0,01
	keine Angabe	0,1	0,1	0,1	--- <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	83,6	82,6	83,1	n.s.
	schlecht	16,0	17,2	16,8	n.s.
	keine Angabe	0,4	0,2	0,1	--- <sup>2</sup>
<b>HEI-NVS II</b>	arithm. MW±SD	80,2±10,7	78,2±10,5	79,2±11,2	<0,05

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und dem HEI-NVS II über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

Im Anteil der Bio-Käufer und Nicht-Bio-Käufer konnten keine Unterschiede zwischen der ersten und vierten Quartile festgestellt werden. Männer, die der vierten Quartile zugeordnet wurden, gehören zu einem höheren Anteil der Altersgruppe der 18- bis 34-Jährigen, der 35- bis 50-Jährigen sowie der mittleren sozialen Schicht an. Sie sind zu einem höheren Anteil

normalgewichtig, zu einem niedrigeren Anteil adipös, weisen einen geringeren Anteil an Personen mit einem zu hohen Taillenumfang und WHR auf, weisen eine höhere Energiezufuhr auf, als Männer die der ersten Quartile zugeordnet wurden. Daneben zeigt sich, dass der Anteil an Männern, die gute Ernährungskennnisse aufweisen, die Gruppe der sportlich Aktiven und Raucher höher ist als in der Personengruppe, die dem ersten Quartil zugeordnet wurden. Der Anteil an Männern, die ihren allgemeinen Gesundheitszustand als gut einschätzen, ist hier ebenfalls höher. Zudem weisen sie, gemessen an der HEI-NVS II-Indexsumme, eine Ernährung näher an den Empfehlungen auf.

Die zweite Hauptkomponente beschreibt einen gleichzeitigen hohen Verzehr an Obst, Gemüse, Fisch, Getränke, Nüsse und Samen, sowie einen geringen Verzehr an Fleisch und Streichfett (Tab. 4.12).

**Tab. 4.12: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der zweiten Hauptkomponente**

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev)	2. Hauptkomponente (MW±Std dev)		Signifikanz <sup>1</sup>	Faktorladung der 2. Hauptkomponente	Interpretation der 2. Hauptkomponente
		1. Quartile	4. Quartile			
Obst	249±223	104±96	465±285	<0,001	<b>0,57</b>	hoher Obstverzehr
Gemüse	239±143	139±79	367±170	<0,001		hoher Gemüseverzehr
Getreide	303±127	312±139	304±129	n.s.		
Kartoffeln	92±57	99±63	87±56	<0,001	<b>0,21</b>	hoher Nuss- und Samenverzehr
Nüsse und Samen	5±13	2±6	9±21	<0,001		
Milch	249±258	207±230	289±274	<0,001	<b>0,35</b>	hoher Fischverzehr
Eier	16±18	19±22	15±17	<0,001		niedriger Fleischverzehr
Fisch	29±29	16±16	46±40	<0,001	<b>-0,21</b>	niedriger Streichfettverzehr
Fleisch	133±83	166±97	111±73	<0,001		hoher Getränkeverzehr
Streichfett	30±27	43±35	22±18	<0,001	<b>-0,25</b>	
Getränke	2332±981	1967±880	2704±1071	<0,001	<b>0,23</b>	
Alkohol <sup>2</sup>	16±18	20±23	13±15	<0,001		
Süßwaren	68±63	90±79	55±53	<0,001		

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

Die weitere Charakterisierung zeigt, dass Männer, die dem vierten Quartil der zweiten Hauptkomponente zugeordnet werden konnten, zu einem höheren Anteil Bio-Käufer sind. Zudem ist der Anteil der Männer aus der Altersgruppe der 51- bis 64-Jährigen sowie der 65- bis 80-Jährigen und insbesondere der oberen sozialen Schicht höher als in der Personengruppe, die dem ersten Quartil zugeordnet wurde. Daneben unterscheiden sich Männer aus dem vierten Quartil durch einen geringeren Anteil Untergewichtiger, Normalgewichtiger, einen höheren Anteil Übergewichtiger und Adipöser. Männer aus dem vierten Quartil können zu einem höheren Anteil der Gruppe der sportlich Aktiven und zu einem geringeren Anteil der Gruppe der Raucher zugeordnet werden. Zudem schätzt ein höherer Personenanteil die Schlafqualität als gut ein. Gemessen an der HEI-NVS II-Indexsumme weisen sie im Vergleich zu Personen der ersten Quartile und der Gesamtgruppe eine insgesamt günstigere Ernährung auf (Tab. 4.13).

Die Ergebnisse der dritten, vierten und fünften Hauptkomponenten der Männer sind im Anhang dargestellt (Kapitel 0).

**Tab. 4.13: Charakterisierung der Gesamtgruppe sowie Männer der ersten und vierten Quartile der zweiten Hauptkomponente**

		Gesamtgruppe (%)	2. Hauptkomponente		Signifikanzen <sup>1</sup>
			1. Quartile (%)	4. Quartile (%)	
<b>Bio-Kauf</b>	ja	38,6	26,7	52,0	<0,001
	nein	61,5	73,3	48,0	<0,001
<b>Altersgruppe</b>	18–34 Jahre	19,7	28,6	12,8	<0,001
	35–50 Jahre	32,0	33,6	30,0	<0,05
	51–64 Jahre	25,3	21,1	29,9	<0,001
	65–80 Jahre	23,0	16,7	27,3	<0,001
<b>Soziale Schicht<sup>3</sup></b>	untere Schicht	17,6	22,1	15,5	<0,001
	mittlere Schicht	28,4	30,7	24,6	<0,001
	obere Schicht	54,1	47,2	59,9	<0,001
<b>BMI</b>	Untergewicht	0,7	1,5	0,3	<0,001
	Normalgewicht	30,0	36,8	24,8	<0,001
	Übergewicht	45,0	40,3	46,4	<0,001
	Adipositas	20,1	17,3	23,7	<0,001
	keine Angabe	4,3	4,2	4,9	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 102 cm	68,6	71,3	66,9	<0,01
	über 102 cm	28,0	25,0	29,6	<0,01
	keine Angabe	3,4	3,8	3,6	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüft-umfang)</b>	bis 1,0	81,2	81,5	80,2	n.s.
	über 1,0	15,4	14,7	16,2	n.s.
	keine Angabe	3,4	3,8	3,6	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	2569±850	2766±996	2578±791	<0,001
<b>Ernährungskennnisse</b>	gut	17,1	11,1	24,7	<0,001
	mittel	38,2	36,6	37,6	n.s.
	schlecht	43,8	51,0	36,9	<0,001
	keine Angabe	0,9	1,3	0,8	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	56,2	47,7	63,4	<0,001
	nein	43,4	52,0	36,1	<0,001
	keine Angabe	0,5	0,3	0,5	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	28,1	37,3	21,2	<0,001
	nein	71,6	62,5	78,6	<0,001
	keine Angabe	0,2	0,2	0,2	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	75,2	73,7	75,0	n.s.
	mittel	20,7	21,8	21,6	n.s.
	schlecht	3,9	4,4	3,3	n.s.
	keine Angabe	0,1	0,1	0,1	--- <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	83,6	81,6	84,5	<0,05
	schlecht	16,0	18,1	15,2	<0,05
	keine Angabe	0,4	0,3	0,3	--- <sup>2</sup>
<b>HEI-NVS II</b>	arithm. MW±SD	80,2±10,7	69,2±8,1	89,7±8,0	<0,001

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und dem HEI-NVS II über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

#### **4.1.6 Methodische Ergebnisse und Diskussion**

Für das Ziel, die Eignung verschiedener Methoden zur umfassenden Darstellung des Lebensmittelverzehrs zu prüfen, wurde die Hauptkomponentenanalyse unter Anwendung einer Varimax-Rotation sowie des Verfahrens Simple Structure durchgeführt. In einer Hauptkomponentenanalyse werden die Verzehrsmengen der einbezogenen Lebensmittelvariablen auf Basis ihrer Korrelationen zusammengefasst (Newby und Tucker 2004 S 177). Der Lebensmittelverzehr wird durch die extrahierten Hauptkomponenten, die eine charakteristische Kombination an Lebensmittelgruppen und ihren Verzehrsmengen widerspiegeln, dargestellt.

##### **Kriterium: Berücksichtigung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen bei Anwendung der Methode**

Da sich explorative Methoden wie die Hauptkomponentenanalyse nicht auf bisheriges Wissen stützen (Schulze und Hoffmann 2006 S 860, Newby und Tucker 2004 S 177), und das Verfahren keine Einschränkung über die Anzahl der Variablen, die bei der Berechnung einbezogen werden können, gibt, konnten alle 13 für das Untersuchungsziel relevanten Lebensmittelgruppen mit den entsprechenden Verzehrsmengen bei der Anwendung der Hauptkomponentenanalyse einbezogen werden.

Das Kriterium, möglichst alle Lebensmittelgruppen mit den entsprechenden Verzehrsmengen bei der Anwendung der Methode zu berücksichtigen, wird erfüllt.

##### **Kriterium: Darstellung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse**

Der Lebensmittelverzehr wird durch die extrahierten Hauptkomponenten dargestellt (Schulze et al. 2003a S 410). Eine Hauptkomponente ist eine Linearkombination aller Ursprungsvariablen (Kessler 2006 S 22). Das bedeutet, dass für jede in der Berechnung einbezogene Variable eine Faktorladung in jeder Hauptkomponente berechnet wird (Bortz und Schuster 2010 S 393). Die Lebensmittelverzehrsmengen werden durch die Faktorladungen in einer Hauptkomponente beschrieben. Eine Ladung um Null bedeutet, dass die entsprechende Variable nicht charakteristisch für die Hauptkomponente ist, sondern dem Durchschnitt der Gesamtgruppe entspricht. Für die leichtere Interpretation der Ergebnisse werden üblicherweise nur die für eine Hauptkomponente charakteristischen Lebensmittelgruppen mit einer höheren Faktorladung dargestellt (siehe Kapitel 4.1.2). Für die vorliegende Auswertung bedeutet dies, dass durch eine Hauptkomponente alle 13 einbezogenen Lebensmittelgruppen, die bei der Analyse berücksichtigt wurden, dargestellt werden können, üblicherweise aber nur die charakteristischen Lebensmittelgruppen dargestellt und für

weitere Auswertungen herangezogen werden. In der vorliegenden Auswertung wurden bei den Frauen für die erste Hauptkomponente 6 und bei der zweiten Hauptkomponente 5 der 13 einbezogenen Lebensmittelgruppen als charakteristisch angesehen. Bei den Männern waren dies bei der ersten Hauptkomponente 8 und bei der zweiten Hauptkomponente 7 der 13 einbezogenen Lebensmittelgruppen. Damit werden bei den Frauen und Männern nur ein Teil der relevanten Lebensmittelgruppen über die extrahierten Hauptkomponenten dargestellt.

Die Hauptkomponentenanalyse wird insbesondere verwendet, um Ursache-Wirkungs-Zusammenhänge zwischen der Ernährung und Bioparametern oder Krankheiten zu untersuchen (Beispiele: Guallar-Castillon et al. 2012, Daniel et al. 2011, Butler et al. 2010, Edefonti et al. 2010, Erber et al. 2010, Hamer und Mishra 2010, Shaheen et al. 2010, Bertuccio et al. 2009, de Stefani et al. 2008a, de Stefani et al. 2008b, Edefonti et al. 2008, Fung et al. 2001, Palli et al. 2001). Für diese Untersuchungen werden auf Basis des Kaisers-Kriteriums (Newby und Tucker 2004 S 192) nur wenige, besonders aussagekräftige Hauptkomponenten herangezogen, die einen möglichst großen Anteil der Varianz erklären (Schulze 2002 S 51). Es wird untersucht, ob Personen, die sich eher nach einer der ersten beiden Hauptkomponenten, durch die der Großteil der Varianz erklärt wird, ernähren, ein höheres oder niedrigeres Risiko für die Entwicklung einer bestimmten Erkrankung haben.

Die Anwendung der Hauptkomponentenanalyse zur Darstellung der Ernährung ohne Untersuchung von Ursache-Wirkungs-Zusammenhängen findet seltener statt (Beispiele: del Mar Bibiloni et al. 2012, Richter et al. 2012, Torjusen et al. 2012, Charreire et al. 2011, Cutler et al. 2011, Hare-Bruun et al. 2011, Olinto et al. 2011, Northstone und Emmett 2010, Arkkola et al. 2008). Auch bei diesen Analysen werden zur Interpretation der Ergebnisse nur die aussagekräftigsten Hauptkomponenten herangezogen, um den Lebensmittelverzehr einer bestimmten Personengruppe darzustellen (Reedy et al. 2010 S 480).

In der vorliegenden Auswertung wurden bei beiden Geschlechtern jeweils 13 Hauptkomponenten extrahiert, von denen auf Basis des Kaisers-Kriteriums bei den Frauen vier und bei den Männern fünf Hauptkomponenten ausgewählt wurden. Die mit diesen Hauptkomponenten erklärte Varianz des Lebensmittelverzehrs liegt bei den Frauen bei 43,5 % und bei den Männern bei 52,5 %. In den meisten durch die von Brennan et al. (2010), Edefonti et al. (2009) und Newby und Tucker (2004) zitierten Studien werden weniger als 50 % der Varianz erklärt.

Allerdings geht durch die Auswahl einer bestimmten Anzahl an Hauptkomponenten, wie von Moosbrugger und Kelava (2007 S 309f) beschrieben, ein Teil der Gesamtvarianz verloren. Hoffmann et al. (2004 S 936) beschreiben, dass die Anwendung der Hauptkomponentenanalyse zur Untersuchung von Zusammenhängen zwischen der Ernährung und Erkrankung nicht immer zu den erwarteten Ergebnissen führt. Die Ergebnisse stimmten nicht immer mit

denen in der Literatur bekannten Ergebnissen überein. Als Grund hierfür wird von den Autoren diskutiert, dass die Anzahl der zur Auswertung herangezogenen Hauptkomponenten und damit der erklärte Varianzanteil für den untersuchten Ursache-Wirkungs-Zusammenhang möglicherweise nicht ausreichen. Newby und Tucker (2004 S 197) gehen darauf ein, dass weitere Untersuchungen zur Bedeutung von Eigenwerten wie dem Kaisers-Kriterium, durch welche die Anzahl der Hauptkomponenten reduziert werden, mit ihrem Einfluss auf die Ergebnisse notwendig sind.

Für die vorliegende Auswertung bedeutet dies, dass durch diese klassische Herangehensweise der Datenreduktion ein Teil der Information des Lebensmittelverzehr unberücksichtigt bleibt und damit keine umfassende Darstellung des Lebensmittelverzehr erfolgt.

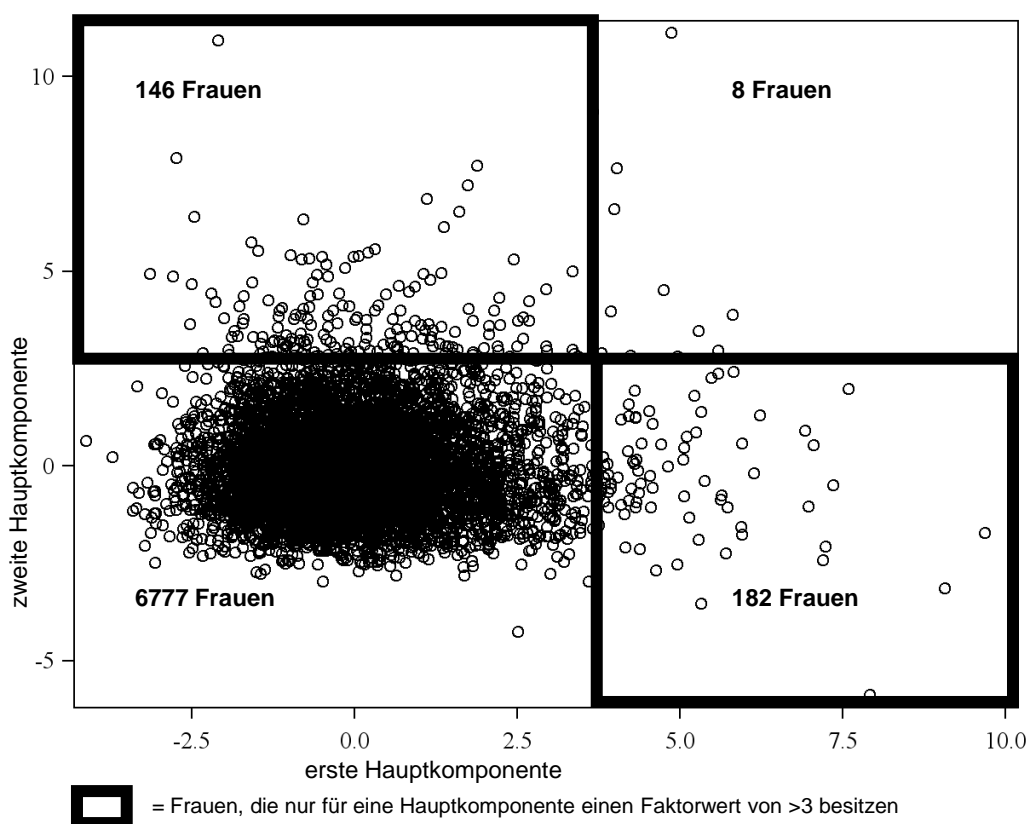
Für die Darstellung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse bedeutet dies, dass durch eine Hauptkomponenten alle Lebensmittelgruppen, die bei der Analyse berücksichtigt werden mit den dazugehörigen Verzehrsmengen, dargestellt werden können. Üblicherweise werden aber nur die charakteristischen Lebensmittelgruppen für weitere Auswertungen herangezogen. Zudem geht durch die Reduktion aller extrahierten Hauptkomponenten auf einige wenige aussagekräftige Hauptkomponenten ein Teil der Information über den Lebensmittelverzehr verloren.

### **Kriterium: Identifizierung von Personengruppen mit ähnlichem Lebensmittelverzehr**

Bei der Hauptkomponentenanalyse werden Linearkombinationen der Ausgangsdaten, die Hauptkomponenten, extrahiert (Schendera 2010 S Viff), durch die eine charakteristische Kombination an Lebensmittelgruppen dargestellt werden. Der Lebensmittelverzehr einer Person wird, wie durch Schulze und Hoffmann (2006 S 861) beschrieben, durch die Faktorwerte zu den Hauptkomponenten ausgedrückt. Durch die Faktorwerte kann jede Person jeder Hauptkomponente zugordnet werden (Hearty und Gibney 2009 S 598f, Newby und Tucker 2004 S 197). Je höher der Faktorwert zu einer Hauptkomponente, desto besser lässt sich der Lebensmittelverzehr einer Person durch eine Hauptkomponente darstellen. Allerdings ist es möglich, dass eine Person für mehrere Hauptkomponenten einen hohen Faktorwert besitzt (Newby und Tucker 2004 S 197, Hu 2002 S 5). Aus diesem Grund sollte, wie von Newby und Tucker (2004 S 197) beschreiben, die Zuordnung der Personen zu einzelnen Hauptkomponenten vorsichtig erfolgen. Weist eine Person für mehrere Hauptkomponenten einen hohen Faktorwert auf, lässt sich der Lebensmittelverzehr dieser Person nur durch die Kombination mehrerer Hauptkomponenten beschreiben (Newby und Tucker 2004 S 197). Eine spezifische Zuordnung zu einer einzigen Hauptkomponente und damit die Darstellung des Lebensmittelverzehr durch eine Hauptkomponente ist für diese Person nicht möglich.

Wie spezifisch die Studienteilnehmenden den extrahierten Hauptkomponenten zugeordnet wurden, und damit die Prüfung, ob Personen mit einem ähnlichen Lebensmittelverzehr identifiziert werden können, erfolgt durch die grafische Gegenüberstellung der ersten beiden Hauptkomponenten, welche die meiste Varianz des Lebensmittelverzehrs erklären. Zur Identifizierung von Personengruppen mit einem ähnlichen Lebensmittelverzehr sollte erkennbar sein, dass viele Personen entweder für die erste **oder** für die zweite Hauptkomponente, nicht aber für beide Hauptkomponenten einen hohen Faktorwert besitzen. Nur in diesem Fall ist eine spezifische Zuordnung zu einer der beiden Hauptkomponenten möglich. Beispielhaft wird dies an der Gruppe der Frauen (Abb. 4.2) dargestellt und erläutert. Die Abbildung der Männer befindet sich im Anhang (Kapitel 9.2).

Frauen, die für beide Hauptkomponenten einen Faktorwert um Null aufweisen (in Abb. 4.2 nicht markierter Bereich links unten), stellen mit 95 % den weitaus größten Teil der Gruppe dar. Ein Wert um Null bedeutet, dass sie weder der ersten noch der zweiten Hauptkomponente spezifisch zugeordnet werden können und damit der Lebensmittelverzehr dieser Personen weder über die erste noch über die zweite Hauptkomponente dargestellt werden kann. Entweder lässt sich der Lebensmittelverzehr dieser Frauen nur durch eine der anderen Hauptkomponente beschreiben oder durch die Kombination mehrerer anderer Hauptkomponenten.



**Abb. 4.2:** Zuordnung der Frauen zu den ersten beiden Hauptkomponenten auf Basis der Faktorwerte

Der Lebensmittelverzehr der Frauen, die entweder für die erste oder zweite Hauptkomponente einen hohen Faktorwert besitzen, weisen einen ähnlichen Lebensmittelverzehr auf, der sich über die entsprechende Hauptkomponente beschreiben lässt. Die grafische Gegenüberstellung der beiden Hauptkomponenten zeigt allerdings, dass es nur wenige Frauen gibt, die für eine der ersten beiden Hauptkomponenten einen hohen Faktorwert (>3) aufweisen (mit Rahmen markierter Bereich). Insgesamt können 182 von 7113 Frauen eindeutig der ersten und 146 Frauen der zweiten Hauptkomponente zugeordnet werden. Diese beiden Personengruppen besitzen einen, durch die jeweilige Hauptkomponente beschriebenen charakteristischen Lebensmittelverzehr (Tab. 4.14).

**Tab. 4.14: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen, die eine hohe Faktorladung für die erste oder zweite Hauptkomponente aufweisen**

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev) (n=7113)	Lebensmittelverzehr (MW±Std dev) (n=182)	Faktorladung	Interpretation der 1. Hauptkomponente
<b>Obst</b>	292±233	266±199		
<b>Gemüse</b>	251±150	300±156		
<b>Getreide</b>	239±99	<b>393±151</b>	<b>0,41</b>	hoher Getreideverzehr
<b>Kartoffeln</b>	72±43	<b>123±64</b>	<b>0,37</b>	hoher Kartoffelverzehr
<b>Nüsse und Samen</b>	4±11	7±18		
<b>Milch</b>	239±211	332±308		
<b>Eier</b>	12±13	<b>33±29</b>	<b>0,34</b>	hoher Eiverzehr
<b>Fisch</b>	23±22	29±27		
<b>Fleisch</b>	75±48	<b>167±94</b>	<b>0,46</b>	hoher Fleischverzehr
<b>Streichfett</b>	21±17	<b>56±34</b>	<b>0,44</b>	hoher Streichfettverzehr
<b>Getränke</b>	2335±889	2366±924		
<b>Alkohol<sup>1</sup></b>	6±9	7±11		
<b>Süßwaren</b>	61±56	<b>146±135</b>	<b>0,34</b>	hoher Süßwarenverzehr
Lebensmittelgruppen	Gesamtgruppe (MW±Std dev) (n=7113)	Lebensmittelverzehr (MW±Std dev) (n=182)	Faktorladung	Interpretation der 2. Hauptkomponente
<b>Obst</b>	292±233	<b>784±575</b>	<b>0,51</b>	hoher Obstverzehr
<b>Gemüse</b>	251±150	<b>603±284</b>	<b>0,54</b>	hoher Gemüseverzehr
<b>Getreide</b>	239±99	203±100		
<b>Kartoffeln</b>	72±43	65±52		
<b>Nüsse und Samen</b>	4±11	<b>18±35</b>	<b>0,20</b>	hoher Nuss- und Samenverzehr
<b>Milch</b>	239±211	<b>374±507</b>	<b>0,20</b>	hoher Milchverzehr
<b>Eier</b>	12±13	11±12		
<b>Fisch</b>	23±22	<b>63±61</b>	<b>0,40</b>	hoher Fischverzehr
<b>Fleisch</b>	75±48	60±54		
<b>Streichfett</b>	21±17	14±13		
<b>Getränke</b>	2335±889	<b>3374±1165</b>	<b>0,38</b>	hoher Getränkeverzehr
<b>Alkohol<sup>1</sup></b>	6±9	6±9		
<b>Süßwaren</b>	61±56	40±47		

<sup>1</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

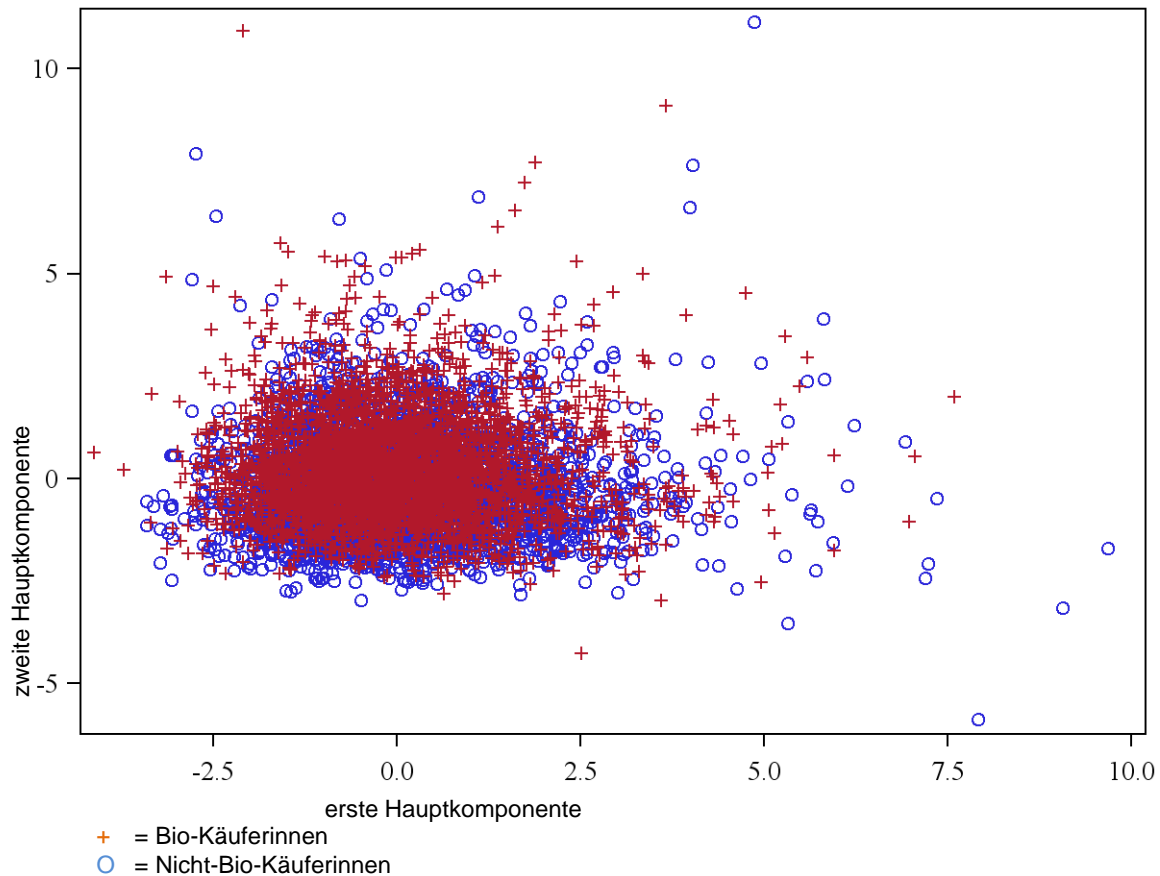
Die für die Hauptkomponente charakteristischen Lebensmittelgruppen sind fett gedruckt.

Frauen, die eindeutig der **ersten Hauptkomponente** zugeordnet wurden, verzehren mehr Getreide, Kartoffeln, Eier und weniger Fleisch, Streichfett sowie Süßwaren als die Gesamtgruppe der Frauen und im Vergleich zu den Frauen, die der vierten Quartile der ersten Hauptkomponente zugeordnet wurden (getestet über Mann-Whitney-U-Test, jeweils:  $p < 0,001$ ). Frauen, die eindeutig der **zweiten Hauptkomponente** zugeordnet wurden, verzehren mehr Obst, Gemüse, Nüsse und Samen, Fisch sowie Getränke als die Gesamtgruppe der Frauen und im Vergleich zu den Frauen, die der vierten Quartile der ersten Hauptkomponente zugeordnet wurden (getestet über Mann-Whitney-U-Test, jeweils:  $p < 0,01$ ). In der Verzehrsmenge von Milch konnten keine Unterschiede festgestellt werden.

Für das Kriterium, Personengruppen mit einem ähnlichen Lebensmittelverzehr zu identifizieren, bedeutet dies, dass ein Großteil der Frauen (und Männer: getestet, aber nicht dargestellt) nicht spezifisch einer der beiden aussagekräftigsten Hauptkomponenten zugeordnet werden kann. Damit ist es in dieser Analyse nicht möglich, Frauen und Männer mit einem ähnlichen Lebensmittelverzehr über die extrahierten Hauptkomponenten zu identifizieren.

#### **Kriterium: Differenzierung von Bio-Käufer/innen und Nicht-Bio-Käufer/innen**

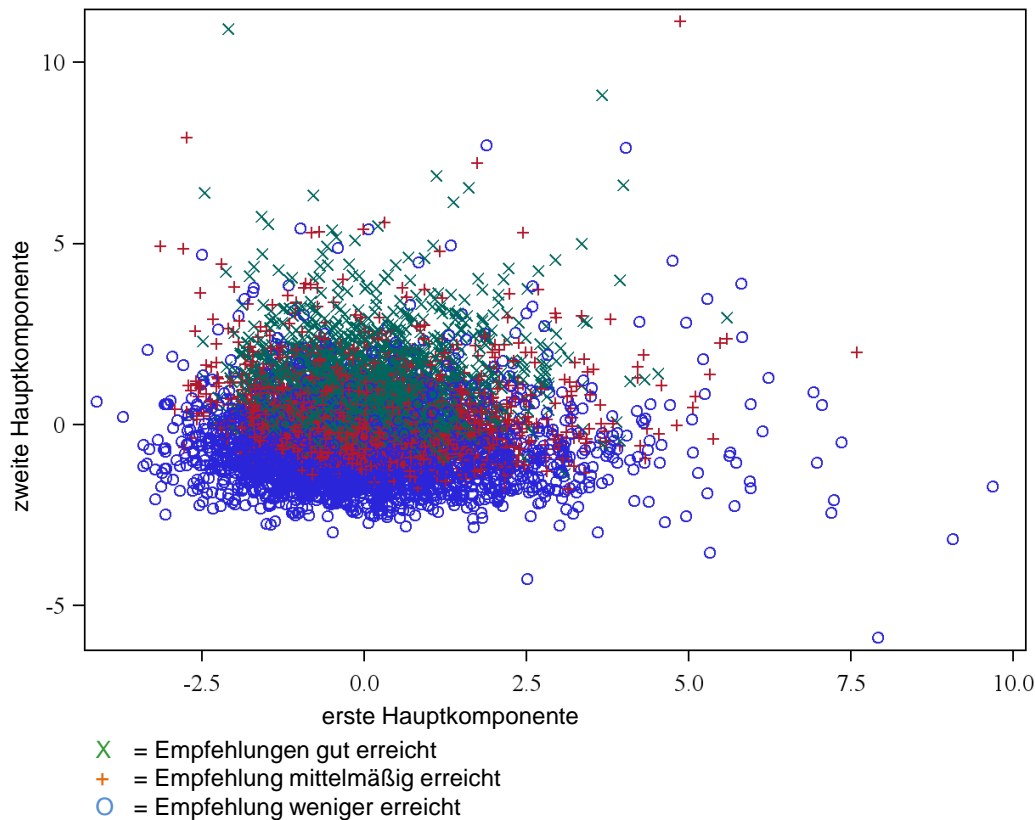
Um zu prüfen, ob eine Differenzierung der Bio-Käufer/innen und Nicht-Bio-Käufer/innen über die ersten beiden Hauptkomponenten möglich ist, werden diese beiden Personengruppen anhand ihres Faktorwertes grafisch gegenübergestellt (Abb. 4.3). Die Abbildung der Männer befindet sich im Anhang (Kapitel 9.2). Die Differenzierung von Personengruppen wie die der Bio-Käufer/innen und Nicht-Bio-Käufer/innen erfolgt damit indirekt nach Durchführung der Hauptkomponentenanalyse.



**Abb. 4.3: Differenzierung der Bio-Käuferinnen und Nicht-Bio-Käuferinnen über die ersten beiden Hauptkomponenten auf Basis der Faktorwerte**

Es zeigt sich, dass sich die Bio-Käuferinnen und Nicht-Bio-Käuferinnen weder der ersten noch der zweiten Hauptkomponente eindeutig zuordnen lassen. Damit ist die Hauptkomponentenanalyse zum Differenzieren der Bio-Käuferinnen und Nicht-Bio-Käuferinnen nicht geeignet.

Dass durch die Hauptkomponenten grundsätzlich unterschiedliche Personengruppen differenziert werden können, wird anhand der Ergebnisse der Indexsummen beispielhaft dargestellt. Hierfür wird die Aufteilung der Frauen anhand der HEI-NVS II-Indexsumme, in die Terzilen die die „Empfehlungen gut erreichen“, die „Empfehlungen mittelmäßig erreichen“ und die „Empfehlungen weniger erreichen“ (Vergleich Kapitel 3.1.4) herangezogen (Abb. 4.4).



**Abb. 4.4: Zuordnung der Frauen zu den ersten beiden Hauptkomponenten auf Basis der Faktorwerte, getrennt nach der HEI-NVS II-Indexsumme**

Frauen, die die „Empfehlungen gut erreichen“, erzielen für die zweite, gesundheitlich günstigere Hauptkomponente einen höheren Faktorwert als Frauen, die die „Empfehlungen mittelmäßig erreichen“ oder die „Empfehlungen wenig erreichen“. Frauen, welche die „Empfehlungen gut erreichen“, ernähren sich eher nach der zweite Hauptkomponente als Frauen, die die „Empfehlungen wenig erreichen“.

Für das Kriterium der Differenzierung von Personengruppen bedeutet dies, dass Bio-Käufer/innen und Nicht-Bio-Käufer/innen zwar nicht differenziert werden können, die Hauptkomponentenanalyse aber dennoch geeignet sein kann, Personengruppen zu differenzieren.

#### **Kriterium: Charakterisierung der identifizierten und differenzierten Personengruppen**

Wie von Newby und Tucker (2004 S 197) beschrieben, sind die Ergebnisse der Hauptkomponentenanalyse im Vergleich zu anderen Methoden prädestiniert, um weitere Analysen auf Basis der extrahierten Hauptkomponenten anzuschließen. Hierfür wird der Faktorwert, der für jede Person für jede Hauptkomponente berechnet wird, herangezogen. Dies wird insbesondere genutzt, um bei der Anwendung der Untersuchung von Ursache-Wirkungs-Zusammenhängen, eine Risikogruppe weiter zu charakterisieren (Beispiele: Edefonti et al. 2010, de Stefani et al. 2008a). Hierfür werden die Studienteilnehmenden üblicherweise auf

Basis ihrer Faktorwerte in gleich große Gruppen wie Quartilen aufgeteilt und die mit dem höchsten (gute Übereinstimmung zur Hauptkomponente) mit denen mit dem niedrigsten Faktorwert (weniger gute Übereinstimmung zur Hauptkomponente) verglichen.

In der vorliegenden Auswertung wurden die Studienteilnehmenden in Quartilen aufgeteilt und mit weiteren ernährungs- und gesundheitsrelevanten Variablen charakterisiert. Hierbei konnten keine eindeutigen Ergebnisse gewonnen werden (siehe inhaltliche Ergebnisse Kapitel 4.1.5). Auf Basis der deskriptiven Auswertungen (Kapitel 2.3.2) sowie der Index-Ergebnisse (Kapitel 3.1.4) wäre zu erwarten, dass Zusammenhänge zwischen einzelnen Hauptkomponenten und den zu beschreibenden Variablen beobachtet werden können. Beispielweise wäre zu erwarten gewesen, dass sich der jeweils ersten Hauptkomponente der Frauen und Männer, die sich unter anderem durch einen hohen Getreide-, Kartoffel-, Eier-, Fleisch- und Streichfettverzehr kennzeichnen, einen höheren Anteil an Personen aus der unteren sozialen Schicht, mehr Raucher und weniger sportlich Aktive vertreten sind. Zudem wurden bei den Frauen, die dem ersten Quartil der zweiten Hauptkomponente zugeordnet wurden, im Gegensatz zur graphischen Gegenüberstellung, signifikant weniger Bio-Käuferinnen als im vierten Quartil ermittelt (Tab. 4.8). Bei den Frauen des ersten und vierten Quartils der ersten Hauptkomponente konnten keine Unterschiede im Anteil der Bio-Käuferinnen festgestellt werden (Tab. 4.6).

Für das vorliegende Kriterium bedeutet dies, dass grundsätzlich identifizierte und differenzierte Personengruppen mit weiteren ernährungs- und gesundheitsrelevanten Variablen charakterisiert werden können. Für die vorliegende Auswertung ist dies allerdings nicht möglich, da die Kriterien des Identifizierens und Differenzierens von Personengruppen nicht erfüllt werden.

### **Fazit**

Es lässt sich festhalten, dass die Kriterien der umfassenden Darstellung des Lebensmittelverzehrs nur zum Teil erfüllt werden. Sowohl bei der Anwendung der Hauptkomponentenanalyse als auch bei der Darstellung der Ergebnisse können grundsätzlich alle 13 Lebensmittelgruppen und die verzehrten Lebensmittelmengen berücksichtigt und damit die ersten beiden Kriterien erfüllt werden. Allerdings werden aufgrund des Ziels der Methode, eine Datenreduktion vorzunehmen, üblicherweise nur die für die extrahierten Hauptkomponenten charakteristischen Lebensmittelgruppen dargestellt. In der vorliegenden Auswertung wurden nur 5 bis 8 der 13 Lebensmittelgruppen als charakteristisch für eine Hauptkomponente angesehen und dargestellt. Zudem wurden nur vier bzw. fünf der jeweils 13 extrahierten Hauptkomponenten der Frauen bzw. Männer als aussagekräftige Hauptkomponenten identifiziert. Durch diese Reduktion geht ein weiterer Teil der Information des Lebensmittelverzehrs verloren.

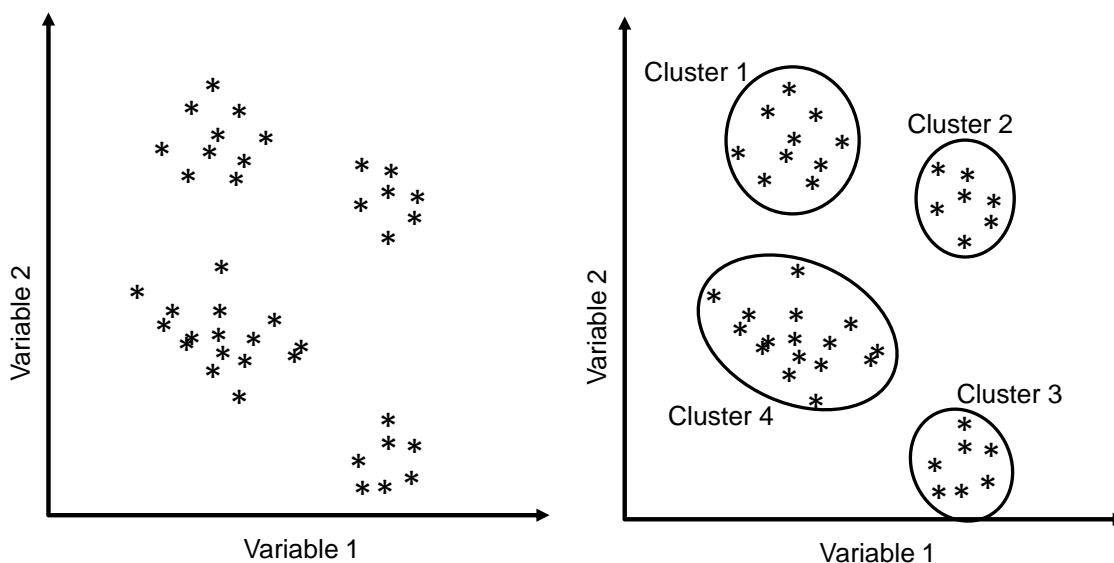
Die Ergebnisse der Hauptkomponentenanalyse sind prädestiniert dafür, weitere Analysen auf Basis der extrahierten Hauptkomponenten anzuschließen. Allerdings war dies in der vorliegenden Auswertung nur eingeschränkt möglich, da auch das Identifizieren und Differenzieren von Personengruppen bei der vorliegenden Studiengruppe nicht möglich war. Die Bio-Käufer/innen und Nicht-Bio-Käufer/innen lassen sich nicht über den Lebensmittelverzehr eindeutig einer extrahierten Hauptkomponenten zuordnen und damit auch nicht weiter charakterisieren.

Die Hauptkomponentenanalyse ist nicht geeignet, um eine umfassende Darstellung des Lebensmittelverzehrs anhand der Daten der NVS II vorzunehmen.

Bei dieser Methode stellt sich grundsätzlich die Frage, ob sie bei anderen Studiengruppen vielversprechendere Ergebnisse erzielen kann. Der geeignete Einsatz dieser Methode ist eher in der klassischen Anwendung der Untersuchung von Ursache-Wirkungs-Zusammenhängen zwischen den extrahierten Hauptkomponenten und beispielsweise Erkrankungen zu sehen.

## 4.2 Clusteranalyse

Clusteranalyse ist ein Sammelbegriff für verschiedene klassenbildende Verfahren. Die Aufteilung von Objekten, Fällen oder Variablen erfolgt auf Basis von Ähnlichkeiten oder Distanzen bezüglich der gemessenen Objekte, Fälle oder Variablen (Christof 2007 S 114). Ziel dabei ist, Gruppen, die sogenannten Cluster, zu identifizieren, die sich bezogen auf die berücksichtigten Merkmale innerhalb des Clusters möglichst ähnlich und zu weiteren Clustern möglichst unähnlich sind (Backhaus et al. 2008 S 391, Ester und Sander 2000 S 4). In Abb. 4.5 ist schematisch eine zu clusternden Datenmenge dargestellt. Auf Basis von zwei Variablen können in diesem idealen Beispiel vier Cluster identifiziert werden.



**Abb. 4.5: Schematische Darstellung einer Clusterbildung (eigene Darstellung)**

Verfahren der Clusteranalyse, die zur Darstellung der Ernährung eingesetzt werden, sind **hierarchische** und **nicht-hierarchische Verfahren** (Schendera 2010 S 9, Backhaus et al. 2008 S 392, Schendera 2004 S 540).

### Hierarchische Verfahren der Clusteranalyse

Bei hierarchischen Verfahren wird eine Vielzahl an Clusterlösungen (unterschiedliche Anzahl an Clustern) zur anschließenden Auswertung berechnet. Der Anwender muss sich nach Berechnung der Clusteranalyse anhand verschiedener Kriterien für eine Clusterlösung, also wie viele Cluster gebildet werden sollen, entscheiden (Backhaus et al. 2008 S 430). Bei dieser Entscheidung fließt eine gewisse Subjektivität mit ein, da die verschiedenen Auswahlkriterien weder eindeutig oder objektiv sind, noch zum gleichen Ergebnis führen (Schendera 2004 S 540). Die hierarchische Clusteranalyse kann verwendet werden, um Variablen (Datenspalten) oder Objekte (Fälle, Datenzeilen) zu clustern (Schendera 2010 S 10ff).

Bei den hierarchischen Verfahren wird zwischen agglomerativen und divisiven Verfahren unterschieden (Schendera 2010 S 23, Backhaus et al. 2008 S 412, Schendera 2004 S 540), wobei in der Regel agglomerative Verfahren angewendet werden (Christof 2007 S 116).

Bei **agglomerativen Verfahren** befinden sich zu Beginn alle Objekte in einem eigenen Cluster, die sukzessiv zusammengefügt werden, bis sich alle Objekte in einem einzigen gemeinsamen Cluster befinden (Schendera 2010 S 23, Backhaus et al. 2008 S 430). Das Zusammenfügen (Fusionieren) von jeweils zwei Objekten oder Clustern erfolgt auf Basis ihrer Distanzen oder Ähnlichkeiten zueinander (Schendera 2010 S 23, Backhaus et al. 2008 S 392f). Die daraus gebildete Distanz- oder Ähnlichkeitsmatrix stellt die Ausgangsdaten zur Fusionierung der Objekte dar. In der Regel wird bei metrischen Daten ein Distanzmaß und bei nominalen und ordinalen Daten ein Ähnlichkeitsmaß verwendet (Backhaus et al. 2008 S 411). Die Wahl, ob ein Ähnlichkeits- oder Distanzmaß verwendet wird, beeinflusst das Ergebnis der Clusteranalyse (Backhaus et al. 2008 S 392f, Christof 2007 S 115).

Beispiele für agglomerative Verfahren sind verschiedene Linkage-Verfahren sowie das Ward-Verfahren (Backhaus et al. 2008 S 416ff, Christof 2007 S 117).

Bei **Single-Linkage-Verfahren (Nearest-Neighbour-Verfahren)** erfolgt die Gruppierung auf Basis der geringsten Distanz zweier Objekte. Dieses Verfahren ist im Gegensatz zum Complete-Linkage Verfahren (siehe nächster Absatz) geeignet, Ausreißer zu entdecken. Jedoch neigt dieses Verfahren zur Kettenbildung, d. h. es werden schnell große Cluster gebildet, durch welche die Objekte nicht ausreichend differenziert werden können (Schendera 2010 S 25, Backhaus et al. 2008 S 416ff).

Bei dem **Verfahren Complete-Linkage (Furthest-Neighbour-Verfahren)** werden (im Gegensatz zum Single-Linkage Verfahren) zur Clusterbildung nicht die geringsten Abstände zweier Objekte, sondern die größten Distanzen zur Bildung herangezogen (Schendera 2010 S 25, Backhaus et al. 2008 S 419f).

Bei dem **Average-Linkage-Verfahren (Mittelwertverfahren)** werden die Distanzen zwischen Clustern durch ihre Zentroiddistanzen berechnet (Deichsel und Trampisch 1985 S 35). Das bedeutet, dass die Bildung eines Clusters durch die kleinsten Abstände auf Basis des Mittelwertes der Abstände aller Elemente (Fälle, Variablen, Cluster) zweier Cluster erfolgt (Schendera 2010 S 25).

Im **Ward-Verfahren** werden im Gegensatz zu den Linkage-Verfahren die Objekte nicht auf Basis von geringsten Distanzen zweier Objekte, sondern auf Basis eines Heterogenitätsmaßes vereint. Das bedeutet, es werden die Objekte, welche die Streuung (Varianz) in einem Cluster möglichst wenig erhöhen, zusammengefügt. Es entstehen möglichst homogene Cluster. Als Heterogenitätsmaß wird die quadrierte euklidische Distanz verwendet (Backhaus et al. 2008 S 420ff).

**Divisive Verfahren** arbeiten umgekehrt zu agglomerativen Verfahren. Zu Beginn gibt es ein einziges gemeinsames Cluster, in dem alle Objekte enthalten sind (Backhaus et al. 2008 S 412, Schendera 2004 S 540). Die Objekte werden durch die Clusteranalyse so lange aufgeteilt (partitioniert), bis sich jedes Objekt in jeweils einem eigenen Cluster befindet (Schendera 2010 S 23, 2004 S 540). Da divisiven Verfahren derzeit eine geringe Bedeutsamkeit beigemessen wird (Backhaus et al. 2008 S 412, Christof 2007 S 116), wird hier im Folgenden nicht weiter darauf eingegangen.

### **Nicht-Hierarchische Verfahren der Clusteranalyse**

Bei nicht-hierarchischen Clusteranalysen (partitionierenden Verfahren) wird von einer zuvor bekannten Gruppierung und einer bereits festgelegten Anzahl an Clustern ausgegangen, die zur Berechnung des Verfahrens angegeben werden müssen (Backhaus et al. 2008 S 412). Eine Möglichkeit, bei nicht-hierarchischen Verfahren die optimale Clusteranzahl zu finden, besteht darin, Testverfahren vorzulagern. Allerdings sind die hierbei zur Verfügung stehenden Kennzahlen nicht immer eindeutig, was bedeutet, dass die Objektivität bei der Wahl der optimalen Clusteranzahl nicht immer gegeben ist (Schendera 2004 S 540).

Bei nicht-hierarchischen Clusteranalysen werden die Objekte durch Algorithmen solange den Clustern zugeordnet, bis ein Optimum erreicht wird. Im Gegensatz zu hierarchischen Verfahren können während des Zusammenführens Objekte zwischen den Clustern getauscht werden (Backhaus et al. 2008 S 412). Die nicht-hierarchische Clusteranalyse kann verwendet werden, um Objekte (Fälle, Datenzeilen) nicht aber Variablen (Datenspalten) zu clustern (Schendera 2010 S 10ff).

Ein Beispiel für nicht-hierarchische Verfahren ist die **Clusterzentralanalyse** (K-Means). Bei der Clusterzentralanalyse wird zunächst eine vorgegebene Anzahl an Elementen (Fälle oder Objekte) ausgewählt. Auf Basis dieser Elemente werden Clusterzentren (Klassenmittelpunkte) berechnet. Anschließend werden die zu clusternden Elemente dem nächstgelegenen Klassenmittelpunkt zugeordnet und so die Cluster gebildet. Danach wird für jedes Cluster ein neuer Klassenmittelpunkt bestimmt. Die letzten beiden Schritte werden so lange durchgeführt, bis stabile Cluster gefunden werden. Dieses Verfahren ist im Gegensatz zur hierarchisch agglomerativen Analyse auch für große Datenmengen geeignet (Christof 2007 S 118).

#### **4.2.1 Wahl eines Verfahrens für die eigene Auswertung**

Für die Entscheidung, ob ein hierarchisches oder nicht-hierarchisches Verfahren zur Auswertung der Daten geeignet ist, muss zunächst geprüft werden, ob Variablen (Datenspalten) oder Objekte (Fälle, Datenzeilen) vorliegen. Während bei der Clusterung von Variablen

lediglich hierarchische Verfahren angewendet werden können, stehen für Objekte sowohl hierarchische als auch nicht-hierarchische Verfahren zur Verfügung.

Daneben muss geprüft werden, ob die Anzahl der zu bildenden Cluster zuvor bekannt ist oder abgeleitet werden muss. Während bei nicht-hierarchischen Verfahren wie der K-Means Clusterung die Anzahl der Cluster vor der Berechnung der Analyse festgelegt werden muss (Schendera 2004 S 540), werden bei den hierarchischen Verfahren die Anzahl der Cluster erst nach der Berechnung der Analyse festgelegt (Backhaus et al. 2008 S 430).

In der vorliegenden Auswertung sollen Personen als Objekte (Datenzeilen) auf Basis ihres Lebensmittelverzehrs geclustert werden. Damit können theoretisch sowohl hierarchische als auch nicht-hierarchische Verfahren für die vorliegende Auswertung angewendet werden. Allerdings ist die Anzahl der zu bildenden Cluster unbekannt, weswegen ein hierarchisches Verfahren für die Auswertung gewählt wird.

Zur Darstellung des Lebensmittelverzehrs wird aus den zur Verfügung stehenden hierarchischen Verfahren das Verfahren Ward mit einem euklidischen Distanzmaß gewählt. Bei diesem Verfahren werden die Cluster so gebildet, dass sukzessive die Objekte zusammengefasst werden, die den geringsten Zuwachs zu einem Heterogenitätsmaß liefern. Das bedeutet, dass die Objekte (hier: Frauen bzw. Männer) gruppiert werden, welche die Streuung (Varianz) (hier: der Verzehrsmengen) in einer Gruppe möglichst wenig erhöhen (Backhaus et al. 2008 S 420ff). Es werden die Personen gruppiert, deren Lebensmittelverzehr sich möglichst ähnlich ist.

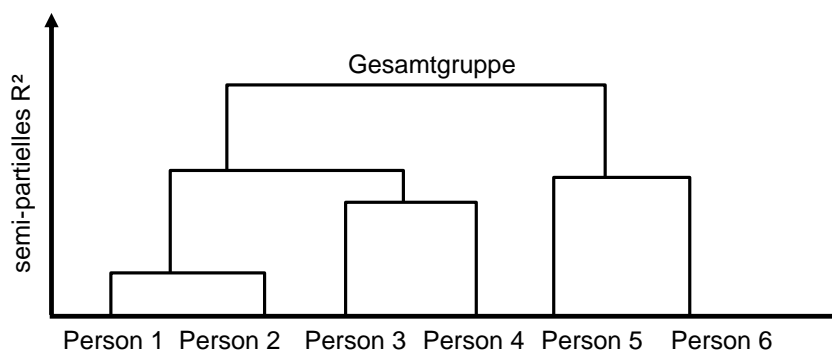
#### **4.2.2 Theoretischer Hintergrund**

Bei der Anwendung der Clusteranalyse werden Personen aufgrund ihres ähnlichen Lebensmittelverzehrs gruppiert. Aus einer heterogenen Personengruppe werden homogene Cluster gebildet, die Personen mit einem ähnlichen Lebensmittelverzehr gruppieren (Bamia et al. 2005 S 103, Hu 2002 S 4). Ergebnis sind die identifizierten Cluster, die in der Literatur als Ernährungsmuster beschrieben werden (Newby und Tucker 2004 S 117). Allen Clusteranalysen liegt die Annahme zugrunde, dass die zu clusternden Objekte in Gruppen aufteilbar sind. Nur wenn eine Intracluster-Homogenität (Ähnlichkeit der Elemente innerhalb eines Clusters) und Intercluster-Heterogenität (Unähnlichkeit der Elemente verschiedener Cluster) vorliegt, ist eine Clusterung sinnvoll (Schendera 2010 S 7f). Es wird empfohlen, mittels Streudiagramm der Variablen zu prüfen, ob überhaupt Zusammenhänge zwischen den Variablen vorliegen (Schendera 2010 S 2f). Ist zu Beginn nicht bekannt, ob tatsächlich Cluster gebildet werden können, muss dies abschließend nach der Berechnung der Clusteranalyse geprüft werden. Es soll ausgeschlossen werden, dass die gebildeten Cluster zufällig zustande gekommen sind (Schendera 2010 S 7f). Beispielsweise kann die inhaltliche Interpretierbarkeit der gewonnenen Cluster geprüft werden. Idealerweise sollten die

Ergebnisse aus der Theorie ableitbar sein und durch die Clusteranalyse bestätigt werden (Schendera 2010 S 18).

### Bestimmung der optimalen Clusterzahl

Das Verfahren Ward setzt im Gegensatz zum Verfahren K-Means nicht voraus, zu Beginn der Analyse die Clusteranzahl bereits zu kennen. Dennoch werden durch die Anwendung des Verfahrens Ward mehrere Clusterlösungen angeboten. Zur Bestimmung der optimalen Clusteranzahl, werden die Kennzahlen CCC (Cubic Clustering Criterion), *pseudo F* (PSF) sowie das *pseudo t<sup>2</sup>* (PST2) empfohlen (SAS 2009a S 1235, Schendera 2004 S 567). Diese Kennzahlen geben Auskunft über die Heterogenität oder Homogenität der Cluster. Daneben steht das Dendrogramm zur Verfügung (Schendera 2010 S 59ff, SAS 2009a S 41235f), eine baumartige Grafik der erzeugten Cluster (Christof 2007 S 118), bei welchem das semi-partielle  $R^2$  dargestellt wird. Das semi-partielle  $R^2$  liefert Auskunft über den Varianzanteil zweier zusammengefügter Objekte oder Cluster (SAS 2009a S 41235f) (Abb. 4.6).



**Abb. 4.6: Beispiel für ein Dendrogramm zur Bestimmung der optimalen Clusteranzahl (eigene Darstellung)**

Das schrittweise Zusammenfassen der Objekte zu den Clustern kann von unten nach oben gelesen werden. Das Dendrogramm erlaubt, die optimale Clusteranzahl anhand der relativen (da standardisierten) Heterogenität zwischen den Clustern zu vergleichen. Je länger die Linien im Dendrogramm desto größer die Zunahme der Heterogenität, womit das Ausmaß der zunehmenden Unähnlichkeit oder abnehmenden Ähnlichkeit wiedergegeben wird. Rückschlüsse auf eine optimale Clusteranzahl sind damit möglich (Schendera 2010 S 59ff). So ist in Abb. 4.6 ersichtlich, dass die Zunahme der Heterogenität durch das Gruppieren von Person 3 und 4 höher ist, als die Zunahme der Heterogenität durch das Gruppieren von Person 1 und 2.

Neben den oben genannten Kennzahlen können der *F-Wert* und *T-Wert* herangezogen werden. Durch den F-Wert wird die Varianz jeder einzelnen Variablen innerhalb eines Clusters im Verhältnis zur Varianz der ungruppierten Ausgangsdaten berechnet (Schendera 2010 S 65f). Je kleiner der Wert, desto homogener ist die Streuung innerhalb eines Clusters

im Vergleich zu den ungeclusterten Ausgangsdaten. Die Ausgangsdaten besitzen einen F-Wert von eins. Idealerweise liegen möglichst viele F-Werte innerhalb eines Clusters unter eins vor. Liegen alle Variablen eines Clusters mit einem F-Wert unter eins, so handelt es sich um ein homogenes Cluster (Schendera 2010 S 65f).

Durch den T-Wert wird der Mittelwert (MW) der Variablen im Cluster mit dem Mittelwert der Variablen aus den Ausgangsdaten subtrahiert und mit der Standardabweichung (Std dev) der einzelnen Variablen innerhalb eines Clusters verglichen (Schendera 2010 S 65f). Ein positiver T-Wert einer Variablen lässt auf einen Wert über dem Durchschnitt und ein negativer Wert auf einen Wert unter dem Durchschnitt der Gesamtgruppe schließen (Schendera 2010 S 65f).

Bei der Wahl der geeigneten Clusteranzahl handelt es sich, wie von Backhaus et al. (2008 S 393f, 430) beschrieben, um einen Zielkonflikt zwischen Handhabbarkeit (geringe Clusterzahl) und Homogenitätsanforderung (große Clusterzahl).

### **4.2.3 Anwendung der Clusteranalyse im Bereich Ernährung**

Über die Clusteranalyse wird die ursprüngliche Gesamtgruppe auf Basis ihrer Nährstoffzufuhr- oder ihres Lebensmittelverzehrs in mehrere Untergruppen, den sogenannten Clustern, aufgegliedert. Die den Clustern zugeordneten Objekte können anhand aller in der Clusterung einbezogenen Variablen beschrieben werden. Auch Variablen, welche bei der Clusteranalyse keine Berücksichtigung fanden, können zur anschließenden Charakterisierung der in den Clustern eingeordneten Objekte herangezogen werden.

Clusteranalysen zur Darstellung der Ernährung werden seit den 1980er Jahren angewendet. Beispielsweise haben Bodenstedt et al. (1983) basierend auf den Daten der EMSIG-Studie (Ernährungsmodellstudie in Gießen), die der Untersuchung des Ernährungsverhaltens diente, eine Clusteranalyse zur Identifizierung sogenannter "Nahrungsmustern" angewendet. In Abhängigkeit vom Untersuchungsziel können Clusteranalysen im Bereich Ernährung in vier Kategorien eingeteilt werden (siehe Kapitel 3.1.1). In die Clusteranalyse können (1) **Nährstoffe**, (2) **Lebensmittel und Lebensmittelgruppen**, (3) **Nährstoffe, Lebensmittel und Lebensmittelgruppen** sowie (4) **Nährstoffe, Lebensmittel und Lebensmittelgruppen kombiniert mit weiteren nährstoff- und lebensmittelunabhängigen Variablen** beispielsweise in Bezug auf den Lebensstil einbezogen werden.

Nach Durchführung der Clusteranalyse können Ursache-Wirkungs-Zusammenhänge zwischen der Ernährung und Bioparametern und Krankheiten vorgenommen sowie Personengruppen mit einer ähnlichen Ernährung weiter charakterisiert werden. Im Folgenden werden die vier Kategorien der Clusteranalyse beschrieben und Beispiele für die Untersuchung von Zusammenhängen sowie Charakterisierung von Personengruppen mit weiteren Variablen genannt.

- (1) Ein Beispiel für die Anwendung einer **nährstoffbasierten Clusteranalyse** ist die Längsschnittstudie von Pedone et al. (2011), in welcher der Zusammenhang zwischen der Nährstoffzufuhr und der Knochendichte untersucht wurde. In die Clusteranalyse wurden Nährstoffe, die im Zusammenhang mit der Gesamtqualität der Ernährung oder dem Knochenstoffwechsel stehen, wie Calcium, Phosphor, Vitamin D und Magnesium, aufgenommen. Anschließend wurden die den gebildeten Clustern zugeordneten Personen mit Variablen wie dem Alter, Geschlecht und Raucherstatus charakterisiert (Pedone et al. 2011).
- (2) **Lebensmittel- und lebensmittelgruppenbasierte Analysen** werden eingesetzt, um z. B. Zusammenhänge zwischen dem Lebensmittelverzehr und Krankheiten zu untersuchen (Beispiele: Hartmann et al. 2012, Nyholm et al. 2012, Anderson et al. 2011, de Kort et al. 2011, Greenwood et al. 2000). Seltener findet die Clusteranalyse ohne Untersuchungen von Ursache-Wirkungs-Zusammenhängen statt. Beispielsweise wird die Clusteranalyse verwendet, um den Lebensmittelverzehr von Schwangeren darzustellen (Beispiel: McGowan und McAuliffe 2012, Okubo et al. 2011). Anschließend werden die den gebildeten Clustern zugeordneten Personen häufig mit weiteren Variablen verknüpft, um die jeweiligen Personengruppen weiter zu charakterisieren oder um die untersuchten Zusammenhänge weiter zu analysieren. Zu den Variablen, die zur Charakterisierung verwendet werden, gehören der sozioökonomische Status (Beispiele: Nyholm et al. 2012, de Kort et al. 2011), Bildung (Beispiele: Hartmann et al. 2012, Anderson et al. 2011), Alter oder Geschlecht (Beispiele: Hartmann et al. 2012, Nyholm et al. 2012, Anderson et al. 2011) sowie gesundheitsrelevante Variablen wie Raucherstatus (Beispiele: Nyholm et al. 2012, de Kort et al. 2011), BMI und sportliche Aktivität (Beispiele: Hartmann et al. 2012, Anderson et al. 2011).
- (3) **Nährstoff- und gleichzeitig lebensmittel/lebensmittelgruppenbasierte Analysen** werden eingesetzt, um unter anderem den Zusammenhang zwischen der Ernährung und dem Gesundheits- bzw. Krankheitszustand (Beispiele: Rodriguez-Ramirez et al. 2011, Millen et al. 2005) sowie dem Risiko, an verschiedenen Krankheiten zu erkranken (Beispiel: Millen et al. 2005), zu untersuchen. Daneben werden diese Analysen verwendet, um die Ernährung als solche, beispielsweise von Kindern und ihren Eltern zu untersuchen (Beispiel: Ovaskainen et al. 2009). Ebenso wie die Ergebnisse anderer Clusteranalysen werden auch die hier geclusterten Personen häufig mit weiteren Variablen wie dem Alter oder Geschlecht (Beispiele: Rodriguez-Ramirez et al. 2011, Millen et al. 2005) sowie gesundheitsrelevanten Variablen wie Raucherstatus, BMI und sportlicher Aktivität (Beispiel: Millen et al. 2005) charakterisiert.

- (4) Bei einer Clusteranalyse auf Basis von **Nährstoffen, Lebensmittel/Lebensmittelgruppen kombiniert mit weiteren nährstoff- und lebensmittelunabhängigen Variablen** werden im Gegensatz zu den letzten drei Varianten nicht nur Variablen aus dem Bereich Ernährung, sondern auch Variablen aus anderen Bereichen direkt in die Clusteranalyse einbezogen. So wurden z. B. von Patterson et al. (1994b) neben einem Ernährungsindex die Variablen körperliche Aktivität und Rauchverhalten in die Clusteranalyse einbezogen, um gesunde Lebensstilmuster zu identifizieren.

Die Anzahl der in den genannten Beispielen identifizierten Cluster variiert von zwei bis sieben.

Einen weiteren Überblick über die Anwendung von Clusteranalysen wird durch eine Übersichtsarbeit von Devlin et al. (2012) gegeben. Knapp 30 Veröffentlichungen sind in dieser Arbeit beschrieben. Daneben haben Newby und Tucker (2004) Studien, bei welchen die Cluster- oder Faktorenanalyse zur Bestimmung von Ernährungsmustern angewendet wurde, zusammengetragen. Insgesamt konnten ab dem Jahr 1980 bis zum Zeitpunkt der Publikation im Jahr 2004 35 Veröffentlichungen mit der Anwendung von Clusteranalysen identifiziert werden. In den genannten Veröffentlichungen konnten zahlreiche Zusammenhänge zwischen der Ernährung und weiteren Variablen wie Krankheiten oder Biomarkern gefunden werden. Der Einsatz der Clusteranalyse zur Darstellung der Ernährung ohne den Ursache-Wirkungs-Zusammenhang zwischen der Ernährung und weiteren Variablen wie Bioparametern oder Erkrankungen zu untersuchen, fand bisher kaum Anwendung (Beispiele: McGowan und McAuliffe 2012, Okubo et al. 2011, Ovaskainen et al. 2009) und wird in diesen Veröffentlichungen nicht thematisiert.

#### **4.2.4 Anwendung der Clusteranalyse auf den Lebensmittelverzehr der Teilnehmenden der NVS II**

Der Lebensmittelverzehr von Frauen und Männern wird mit der Clusteranalyse mit Hilfe des hierarchischen Verfahrens Ward dargestellt. Wie bereits bei der Hauptkomponentenanalyse beschrieben (Kapitel 4.1.4), können unabhängig von Empfehlungen die 13 Lebensmittelgruppen Obst, Gemüse, Getreide, Kartoffeln, Milch, Fisch, Fleisch, Eier, Alkohol, Streichfett, alkoholfreie Getränke, Süßwaren sowie Nüsse und Samen in die Auswertung einbezogen werden (Gruppierung der Lebensmittelgruppen siehe Anhang Tab. 9.1).

Es wird davon ausgegangen, dass durch eine Clusterung sowohl bei den Frauen als auch bei den Männern auf Basis der verzehrten Lebensmittelmengen Cluster, und damit Personen, die sich in ihren verzehrten Lebensmittelmengen ähneln, identifiziert werden können. Eine weitere Annahme ist, dass sich in diesen identifizierten Clustern der Anteil der Bio-Käufer/innen und Nicht-Bio-Käufer/innen unterscheidet, und zwar dass Bio-Käufer/innen zu einem höheren Anteil in gesundheitlich eher günstigen und Nicht-Bio-Käufer/innen zu

einem höheren Anteil in gesundheitlich eher ungünstigen Clustern wiederzufinden sind. Die letztere Annahme stützt sich auf die bisherigen deskriptiven Ergebnisse (Kapitel 2.3.2), die gezeigt haben, dass sich sowohl bei den Frauen als auch bei den Männern Bio-Käufer/innen und Nicht-Bio-Käufer/innen deutlich in ihren verzehrten Lebensmittelmengen voneinander unterscheiden. Folglich ist davon auszugehen, dass sich sowohl die Frauen als auch die Männer aufgrund ihres unterschiedlichen Lebensmittelverzehr gruppiert lassen.

Vor Beginn der Analyse wird von Schendera (2010 S 2f) empfohlen, nicht nur Hypothesen über die Ergebnisse aus der Theorie abzuleiten, sondern auch die Eignung der Daten zu prüfen, also ob überhaupt Zusammenhänge zwischen den Variablen vorliegen. Hierfür werden beispielhaft die Variablen Obst, Gemüse und Fleisch zunächst mittels Streudiagramm visualisiert. Die Verzehrsmengen der genannten Lebensmittelgruppen werden paarweise in einer zweidimensionalen Grafik gegenübergestellt. Zur Bestimmung der geeigneten Clusteranzahl werden das Dendrogramm, die Kennzahlen CCC, pseudo F, pseudo T sowie die F-Werte herangezogen.

Die Personen der einzelnen Cluster werden anschließend durch die in der Clusterung berücksichtigten Variablen beschrieben. Es wird verglichen, wie sich die durchschnittlich verzehrten Lebensmittelmengen der Personen, die den Clustern zugeordnet wurden, unterscheiden. Anschließend werden die Personen der einzelnen Cluster anhand weiterer Variablen charakterisiert. Zum einen wird untersucht, in welchem Cluster sich die Bio-Käufer/innen und Nicht-Bio-Käufer/innen wiederfinden, zum anderen werden weitere Variablen wie Alter, soziale Schicht sowie gesundheitsrelevante Variablen wie BMI, sportliche Aktivität sowie Raucherstatus zur weiteren Beschreibung der Personengruppen herangezogen.

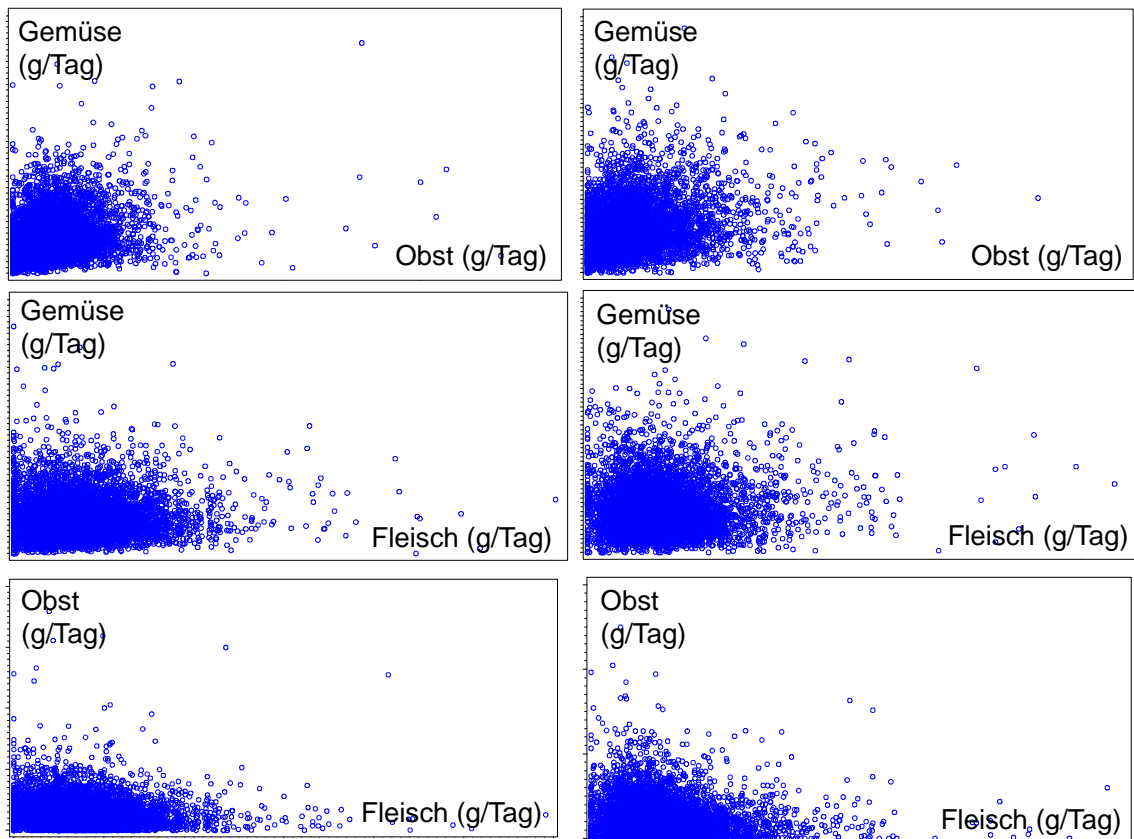
#### **4.2.5 Inhaltliche Ergebnisse**

Zunächst werden die inhaltlichen Ergebnisse der Clusteranalyse präsentiert, um darauf aufbauend zu untersuchen, ob sich die Clusteranalyse zur umfassenden Darstellung der Ernährung eignet, was im nächsten Kapitel anhand der hierfür definierten Kriterien vorgenommen wird.

Zu Beginn werden die Verzehrsmengen einzelner Lebensmittelgruppen grafisch gegenübergestellt, um einen ersten Hinweis über die Zusammenhänge der Variablen zu erhalten. Anschließend werden die inhaltlichen Ergebnisse der Clusteranalyse, getrennt nach Geschlecht, dargestellt. Die den identifizierten Clustern zugeordneten Personen werden im nächsten Schritt mit weiteren Variablen, die nicht in die Analyse einbezogen wurden, charakterisiert.

### Grafische Gegenüberstellung der Lebensmittelgruppen

Es ist nicht möglich, alle 13 Lebensmittelgruppen gemeinsam in einer Grafik darzustellen. Beispielhaft für die zu clusternden Verzehrdaten werden die verzehrten Mengen Obst, Gemüse und Fleisch jeweils gegenübergestellt (Abb. 4.7). Bei diesen drei Lebensmittelgruppen wurden in den bisherigen Ergebnissen (Kapitel 2.3.2 und 3.1.4) die größten Unterschiede in den verzehrten Lebensmittelmengen zwischen Bio-Käufer/innen und Nicht-Bio-Käufer/innen festgestellt.



**Abb. 4.7: Obst-, Gemüse- und Fleischverzehrsmenge (g/Tag) aller Frauen (links) und Männer (rechts) zur Prüfung der Eignung der Daten zur Berechnung einer Clusteranalyse**

Im Gegensatz zur Darstellung der Clusterbildung in Abb. 4.5 lässt Abb. 4.7 auf keine trennbaren Gruppen schließen. Es ist eine dichte Punktwolke zu erkennen. Dieses Ergebnis kann auch nicht durch das weitere Hineinzoomen in die Grafik oder Darstellung einer Teilstichprobe (getestet, aber nicht dargestellt) verbessert werden. Ähnliche Ergebnisse ergaben sich durch die Gegenüberstellung der weiteren Lebensmittelgruppen (getestet, aber nicht dargestellt). In allen Grafiken ist erkennbar, dass es eine große Anzahl an Personen gibt, die einen ähnlichen Lebensmittelverzehr der dargestellten Lebensmittelgruppen aufweisen. Auf Basis der Streudiagramme ist, im Gegensatz zur Annahme basierend auf den Ergebnissen der deskriptiven Auswertung, eine sinnvolle Clusterung noch fraglich. Allerdings wurden in den Streudiagrammen lediglich drei der 13 Lebensmittelgruppen dargestellt.

Möglicherweise lässt sich eine Gruppierung nur auf Basis aller 13 Lebensmittelgruppen erkennen, für die eine gleichzeitige, d. h. 13-dimensionale vorherige visuelle Prüfung nötig wäre. Gewissheit soll die Durchführung und die anschließende inhaltliche Prüfung der identifizierten Cluster liefern.

### Cluster der Frauen

In der Gruppe der Frauen wird ein CCC-Wert über zwei, der auf eine gute Clusterzahl hinweist (SAS 2009a S 1235), bei 76 Clustern erreicht. Der pseudo F-Wert, der möglichst hoch sein sollte (Schendera 2004 S 559), weist auf zwei oder drei Cluster und der pseudo T-Wert, der möglichst niedrig sein sollte (Schendera 2004 S 559), auf drei oder sieben Cluster hin. Das Dendrogramm sowie der F-Wert weisen auf zwei Cluster hin. Im Folgenden wird die zwei-Clusterlösung präsentiert, die sich durch die Ergebnisse des Dendogramms, des pseudo T-Wertes sowie des F-Wertes begründen lassen (siehe Anhang Tab. 9.12).

Durch die Clusteranalyse wird das Ziel verfolgt, aus einer heterogenen Personengruppe homogene Cluster, die Personen mit einem ähnlichen Lebensmittelverzehr gruppieren, zu bilden (Bamia et al. 2005 S 103, Hu 2002 S 4). Die Darstellung des Lebensmittelverzehrs erfolgt durch die Beschreibung der durchschnittlichen Verzehrsmengen aller einbezogenen Lebensmittelgruppen der in einem Cluster gruppierten Personen (Tab. 4.15).

**Tab. 4.15: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen des ersten und zweiten Clusters**

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev)	Cluster (MW±Std dev)		Signifikanzen <sup>1</sup>
		1. Cluster (n=3287)	2. Cluster (n=3826)	
Obst	292±233	255±176	<b>323±268</b>	<0,001
Gemüse	251±150	246±137	255±161	n.s.
Getreide	239±99	<b>280±102</b>	203±82	<0,001
Kartoffeln	72±43	<b>77±44</b>	68±42	<0,001
Nüsse und Samen	4±11	2±5	<b>5±14</b>	<0,001
Milch	239±211	215±155	<b>259±247</b>	<0,001
Eier	12±13	<b>15±15</b>	10±9	<0,001
Fisch	23±22	20±16	<b>25±26</b>	<0,001
Fleisch	75±48	<b>87±54</b>	64±40	<0,001
Streichfett	21±17	<b>27±20</b>	15±11	<0,001
Getränke	2335±889	2288±760	2376±985	n.s.
Alkohol <sup>2</sup>	6±9	4±5	<b>7±10</b>	<0,001
Süßwaren	61±56	<b>78±68</b>	46±37	<0,001

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen dem 1. und 2. Cluster vorliegt (getestet über den Rangsummentest Mann-Whitney-U-Test).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

Fettgedruckt sind die jeweils signifikant höchsten Verzehrsmengen einer Lebensmittelgruppe.

Dem ersten Cluster konnten 46,2 % und dem zweiten Cluster 53,8 % der Frauen zugeordnet werden. Frauen aus dem **ersten Cluster** verzehren im Vergleich zu Frauen aus dem **zweiten Cluster** weniger Obst, Nüsse und Samen, Milch, Fisch sowie Alkohol und mehr

Getreide, Kartoffeln, Eier, Fleisch, Streichfett und Süßwaren. Bei den Lebensmittelgruppen Gemüse und Getränke konnten keine Unterschiede festgestellt werden.

Ergänzend zu den beschriebenen Ergebnissen wird geprüft, ob sich die Bio-Käuferinnen und Nicht-Bio-Käuferinnen den beiden Clustern zuordnen lassen und wie sich die Personen, die den einzelnen Clustern zugeordnet wurden, weiter charakterisieren lassen (Tab. 4.16).

**Tab. 4.16: Charakterisierung der Frauen der Gesamtgruppe sowie der Frauen des ersten und zweiten Clusters**

		Gesamtgruppe (%)	1. Cluster (n=3287)	2. Cluster (n=3826)	Signifikanzen <sup>1</sup>
<b>Bio-Kauf</b>	ja	50,3	49,4	51,1	n.s.
	nein	49,7	50,6	49,0	n.s.
<b>Altersgruppe</b>	18–34 Jahre	20,7	22,9	18,8	<0,001
	35–50 Jahre	35,3	38,0	32,9	<0,001
	51–64 Jahre	24,2	21,5	26,5	<0,001
	65–80 Jahre	19,8	17,6	21,8	<0,001
<b>Soziale Schicht<sup>3</sup></b>	untere Schicht	21,0	21,5	20,6	n.s.
	mittlere Schicht	29,9	29,8	30,0	n.s.
	obere Schicht	49,1	48,7	49,4	n.s.
<b>BMI</b>	Untergewicht	1,6	1,9	1,4	n.s.
	Normalgewicht	46,0	47,6	44,5	<0,01
	Übergewicht	27,9	26,4	29,3	<0,01
	Adipositas	18,9	18,9	18,8	n.s.
	keine Angabe	5,6	5,1	6,0	... <sup>2</sup>
<b>Tailenumfang</b>	bis 88 cm	65,8	65,9	65,7	n.s.
	über 88 cm	29,4	29,4	29,4	n.s.
	keine Angabe	4,8	4,7	4,9	... <sup>2</sup>
<b>WHR (Taillen- zu Hüft-umfang)</b>	bis 0,85	74,2	75,3	73,2	<0,05
	über 0,85	21,0	20,0	21,9	<0,05
	keine Angabe	4,8	4,7	4,9	... <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	1968±589	2158±574	1804±552	<0,001
<b>Ernährungskennntnisse</b>	gut	40,9	40,7	41,0	n.s.
	mittel	33,9	34,3	33,5	n.s.
	schlecht	24,2	24,2	24,3	n.s.
	keine Angabe	1,0	0,8	1,2	... <sup>2</sup>
<b>Sportlich aktiv</b>	ja	58,9	56,2	61,2	<0,001
	nein	40,6	43,4	38,1	<0,001
	keine Angabe	0,6	0,5	0,7	... <sup>2</sup>
<b>Raucherstatus</b>	ja	21,7	21,5	21,9	n.s.
	nein	78,0	78,3	77,8	n.s.
	keine Angabe	0,3	0,2	0,4	... <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	76,2	76,8	75,6	n.s.
	mittel	20,2	19,9	20,5	n.s.
	schlecht	3,5	3,2	3,8	n.s.
	keine Angabe	0,1	0,2	0,1	... <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	78,8	78,8	78,9	n.s.
	schlecht	20,7	20,9	20,6	n.s.
	keine Angabe	0,5	0,4	0,6	... <sup>2</sup>
<b>HEI-NVS II</b>	MW±Std dev	84,8±9,8	84,7±9,9	84,9±9,8	n.s.

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen dem 1. und 2. Cluster vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und HEI-NVS II über den Rangsummentest Mann-Whitney-U-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

Der Anteil der Bio-Käuferinnen und Nicht-Bio-Käuferinnen unterscheidet sich nicht in den beiden Clustern. Frauen aus dem ersten Cluster gehören zu einem höheren Anteil der Altersgruppe der 18- bis 34- und 35- bis 50-Jährigen und zu einem geringeren Anteil der Altersgruppe der 51- bis 64-Jährigen und 65- bis 80-Jährigen an. Bei der Schichtzugehörigkeit konnten keine Unterschiede zwischen den beiden Clustern festgestellt werden. Frauen aus dem ersten Cluster sind zu einem höheren Anteil normalgewichtig, zu einem geringeren Anteil übergewichtig und zu einem höheren Anteil in der Gruppe mit einem wünschenswerten WHR bis 0,85 vertreten. Im Taillenumfang gibt es keine Unterschiede zwischen den beiden Clustern. Allerdings weisen Frauen des ersten Clusters eine höhere Energiezufuhr auf als Frauen aus dem zweiten Cluster. Daneben lassen sie sich durch einen geringeren Anteil an sportlich aktiven Frauen kennzeichnen. Unterschiede in den Ernährungskennntnissen, dem Raucherstatus, der Selbsteinschätzung des allgemeinen Gesundheitszustands und der Schlafqualität sowie in der HEI-NVS II-Indexsumme konnten nicht festgestellt werden.

### **Cluster der Männer**

In der Gruppe der Männer wird ein CCC-Wert über zwei bei 79 Clustern erreicht. Der pseudo F-Wert sowie das Dendrogramm weisen auf drei Cluster und der pseudo T-Wert auf drei, sieben oder zehn Cluster hin. Darauf beziehend, werden für die Männer drei Cluster präsentiert (siehe Anhang Tab. 9.13).

Ebenso wie bei den Frauen erfolgt die Darstellung des Lebensmittelverzehrs durch die Gruppierung der Männer, die einen ähnlichen Lebensmittelverzehr aufweisen (Tab. 4.17).

**Tab. 4.17: Lebensmittelverzehr der Gesamtgruppe sowie der Männer des ersten, zweiten und dritten Clusters**

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev)	Cluster (MW±Std dev)			Signifikanzen <sup>1</sup>
		1. Cluster (n=5364)	2. Cluster (n=502)	3. Cluster (n=95)	
Obst	249±223	250±225	221±161	<b>357±329</b>	1/2 <sup>n.s.</sup> ; 1/3 <sup>***</sup> ; 2/3 <sup>***</sup>
Gemüse	239±143	238±142	242±154	<b>276±161</b>	1/2 <sup>n.s.</sup> ; 1/3 <sup>*</sup> ; 2/3 <sup>*</sup>
Getreide	303±127	<b>292±120</b>	<b>409±149</b>	340±127	1/2 <sup>***</sup> ; 1/3 <sup>***</sup> ; 2/3 <sup>***</sup>
Kartoffeln	92±57	90±55	<b>125±67</b>	<b>80±57</b>	1/2 <sup>***</sup> ; 1/3 <sup>*</sup> ; 2/3 <sup>***</sup>
Nüsse und Samen	5±13	3±8	4±7	<b>82±33</b>	1/2 <sup>n.s.</sup> ; 1/3 <sup>***</sup> ; 2/3 <sup>***</sup>
Milch	249±258	<b>248±261</b>	255±235	278±218	1/2 <sup>**</sup> ; 1/3 <sup>*</sup> ; 2/3 <sup>n.s.</sup>
Eier	16±18	16±18	<b>23±21</b>	18±18	1/2 <sup>***</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>*</sup>
Fisch	29±29	29±29	29±26	29±23	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
Fleisch	133±83	126±71	<b>216±139</b>	121±78	1/2 <sup>***</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>***</sup>
Streichfett	30±27	26±19	<b>74±45</b>	26±26	1/2 <sup>***</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>***</sup>
Getränke	2332±981	2310±971	2545±1046	2409±1065	1/2 <sup>***</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
Alkohol <sup>2</sup>	16±18	16±19	15±17	17±22	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
Süßwaren	68±63	<b>61±50</b>	<b>146±114</b>	75±60	1/2 <sup>***</sup> ; 1/3 <sup>*</sup> ; 2/3 <sup>***</sup>

<sup>1</sup>Die Signifikanzen geben an, ob ein statistischer Unterschied zwischen dem 1. und 2. Cluster (1/2), 1. und 3. Cluster (1/3) sowie 2. und 3. Cluster (2/3) vorliegt (getestet über den Rangsummentest Kruskal-Wallis; n. s. = nicht signifikant; \*p < 0,05; \*\*p < 0,01; \*\*\*p < 0,001).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

Die jeweils signifikant höchsten und niedrigsten Verzehrsmengen einer Lebensmittelgruppe sind fettgedruckt.

Dem ersten Cluster wurden 90,0 %, dem zweiten Cluster 8,4 % und dem dritten Cluster 1,6 % der Männer zugeordnet. Der Anteil der Männer, die den einzelnen Clustern zugeordnet wurden, ist damit ungleich verteilt.

Männer des **ersten Clusters** sind durch den geringsten Verzehr an Getreide, Milch und Süßwaren gekennzeichnet. Daneben verzehren Männer dieses Clusters weniger Kartoffeln, Eier, Fleisch, Getränke und Streichfett als Männer aus dem zweiten Cluster und im Vergleich zu Männern aus dem dritten Cluster weniger Obst, Gemüse, Nüsse und Samen sowie mehr Kartoffeln. Trotz der großen Anzahl an Männern, die diesem Cluster zugeordnet wurden, konnte aufgezeigt werden, dass Männer des ersten Clusters im Vergleich zur Gesamtgruppe weniger Getreide, Kartoffeln, Eier, Fleisch, Streichfett und Süßwaren verzehren (nicht dargestellt).

Männer aus dem **zweiten Cluster** zeichnen sich durch den höchsten Verzehr an Getreide, Kartoffeln, Eier, Fleisch, Streichfett und Süßwaren aus. Zudem verzehren Männer dieses Clusters im Vergleich zu denen aus dem ersten Cluster mehr Milch und Getränke sowie im Vergleich zu Männern aus dem dritten Cluster weniger Obst, Gemüse, Nüsse und Samen.

Das **dritte Cluster**, das kleinste Cluster, gruppiert Männer, die mehr Obst, Gemüse, Nüsse und Samen und weniger Kartoffeln als Männer der anderen beiden Cluster verzehren. Daneben zeichnen sie sich im Vergleich zu Männern aus dem zweiten Cluster durch einen geringeren Verzehr an Getreide, Eier, Fleisch, Streichfett und Süßwaren aus. Im Vergleich zu Männern aus dem ersten Cluster verzehren sie mehr Getreide, Milch und Süßwaren. Im

Verzehr von Fisch und Alkohol konnten zwischen den Clustern keine Unterschiede festgestellt werden.

Die in den drei Clustern gruppierten Männer, werden im Folgenden mit weiterer Variablen wie dem Bio-Kauf und gesundheitsrelevanten Variablen charakterisiert und verglichen (Tab. 4.18).

**Tab. 4.18: Charakterisierung der Gesamtgruppe sowie der Männer des ersten, zweiten und dritten Clusters**

		Gesamtgruppe (%)	1. Cluster (n=5364)	2. Cluster (n=502)	3. Cluster (n=95)	Signifikanzen <sup>1</sup>
<b>Bio-Kauf</b>	ja	38,6	39,2	29,7	50,5	1/2***; 1/3*; 2/3***
	nein	61,5	60,8	70,3	49,5	1/2***; 1/3*; 2/3***
<b>Altersgruppe</b>	18–34 Jahre	19,7	19,5	23,9	11,6	1/2*; 1/3 <sup>n.s.</sup> ; 2/3**
	35–50 Jahre	32,0	31,4	37,1	40,0	1/2**; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	51–64 Jahre	25,3	25,2	23,5	36,8	1/2 <sup>n.s.</sup> ; 1/3**; 2/3**
	65–80 Jahre	23,0	23,9	15,5	11,6	1/2***; 1/3**; 2/3 <sup>n.s.</sup>
<b>Soziale Schicht<sup>3</sup></b>	untere Schicht	17,6	17,0	24,3	15,8	1/2***; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	mittlere Schicht	28,4	28,2	31,9	20,0	1/2*; 1/3 <sup>n.s.</sup> ; 2/3*
	obere Schicht	54,1	54,8	43,8	64,2	1/2***; 1/3 <sup>n.s.</sup> ; 2/3***
<b>BMI</b>	Untergewicht	0,7	0,7	1,0	0,0	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	Normalgewicht	30,0	29,2	37,5	35,8	1/2***; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	Übergewicht	45,0	45,3	41,2	47,4	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	Adipositas	20,1	20,4	17,9	12,6	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	keine Angabe	4,3	4,5	2,4	4,2	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 102 cm	68,6	68,1	73,3	73,7	1/2*; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	über 102 cm	28,0	28,5	24,1	22,1	1/2*; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	keine Angabe	3,4	3,4	2,6	4,2	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüft-umfang)</b>	bis 1,0	81,2	80,9	83,5	84,2	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	über 1,0	15,4	15,6	13,9	11,6	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	keine Angabe	3,4	3,4	2,6	4,2	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	2569±850	2449±733	3736±1044	3211±914	1/2***; 1/3***; 2/3***
<b>Ernährungskennnisse</b>	gut	17,1	17,0	15,1	28,4	1/2 <sup>n.s.</sup> ; 1/3**; 2/3**
	mittel	38,2	38,2	39,4	33,7	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	schlecht	43,8	43,9	44,2	36,8	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	keine Angabe	0,9	0,9	1,2	1,1	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	56,2	56,5	51,2	61,1	1/2*; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	nein	43,4	43,0	48,2	39,0	1/2*; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	keine Angabe	0,5	0,5	0,6	0,0	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	28,1	27,2	37,7	32,6	1/2***; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	nein	71,6	72,6	62,4	67,4	1/2***; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	keine Angabe	0,2	0,3	0,0	0,0	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	75,2	75,2	75,3	77,9	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	mittel	20,7	20,9	19,9	16,8	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	schlecht	3,9	3,8	4,8	4,2	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	keine Angabe	0,1	0,1	0,0	1,1	--- <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	83,6	83,7	83,1	82,1	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	schlecht	16,0	15,9	16,9	17,9	1/2 <sup>n.s.</sup> ; 1/3 <sup>n.s.</sup> ; 2/3 <sup>n.s.</sup>
	keine Angabe	0,4	0,4	0,0	0,0	--- <sup>2</sup>
<b>HEI-NVS II</b>	MW±Std dev	80,2±10,7	80,7±10,7	75,4±9,6	84±10,6	1/2***; 1/3**; 2/3***

<sup>1</sup>Die Signifikanzen geben an, ob ein statistischer Unterschied zwischen dem 1. und 2. Cluster (1/2), 1. und 3. Cluster (1/3) sowie 2. und 3. Cluster (2/3) vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und HEI-NVS II über den Rangsummentest Kruskal-Wallis; n. s. = nicht signifikant; \*p <0,05; \*\*p <0,01; \*\*\*p <0,001).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

Männer aus dem dritten Cluster weisen mit 50,5 % den höchsten Anteil an Bio-Käufern auf. Im ersten Cluster ist der Anteil mit 39,2 % am zweithöchsten und im zweiten Cluster mit 29,7 % am niedrigsten.

Männer aus dem **ersten Cluster** weisen im Vergleich zu Männern der anderen beiden Clustern den höchsten Anteil in der Altersgruppe der 65- bis 80-Jährigen auf und sind im Vergleich zu Männern aus dem zweiten Cluster zu einem höheren Anteil in der oberen sozialen Schicht vertreten. Zudem wurden sie im Vergleich zu Männern aus dem zweiten Cluster zu einem geringeren Anteil in die Gruppe der Normalgewichtigen und zu einem höheren Anteil in die Gruppe derjenigen, die einen erhöhten Taillenumfang aufweisen, zugeordnet. Im Anteil Übergewichtiger, Adipöser sowie im Anteil an Personen mit einem zu hohen WHR konnte bei diesen Männern im Vergleich zu Männern der anderen beiden Cluster keine Unterschiede festgestellt werden. Zudem weisen sie im Vergleich zu den anderen beiden Clustern eine geringere Energiezufuhr auf und sind im Vergleich zu Männern aus dem zweiten Cluster zu einem höheren Anteil der Gruppe der sportlich Aktiven und zu einem geringeren Anteil der Gruppe der Nichtraucher zugeordnet. Bei den Ernährungskennnissen sowie der Selbsteinschätzung des allgemeinen Gesundheitszustands und der Schlafqualität konnten keine Unterschiede festgestellt werden. Männer aus dem ersten Cluster weisen im Vergleich zu Männern aus dem zweiten Cluster eine höhere und im Vergleich zu Männern aus dem dritten Cluster eine niedrigere HEI-NVS II-Indexsumme auf.

Männer, aus dem **zweiten Cluster** weisen mit über 70 % den höchsten Anteil an Nicht-Bio-Käufern auf. Sie gehören im Vergleich zu Männern der anderen beiden Clustern zu einem höheren Anteil der jüngsten Altersgruppe und der mittleren sozialen Schicht an. Männer dieses Clusters besitzen die höchste Energiezufuhr, sind zu einem höheren Anteil normalgewichtig als Männer aus dem ersten Cluster und gehören zu einem höheren Anteil der Gruppe mit einem Taillenumfang unter 102 cm an. Zudem weisen sie einen geringeren Anteil sportlich Aktiver und Raucher als im ersten Cluster auf. Im Vergleich zum dritten Cluster konnte bei diesen beiden Variablen keine Unterschiede festgestellt werden. Von allen Clustern weisen sie die niedrigste HEI-NVS II-Indexsumme auf und ernähren sich demnach im Vergleich zu den anderen Clustern gesundheitlich eher ungünstig.

Männer, die gemessen an der HEI-NVS II-Indexsumme dem gesundheitlich eher günstigen **dritten Cluster** zugeordnet wurden, zeichnen sich durch den höchsten Anteil an Bio-Käufern und Männern der Altersgruppe der 51- bis 64-Jährigen aus. Sie sind zu einem höheren Anteil der oberen sozialen Schicht zugeordnet als Männer aus dem zweiten Cluster. Zudem weisen sie im Vergleich zu Männern der anderen beiden Cluster den höchsten Anteil an Männern mit guten Ernährungskennnissen auf. In den weiteren Variablen konnten keine Unterschiede zwischen dem dritten Cluster und den anderen beiden Clustern festgestellt werden.

#### **4.2.6 Methodische Ergebnisse und Diskussion**

Für die Zielsetzung der Arbeit, die Eignung verschiedener Methoden zur umfassenden Darstellung des Lebensmittelverzehrs zu prüfen, wurde eine Clusteranalyse unter Anwendung des Verfahrens Ward durchgeführt. Die Clusteranalyse wurde getrennt nach Geschlecht auf Basis von 13 Lebensmittelgruppen berechnet. Über die Clusteranalyse werden Personen mit einem ähnlichen Lebensmittelverzehr gruppiert. Der Lebensmittelverzehr wird durch die Mittelwerte der Verzehrsmengen der Personen, die einem Cluster zugeordnet wurden dargestellt.

##### **Kriterium: Berücksichtigung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen bei Anwendung der Methode**

Die Clusteranalyse stützt sich wie die Hauptkomponentenanalyse nicht auf bisheriges Wissen (Schulze und Hoffmann 2006 S 860, Newby und Tucker 2004 S 177). Zudem wird durch das Verfahren keine Einschränkung über die Anzahl der Variablen, die bei der Berechnung einbezogen werden können, geben. Daher konnten, bei der vorliegenden Auswertung, unabhängig von quantifizierbaren Empfehlungen alle 13 für das Untersuchungsziel relevanten Lebensmittelgruppen bei der Anwendung der Clusteranalyse einbezogen werden.

Das Kriterium, möglichst alle Lebensmittelgruppen mit den dazugehörigen verzehrten Lebensmittelmengen bei der Anwendung der Methode zu berücksichtigen, wird erfüllt.

##### **Kriterium: Darstellung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse**

Über die Clusteranalyse werden in den gebildeten Clustern Personen mit einem ähnlichen Lebensmittelverzehr gruppiert (Bamia et al. 2005 S 103, Hu 2002 S 4). Die Darstellung des Lebensmittelverzehrs erfolgt durch die Mittelwerte der Verzehrsmengen aller 13 einbezogenen Lebensmittelgruppen der Personen, die einem Cluster zugeordnet wurden.

Das Kriterium, möglichst alle Lebensmittelgruppen mit den dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse darzustellen, wird erfüllt.

##### **Kriterium: Identifizieren von Personengruppen mit ähnlichem Lebensmittelverzehr**

Ziel der Clusteranalyse ist es, Personengruppen mit einem ähnlichen Lebensmittelverzehr mit Hilfe der gebildeten Cluster zu identifizieren. Hierfür soll wie von Schendera (2010) empfohlen, die inhaltliche Interpretierbarkeit der Ergebnisse geprüft werden, um auszuschließen, dass die gebildeten Cluster zufällig zustande gekommen sind (Schendera 2010 S 7f, 18).

Bei den Frauen wurden zwei und bei den Männern drei Cluster gebildet. Die Anzahl der gebildeten Cluster entspricht denen anderer Studien. In den in der Arbeit zitierten Studien in denen eine Clusteranalyse angewendet wurde, variiert die Anzahl der identifizierten Cluster zwischen zwei und acht Clustern.

Dem ersten und zweiten Cluster der Frauen wurden mit 46,2 % und 53,8 % ähnlich viele Frauen zugeordnet. Bei den Männern ist die sehr ungleiche Personenanzahl in den drei Clustern auffällig. Das erste Cluster enthält 90 %, das zweite Cluster 8,4 % und das dritte Cluster 1,6 % der Männer. Um der Frage nachzugehen, ob bei den Männern eine andere Clusterzahl zu einer gleichmäßigeren Aufteilung führt, wurde die aufgrund des pseudo T-Wertes angebotene Anzahl von sieben Clustern diesbezüglich geprüft. Auch bei sieben Clustern gab es ein Cluster mit etwa der Hälfte der Männer, sodass bei dieser Variante ebenfalls eine sehr ungleiche Verteilung der Männer zu den Clustern vorlag. Es ist davon auszugehen, dass durch die Clusteranalyse keine bessere Verteilung der Personen in die Cluster auf Basis von Ähnlichkeiten im Lebensmittelverzehr erzielt werden kann.

Über die Clusteranalyse werden alle Objekte bzw. Personen über Cluster gruppiert. Zwar wurde der Datensatz zu Beginn der Auswertungen auf Plausibilität geprüft und Außerer beseitigt, dennoch kann bei einer Clusteranalyse nicht ausgeschlossen werden, dass Objekte nur zufällig durch das Verfahren einem Cluster zugeordnet wurden und möglicherweise nicht in das Cluster passen.

Da die Clusteranalyse nicht frei von subjektiven Entscheidungen, wie der Wahl der optimalen Clusteranzahl (Schendera 2004 S 540f) ist, ist nicht auszuschließen, dass ein anderes Clusterverfahren, eine andere Auswahl an Lebensmittelgruppen oder andere Kriterien zur Wahl der Clusteranzahl andere Cluster gebildet hätte. Beispielsweise haben Hearty und Gibney (2009 S 599f) in ihrer Studie festgestellt, dass bereits die Auswahl der einbezogenen Variablen Einfluss auf die Clusterbildung haben kann. Zum einen wurden Verzehrsmengen von Lebensmitteln als Gramm pro Tag und zum anderen Verzehrsmengen von Lebensmitteln ausgedrückt als Energieprozent der Gesamtenergiezufuhr in die Clusteranalyse einbezogen. In der erstgenannten Variante wurde ein Cluster mit einem, im Vergleich zu den anderen Clustern, hohen Personenanteil (40 % der Teilnehmenden) gebildet, bei der zweiten Variante waren die Gruppengrößen der einzelnen Cluster gleichmäßig verteilt.

Wie durch die Literatur ersichtlich, ist die Clusteranalyse durchaus geeignet, Personengruppen mit einer ähnlichen Ernährung zu identifizieren. So haben McGowan und McAuliffe (2012) die Ernährung von schwangeren Frauen und Ovaskainen et al. (2009) den Zusammenhang zwischen der Ernährung von Kindern und ihren Eltern untersucht und konnten auf diesem Weg Personengruppen mit einem ähnlichen Lebensmittelverzehr identifizieren.

Für das vorliegende Kriterium bedeutet dies, dass grundsätzlich Personen mit einem ähnlichen Lebensmittelverzehr über die Cluster identifiziert werden können. In der vorliegenden Analyse ist dies nur für die Gruppe der Frauen möglich. Bei den Männern kann die vorliegende Analyse aufgrund der großen Gruppengrößenunterschiede in den gebildeten Clustern nicht zur Identifizierung von Personengruppen mit einem ähnlichen Lebensmittelverzehr beitragen.

### **Kriterium: Differenzierung von Bio-Käufer/innen und Nicht-Bio-Käufer/innen**

Wie bei der Hauptkomponentenanalyse ist es möglich, indirekt und damit nach Berechnung der Clusteranalyse zu prüfen, ob sich die Bio-Käufer/innen und Nicht-Bio-Käufer/innen durch die gebildeten Clustern differenzieren lassen.

In der Gruppe der Frauen konnten im Anteil der Bio-Käuferinnen und Nicht-Bio-Käuferinnen keine Unterschiede zwischen den beiden Clustern festgestellt werden. In der Gruppe der Männer zeigt sich dagegen, dass sich der Anteil der Bio-Käufer und Nicht-Bio-Käufer zwischen den drei Clustern unterscheidet. Allerdings sind die Ergebnisse in der Gruppe der Männer, aufgrund der ungleichen Clusteraufteilung nicht sehr aussagekräftig. Dem ersten Cluster, dem 90 % der Männer zugeordnet wurden, enthält 60 % Nicht-Bio-Käufer und 40 % Bio-Käufer und entspricht damit dem Durchschnitt der Gesamtgruppe der Männer.

Auf Basis der deskriptiven Ergebnisse (Kapitel 2.3.2) sowie der Ergebnisse des Indexes (Kapitel 3.1.4) wäre zu erwarten, dass sich die Frauen und Männer bezüglich des Lebensmittelverzehrs besser differenzieren lassen und die Bio-Käufer/innen, bezogen auf die verzehrten Lebensmittelmengen, zu einem höheren Anteil einem Cluster mit einer günstigeren Lebensmittelauswahl und die Nicht-Bio-Käufer/innen zu einem höheren Anteil einem Cluster mit ungünstiger Lebensmittelauswahl zugeordnet werden können.

Für das Kriterium, Personengruppen zu differenzieren bedeutet dies, dass die Clusteranalyse zwar grundsätzlich dazu beitragen kann, eine Personengruppe zu differenzieren, dies anhand der vorliegenden Studiendaten allerdings nicht möglich ist.

### **Kriterium: Charakterisierung der identifizierten und differenzierten Personengruppen**

Der Vorteil explorativer Methoden gegenüber der deskriptiven Auswertung ist, dass die gebildeten Cluster für verschiedene Untersuchungen von Zusammenhängen zwischen Ernährung und weiteren Variablen genutzt werden können. So können nach Berechnung der Clusteranalyse weitere Analysen angeschlossen werden, um die den gebildeten Clustern zugeordneten Personen zu charakterisieren (Hu 2002 S 4). Beispielsweise haben McGowan und McAuliffe (2012) sowie Okubo et al. (2011), welche die Ernährung von Schwangeren

untersucht haben, Frauen, die den einzelnen Clustern zugeordnet wurden, mit weiteren Variablen wie Alter und Bildung verknüpft, um die Studiengruppen weiter zu charakterisieren. Auch in der vorliegenden Auswertung wurden die in den Clustern gruppierten Personen, die einen ähnlichen Lebensmittelverzehr aufweisen, durch weitere Variablen charakterisiert. In der Gruppe der Frauen weisen die in den zwei gebildeten Clustern gruppierten Frauen Unterschiede in den verzehrten Lebensmittelmengen auf. Bei den weiteren beschreibenden Variablen konnten allerdings nur wenige Unterschiede, bei den Männern mehr Unterschiede ermittelt werden (Kapitel 4.2.5).

Auf Basis der deskriptiven Ergebnisse (Kapitel 2.3.2) sowie der Ergebnisse des Indexes (Kapitel 3.1.4) wäre zu erwarten, dass sich die Frauen und Männer bezüglich des Lebensmittelverzehrs besser mit weiteren Variablen charakterisieren lassen. Beispielsweise wäre zu erwarten gewesen, dass Frauen und Männer, die einem, bezogen auf den Lebensmittelverzehr, günstigeren Cluster zugeordnet wurden auch ein gesundheitsbewussteres Verhalten aufweisen. Die Gründe, warum diese Ergebnisse nicht festgestellt werden konnten, sind möglicherweise in den Limitierungen der Clusteranalyse zu finden. Wie alle hier vorgestellten Methoden ist auch die Clusteranalyse nicht frei von subjektiven Entscheidungen. Neben der Wahl des geeigneten Verfahrens der Clusteranalyse (Bamia et al. 2005 S 111, Schendera 2004 S 540f) werden die Ergebnisse von der Wahl der einbezogenen Variablen sowie der Festlegung der optimalen Clusteranzahl beeinflusst (Schendera 2004 S 540f). Beispielsweise wurde auch in der vorliegenden Auswertung die Auswirkung der Clusteranzahl getestet, die allerdings keinen wesentlichen Einfluss auf die Ergebnisse hatte.

Für das vorliegende Kriterium bedeutet dies, dass zwar grundsätzlich Personengruppen mit weiteren, beispielsweise ernährungs- und gesundheitsrelevanten Variablen, charakterisiert werden können, dies in der vorliegenden Auswertung allerdings nicht möglich ist.

## **Fazit**

Die Kriterien, bei der Anwendung der Clusteranalyse möglichst alle Lebensmittelgruppen mit den dazugehörigen verzehrten Lebensmittelmengen zu berücksichtigen und über die Ergebnissen darzustellen, werden erfüllt, da unabhängig von Empfehlungen alle 13 für das Untersuchungsziel relevanten Lebensmittelgruppen einbezogen und über die Ergebnisse dargestellt werden können. Auch das Identifizieren und Differenzieren von Personengruppen ist grundsätzlich mit der Methode möglich. Allerdings werden diese beiden Kriterien mit den vorliegenden Daten nur zum Teil erfüllt. Eine Identifizierung von Personengruppen mit einem ähnlichen Lebensmittelverzehr ist in der Gruppe der Frauen, nicht aber in der Gruppe der Männer möglich. Das anschließende Differenzieren der Bio-Käufer/innen und Nicht-Bio-Käufer/innen über die gebildeten Cluster ist ebenfalls, mit den vorliegenden Studiendaten nur

zum Teil möglich. Eine Charakterisierung der Identifizierten und Differenzierten Personengruppen ist ebenfalls möglich, konnte bei dieser Auswertung allerdings nicht sinnvoll umgesetzt werden.

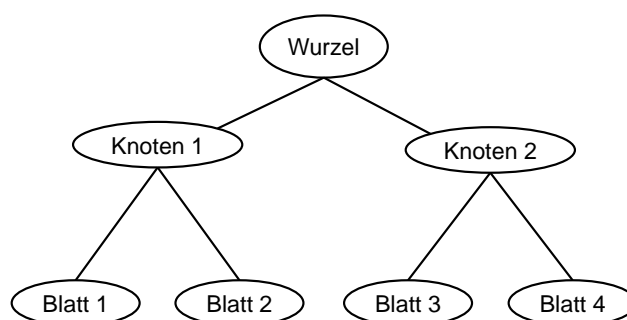
Die Anwendung der Clusteranalyse zur umfassenden Darstellung des Lebensmittelverzehrs anhand der Daten der NVS II ist von den gegebenen Voraussetzungen her geeignet, erfüllt jedoch für die vorliegende Studiengruppe der NVS II nur zum Teil die Kriterien der umfassenden Darstellung des Lebensmittelverzehrs. Dennoch ist es möglich, dass, wie durch Literaturbeispielen gezeigt, die Clusteranalyse bei anderen Studiengruppen zur umfassenden Darstellung des Lebensmittelverzehrs geeignet ist.

Damit kann der Lebensmittelverzehr durch die Ergebnisse der Clusteranalyse umfassender dargestellt werden, als dies über rein deskriptive Vergleiche einzelner Lebensmittelgruppen möglich ist.

### 4.3 Entscheidungsbaum

Über einen Entscheidungs- oder Klassifikationsbaum können Entscheidungen (Beierle und Kern-Isberner 2006 S 105, Quinlan 1993 S 5), Klassifizierungen (Christof 2007 S 127, Petersohn 2005 S 136, Rokach und Maimon 2005 S 476, Ester und Sander 2000 S 126, Quinlan 1993 S 5, 1986 S 83) sowie Alternativen und ihre Konsequenzen (Eisenführ et al. 2010 S 48) dargestellt werden. Ebenso können auf Basis von konkreten Beispieldaten allgemeine Regeln, im Sinne von wenn x vorliegt folgt y, abgeleitet werden (von Rimscha 2010 S 113), da sich der Pfad von der Wurzel zu den Blättern als logische Wenn-Dann-Regel lesen lässt (Beierle und Kern-Isberner 2006 S 105). Entscheidungsbäume werden häufig unter Maschinellern Lernen (ML) geführt (von Rimscha 2010 S 113, Beierle und Kern-Isberner 2006 S 97ff). Wie der Name impliziert, können die Ergebnisse grafisch als Baum dargestellt werden.

Bei Entscheidungsbäumen handelt es sich um sogenannte Top-Down Induction of Decision Tree (TDIDT) Verfahren (Beierle und Kern-Isberner 2006 S 115). Die Bildung eines Baums beginnt oben mit einer gemeinsamen Wurzel und verläuft über Knoten nach unten zu den sogenannten Blättern (Christof 2007 S 127f, Quinlan 1986 S 83) (Beispiel siehe Abb. 4.8).



**Abb. 4.8: Schematische Darstellung eines Entscheidungsbaums (eigene Darstellung)**

Die Objekte, die zu klassifizierende Datenmenge, wird sukzessive über zahlreiche Knoten zu Blättern aufgeteilt. Zur Bildung der Knoten wird eine Klassifizierungsvariable, die gleichzeitig eine abhängige Variable sein muss, verwendet. Ziel ist, dass die in den Blättern gruppierten Objekte untereinander möglichst homogen und zu den Objekten der weiteren Blätter möglichst heterogen sind (Christof 2007 S 127f).

Bei diesem Verfahren handelt es sich im Gegensatz zu Hauptkomponenten- und Clusteranalyse nicht um ein klassisches multivariates Verfahren, da zur Klassifizierung der Objekte des Datensatzes in Knoten und Blätter nicht mehrere Variablen gleichzeitig, sondern immer nur eine der vielen einbezogenen Variablen herangezogen wird. Welche der zur Verfügung stehenden Variablen zur Klassifizierung der Objekte verwendet wird, ist von den an einem Knoten enthaltenen Objekten und von dem verwendeten Verfahren abhängig. Das bedeutet,

dass über den gewählten Algorithmus an jedem Knoten neu geprüft wird, welche Variable sich zur Klassifizierung der dort enthaltenen Objekte am besten eignet.

### **Verfahren des Entscheidungsbaums**

Entscheidungsbaume können über unterschiedliche Verfahren gebildet werden. In aller Regel beginnt die Bildung eines Entscheidungsbaums mit einem sogenannten Trainingsdatensatz. Hierfür wird ein Teil des Datensatzes, der sogenannte Testdatensatz (bestimmt über Zufallsgenerierung), zurückgehalten und mit den übrigen Daten, den Trainingsdaten, der Baum trainiert bzw. generiert. Anschließend wird der entstandene Entscheidungsbaum mit dem Testdatensatz geprüft. Durch die Anwendung des Trainingsdatensatzes wird das Ziel verfolgt, nach der Generierung des Baums auch neu erhobene Daten, die nicht aus dem vorliegenden Datensatz stammen, über den Entscheidungsbaum klassifizieren zu können (Beierle und Kern-Isberner 2006 S 108, Ester und Sander 2000 S 107).

Je nach angewendetem Verfahren stehen nach Berechnung des Entscheidungsbaums verschiedene Optionen wie das Pruning, mit welchem überflüssigen Äste entfernt werden können, zur Verfügung (Runkler 2010 S 103, Christof 2007 S 131, Ester und Sander 2000 S 133).

Die wesentliche Grundlage bzw. Idee der unterschiedlichen Entscheidungsbaum-Verfahren werden auf konzeptionelle Lernsysteme (CLS) von Hunt et al. (1966) zurückgeführt (Quinlan 1993 S 17). Über CLS wurden erste Bäume zum Klassifizieren von Objekten erzeugt. Darauf aufbauend wurden Ende der 1970er Jahre von Ross Quinlan weitere Verfahren getestet (Quinlan 1979 S 171) und ID3 entwickelt (Quinlan 1986 S 84). ID3 wird als Grundbaustein für viele Weiterentwicklungen wie C4.5 (Petersohn 2005 S 143, Quinlan 1993 S 2) und C5.0 beschrieben (Petersohn 2005 S 143). Daneben haben sich eine Reihe weiterer Verfahren wie CHAID (CHi-square based Automatic Interaction Detector) und CART (Classification And Regression Tree) etabliert.

Das Verfahren **CHAID** wurde erstmals in den 1960er Jahren von Sonquist und Morgan (1964) beschrieben. Im Gegensatz zu den anderen genannten Verfahren arbeitet CHAID ausschließlich mit kategorialen Variablen. Metrische Variablen müssen zuvor in kategoriale Variablen umgewandelt werden (Christof 2007 S 130). Wenn durch den Nutzer nicht anders vorgegeben, wird die Variable, durch welche die Objekte an einem Knoten aufgeteilt werden, durch einen Chi-Quadrat-Test ermittelt. Es wird immer die Variable verwendet, die einen signifikanten Beitrag zur Klassifizierung der Objekte leisten kann (Christof 2007 S 130).

**CART** wurde erstmals von Breiman et al. (1984) beschrieben. Wie der Name sagt, können mit Hilfe dieses Verfahrens sowohl Klassifikationsbäume als auch Regressionsbäume erstellt werden (Breiman et al. 1984). Im Gegensatz zu CHAID können bei diesem Verfahren sowohl kategoriale als auch metrische Variablen verwendet werden (Christof 2007 S 131). Bei

kategorialen Variablen wird ein Klassifikationsbaum und bei metrischen Variablen ein Regressionsbaum erzeugt (Tutz 2000 S 318). Bei kategorialen Variablen wird dies durch eine geeignete Zusammenfassung der Kategorien und bei metrischen Variablen durch die Bestimmung eines Trennwertes (Split-Punkts) erreicht. Die durch die Aufteilung erreichte Reduktion, die sogenannte Unreinheit, wird als Kriterium herangezogen: Enthält ein Knoten nur Variablen aus einer Klasse, bezogen auf die Klassifizierungsvariable (wünschenswert), ist die Unreinheit null. Enthält ein Knoten gleich viele Variablen beider Klassen, ist die Unreinheit am größten. Als Maß für die Unreinheit wird die Entropie (Informationsgehalt an einem Knoten) oder der Gini-Koeffizient (Maß zur Darstellung der Ungleichverteilung, bezogen auf die Varianz) herangezogen (Christof 2007 S 131). Daneben können bei diesem Verfahren über das sogenannte Pruning überflüssige Äste (Runkler 2010 S 103), die wenig zur Klassifikationsqualität beitragen, entfernt werden (Christof 2007 S 131).

**ID3** und seine Weiterentwicklungen werden zu den bekanntesten maschinellen Lernsystemen gezählt (Beierle und Kern-Isberner 2006 S 115). Sie können wie CART mit kategorialen (Petersohn 2005 S 144) und metrischen Variablen (Runkler 2010 S 103) arbeiten. Die Auswahl der Variablen erfolgt auf Basis des Informationsgehalts an einem Knoten. Ziel dabei ist, den Entscheidungsbaum möglichst klein zu halten (wenige Knoten) (Beierle und Kern-Isberner 2006 S 115). Ebenso kann der Entscheidungsbaum ähnlich wie bei CART über verschiedene Optionen wie durch das Entfernen überflüssiger Äste optimiert werden.

Zusätzlich zu ID3 können die Verfahren **C4.5** und **C5.0** z. B. mit fehlenden Einträgen arbeiten (Runkler 2010 S 103). Daneben wurde mit C4.5 eine Optimierung der Auswahl der geeigneten Attribute, der Pruningmöglichkeiten (z. B. das Abschneiden von Ästen) sowie eine Optimierung der Anwendung des Testdatensatzes, das sogenannte „Windowing“, vorgenommen (Petersohn 2005 S 156). Für das Windowing wird ein Teil des Trainingsdatensatzes zur Generierung des Baums, der andere Teil zur Verifizierung verwendet. Wurden Objekte bei der Verifizierung falsch zugeordnet, werden sie dem ersten Trainingsdatensatz zugefügt und der Baum neu erstellt, mit dem Ziel möglichst viele Objekte richtig zu klassifizieren (Quinlan 1993 S 57).

### **4.3.1 Wahl eines Verfahrens für die eigene Anwendung**

In der vorliegenden Arbeit sollen Personen auf Basis ihres Lebensmittelverzehrs anhand der Variable Bio-Käufer/in versus Nicht-Bio-Käufer/in voneinander differenziert werden. Da es sich bei den Lebensmittelverzehrsdaten um metrische Daten handelt, können die Verfahren CART, ID3 und deren Weiterentwicklungen verwendet werden. Das Verfahren C4.5 von Ross Quinlan gehört zu den derzeit aktuellsten Weiterentwicklungen der Entscheidungs-

bäume. Es wurde, wie von Hornik et al. (2009) beschrieben, unter dem Namen J48 funktionsidentisch in die plattformunabhängige Programmiersprache JAVA umgeschrieben. Um zu analysieren, ob der Lebensmittelverzehr von Bio-Käufer/innen und Nicht-Bio-Käufer/innen mit einem Entscheidungsbaum umfassend dargestellt werden kann, wird das Verfahren J48 verwendet. Dabei wird kein Trainingsdatensatz berechnet, da in der vorliegenden Auswertung nicht das Ziel verfolgt wird, den Baum zu generalisieren, um diesen auf weitere neue Studiendaten anzuwenden.

### **4.3.2 Theoretischer Hintergrund**

Ebenso wie bei der Hauptkomponentenanalyse und der Clusteranalyse, kann der Entscheidungsbaum auf Basis von Nährstoffen, Lebensmitteln und Lebensmittelgruppen aufgebaut werden. Personen können z. B. auf Basis ihres Lebensmittelverzehrs über eine Klassifizierungsvariable, z. B. Bio-Käufer/innen versus Nicht-Bio-Käufer/innen, voneinander getrennt werden.

Zur Generierung des Entscheidungsbaums wird bei jedem Schritt die Variable (hier: Lebensmittelgruppe) gesucht, durch welche sich die Objekte (hier: Personen) an einem Knoten am besten über eine Klassifizierungsvariable (hier: Bio-Käufer/innen versus Nicht-Bio-Käufer/innen) differenzieren lassen (Kokai 2002 S 98). Das bedeutet, dass an jedem Knoten immer *alle* Variablen (hier: Lebensmittelgruppen) zur Teilung der Objekte (hier: Personen-Gruppe) zur Verfügung stehen. Damit kann eine Lebensmittelgruppe an mehreren Knoten verwendet werden.

Die Aufteilung endet, wenn keine weitere Variable zur Aufteilung zur Verfügung steht, eine zuvor festgelegte Aufteilungsgrenze (z. B. maximal vier Knoten) erreicht wird oder der Knoten nicht genügend Objekte enthält, um weiter aufgegliedert zu werden. Die entsprechenden Abbruchkriterien können zu Beginn definiert werden (Christof 2007 S 127ff). Kann ein Knoten nicht weiter aufgeteilt werden, entsteht ein Blatt, in dem Objekte mit ähnlichen Eigenschaften klassifiziert sind. Durch den Pfad des Entscheidungsbaums (von der Wurzel über die Knoten zum Blatt) können die charakteristischen Eigenschaften der in einem Blatt klassifizierten Objekte beschrieben werden. Als Ergebnis wird auch angegeben, welche und wie viele Objekte, bezogen auf die Klassifizierungsvariable, richtig oder falsch zugeordnet wurden. Daneben ist erkennbar, durch welche Variable sich die Objekte am besten differenzieren lassen. Die am besten klassifizierende Variable stellt den ersten Knoten dar.

Ziel eines generierten Entscheidungsbaums ist, möglichst alle Objekte (hier: Bio-Käufer/innen und Nicht-Bio-Käufer/innen) richtig zu klassifizieren (Petersohn 2005 S 145). Unterschiedliche Klassifikationskriterien, wie die Korrekt- und Falschklassifikationsrate, liefern Auskunft darüber, wie gut oder schlecht sich die Objekte differenzieren lassen. Die

Klassifikationsraten werden mit der Erzeugung des Baums berechnet. Ihre Berechnung wird im Folgenden an dem Beispiel von gesunden und kranken Personen einer gegebenen Studiengruppe aufgezeigt. Diese Personen sollen anhand von dem Vorhandensein oder Nicht-Vorhandensein verschiedener Symptome als gesund oder krank klassifiziert werden. Idealerweise sollten alle gesunden Personen als gesund und alle kranken Personen als krank identifiziert werden. Zur Einschätzung der **Klassifikationsgüte** stehen folgende Kriterien zur Auswahl (Runkler 2010 S 86):

- Richtig positiv (true positiv = TP): Eine kranke Person wird als krank klassifiziert
- Richtig negativ (true negative = TN): Eine gesunde Person wird als gesund klassifiziert
- Falsch positiv (false positive = FP): Eine gesunde Person wird als krank klassifiziert
- Falsch negativ (false negative = FN): Eine kranke Person wird als gesund klassifiziert

Anhand dieser Kriterien berechnet sich die Korrektklassifikation als  $T = TP + TN$  (Anzahl der richtig klassifizierten Personen) und die Falschklassifikation als  $F = FP + FN$  (Anzahl der falsch klassifizierten Personen).

Daneben können in Abhängigkeit vom Untersuchungsziel auf Basis der bereits genannten Kriterien weitere Gütekriterien herangezogen werden (Tsien et al. 1998 S 494):

- **Sensitivität** ist die Wahrscheinlichkeit, dass eine kranke Person als krank klassifiziert wird: Berechnet aus der Anzahl der korrekt klassifizierten erkrankten Personen (TP) dividiert durch die Anzahl der tatsächlich kranken Personen (TP + FN).
- **Spezifität** ist die Wahrscheinlichkeit, dass eine gesunde Person als gesund klassifiziert wird: Berechnet aus der Anzahl der korrekt als gesund klassifizierten Personen (TN) dividiert durch die Anzahl der tatsächlich gesunden Patienten (TN + FP).
- **Positiver Vorhersagewert** entspricht der Anzahl der korrekt klassifizierten kranken Personen (TP) dividiert durch die Anzahl der korrekt und falsch klassifizierten kranken Personen (TP + FP).

Die Genauigkeit des gebildeten Entscheidungsbaums kann über eine *m*-fache-Kreuzvalidierung (cross-validation) geprüft werden. Hierfür werden die Ausgangsdaten in *m* gleichgroße Teilmengen aufgeteilt, mit diesen jeweils ein Entscheidungsbaum erzeugt und die Klassifikationsgüte berechnet (Ester und Sander 2000 S 109). Entspricht das Ergebnis der *m*-fachen Kreuzvalidierung der des Baums ohne Validierung, ist die Wahrscheinlichkeit eines zufällig zustande gekommen Ergebnisses gering.

### Bildung eines Knotens

Zur Teilung der Objekte (hier: Personen) wird die am betrachteten Knoten (T) enthaltene Information  $Info(T)$  der dort vorliegenden Variablen (hier: Lebensmittelgruppen) herangezogen. Die  $Info(T)$  wird auch als **Entropie** bezeichnet (Runkler 2010 S 99ff, Quinlan 1993 S 20ff). Diese Information lässt sich logarithmisch aus der Auftrittswahrscheinlichkeit ( $P_i$ ) aller Variablen ( $n$ ) mit einer bestimmten Ausprägung berechnen (Beierle und Kern-Isberner 2006 S 115):

$$Info(T) = - \sum_{i=1}^n p_i \log_2 p_i$$

Der informationstheoretisch beste Entscheidungsbaum ist derjenige, der an jedem Knoten die Variable verwendet, die zum maximalen **Informationsgewinn (Gain)** an einem Knoten führt (Runkler 2010 S 99ff). Der Informationsgewinn berechnet sich aus der Differenz der Entropie an einem Knoten abzüglich der Information durch eine einzelnen Variablen (z. B. Aufteilung nach Obstverzehr). Auf diesem Weg wird der Informationsgewinn aller Variablen berechnet und verglichen. Die Variable die den höchsten Informationsgewinn liefert wird verwendet (Quinlan 1993 S 23).

Zur Berechnung des Gain-Ratios wird zunächst die sogenannte Split-Information, die Entropie jeder einzelnen Variable berechnet (Quinlan 1993 S 23):

$$Split\ Info(x) = - \sum_{i=1}^n \frac{|T_i|}{|T|} \log_2 \left( \frac{|T_i|}{|T|} \right).$$

Daraus lässt sich das Gain-Ratio wie folgt berechnen:

$$Gain\ ratio(x) = \frac{Gain(x)}{Split\ Info(x)}$$

Die Variable, durch welche der höchste Informationsgewinn (Gain Ratio) erzielt wird, wird zur Klassifikation (hier: Teilung der Personen anhand der Klassifizierungsvariable Bio-Käufer/innen versus Nicht-Bio-Käufer/innen) an dem entsprechenden Knoten herangezogen. Das bedeutet, dass bei jedem Knoten neu entschieden wird, welche Variable sich am besten zur Klassifizierung eignet (Quinlan 1993 S 22). Die Anzahl der in den Entscheidungsbaum einzubeziehenden Variablen muss daher zuvor nicht beschränkt werden. Eine Variable wird an einem Knoten nur verwendet, wenn sie zur Klassifizierung beiträgt (Beierle und Kern-Isberner 2006 S 106, Petersohn 2005 S 146).

Bei dem in der vorliegenden Arbeit angewendeten Verfahren J48 endet die Aufteilung in Knoten, wenn ein neuer Knoten keinen signifikanten Einfluss auf die Klassifizierung hat (Petersohn 2005 S 158). Daneben können weitere Abbruchkriterien, wie die Definition einer Mindestanzahl von Objekten pro Knoten, festgelegt werden (Christof 2007 S 127ff).

Der Entscheidungsbaum wird von oben, der Wurzel, nach unten, entlang der Knoten zu den Blättern gelesen (Christof 2007 S 127f). Sowohl die Knoten als auch die Blätter stehen als Ergebnis des Baums zur Verfügung.

Die Pfade entlang der Knoten entsprechen einer logischen Formel in Form einer Wenn-Dann-Regel (Beierle und Kern-Isberner 2006 S 107). Der jeweils linke Teil des Baums entspricht einer „kleiner-gleich ( $\leq$ )“ und der rechte Teil einer „größer-als ( $>$ )“ Aussage, bezogen auf den jeweils darüber liegenden Knoten (Ester und Sander 2000 S 23). Für die vorliegende Auswertung bedeutet dies, dass die Studienteilnehmenden anhand ihrer Lebensmittelverzehrsmenge in „größer als“ oder „kleiner/gleich“ aufgeteilt werden.

Neben der Aufteilung der Datenmenge in Blätter lässt sich aus der Baumstruktur die Wichtigkeit der Variable ablesen. Bei der Erstellung der Pfade wird die Variable, die am besten zur Klassifizierung beiträgt, oben im Entscheidungsbaum platziert. Die für die Klassifizierung wichtigen Variablen können damit von weniger wichtigen unterschieden werden (von Rimscha 2010 S 116).

Die gebildeten Blätter stellen die klassifizierten Gruppen dar (Rokach und Maimon 2005 S 477), die anhand weiterer, im Entscheidungsbaum nicht berücksichtigter Variablen wie soziodemographische Variablen, beschrieben werden können.

### 4.3.3 Anwendung von Entscheidungsbäumen

Entscheidungsbäume werden in vielen verschiedenen Wissenschaftsbereichen wie der Statistik, dem maschinellen Lernen und Data Mining angewendet (Rokach und Maimon 2005 S 476). Mit dem Einsatz von Entscheidungsbäumen werden sehr unterschiedliche Ziele verfolgt (Rokach und Maimon 2005 S 476). Die Anwendung der Entscheidungsbäume lässt sich wie folgt kategorisieren: (1) **automatische Klassifizierung von Datenobjekten/maschinelles Lernen**, (2) **Entscheidungstheorie** und (3) **Veranschaulichen von Konsequenzen** (Beierle und Kern-Isberner 2006 S 105).

- (1) Das **automatische Klassifizieren von Datenobjekten/maschinelles Lernen** kann z.B. verwendet werden, um aus Merkmalen wie der Anzahl der Beine und der Körpergröße Lebewesen wie Fische, Vögel, Menschen, Katzen und Pferde automatisch zu den verschiedenen Arten zu klassifizieren (Runkler 2010 S 99).

Ein weiteres Beispiel für die Anwendung des Entscheidungsbaums ist die Einteilung von Personen, die eine Hypothek aufgenommen haben, in die Gruppen derjenigen, die ihre Schuld vollständig und pünktlich beglichen haben, und derjenigen, die ihre Schuld verspätet beglichen haben (Rokach und Maimon 2005 S 476).

- (2) Als **Entscheidungstheorie** beim Treffen von Entscheidungen (Runkler 2010 S 99, von Rimscha 2010 S 113ff, Beierle und Kern-Isberner 2006 S 105, Rokach und Maimon 2005 S 476) kann der Entscheidungsbaum bei der Diagnose unter-

schiedlicher Krankheiten unterstützen (Beispiele: Hische et al. 2010, Su et al. 2007, Pavlopoulos et al. 2004, Tsien et al. 1998, Mair et al. 1995). Auf Basis von grundlegenden Untersuchungen wie Temperatur- oder Blutdruckmessungen können bereits viele Krankheiten ausgeschlossen werden, womit sich je nachdem weitere Untersuchungen erübrigen. Ein Entscheidungsbaum ermöglicht folglich die Prüfung einzelner Merkmale in der Reihenfolge der Wichtigkeit (Runkler 2010 S 99).

Beispielsweise haben Su et al. (2007) mit Hilfe des Entscheidungsbaums Biomarker zur Diagnose von Magenkrebs identifiziert. Daneben haben Mair et al. (1995) über die Anwendung des Entscheidungsbaums aufzeigen können, dass bei einer bestimmten Personengruppe mit Brustschmerzen die Ergebnisse des Elektrokardiogramms die am besten diskriminierende Variable zwischen Personen mit und ohne Myokardinfarkt darstellt. Hische et al. (2010) konnten durch die Ergebnisse des Entscheidungsbaums aufzeigen, dass der Nüchternblutglukosewert bedeutend für die Identifizierung von Personen mit einem erhöhten Risiko, an Diabetes Mellitus zu erkranken ist, sowie für Personen, bei denen Diabetes Mellitus bereits vorliegt.

- (3) Der Entscheidungsbaum kann auch verwendet werden, um **Konsequenzen zu veranschaulichen**. Beispielsweise stellt sich ein Unternehmer bei einer Produktneuentwicklung die Frage, ob die Weiterentwicklung fortgesetzt oder abgebrochen werden soll. Über den Baum kann dargestellt werden, mit welcher Wahrscheinlichkeit ein erfolgreicher Abschluss erwartet wird, und welche Konsequenzen, wie die Wahrscheinlichkeit einer hohen versus niedrigen Nachfrage, sich daraus ergeben könnten (Eisenführ et al. 2010 S 48).

Aus den generierten Entscheidungsbäumen können auch Regeln in Form einer Wenn-Dann-Formel erzeugt werden, da sich der Pfad von der Wurzel zu den einzelnen Blättern als logische Wenn-Dann-Regel lesen lässt (Beierle und Kern-Isberner 2006 S 107).

Zur Darstellung des Lebensmittelverzehr wurden Entscheidungsbäume in der Literatur bisher nicht beschrieben. Die einzige bekannte Anwendung eines Entscheidungsbaums im Bereich Ernährung ist eine Untersuchung von Hearty und Gibney (2008). Hier wurde der Lebensmittelverzehr bei einzelnen Mahlzeiten wie beim Frühstück untersucht. Zur Klassifizierung wurden die Studienteilnehmenden auf Basis eines Healthy Eating Indexes in Quintile eingeteilt. Die Gruppierung der Personen der ersten und fünften Quintile wurde als Klassifizierungsvariable der einzelnen Mahlzeiten herangezogen. Zur Trennung dieser beiden Gruppen diente der Verzehr oder Nicht-Verzehr verschiedener Lebensmittelgruppen wie „frisches Obst“ oder „Fruchtsaft“ (Hearty und Gibney 2008).

#### **4.3.4 Anwendung des Entscheidungsbaums auf den Lebensmittelverzehr der Teilnehmenden der NVS II**

Mit Hilfe des Verfahrens J48 wird in der vorliegenden Arbeit ein Entscheidungsbaum erstellt und analysiert, ob er sich zur umfassenden Darstellung des Lebensmittelverzehrs eignet. Als Klassifizierungsvariable dient die Einteilung der Studienteilnehmenden in die Gruppen Bio-Käufer/innen versus Nicht-Bio-Käufer/innen. Ebenso wie bei der Hauptkomponenten- und Clusteranalyse wurden die 13 Lebensmittelgruppen Obst, Gemüse, Getreide, Kartoffeln, Milch, Fisch, Fleisch, Eier, Alkohol, Streichfett, alkoholfreie Getränke, Süßwaren sowie Nüsse und Samen in die Auswertung einbezogen.

Über die Option „M = 20“ wurde festgelegt, dass bei der Bildung des Baums mindestens 20 Personen einem Blatt zugeordnet sein müssen, um eine sinnvolle Interpretation der Ergebnisse zu ermöglichen. Zur Prüfung der Entscheidungsbauergebnisse werden die folgenden Kriterien herangezogen:

- Richtig positiv (TP): Bio-Käufer/innen werden als Bio-Käufer/innen klassifiziert
- Richtig negativ (TN): Nicht-Bio-Käufer/innen werden als Nicht-Bio-Käufer/innen klassifiziert
- Falsch positiv (FP): Bio-Käufer/innen werden als Nicht-Bio-Käufer/innen klassifiziert
- Falsch negativ (FN): Nicht-Bio-Käufer/innen werden als Bio-Käufer/innen klassifiziert

Zudem wird die Korrektklassifikation ( $T = TP + TN$ ) mit der Anzahl der richtig klassifizierten Bio-Käufer/innen und Nicht-Bio-Käufer/innen sowie die Falschklassifikation ( $F = FP + FN$ ), berechnet aus der Anzahl der falsch klassifizierten Bio-Käufer/innen und Nicht-Bio-Käufer/innen, herangezogen.

Daneben wird das Ergebnis des Entscheidungsbaums über eine 10fach-Kreuzvalidierung geprüft. Hierfür wird die Korrekt- zu Falschklassifikationsrate des berechneten Entscheidungsbaums mit und ohne Kreuzvalidierung verglichen.

In den entstehenden Blättern werden Personen gruppiert, die einen ähnlichen Lebensmittelverzehr aufweisen. Anschließend werden einzelne Blätter miteinander verglichen und analog zu den bisher beschriebenen drei Methoden, die dort zugeordneten Personen anhand weiterer beschreibender Variablen wie Alter, soziale Schicht sowie gesundheitsrelevanter Variablen wie BMI, sportliche Aktivität und Raucherstatus charakterisiert.

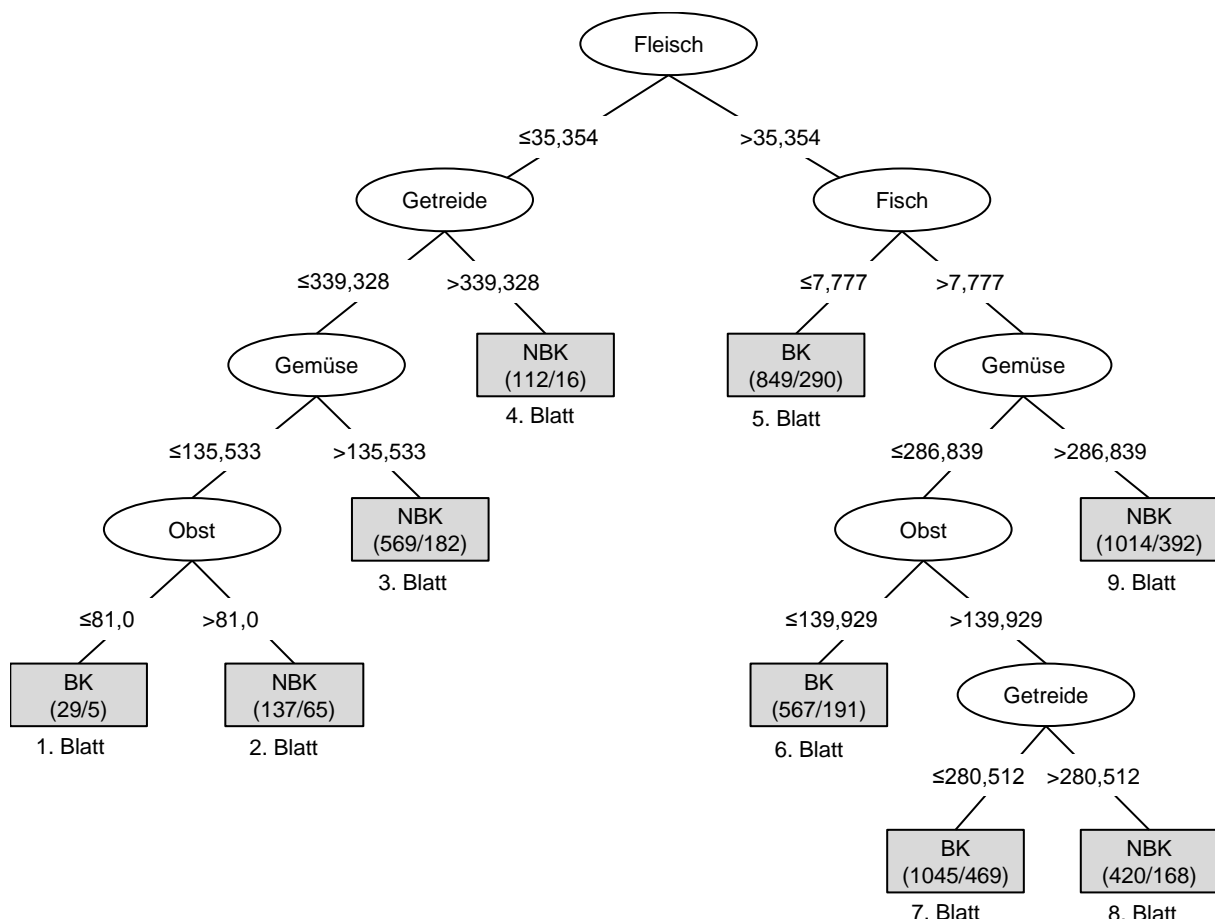
#### **4.3.5 Inhaltliche Ergebnisse**

Zur Darstellung des Lebensmittelverzehrs werden im Folgenden, getrennt für Frauen und Männer, die inhaltlichen Ergebnisse des Entscheidungsbaums präsentiert. Diese inhaltliche Auswertung dient als Basis für die Prüfung der Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs, die im nächsten Kapitel vorgenommen wird.

Zur Darstellung des Lebensmittelverzehrs der Personen, die den unterschiedlichen Blättern zugeordnet wurden, werden die Pfade von der Wurzel über die einzelnen Knoten zu den gebildeten Blättern beschrieben. Diese sogenannten Trennwerte (Split-Punkte) werden gemäß der Berechnung an einem Knoten mit drei Nachkommastellen angegeben. Zusätzlich wird der Lebensmittelverzehr der Frauen und Männer (in Gramm pro Tag) durch den arithmetischen Mittelwert (MW) mit dem dazugehörigen 95. Konfidenzintervall in den Ergebnistabellen aufgeführt. Die Konfidenzintervalle dienen dem Vergleich des Lebensmittelverzehrs zwischen den einzelnen Blättern. Die Unterschiede gelten als signifikant, wenn sich die Konfidenzintervalle nicht überschneiden.

### Entscheidungsbaum der Frauen

Die Gruppe der Frauen konnte über den Lebensmittelverzehr von 13 Lebensmittelgruppen anhand der Klassifizierungsvariable Bio-Käuferinnen versus Nicht-Bio-Käuferinnen in insgesamt neun Blätter aufgeteilt werden (Abb. 4.9).



Die Lebensmittelverzehrmengen auf dem Pfad sind als kleiner/gleich und größer als Angaben in Gramm pro Tag angegeben. Die Angabe BK = Bio-Käuferinnen und NBK = Nicht-Bio-Käuferinnen weisen auf ein von Bio-Käuferinnen- oder Nicht-Bio-Käuferinnen dominierendes Blatt hin. Die Zahlen in Klammer geben die Anzahl der Bio-Käuferinnen und Nicht-Bio-Käuferinnen an, wobei immer der größere Anteil zuerst genannt ist.

**Abb. 4.9: Entscheidungsbaum der Bio-Käuferinnen und Nicht-Bio-Käuferinnen auf Basis der Verzehrsmenge von 13 Lebensmittelgruppen (Frauen)**

Die Gesamtgruppe der Frauen wird zunächst auf Basis eines Fleischverzehr von kleiner/gleich oder größer als 35,354 g/Tag voneinander getrennt. Das bedeutet, dass sich die Frauen bezogen auf die Klassifizierungsvariable Bio-Käuferinnen versus Nicht-Bio-Käuferinnen am besten über die Lebensmittelgruppe Fleisch differenzieren lassen. Dem Ast mit dem höheren Fleischverzehr wurden mit über 80 % die meisten Frauen zugeordnet, von denen 47 % Bio-Käuferinnen und 53 % Nicht-Bio-Käuferinnen sind. Dieses Ergebnis bedeutet, dass, wenn eine Frau mehr als 35,354 g Fleisch pro Tag verzehrt, sie mit einer 53 %igen Wahrscheinlichkeit eine Nicht-Bio-Käuferin ist.

Um die Verzehrsmenge der einzelnen Blätter miteinander zu vergleichen, wurde zusätzlich zu den im Entscheidungsbaum verwendeten Trennwerten (kleiner/gleich und größer als) an den einzelnen Knoten die durchschnittlich verzehrten Lebensmittelmengen mit den entsprechenden 95. Konfidenzintervallen herangezogen (Tab. 4.19). Der Anteil der Bio-Käuferinnen und Nicht-Bio-Käuferinnen der einzelnen Blätter des Entscheidungsbaums sind in Tab. 4.20 dargestellt.

**Tab. 4.19: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen, die den einzelnen Blättern des Entscheidungsbaums zugeordnet wurden**

Lebensmittelgruppen	Gesamtgruppe MW (95. KI des MW)	Blätter			
		1. Blatt (n=46)	2. Blatt (n=223)	3. Blatt (n=871)	4. Blatt (n=173)
Obst	292(286–297)	<b>31(23–38)</b>	<b>333(298–367)</b>	358(340–376)	337(308–366)
Gemüse	251(247–254)	<b>70(58–82)</b>	<b>89(84–93)</b>	<b>316(305–328)</b>	321(297–346)
Getreide	239(236–241)	<b>140(114–167)</b>	<b>174(164–184)</b>	<b>202(197–207)</b>	<b>424(410–437)</b>
Kartoffeln	72(71–73)	48(36–61)	49(44–54)	57(54–59)	57(51–64)
Nüsse und Samen	4(3–4)	2(0–4)	4(2–5)	6(4–7)	6(4–8)
Milch	239(234–243)	161(109–213)	237(209–265)	256(243–270)	255(221–288)
Eier	12(12–13)	6(4–9)	8(7–10)	8(8–9)	9(8–11)
Fisch	23(22–23)	9(5–12)	14(12–16)	24(22–26)	17(13–21)
Fleisch	75(74–76)	<b>19(15–23)</b>	<b>20(19–21)</b>	<b>19(19–20)</b>	<b>13(12–15)</b>
Streichfett	21(20–21)	16(11–21)	14(13–16)	15(14–16)	25(21–28)
Getränke	2335(2315–2356)	2350(1930–2770)	2171(2057–2285)	2459(2398–2519)	2566(2427–2705)
Alkohol <sup>1</sup>	6(5–6)	5(2–8)	4(3–5)	5(4–5)	5(4–6)
Süßwaren	61(59–62)	79(34–124)	59(50–67)	48(45–51)	66(57–76)
	<b>5. Blatt (n=1292)</b>	<b>6. Blatt (n=821)</b>	<b>7. Blatt (n=1578)</b>	<b>8. Blatt (n=631)</b>	<b>9. Blatt (n=1478)</b>
Obst	243(230–255)	<b>77(74–80)</b>	<b>336(327–345)</b>	<b>323(310–336)</b>	350(336–363)
Gemüse	217(209–224)	<b>160(155–164)</b>	<b>183(180–186)</b>	<b>188(184–193)</b>	<b>414(408–421)</b>
Getreide	240(234–245)	239(232–245)	<b>194(192–197)</b>	<b>346(341–352)</b>	252(247–258)
Kartoffeln	68(66–71)	74(71–77)	78(76–80)	79(76–82)	80(77–82)
Nüsse und Samen	3(2–3)	3(2–3)	3(3–4)	4(3–5)	4(4–5)
Milch	225(215–236)	220(203–237)	234(224–244)	246(231–260)	254(243–266)
Eier	12(11–12)	13(12–14)	13(12–14)	13(13–14)	15(14–15)
Fisch	<b>3(2–3)</b>	<b>24(23–26)</b>	<b>27(27–28)</b>	<b>28(26–29)</b>	<b>33(32–35)</b>
Fleisch	<b>87(84–89)</b>	<b>94(91–97)</b>	<b>80(78–82)</b>	<b>90(86–93)</b>	<b>91(89–94)</b>
Streichfett	21(20–22)	21(20–23)	21(20–22)	27(25–28)	21(20–21)
Getränke	2266(2216–2316)	2133(2076–2189)	2240(2199–2281)	2376(2314–2438)	2518(2471–2565)
Alkohol <sup>1</sup>	5(5–6)	6(6–7)	5(5–6)	5(5–5)	7(6–7)
Süßwaren	67(64–71)	65(61–69)	56(54–58)	73(69–77)	58(55–60)

Fettgedruckt sind die verzehrten Lebensmittelmengen der Gruppen, die zur Bildung der Blätter geführt haben.

<sup>1</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

**Tab. 4.20: Prozentuale Verteilung der Bio-Käuferinnen und Nicht-Bio-Käuferinnen auf die einzelnen Blätter des Entscheidungsbaums (Frauen)**

	1. Blatt	2. Blatt	3. Blatt	4. Blatt	5. Blatt	6. Blatt	7. Blatt	8. Blatt	9. Blatt
<b>Bio-Käuferinnen</b>	21,1	54,7	68,3	82,7	36,8	33,7	45,5	57,1	59,3
<b>Nicht-Bio-Käuferinnen</b>	78,3	45,3	31,7	17,3	63,2	66,3	54,5	43,0	40,7

Das **erste Blatt** ist mit 46 Frauen von allen gebildeten Blättern das kleinste Blatt. Frauen, die hier zugeordnet wurden, sind durch einen Fleischverzehr von maximal 35,354 g/Tag (MW = 19 g/Tag), einen Getreideverzehr von maximal 339,328 g/Tag (MW = 140 g/Tag), einen Gemüseverzehr von maximal 135,533 g/Tag (MW = 70 g/Tag) sowie einen Obstverzehr von maximal 81,000 g/Tag (MW = 31 g/Tag) gekennzeichnet. Frauen dieses Blatts verzehren im Vergleich zu den Frauen der weiteren Blätter am wenigsten Obst und Gemüse. Die Lebensmittelgruppen Kartoffeln, Nüsse und Samen, Milch, Eier, Fisch, Streichfett, Getränke, Alkohol und Süßwaren spielten bei der Bildung dieses Blatts keine Rolle (Tab. 4.19, Abb. 4.9). Diesem Blatt sind mit ca. 78 % die meisten Nicht-Bio-Käuferinnen zugeordnet (Tab. 4.20).

Frauen des **zweiten Blatts** verzehren maximal 35,354 g Fleisch/Tag (MW = 20 g/Tag), maximal 339,328 g Getreide/Tag (MW = 174 g/Tag), maximal 135,533 g Gemüse/Tag (MW = 89 g/Tag) und in Abgrenzung zur Gruppe des ersten Blatts mindestens 81,000 g Obst/Tag (MW = 333 g/Tag) (Tab. 4.19, Abb. 4.9). Diesem Blatt sind 55 % Bio-Käuferinnen und 45 % Nicht-Bio-Käuferinnen zugeordnet (Tab. 4.20).

Das **dritte Blatt** gruppiert Frauen, die maximal 35,354 g Fleisch/Tag (MW = 19 g/Tag), maximal 339,328 g Getreide/Tag (MW = 202 g/Tag), sowie mindestens 135,533 g Gemüse/Tag (MW = 316 g/Tag) verzehren (Tab. 4.19, Abb. 4.9). Dieses Blatt enthält 68 % Bio-Käuferinnen und 32 % Nicht-Bio-Käuferinnen (Tab. 4.20).

Frauen des **vierten Blatts** verzehren im Vergleich zu Frauen der weiteren Blätter mit maximal 35,354 g Fleisch/Tag (MW = 13 g/Tag) am wenigsten Fleisch und mit mindestens 339,328 g/Tag (MW = 424 g/Tag) am meisten Getreide (Tab. 4.19, Abb. 4.9). Diesem Blatt sind mit knapp 83 % die meisten Bio-Käuferinnen zugeordnet (Tab. 4.20).

Den Frauen aus den weiteren fünf Blättern ist gemeinsam, dass sie mehr als 35,354 g Fleisch/Tag verzehren.

Frauen aus dem **fünften Blatt** verzehren durchschnittlich 87 g Fleisch/Tag sowie maximal 7,777 g Fisch/Tag (MW = 3 g/Tag) (Tab. 4.19, Abb. 4.9). Von allen Blättern weisen Frauen dieses Blatts den geringsten Fischverzehr auf (Tab. 4.19). In diesem Blatt sind 37 % Bio-Käuferinnen und 63 % Nicht-Bio-Käuferinnen vertreten (Tab. 4.20).

Frauen, die dem **sechsten Blatt** zugeordnet wurden, verzehren durchschnittlich 94 g Fleisch/Tag, mehr als 7,777 g Fisch/Tag (MW = 24 g/Tag), maximal 286,839 g Gemüse/Tag

(MW = 160 g/Tag), sowie maximal 139,929 g Obst/Tag (MW = 77 g/Tag) (Tab. 4.19, Abb. 4.9). Dieses Blatt enthält 34 % Bio-Käuferinnen und 66 % Nicht-Bio-Käuferinnen (Tab. 4.20).

Das **siebte Blatt** ist mit 1578 Frauen das größte Blatt. Frauen, die hier zugeordnet wurden, verzehren durchschnittlich 80 g Fleisch/Tag, mindestens 7,777 g Fisch/Tag (MW = 27 g/Tag), maximal 286,839 g Gemüse/Tag (MW = 183 g/Tag), mindestens 139,929 g Obst/Tag (MW = 336 g/Tag) sowie maximal 280,512 g Getreide/Tag (MW = 194 g/Tag) (Tab. 4.19, Abb. 4.9). Dieses Blatt enthält 46 % Bio-Käuferinnen und 55 % Nicht-Bio-Käuferinnen (Tab. 4.20).

Frauen, die dem **achten Blatt** zugeordnet wurden, verzehren durchschnittlich 90 g Fleisch/Tag, mindestens 7,777 g Fisch/Tag (MW = 28 g/Tag), maximal 286,839 g Gemüse/Tag (MW = 188 g/Tag), mindestens 139,929 g Obst/Tag (MW = 323 g/Tag) sowie mindestens 280,512 g Getreide/Tag (MW = 346 g/Tag) (Tab. 4.19, Abb. 4.9). Dieses Blatt enthält 57 % Bio-Käuferinnen und 43 % Nicht-Bio-Käuferinnen (Tab. 4.20).

Frauen, die dem **neunten Blatt** zugeordnet wurden, verzehren durchschnittlich 91 g Fleisch/Tag, mehr als 7,777 g Fisch/Tag (MW = 33 g/Tag) sowie mindestens 286,839 g Gemüse/Tag (MW = 414 g/Tag) (Tab. 4.19, Abb. 4.9). Frauen, die diesem Blatt zugeordnet wurden, verzehren im Vergleich zu den Frauen der weiteren Blätter am meisten Gemüse und Fisch (Tab. 4.19). Dieses Blatt enthält 59 % Bio-Käuferinnen und 41 % Nicht-Bio-Käuferinnen (Tab. 4.20).

Zur Darstellung des Lebensmittelverzehr der Frauen wurden bei der Bildung der neun Blätter von den insgesamt 13 zur Verfügung stehenden Lebensmittelgruppen lediglich die fünf Lebensmittelgruppen Fleisch, Getreide, Gemüse, Obst und Fisch herangezogen. Die Lebensmittelgruppen Kartoffeln, Milch, Eier, Streichfett, Getränke, Alkohol, Süßwaren, Nüsse und Samen spielen bei der Aufteilung der Frauen in die Gruppen der Bio-Käuferinnen und Nicht-Bio-Käuferinnen bei dieser Methode keine Rolle. Das bedeutet, dass sich der Lebensmittelverzehr der Bio-Käuferinnen und Nicht-Bio-Käuferinnen dieser Lebensmittelgruppen nicht ausreichend unterscheidet, um die beiden Gruppen mit der Methode des Entscheidungsbaums daraufhin zu differenzieren.

Die Klassifizierungsrate des gesamten Entscheidungsbaums weist eine Korrekt- zu Falschklassifikationsrate von 61,2 % zu 38,8 % auf, wobei der Anteil der richtig klassifizierten Nicht-Bio-Käuferinnen mit 63,8 % höher ist als der Anteil der richtig klassifizierten Bio-Käuferinnen mit 58,6 % (Tab. 4.21).

**Tab. 4.21: Korrekt- zu Falschklassifikationsrate des Entscheidungsbaums der Frauen in die Gruppe der Bio-Käuferinnen und Nicht-Bio-Käuferinnen auf Basis der 13 Lebensmittelgruppen**

	<b>Nicht-Bio-Käuferinnen</b>	<b>Bio-Käuferinnen</b>
<b>Nicht-Bio-Käuferinnen</b>	TN: 2256 (63,8 %)	FN: 1280 (36,2 %)
<b>Bio-Käuferinnen</b>	FP: 1481 (41,4 %)	TP: 2096 (58,6 %)

Die Korrektklassifikationsrate des Entscheidungsbaums unter 10facher Kreuzvalidierung und ohne Validierung ist vergleichbar und liegt bei 57,1 % und 61,2 %. Daher ist davon auszugehen, dass der Entscheidungsbaum ohne Validierung nicht zufällig zustande gekommen ist (Tab. 4.22).

**Tab. 4.22: Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Bio-Käuferinnen und Nicht-Bio-Käuferinnen mit und ohne 10fache Kreuzvalidierung des Entscheidungsbaums (Frauen)**

	<b>ohne Validierung</b>	<b>10fach Validierung</b>
<b>Korrektklassifikation</b>	61,2 %	57,1 %
<b>Falschklassifikation</b>	38,8 %	42,9 %

Ebenso wie bei den Ergebnissen der bisher beschriebenen Methoden können auch diese Ergebnisse mit weiteren Variablen verknüpft werden, um die den gebildeten Blättern zugeordneten Frauen weiter zu charakterisieren. Beispielhaft wird dies anhand der beiden Blätter aufgezeigt, denen die meisten Bio-Käuferinnen bzw. Nicht-Bio-Käuferinnen richtig zugeordnet wurden. Die beste Zuordnung der Bio-Käuferinnen wurde im vierten Blatt und die der Nicht-Bio-Käuferinnen im ersten Blatt erzielt (Tab. 4.23).

**Tab. 4.23: Charakterisierung der Gesamtgruppe sowie der Frauen, die dem ersten und vierten Blatt des Entscheidungsbaums zugeordnet wurden**

		Gesamtgruppe (%)	1. Blatt (%)	4. Blatt (%)	Signifikanzen <sup>4</sup>
<b>Bio-Kauf</b>	ja	50,3	21,7	82,7	<0,001
	nein	49,7	78,3	17,3	<0,001
<b>Altersgruppe</b>	18–34 Jahre	20,7	41,3	32,4	n. s.
	35–50 Jahre	35,3	28,3	41,6	n. s.
	51–64 Jahre	24,2	15,2	16,2	n. s.
	65–80 Jahre	19,8	15,2	9,8	n. s.
<b>Soziale Schicht<sup>1</sup></b>	untere Schicht	21,0	41,3	11,6	<0,001
	mittlere Schicht	29,9	41,3	23,7	<0,05
	obere Schicht	49,1	17,4	64,7	<0,001
<b>BMI</b>	Untergewicht	1,6	0,0	5,2	n. s.
	Normalgewicht	46,0	54,4	71,7	<0,05
	Übergewicht	27,9	28,3	13,9	<0,05
	Adipositas	18,9	10,9	5,2	n. s.
	keine Angabe	5,6	6,5	4,1	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 88 cm	65,8	71,7	85,6	<0,05
	über 88 cm	29,4	23,9	9,3	<0,01
	keine Angabe	4,8	4,4	5,2	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüftumfang)</b>	bis 0,85	74,2	80,4	87,3	n. s.
	über 0,85	21,0	15,2	7,5	n. s.
	keine Angabe	4,8	4,4	5,2	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW (95. KI des MW))	1968 (1954–1981)	1286 (1146–1426)	2370 (2286–2455)	signifikant
<b>Ernährungs-kenntnisse</b>	gut	40,9	26,1	54,3	<0,001
	mittel	33,9	32,6	31,2	n. s.
	schlecht	24,2	41,3	14,5	<0,001
	keine Angabe	1,0	0,0	0,0	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	58,9	32,6	76,3	<0,001
	nein	40,6	67,4	23,7	<0,001
	keine Angabe	0,6	0,0	0,0	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	21,7	45,7	12,1	<0,001
	nein	78,0	54,4	86,7	<0,001
	keine Angabe	0,3	0,0	1,2	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>3</sup></b>	gut	76,2	60,9	85,0	<0,001
	mittel	20,2	28,3	9,8	<0,01
	schlecht	3,5	10,9	4,6	n. s.
	keine Angabe	0,1	0,0	0,6	--- <sup>2</sup>
<b>Schlafqualität<sup>3</sup></b>	gut	78,8	76,1	78,6	n. s.
	schlecht	20,7	23,9	20,8	n. s.
	keine Angabe	0,5	0,0	0,6	--- <sup>2</sup>
<b>HEI-NVS II</b>	MW(95. KI des MW)	84,8(84,6–85,1)	65,0(62,6–67,3)	89,4(88,1–90,8)	signifikant

<sup>1</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Selbsteinschätzung der Teilnehmenden

<sup>4</sup>Das Signifikanzniveau p gibt an, ob ein Unterschied zwischen dem 1. und 4. Blatt vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Energiezufuhr und HEI-NVS II über die Konfidenzintervalle).

Ein Vergleich der Frauen, die dem ersten und vierten Blatt zugeordnet wurden zeigt, dass Frauen aus dem ersten Blatt, dem 78 % Nicht-Bio-Käufer angehören, im Vergleich zu Frauen aus dem vierten Blatt, dem knapp 83 % Bio-Käufer zugeordnet wurden, zu einem höheren Anteil der unteren und mittleren sozialen Schicht und zu einem geringeren Anteil der oberen sozialen Schicht angehören. Frauen aus dem ersten Blatt weisen, im Vergleich zu Frauen

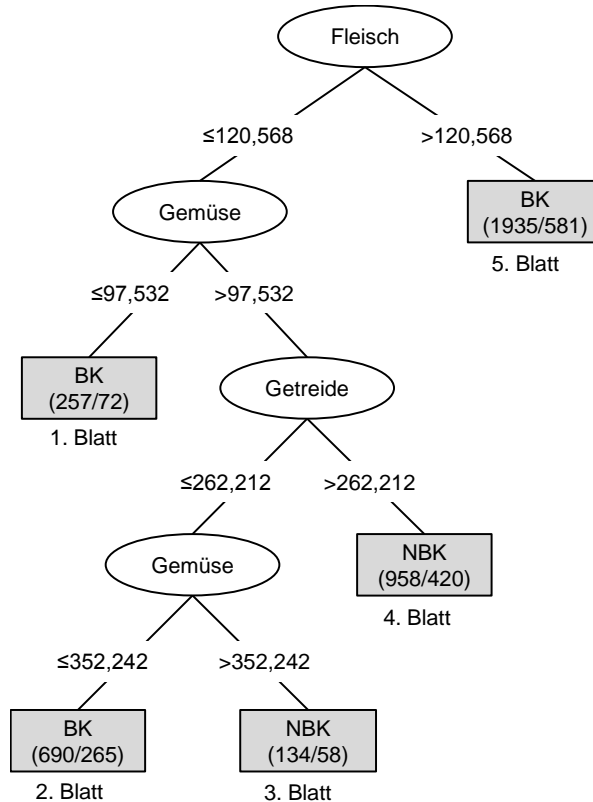
aus dem vierten Blatt, einen geringeren Anteil an normalgewichtigen, einen höheren Anteil an übergewichtigen Personen auf und gehören zu einem höheren Anteil der Gruppe mit einem erhöhten Taillenumfang an. Frauen, die dem ersten Blatt zugeordnet wurden, nehmen weniger Kalorien auf als Frauen, die dem vierten Blatt zugeordnet wurden, und weisen einen geringeren Anteil an Personen mit guten und einen höheren Anteil an Personen mit schlechten Ernährungskennnissen auf. Frauen aus Blatt eins geben zu einem höheren Anteil an, sportlich nicht aktiv zu sein, Raucher zu sein, und zu einem geringeren Anteil gute und zu einem höheren Anteil einen mittleren Gesundheitszustand zu besitzen. Frauen aus dem ersten Blatt weisen im Vergleich zu allen weiteren Blättern mit durchschnittlich 65 Punkten die geringste HEI-NVS II-Indexsumme auf.

Bei den weiteren Faktoren konnten keine Unterschiede zwischen Frauen, die dem ersten und Frauen, die dem vierten Blatt zugeordnet wurden, festgestellt werden.

Die Charakterisierung der Frauen, die den weiteren Blättern zugeordnet wurden, sind im Anhang dargestellt (Kapitel 9.3).

### Entscheidungsbaum der Männer

Die Gruppe der Männer konnte über den Lebensmittelverzehr von 13 Lebensmittelgruppen anhand der Klassifizierungsvariable Bio-Käufer versus Nicht-Bio-Käufer in insgesamt fünf Blätter aufgeteilt werden (Abb. 4.10).



Die Lebensmittelverzehrsmengen auf dem Pfad sind als kleiner/gleich und größer als Angaben in Gramm pro Tag angegeben. Die Angabe BK = Bio-Käufer und NBK = Nicht-Bio-Käufer weisen auf ein von Bio-Käufern oder Nicht-Bio-Käufern dominiertes Blatt hin. Die Zahlen in Klammer geben die Anzahl der Bio-Käufer und Nicht-Bio-Käufer an, wobei immer der größere Anteil zuerst genannt ist.

**Abb. 4.10: Entscheidungsbaum der Bio-Käufer und Nicht-Bio-Käufer auf Basis der Verzehrsmenge von 13 Lebensmittelgruppen (Männer)**

Wie bei den Frauen zeigt sich, dass sich die Männer bezogen auf die Klassifizierungsvariable Bio-Kauf am besten über die Lebensmittelgruppe Fleisch differenzieren lassen. Die Gesamtgruppe der Männer lässt sich auf Basis des Trennwertes von 120,568 g Fleisch/Tag in zwei fast gleich große Gruppen einteilen. Der Gruppe mit dem höheren Fleischverzehr wurden 31 % Bio-Käufer und 69 % Nicht-Bio-Käufer zugeordnet. D. h. wenn ein Mann mehr als 120,568 g Fleisch/Tag verzehrt, ist er mit einer knapp 70 %igen Wahrscheinlichkeit ein Nicht-Bio-Käufer.

Die Darstellung des Lebensmittelverzehrs der in einem Blatt gruppierten Männer erfolgt durch die Beschreibung des Pfades von der Wurzel zu den gebildeten Blättern. Hierfür wurde zusätzlich zu den im Entscheidungsbaum verwendeten Trennwerten (kleiner/gleich und größer als) an den einzelnen Knoten die durchschnittlich verzehrten Lebensmittel-

mengen mit den entsprechenden 95. Konfidenzintervallen herangezogen, um die einzelnen Blätter miteinander zu vergleichen (Tab. 4.24). Der Anteil der Bio-Käufer und Nicht-Bio-Käufer der einzelnen Blätter des Entscheidungsbaums sind in Tab. 4.25 dargestellt.

**Tab. 4.24: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer, die den einzelnen Blättern des Entscheidungsbaums zugeordnet wurden**

Lebensmittelgruppen	Gesamtgruppe MW (95. KI des MW)		Blätter MW (95. KI des MW)			
	1. Blatt (n=408)	2. Blatt (n=1036)	3. Blatt (n=203)	4. Blatt (n=1407)	5. Blatt (n=2907)	
Obst	249(243–255)	181(162–200)	268(254–281)	373(335–411)	294(281–307)	221(214–228)
Gemüse	239(235–242)	<b>65(63–68)</b>	<b>206(202–210)</b>	<b>480(463–497)</b>	<b>267(260–274)</b>	244(239–249)
Getreide	303(299–306)	257(245–270)	<b>192(189–195)</b>	<b>194(187–200)</b>	<b>369(364–375)</b>	324(319–329)
Kartoffeln	92(91–94)	64(60–69)	83(80–86)	84(77–92)	82(80–85)	105(103–107)
Nüsse und Samen	5(4–5)	3(2–5)	3(3–4)	7(5–9)	5(4–6)	5(4–5)
Milch	249(243–256)	238(211–264)	229(214–243)	264(226–303)	264(251–276)	250(240–260)
Eier	16(16–17)	13(11–15)	13(12–14)	14(12–16)	14(13–14)	19(18–20)
Fisch	29(28–30)	19(17–22)	28(26–29)	40(35–45)	30(28–31)	30(29–31)
Fleisch	133(131–135)	<b>73(70–76)</b>	<b>76(75–78)</b>	<b>71(67–76)</b>	<b>78(77–80)</b>	<b>193(190–196)</b>
Streichfett	30(29–30)	28(26–31)	22(21–23)	21(18–23)	29(28–30)	34(33–35)
Getränke	2332(2307–2357)	1967(1873–2061)	2124(2067–2182)	2369(2235–2502)	2365(2316–2413)	2438(2402–2474)
Alkohol <sup>1</sup>	16(15–16)	13(11–15)	15(14–16)	16(14–19)	12(11–13)	18(18–19)
Süßwaren	68(67–70)	66(61–72)	54(50–57)	46(40–53)	68(65–71)	76(73–78)

Fettgedruckt sind die verzehrten Lebensmittelmengen der Gruppen, die zur Bildung der Blätter geführt haben.

<sup>1</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

**Tab. 4.25: Prozentuale Verteilung der Bio-Käufer und Nicht-Bio-Käufer auf die einzelnen Blätter des Entscheidungsbaums (Männer)**

	1. Blatt	2. Blatt	3. Blatt	4. Blatt	5. Blatt
<b>Bio-Käufer</b>	31,1	37,1	55,7	54,7	31,1
<b>Nicht-Bio-Käufer</b>	68,9	62,9	44,3	45,3	68,9

Männer der ersten vier Blätter sind alle durch einen Fleischverzehr von weniger als 120,568 g/Tag gekennzeichnet. Daneben lassen sie sich wie folgt beschreiben:

Männer des **ersten Blatts** verzehren durchschnittlich 73 g Fleisch/Tag und weisen einen Gemüseverzehr von maximal 97,532 g/Tag (MW = 65 g/Tag) auf (Abb. 4.10, Tab. 4.24). Von allen Blättern lassen sie sich durch den geringsten Gemüseverzehr kennzeichnen (Tab. 4.24). Diesem Blatt sind 69 % Bio-Käufer und 31 % Nicht-Bio-Käufer zugeordnet (Tab. 4.25).

Männer des **zweiten Blatts** verzehren durchschnittlich 76 g Fleisch/Tag, maximal 262,212 g Getreide/Tag (MW = 192 g/Tag) und weisen einen Gemüseverzehr zwischen 97,532 g und 352,242 g/Tag (MW = 206 g/Tag) auf (Abb. 4.10, Tab. 4.24). Dieses Blatt enthält 37 % Bio-Käufer und 63 % Nicht-Bio-Käufer (Tab. 4.25).

Bei dem **dritten Blatt** handelt es sich mit 203 Männern um das kleinste Blatt. Männer dieses Blatts zeichnen sich durch einen durchschnittlichen Fleischverzehr von 71 g/Tag, einen Getreideverzehr von maximal 262,212 g/Tag (MW = 194 g/Tag) sowie einen Gemüseverzehr von mindestens 352,242 g/Tag (MW = 480 g/Tag) aus (Abb. 4.10, Tab. 4.24). Von allen

Blättern verzehren diese Männer am meisten Gemüse und weisen mit einem Anteil von 56 % den höchsten Bio-Käuferanteil auf (Tab. 4.25).

Das **vierte Blatt** gruppiert Männer, die ebenso wie Männer aus Blatt zwei und drei maximal 120,568 g Fleisch/Tag verzehren (MW = 78 g/Tag) und einen Gemüseverzehr von mindestens 97,532 g/Tag (MW = 267 g/Tag) aufweisen. Daneben lassen sie sich durch einen Getreideverzehr von mindestens 262,212 g/Tag (MW = 369 g/Tag) kennzeichnen (Abb. 4.10, Tab. 4.24). Von allen Blättern verzehren diese Männer am meisten Getreide (Tab. 4.24). Diesem Blatt wurden 55 % Bio-Käufer und 45 % Nicht-Bio-Käufer zugeordnet (Tab. 4.25).

Das **fünfte Blatt** ist mit 2907 Männern, und damit mit knapp der Hälfte der Männer, das größte Blatt. Diese Männer lassen sich ausschließlich über einen Fleischverzehr von mindestens 120,568 g/Tag (MW = 193 g/Tag) charakterisieren. Sie weisen von allen Blättern den höchsten Fleischverzehr auf (Abb. 4.10, Tab. 4.24). Diese Personengruppe kann durch keine weitere Lebensmittelgruppe aufgeteilt werden. Diesem Blatt sind mit 31 % die wenigsten Bio-Käufer und mit 69 % die meisten Nicht-Bio-Käufer zugeordnet (Tab. 4.25).

Zur Darstellung des Lebensmittelverzehrs der Männer wurden bei der Bildung der fünf Blätter von den 13 einbezogenen Lebensmittelgruppen lediglich Fleisch, Gemüse und Getreide herangezogen. Die neun Lebensmittelgruppen Obst, Kartoffeln, Fisch, Milch, Eier, Streichfett, Getränke, Alkohol, Süßwaren, Nüsse und Samen haben zur Klassifizierung der Männer in die Gruppen der Bio-Käufer und Nicht-Bio-Käufer bei dieser Methode nicht beigetragen. Das bedeutet, dass sich der Lebensmittelverzehr der Bio-Käufer und Nicht-Bio-Käufer dieser Lebensmittelgruppen nicht ausreichend unterscheidet, um die beiden Gruppen mit der Methode des Entscheidungsbaums zu differenzieren.

Die Klassifizierungsrate des Entscheidungsbaums weist eine Korrekt- zu Falschklassifikationsrate von 64,1 % zu 35,9 % auf und liegt damit über der Klassifikationsrate der Frauen. Wie bei den Frauen ist der Anteil der richtig klassifizierten Nicht-Bio-Käufer mit 80,2 % höher als der Anteil der richtig klassifizierten Bio-Käufer mit 38,4 % (Tab. 4.26).

**Tab. 4.26: Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Männer in die Gruppe der Bio-Käufer und Nicht-Bio-Käufer auf Basis der 13 Lebensmittelgruppen**

	<b>Nicht-Bio-Käufer</b>	<b>Bio-Käufer</b>
<b>Nicht-Bio-Käufer</b>	TN: 2936 (80,2 %)	FN: 727 (19,8 %)
<b>Bio-Käufer</b>	FP: 1415 (61,6 %)	TP: 883 (38,4 %)

Die Korrektklassifikationsrate des Entscheidungsbaums unter 10facher Kreuzvalidierung und ohne Validierung ist vergleichbar und liegt bei 62,8 % bzw. 64,1 % Es ist also davon auszugehen, dass der Entscheidungsbaum ohne Validierung nicht zufällig zustande gekommen ist (Tab. 4.27).

**Tab. 4.27: Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Bio-Käufer und Nicht-Bio-Käufer mit und ohne 10fache Kreuzvalidierung des Entscheidungsbaums (Männer)**

	ohne Validierung	10fach Validierung
<b>Korrektklassifikation</b>	64,1 %	62,8 %
<b>Falschklassifikation</b>	35,9 %	37,2 %

Wie bei den Frauen werden im Folgenden zwei Blätter ausgewählt, um die dort zugeordneten Männer näher zu charakterisieren. Beispielhaft wird dies anhand der beiden Blätter aufgezeigt, denen die meisten Bio-Käufer oder Nicht-Bio-Käufer richtig zugeordnet wurden. Die beste Zuordnung der Bio-Käufer und Nicht-Bio-Käufer erfolgte im ersten bzw. fünften Blatt und im dritten Blatt (Tab. 4.28). Das erste und fünfte Blatt enthalten mit 69 % den gleichen Bio-Käuferanteil. Für den Vergleich der Blätter wird im folgenden Blatt, das absolut mehr Bio-Käufer enthält herangezogen.

**Tab. 4.28: Charakterisierung der Gesamtgruppe sowie der Männer, die dem dritten und fünften Blatt des Entscheidungsbaums zugeordnet wurden**

		Gesamtgruppe (%)	3. Blatt (%)	5. Blatt (%)	Signifikanzen <sup>4</sup>
<b>Bio-Kauf</b>	ja	38,6	55,7	31,1	<0,001
	nein	61,5	44,3	68,9	<0,001
<b>Altersgruppe</b>	18–34 Jahre	19,7	10,4	25,8	<0,001
	35–50 Jahre	32,0	28,1	34,7	n. s.
	51–64 Jahre	25,3	26,1	23,7	n. s.
	65–80 Jahre	23,0	35,5	15,8	<0,001
<b>Soziale Schicht<sup>1</sup></b>	untere Schicht	17,6	14,3	19,3	n. s.
	mittlere Schicht	28,4	28,6	29,1	n. s.
	obere Schicht	54,1	57,1	51,6	n. s.
<b>BMI</b>	Untergewicht	0,7	0,0	0,8	n. s.
	Normalgewicht	30,0	18,2	31,2	<0,001
	Übergewicht	45,0	48,3	43,4	n. s.
	Adipositas	20,1	26,1	21,4	n. s.
	keine Angabe	4,3	7,4	3,2	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 102 cm	68,6	62,6	68,8	n. s.
	über 102 cm	28,0	32,5	28,4	n. s.
	keine Angabe	3,4	4,9	2,9	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüft-umfang)</b>	bis 1,0	81,2	79,3	81,8	n. s.
	über 1,0	15,4	15,8	15,3	n. s.
	keine Angabe	3,4	4,9	2,9	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW (95. KI des MW))	2569 (2548–2591)	2095 (2018–2173)	2914 (2881–2946)	signifikant
<b>Ernährungs-kenntnisse</b>	gut	17,1	21,2	15,8	<0,05
	mittel	38,2	35,0	39,4	n. s.
	schlecht	43,8	41,9	44,2	n. s.
	keine Angabe	0,9	2,0	0,6	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	56,2	61,1	55,1	n. s.
	nein	43,4	37,9	44,6	n. s.
	keine Angabe	0,5	1,0	0,3	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	28,1	22,7	33,4	<0,01
	nein	71,6	76,8	66,3	<0,01
	keine Angabe	0,2	0,5	0,3	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>3</sup></b>	gut	75,2	70,5	75,6	n. s.
	mittel	20,7	22,2	21,3	n. s.
	schlecht	3,9	6,9	2,9	<0,01
	keine Angabe	0,1	0,5	0,1	--- <sup>2</sup>
<b>Schlafqualität<sup>3</sup></b>	gut	83,6	78,8	82,9	n. s.
	schlecht	16,0	20,7	16,8	n. s.
	keine Angabe	0,4	0,5	0,3	--- <sup>2</sup>
<b>HEI-NVS II</b>	MW(95. KI des MW)	80,3(80,0–80,5)	89,9(88,7–91,2)	77,1(76,7–77,5)	signifikant

<sup>1</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Selbsteinschätzung der Teilnehmenden

<sup>4</sup>Das Signifikanzniveau p gibt an, ob ein Unterschied zwischen dem 3. und 5. Blatt vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Energiezufuhr und HEI-NVS II über die Konfidenzintervalle).

Ein Vergleich der Männer, die dem dritten und fünften Blatt zugeordnet wurden, zeigt, dass Männer aus dem dritten Blatt, dem 56 % Bio-Käufer angehören, im Vergleich zu Männern aus dem fünften Blatt, dem knapp 69 % Nicht-Bio-Käufer zugeordnet wurden, zu einem geringeren Anteil der Altersgruppe der 18- bis 34-Jährigen und zu einem höheren Anteil der 65- bis 80-Jährigen angehören. Zudem weisen Männer aus dem dritten Blatt einen geringeren Anteil an normalgewichtigen Personen und eine geringere Energiezufuhr auf.

Daneben zeichnen sich Männer aus dem dritten Blatt im Vergleich zu Männern aus dem fünften Blatt durch einen höheren Anteil an Personen, die gute Ernährungskennntnisse besitzen aus, gehören zu einem höheren Anteil der Gruppe der Nichtraucher und der Gruppe, die angeben einen schlechten Gesundheitszustand zu haben an, als Männer, die dem fünften Blatt zugeordnet wurden. Männer aus dem dritten Blatt erreichen im Vergleich zu den weiteren Blättern die höchste HEI-NVS II-Indexsumme. Bei allen weiteren Variablen konnten keine Unterschiede zwischen den Männern des dritten und fünften Blatts festgestellt werden.

Die Charakterisierung der Männer, die den weiteren Blättern zugeordnet wurden, sind im Anhang dargestellt (Kapitel 9.3).

### 4.3.6 Methodische Ergebnisse und Diskussion

Es wurde untersucht, ob sich der Entscheidungsbaum unter Anwendung des Verfahrens J48 eignet, den Lebensmittelverzehr umfassend darzustellen. Hierbei wurden Frauen und Männer jeweils in Abhängigkeit der Klassifizierungsvariable Bio-Käufer/innen versus Nicht-Bio-Käufer/innen auf Basis ihrer Unterschiede im Lebensmittelverzehr von 13 einbezogenen Lebensmittelgruppen voneinander getrennt. Der Lebensmittelverzehr wird durch die entstandenen Blätter dargestellt. Hierbei werden Personen mit einem ähnlichen Lebensmittelverzehr gruppiert.

#### **Kriterium: Berücksichtigung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen bei Anwendung der Methode**

Bei der Bildung des Entscheidungsbaums können eine Vielzahl an Variablen einbezogen werden, da bei jedem Knoten über das Verfahren neu entschieden wird, welche Variable sich am besten zur Klassifizierung eignet (Quinlan 1993 S 22). Wird eine Variable nicht verwendet, trägt sie nicht zur Klassifizierung bei (Beierle und Kern-Isberner 2006 S 106, Petersohn 2005 S 146). Für die vorliegende Auswertung konnten alle 13 für das Untersuchungsziel relevanten Lebensmittelgruppen bei der Anwendung des Entscheidungsbaums berücksichtigt werden.

Das Kriterium, möglichst alle Lebensmittelgruppen mit den dazugehörigen verzehrten Lebensmittelmengen bei der Anwendung der Methode zu berücksichtigen, wird erfüllt.

#### **Kriterium: Darstellung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse**

Die Darstellung des Lebensmittelverzehrs erfolgt durch die entstandenen Blätter, in denen Personen mit einem ähnlichen Lebensmittelverzehr gruppiert werden. Hierfür werden die Personen auf Basis von Unterschieden im Lebensmittelverzehr, in Abhängigkeit einer Klassifizierungsvariablen voneinander getrennt. An jedem Knoten des Baums wird neu ermittelt, welche Lebensmittelgruppe sich am besten für die Differenzierung der dort vorliegenden Studiengruppe eignet. Damit werden, bedingt durch das Berechnungsverfahren, bei der Bildung des Entscheidungsbaums, nur Lebensmittelgruppen herangezogen, die einen Beitrag zur Differenzierung der Studiengruppe leisten. Ist eine Lebensmittelgruppe zur Differenzierung der Personengruppe an einem Knoten nicht relevant, wird sie nicht verwendet. Damit wurden in der vorliegenden Auswertung zwar bei der Berechnung des Entscheidungsbaums alle 13 in dem Verfahren berücksichtigten Lebensmittelgruppen mit den entsprechenden Verzehrsmengen einbezogen, aber nur die, die für die Differenzierung relevant waren, dargestellt. Bei beiden Geschlechtern wurde nur ein kleiner

Teil der 13 Lebensmittelgruppen zur Differenzierung verwendet. Dazu gehören Fleisch, Gemüse und Getreide und zusätzlich bei den Frauen die Lebensmittelgruppen Obst und Fisch. Alle weiteren Lebensmittelgruppen wurden nicht verwendet und sind damit als nicht relevant für die Differenzierung der Bio-Käufer/innen und Nicht-Bio-Käufer/innen anzusehen.

Das Kriterium, möglichst alle Lebensmittelgruppen mit den dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse darzustellen, wird nicht erfüllt.

### **Kriterium: Identifizierung von Personengruppen mit ähnlichem Lebensmittelverzehr**

Durch den Entscheidungsbaum werden über die Blätter Objekte (hier: Studienteilnehmende) gruppiert, die untereinander möglichst homogen und zu den Objekten der weiteren Blätter möglichst heterogen sind (Christof 2007 S 127f). Das bedeutet, dass die in den gebildeten Blättern gruppierten Personen einen ähnlichen Lebensmittelverzehr aufweisen. Allerdings können hierfür nur die Lebensmittelgruppen, die zur Differenzierung der in einem Blatt gruppieren Personen beigetragen haben, zur Beschreibung des ähnlichen Lebensmittelverzehrs herangezogen werden.

Das Kriterium, Personengruppen mit einem ähnlichen Lebensmittelverzehr zu identifizieren, wird erfüllt.

### **Kriterium: Differenzierung von Bio-Käufer/innen und Nicht-Bio-Käufer/innen**

Das Prinzip des Entscheidungsbaums ist, Objekte auf Basis einer Klassifizierungsvariable voneinander zu trennen bzw. zu differenzieren (Kokai 2002 S 98). Dabei wird das Ziel verfolgt, möglichst alle Objekte (hier: Bio-Käufer/innen und Nicht-Bio-Käufer/innen) richtig zu klassifizieren (Petersohn 2005 S 145). Klassifikationskriterien, wie die Korrekt- und Falschklassifikationsrate liefern Auskunft darüber, wie gut oder schlecht sich die Objekte differenzieren lassen (Runkler 2010 S 86):

Die Korrekt- zu Falschklassifikationsrate ist sowohl bei den Frauen als auch bei den Männern mit 61,2 % zu 38,8 % bzw. 64,1 % zu 35,9 % sehr gering. Zudem ist auffällig, dass bei beiden Geschlechtern mehr Nicht-Bio-Käufer/innen als Bio-Käufer/innen korrekt klassifiziert wurden. Der Anteil richtig klassifizierter Bio-Käuferinnen beträgt 58,6 %, der der Nicht-Bio-Käuferinnen 63,8 %. Bei den Männern ist der Anteil richtig klassifizierter Bio-Käufer mit 38,4 % noch geringer als bei den Frauen. Der Anteil richtig klassifizierter Nicht-Bio-Käufer beträgt 80,2 %. Allerdings muss hier darauf aufmerksam gemacht werden, dass in der Gruppe der Männer der Anteil der Bio-Käufer zu Nicht-Bio-Käufern 39 % zu 61 % beträgt (Tab. 2.2). Die Ungleichverteilung dieser Gruppen beeinflusst die Zuordnungsquote. Dagegen ist in der Gruppe der Frauen die Verteilung der Bio-Käuferinnen zu Nicht-Bio-

Käuferinnen mit 49,7 % zu 50,4 % etwa gleich groß und hat damit keinen Einfluss auf die Zuordnungsquote (Tab. 2.2).

Auf Basis der Ergebnisse des HEI-NVS II (siehe Kapitel 3.1.4), bei denen deutlich mehr Unterschiede im Lebensmittelverzehr festgestellt wurden, wäre zu erwarten gewesen, dass höhere Klassifikationsraten erreicht werden. Möglicherweise lassen sich die Bio-Käufer/innen und Nicht-Bio-Käufer/innen nicht besser differenzieren.

Ein Vergleich dieser Ergebnisse mit anderen Studien ist nicht möglich. Entscheidungsbäume werden zwar im medizinischen Bereich häufig für die Diagnostik verschiedener Erkrankungen angewendet (Beispiele: Hische et al. 2010, Su et al. 2007, Pavlopoulos et al. 2004, Tsien et al. 1998, Mair et al. 1995), nicht aber zur Darstellung des Lebensmittelverzehrs.

Neben den genannten Ergebnissen liefert der Entscheidungsbaum auch die Lebensmittelgruppen, durch welche sich eine Personengruppe am besten differenzieren lässt. Der Aufbau des Baums spiegelt stufenweise die Bedeutung der Variablen (hier: Lebensmittelgruppen) wider, wodurch die für die Klassifizierung wesentlichen Variablen identifiziert werden können (von Rimscha 2010 S 116). Sowohl in der Gruppe der Frauen als auch in der Gruppe der Männer hat sich die Lebensmittelgruppe Fleisch als die am besten differenzierende Variable bezüglich des Bio-Kaufs herausgestellt. Auch wenn die Differenzierung der Bio-Käufer/innen und Nicht-Bio-Käufer/innen nicht erfolgreich war, bestätigt zumindest dieses Teilergebnis die bisherigen Ergebnisse der deskriptiven Auswertung (Kapitel 2.3.2) und der Ergebnisse des HEI-NVS II (Kapitel 3.1.4). Auch diese Ergebnisse zeigen, dass sich sowohl in der Gruppe der Frauen als auch in der Gruppe der Männer die Verzehrsmengen von Fleisch zwischen den Bio-Käufer/innen und Nicht-Bio-Käufer/innen stark unterscheiden. Es lässt sich festhalten, dass die in der vorliegenden Auswertung ermittelten Klassifikationsraten zu gering sind, um die Bio-Käufer/innen und Nicht-Bio-Käufer/innen zu differenzieren.

In diesem Zusammenhang ist zu erwähnen, dass auch ein Entscheidungsbaum nicht frei von subjektiven Entscheidungen ist und das Ergebnis durch unterschiedliche Herangehensweisen beeinflusst werden kann. Neben der Anzahl der zur Klassifizierung verwendeten Lebensmittelgruppen können bei diesem Verfahren durch zahlreiche Optimierungsmöglichkeiten wie dem Beschneiden von Ästen sowie das Festlegen einer Mindestanzahl von Objekten in einem Blatt das Ergebnis des Entscheidungsbaums beeinflussen. Verschiedene Herangehensweisen wie das Festlegen einer unterschiedlichen Mindestanzahl an Objekten sowie das Beschneiden von Ästen wurde getestet, führte allerdings nicht zu wesentlichen Unterschieden in der Korrekt- zu Falschklassifikationsrate von Bio-Käufer/innen und Nicht-Bio-Käufer/innen. Um grundsätzlich die Anwendung des Entscheidungsbaums zur Differenzierung von Personengruppen auf Basis von Lebensmittelverzehrdaten zu prüfen,

werden neben den beschriebenen Ergebnissen, bei denen die Klassifizierungsvariable **Bio-Käufer/innen versus Nicht-Bio-Käufer/innen** verwendet wurde, im Folgenden weitere Klassifizierungsvariablen herangezogen.

Zum einen wird die Gesamtgruppe der Bio-Käufer/innen auf Basis der Klassifizierungsvariable **Frauen versus Männer** voneinander getrennt. Bei dieser Variante wird eine Korrekt- zu Falschklassifizierung von 74 % zu 26 % erzielt, wobei Frauen mit einer Quote von 85,5 % häufiger als Männer mit einer Quote von 57,2 % richtig zugeordnet werden (Tab. 4.29).

**Tab. 4.29: Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Bio-Käufer/innen in die Gruppen der Frauen und Männer auf Basis von 13 Lebensmittelgruppen**

	Männer	Frauen
Männer	TN: 1315 (57,2 %)	FN: 983 (42,8 %)
Frauen	FP: 520 (14,5 %)	TP: 3057 (85,5 %)

TN = Richtig negativ: Männer werden als Männer klassifiziert  
 TP = Richtig positiv: Frauen werden als Frauen klassifiziert  
 FP = Falsch positiv: Frauen werden als Männer klassifiziert  
 FN = Falsch negativ: Männer werden als Frauen klassifiziert

Ebenso wie bei der Klassifikation nach Bio-Käufer/innen und Nicht-Bio-Käufer/innen der Frauen und Männer zeichnet sich bei diesem Entscheidungsbaum die Lebensmittelgruppe Fleisch als die am besten diskriminierende Variable ab, um die Gruppe der Bio-Käufer/innen auf Basis der Klassifizierungsvariable Frauen versus Männer zu differenzieren (siehe Anhang Abb. 9.3).

Daneben werden die Bio-Käuferinnen auf Basis der Klassifizierungsvariable **„Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“**, gemessen an der HEI-NVS II-Indexsumme (Aufteilung siehe Kapitel 3, Tab. 3.8), über einen Entscheidungsbaum voneinander getrennt. Bei diesem Entscheidungsbaum werden 92 % der Frauen auf Basis der Klassifizierungsvariablen „Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“ korrekt zugeordnet (Tab. 4.30). Der Entscheidungsbaum ist im Anhang (Abb. 9.4) dargestellt.

**Tab. 4.30: Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Bio-Käuferinnen in die Gruppen „Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“, gemessen an der HEI-NVS II-Indexsumme, auf Basis von 13 Lebensmittelgruppen**

	Empfehlungen gut erreicht	Empfehlungen weniger erreicht
Empfehlungen gut erreicht	TN: 1104 (92,6 %)	FN: 89 (7,4 %)
Empfehlungen weniger erreicht	FP: 98 (8,3 %)	TP: 1094 (91,7 %)

TN = Richtig negativ: Bio-Käuferinnen, die die Empfehlungen gut erreichen werden als solche klassifiziert  
 TP = Richtig positiv: Bio-Käuferinnen, die die Empfehlungen weniger erreichen werden als solche klassifiziert  
 FP = Falsch positiv: Bio-Käuferinnen, die die Empfehlungen weniger erreichen werden als Bio-Käuferinnen, die die Empfehlungen gut erreichen klassifiziert  
 FN = Falsch negativ: Bio-Käuferinnen, die die Empfehlung gut erreichen werden als Bio-Käuferinnen, die die Empfehlung weniger erreichen klassifiziert

Ein ähnliches Ergebnis wird für die Gruppe der Männer für die Bio-Käufer erzielt. Auf Basis der Klassifizierungsvariable „Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“ werden 92 % der Männer korrekt zugeordnet (Tab. 4.31). Der Entscheidungsbaum ist im Anhang (Abb. 9.5) dargestellt.

**Tab. 4.31: Korrekt- und Falschklassifikationsrate des Entscheidungsbaums der Bio-Käufer, in die Gruppen „Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“, gemessen an der HEI-NVS II-Indexsumme, auf Basis von 13 Lebensmittelgruppen**

	Empfehlungen gut erreicht	Empfehlungen weniger erreicht
Empfehlungen gut erreicht	TN: 730 (95,3 %)	FN: 36 (4,7 %)
Empfehlungen weniger erreicht	FP: 93 (12,1 %)	TP: 673 (87,9%)

TN = Richtig negativ: Bio-Käufer, die die Empfehlungen gut erreichen werden als solche klassifiziert  
 TP = Richtig positiv: Bio-Käufer, die die Empfehlungen weniger erreichen werden als solche klassifiziert  
 FP = Falsch positiv: Bio-Käufer, die die Empfehlungen weniger erreichen werden als Bio-Käufer, die die Empfehlungen gut erreichen klassifiziert  
 FN = Falsch negativ: Bio-Käufer, die die Empfehlungen gut erreichen werden als Bio-Käufer, die die Empfehlungen weniger erreichen klassifiziert

Das eigentliche Ziel des Entscheidungsbaums ist, eine direkte Differenzierung von Objekten bzw. in der vorliegenden Auswertung, eine Differenzierung von Personen vorzunehmen. Auf Basis der Klassifizierungsvariable Bio-Kauf ist dies nicht möglich. Durch die weiteren Auswertungen auf Basis der Daten der NVS II konnte aufgezeigt werden, dass sich die in diesem Bereich bisher nicht angewendete Methode des Entscheidungsbaums durchaus eignet, Personengruppen auf Basis ihres unterschiedlichen Lebensmittelverzehr zu differenzieren.

Für das vorliegende Kriterium bedeutet dies, dass der Entscheidungsbaum zwar grundsätzlich geeignet ist, Personengruppen auf Basis ihres Lebensmittelverzehr zu differenzieren, dies in der vorliegenden Auswertung auf Basis der Klassifizierungsvariable Bio-Käufer/innen und Nicht-Bio-Käufer/innen allerdings nicht möglich ist.

**Kriterium: Charakterisierung der identifizierten und differenzierten Personengruppen**

Wie bei den bereits beschriebenen drei Methoden kann auch hier der Vorteil genutzt werden, Personengruppen mit weiteren Variablen umfangreicher zu charakterisieren als dies über die

deskriptive Auswertung möglich ist. Die über den Lebensmittelverzehr identifizierten und differenzierten Personengruppen können weiter charakterisiert werden. Hierfür wurden beispielhaft bei den Frauen und Männern jeweils die beiden Blätter miteinander verglichen, denen am meisten Bio-Käufer/innen und Nicht-Bio-Käufer/innen zugeordnet wurden. Die Verknüpfung mit weiteren Variablen und der Vergleich der beiden Blätter zeigt, dass sich die beiden beschriebenen Blätter der Frauen deutlich voneinander unterscheiden. Die Ergebnisse des Entscheidungsbaums zeigen Zusammenhänge zwischen einer ungünstigeren Lebensmittelauswahl und gleichzeitig einem ungünstigeren Gesundheitsverhalten. Bei der Charakterisierung der beiden Blätter der Männer konnten deutlich weniger Zusammenhänge festgestellt werden. Allerdings muss in der Gruppe der Männer berücksichtigt werden, dass der Anteil der Bio-Käufer in Blatt drei mit 56 % nicht sehr hoch ist.

Für das vorliegende Kriterium bedeutet dies, dass auch mit dem Entscheidungsbaum eine anschließende Charakterisierung der identifizierten und differenzierten Personengruppen grundsätzlich möglich ist. Die Charakterisierung wurde beispielhaft anhand der Blätter, die am meisten Bio-Käufer/innen und Nicht-Bio-Käufer/innen enthalten, aufgezeigt. Allerdings konnten hierbei nur teilweise sinnvolle Ergebnisse aufgezeigt werden.

### **Fazit**

Bei der Erstellung eines Entscheidungsbaums werden alle 13 Lebensmittelgruppen berücksichtigt. Das Kriterium möglichst aller Lebensmittelgruppen und -menge über die Ergebnisse darzustellen, wird allerdings nicht erfüllt, da nur die Lebensmittelgruppen, die zur Differenzierung einer Personengruppe beitragen, über den Baum dargestellt werden.

Das eigentliche Ziel des Entscheidungsbaums ist, Objekte zu differenzieren. Allerdings war dies, wie die Korrekt- und Falschklassifikationsraten zeigen, für die Bio-Käufer/innen und Nicht-Bio-Käufer/innen nicht möglich. Um zu prüfen, ob die Methode des Entscheidungsbaums, die in der Literatur in diesem Bereich bisher nicht beschrieben wurde, grundsätzlich geeignet oder nicht geeignet ist, Personengruppen mit einem ähnlichen Lebensmittelverzehr zu identifizieren und eine Differenzierung auf Basis des Lebensmittelverzehrs vorzunehmen, wurden weitere Klassifizierungsvariablen herangezogen. Anhand dieser Beispiele konnte aufgezeigt werden, dass sich der Entscheidungsbaum durchaus eignet, die Studienteilnehmenden der NVS II anhand ihres Lebensmittelverzehrs zu differenzieren.

Die Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs am Beispiel der Daten der NVS II werden nur zum Teil erfüllt. Dennoch lässt sich festhalten, dass sich der Entscheidungsbaum durchaus zur Identifizierung und Differenzierung von Personengruppen eignet kann. Zudem kann der Lebensmittelverzehr umfassender dargestellt werden, als dies über rein deskriptive Vergleiche einzelner Lebensmittelgruppen möglich ist, und liefert durch

beispielsweise die am besten diskriminierenden Lebensmittelvariablen zusätzliche Informationen über Ernährung ausgewählter Personengruppen.

## 5 Übergreifende Diskussion und Fazit

Ziel der vorliegenden Arbeit ist es, Methoden auf ihre Eignung zur umfassenden Darstellung des Lebensmittelverzehrs am Beispiel von Bio-Käufer/innen und Nicht-Bio-Käufer/innen zu prüfen. Die umfassende Darstellung des Lebensmittelverzehrs durch möglichst alle Lebensmittelgruppen und den entsprechenden Verzehrsmengen ist bedeutend, da Menschen weder isoliert einzelne Nährstoffe noch Lebensmittel oder Lebensmittel einer Gruppe verzehren, sondern eine Vielzahl an Lebensmitteln und den darin enthaltenen Substanzen (Schulze und Hoffmann 2006 S 860), die in dieser Kombination wirken (Hoffmann et al. 2004 S 935).

In der vorliegenden Arbeit wurden verschiedene Methoden daraufhin untersucht, ob diese sich eignen, möglichst alle Lebensmittelgruppen mit den entsprechenden Verzehrsmengen darzustellen, Personengruppen, die sich ähnlich ernähren zu identifizieren und spezifische Personengruppen wie Bio-Käufer/innen und Nicht-Bio-Käufer/innen über den Lebensmittelverzehr zu differenzieren und diese Personengruppen anschließend mit weiteren ernährungs- und gesundheitsrelevanten Variablen zu charakterisieren.

Um einzelne Methoden auf ihre Eignung hierfür zu prüfen, wurden die fünf Kriterien (1) Berücksichtigung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen bei Anwendung der Methode, (2) Darstellung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse, (3) Identifizierung von Personengruppen mit ähnlichem Lebensmittelverzehr (4) Differenzierung von Bio-Käufer/innen und Nicht-Bio-Käufer/innen und (5) Charakterisierung der identifizierten und differenzierten Personengruppen herangezogen. Inwiefern diese fünf Kriterien gleichzeitig erfüllt werden, wird als entscheidend dafür angesehen, ob eine Methode zur umfassenden Darstellung des Lebensmittelverzehrs verschiedener Personengruppen geeignet ist.

Für diese Zielsetzung wurde der HEI-NVS II, die Hauptkomponentenanalyse, die Clusteranalyse und der Entscheidungsbaum angewendet (Tab. 5.1).

**Tab. 5.1: Vergleich der angewendeten Methoden zur umfassenden Darstellung des Lebensmittelverzehrs**

Methode	Prinzip	Ergebnis	Aussage
<b>Index</b>	Über einen Index werden die verzehrten Lebensmittelmengen über ein Punktesystem mit Empfehlungen verglichen und bewertet (Soll-Ist-Vergleich).	Indexwerte und Indexsumme	Die Indexwerte und Indexsumme liefern Auskunft darüber, wie sich Personen im Vergleich zu den Empfehlungen ernähren.
<b>Hauptkomponentenanalyse</b>	Es werden Linearkombinationen (Hauptkomponenten) auf Basis ursprünglicher Lebensmittelverzehrsdaten extrahiert, durch die eine charakteristische Kombination von Lebensmittelgruppen und -menge dargestellt wird.	Hauptkomponenten	Weist eine Person einen hohen Faktorwert für eine Hauptkomponente auf, ernährt sie sich ähnlich wie über die Hauptkomponente beschrieben.
<b>Clusteranalyse</b>	Es werden homogene Personengruppen auf Basis von Gemeinsamkeiten im Lebensmittelverzehr identifiziert.	Cluster	Personen eines Clusters weisen einen ähnlichen Lebensmittelverzehr auf.
<b>Entscheidungsbaum</b>	Es werden homogene Personengruppen auf Basis von Unterschieden im Lebensmittelverzehr identifiziert.	Diskriminierende Lebensmittelvariablen und Blätter	Die am besten diskriminierenden Lebensmittelgruppen werden hierarchisch identifiziert. Personen eines Blatts weisen einen ähnlichen Lebensmittelverzehr auf. Die Korrekt- und Falschklassifikationsraten geben Auskunft darüber, wie gut die Personen differenziert werden können.

Bei einem **Index**, wie dem HEI-NVS II wird der Lebensmittelverzehr mit Empfehlungen verglichen (Schulze und Hoffmann 2006 S 861, Hu 2002 S 4) und über ein Punktesystem bewertet. Ein Index spiegelt eine zuvor definierte ideale Ernährung wider (Miller et al. 2010 S 1272). Wie für den bekanntesten Index, den HEI (Kennedy et al. 1995 S 1104, USDA 1995 S 7), gilt für den HEI-NVS II, dass je mehr die Mengen der verzehrten Lebensmittel den Empfehlungen entsprechen, umso höher ist die erreichte Indexsumme. Als Ergebnis stehen die Indexwerte zur Verfügung, die Rückschlüsse auf den Lebensmittelverzehr einzelner Lebensmittelgruppen zulassen, sowie die Indexsumme, die eine zusammengefasste Bewertung über die Ernährung im Vergleich zu Empfehlungen gibt.

Bei der **Hauptkomponentenanalyse** erfolgt die Darstellung des Lebensmittelverzehrs durch die extrahierten Hauptkomponenten (Schulze und Hoffmann 2006 S 861), die eine charakteristische Kombination an Lebensmittelgruppen und ihren Verzehrsmengen ausdrücken. Jeder Person wird für jede extrahierte Hauptkomponente ein Faktorwert berechnet, der Auskunft darüber gibt, wie gut eine Hauptkomponente den Lebensmittelverzehr einer Person widerspiegelt.

Bei der **Clusteranalyse** werden auf Basis von Korrelationen (Schendera 2010 S Viff) des Lebensmittelverzehrs homogene Personengruppen gebildet (Hu 2002 S 4). Personen, die

einem Cluster zugeordnet werden, weisen einen ähnlichen Lebensmittelverzehr auf (Bamia et al. 2005 S 101). Die Darstellung des Lebensmittelverzehrs erfolgt anhand der Mittelwerte der Verzehrsmengen der Personen, die einem Cluster zugeordnet werden.

Über den **Entscheidungsbaum** werden Objekte (hier: Frauen bzw. Männer) in Abhängigkeit einer Klassifizierungsvariable (hier: Bio-Käufer/innen versus Nicht-Bio-Käufer/innen) aufgrund ihrer Unterschiede ausgewählter Variablen (hier: verzehrten Lebensmittelmengen) getrennt (Kokai 2002 S 98). Die Darstellung des Lebensmittelverzehrs erfolgt durch die gebildeten Blätter, in denen Personen mit einem ähnlichem Lebensmittelverzehr gruppiert werden. Die Lebensmittel, die zur Differenzierung der gruppierten Personen beitragen, können in der Reihenfolge ihrer Bedeutung von der Wurzel entlang über die Knoten abgelesen werden. Wie gut die Objekte auf Basis der Klassifizierungsvariable getrennt werden, wird durch Korrekt- und Falschklassifikationsraten ausgedrückt.

Diese vier Methoden wurden ausgewählt, da sie theoretisch zumindest einen Teil der genannten Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs erfüllen. Aufgrund des unterschiedlichen Prinzips bei der Durchführung der Methode und der unterschiedlichen Schwerpunkte ist davon auszugehen, dass sie alle einen Beitrag zur Darstellung des Lebensmittelverzehrs liefern können.

Zudem sind die Methoden vielversprechend, da Indizes (Beispiele: Bingham et al. 2010, Martinez et al. 2010, Previdelli et al. 2010, Bau et al. 2003, Sell et al. 2003), Faktorenanalysen (Beispiele: del Mar Bibiloni et al. 2012, Richter et al. 2012, Torjusen et al. 2012, Charreire et al. 2011, Cutler et al. 2011, Hare-Bruun et al. 2011, Olinto et al. 2011, Northstone und Emmett 2010, Rezazadeh et al. 2010, Arkkola et al. 2008) und Clusteranalysen (Beispiele: McGowan und McAuliffe 2012, Okubo et al. 2011, Ovaskainen et al. 2009, Knol et al. 2005) bereits eingesetzt wurden, um die Ernährung als solche, ohne Untersuchung von Ursache-Wirkungs-Zusammenhängen darzustellen. Der Entscheidungsbaum wurde in der Literatur in diesem Bereich noch nicht beschrieben.

## **5.1 Gegenüberstellung der Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs**

Ziel der Arbeit ist, zu prüfen, ob sich verschiedene Methoden zur umfassenden Darstellung des Lebensmittelverzehrs eignen. Hierfür wurden ein Index, eine Hauptkomponentenanalyse, eine Clusteranalyse sowie ein Entscheidungsbaum angewendet. Ein Vergleich dieser vier Methoden wurde in der Literatur bislang nicht beschrieben.

In der Literatur sind verschiedene Vergleiche einzelner Methoden wie Indizes, Cluster- und/oder Faktorenanalysen beschrieben, die allerdings überwiegend inhaltlich, d. h. bezogen auf ihre Anwendbarkeit zur Untersuchung von Ursache-Wirkungs-Zusammenhängen

zwischen der Ernährung und Krankheiten oder Bioparametern (Beispiele: Gu und Scarmeas 2011, Miller et al. 2010, Randi et al. 2010, Reedy et al. 2010, Fung et al. 2007, Schulze und Hoffmann 2006, Kant et al. 2004, Newby et al. 2004, Schulze et al. 2003b, Hu 2002, Osler et al. 2001) und weniger methodisch, d. h. bezogen auf die Vor- und Nachteile der Methoden, diskutiert werden. Die Anwendung und die Vergleiche verschiedener Methoden wie Indizes, Cluster- und/oder Faktorenanalysen zur alleinigen Darstellung der Ernährung verschiedener Personengruppen, wie sie in der vorliegenden Arbeit vorgenommen wird, wird selten verfolgt (Beispiele: Mullie et al. 2010, Hearty und Gibney 2009, Costacou et al. 2003).

Die Ergebnisse der vier Methoden sind allerdings aufgrund der unterschiedlichen Prinzipien und Zielsetzungen nicht direkt vergleichbar. Für einen Vergleich der vier Methoden werden daher die hierfür definierten Kriterien für eine umfassende Darstellung des Lebensmittelverzehr herangezogen und geprüft (Tab. 5.2).

**Tab. 5.2: Übersicht über die Erfüllung der Kriterien, die zur Prüfung der Eignung der angewendeten Methoden zur umfassenden Darstellung des Lebensmittelverzehrs herangezogen wurden**

	<b>Berücksichtigung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen bei Anwendung der Methode</b>	<b>Darstellung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse</b>	<b>Identifizierung von Personengruppen mit ähnlichem Lebensmittelverzehr</b>	<b>Differenzierung von Bio-Käufer/innen und Nicht-Bio-Käufer/innen</b>	<b>Charakterisierung der identifizierten und differenzierten Personengruppen</b>
<b>HEI-NVS II</b>	Für 10 der 13 relevanten Lebensmittelgruppen liegen quantifizierbare Empfehlungen vor. Diese konnten im HEI-NVS II einbezogen werden.	Durch die Indexwerte erfolgt eine direkte und durch die Indexsummen eine indirekte Darstellung der Lebensmittelgruppen und den verzehrten Lebensmittelmengen. Dabei werden nur die 10 in den HEI-NVS II einbezogenen Lebensmittelgruppen dargestellt.	Über den HEI-NVS II war keine direkte Identifizierung von Personengruppen mit ähnlichem Lebensmittelverzehr möglich.	Bio-Käufer/innen und Nicht-Bio-Käufer/innen konnten nach Anwendung des HEI-NVS II indirekt anhand der Indexwerte und Indexsummen differenziert werden.	Eine Identifizierung war nicht möglich. Daher konnten nur die differenzierten Personengruppen weiter charakterisiert werden.
<b>Hauptkomponentenanalyse</b>	Unabhängig von quantifizierbaren Empfehlungen konnten alle 13 Lebensmittelgruppen einbezogen werden.	Die Anzahl der dargestellten Lebensmittelgruppen wird i.d.R. auf die für eine Hauptkomponente charakteristischen Lebensmittelgruppen reduziert. Dadurch wurden in der Auswertung nur 5 bis 8 Lebensmittelgruppen für die ersten beiden Hauptkomponenten dargestellt.	Auf Basis der berechneten Faktorwerte konnten keine Personengruppen mit einem ähnlichen Lebensmittelverzehr identifiziert werden, da eine eindeutige Zuordnung zu einer der ersten beiden Hauptkomponenten nicht möglich war.	Eine indirekte Differenzierung der Bio-Käufer/innen und Nicht-Bio-Käufer/innen zu einer der ersten beiden Hauptkomponenten war mit den Daten der NVS II nicht erkennbar. Die Differenzierung von Personengruppen mit anderen Differenzierungsvariablen konnte anhand der Daten der NVS II aufgezeigt werden.	Die Identifizierung und Differenzierung war nicht möglich. Dennoch konnten Personen, mit hohem versus niedrigem Faktorwert für eine Hauptkomponente weiter charakterisiert werden.
<b>Clusteranalyse</b>	Unabhängig von quantifizierbaren Empfehlungen konnten alle 13 Lebensmittelgruppen einbezogen werden.	Wie in der vorliegenden Auswertung werden grundsätzlich alle einbezogenen Lebensmittelgruppen mit der dazugehörigen Verzehrsmenge durch die Mittelwerte der Verzehrsmengen, der über die Cluster gruppierten Personen, dargestellt.	Personen mit ähnlichem Lebensmittelverzehr können grundsätzlich über Cluster identifiziert werden. In der vorliegenden Auswertung war dies nur für die Frauen, nicht aber für die Männer möglich.	Eine indirekte Differenzierung der Bio-Käufer/innen und Nicht-Bio-Käufer/innen über die gebildeten Cluster war nicht erkennbar. Die Literatur zeigt, dass die Methode grundsätzlich zur indirekten Differenzierung von Personengruppen geeignet sein kann.	Grundsätzlich können identifizierte und differenzierte Personengruppen weiter charakterisiert werden. Dies war in der vorliegenden Auswertung allerdings nur teilweise möglich.
<b>Entscheidungsbaum</b>	Unabhängig von quantifizierbaren Empfehlungen konnten alle 13 Lebensmittelgruppen einbezogen werden.	An jedem Knoten des Baums werden auf Basis der Verzehrsmenge zwei Personengruppen gebildet. Zur Bildung der Knoten werden nur die für die Differenzierung relevanten Lebensmittelgruppen herangezogen. In der vorliegenden Auswertung waren dies 3 (Männer) bzw. 5 (Frauen) der 13 Lebensmittelgruppen.	Personen mit ähnlichem Lebensmittelverzehr werden über die Blätter gruppiert.	Durch den Baum wird das Ziel verfolgt, direkt eine Differenzierung vorzunehmen. Die Bio-Käufer/innen und Nicht-Bio-Käufer/innen waren allerdings nicht differenzierbar. Andere Differenzierungsvariablen der NVS II zeigen, dass der Entscheidungsbaum grundsätzlich geeignet ist, Personen auf Basis des Lebensmittelverzehrs zu differenzieren.	Die Differenzierung war nicht möglich, daher konnten nur die identifizierten Personengruppen weiter charakterisiert werden.

**Kriterium: Berücksichtigung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen bei Anwendung der Methode**

Bei einem Index werden Nährstoffe oder Lebensmittel mit den entsprechenden quantifizierbaren Empfehlungen zur Berechnung zu Grunde gelegt (Schulze und Hoffmann 2006 S 861, Newby und Tucker 2004 S 177, Hu 2002 S 4). Da es für die meisten Lebensmittelgruppen quantifizierbare Empfehlungen der DGE gibt, lässt sich festhalten, dass durch den HEI-NVS II 10 der 13 für das Untersuchungsziel relevanten Lebensmittelgruppen mit den entsprechenden Verzehrsmengen berücksichtigt werden konnten. Im Gegensatz dazu können bei der Hauptkomponenten- und Clusteranalyse (Schulze und Hoffmann 2006 S 861, Newby und Tucker 2004 S 177) sowie bei der Anwendung des Entscheidungsbaums unabhängig von quantifizierbaren Empfehlungen alle für das Untersuchungsziel relevanten Lebensmittelgruppen einbezogen werden. Zudem werden durch die letzten drei genannten Methoden vorab keine Begrenzungen zur Anzahl der einbezogenen Lebensmittelgruppen festgelegt. Bei den explorativen Methoden konnten alle 13 Lebensmittelgruppen bei der Auswertung berücksichtigt werden.

Das Kriterium, möglichst alle Lebensmittelgruppen mit den dazugehörigen verzehrten Lebensmittelmengen zu berücksichtigen, wird durch die drei explorativen Methoden Hauptkomponentenanalyse, Clusteranalyse und Entscheidungsbaum, jedoch nicht vollständig durch den HEI-NVS II erfüllt.

**Kriterium: Darstellung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Lebensmittelmengen durch die Ergebnisse**

Als Ergebnis der vier angewendeten Methoden werden die Indexwerte, Indexsumme, Hauptkomponenten, Cluster und Blätter dargestellt.

Durch die Indexwerte erfolgt eine direkte und durch die Indexsumme eine indirekte Darstellung der 10 Lebensmittelgruppen. Die Indexwerte des HEI-NVS II geben Auskunft darüber, wie gut die Empfehlungen der DGE der 10 einzelnen, im Index einbezogenen Lebensmittelgruppen, erreicht werden. Die Indexsumme entsteht additiv und liefert eine zusammengefasste Bewertung über die im HEI-NVS II einbezogenen Lebensmittelgruppen im Vergleich zu den Empfehlungen der DGE. Damit wird durch einen Index, wie auch von Schulze und Hoffmann (2006 S 861) beschrieben, im Gegensatz zur Faktoren- und Clusteranalyse nicht die gesamte Ernährung, sondern nur ausgewählte Aspekte der Ernährung dargestellt.

Bei der Hauptkomponentenanalyse erfolgt die Darstellung des Lebensmittelverzehr durch die extrahierten Hauptkomponenten (Schulze et al. 2003a S 410). Die Lebensmittelverzehrsmengen werden durch die Faktorladungen jeder einbezogenen Variablen in jeder Haupt-

komponente beschrieben. Für die leichtere Interpretation der Hauptkomponenten werden üblicherweise nur die für eine Hauptkomponente charakteristischen Lebensmittelgruppen mit einer höheren Faktorladung dargestellt. In der vorliegenden Auswertung wurden für die ersten beiden Hauptkomponenten der Frauen und Männer zwischen 5 und 8 der 13 Lebensmittelgruppen als charakteristisch angesehen. Damit wird nur ein Teil, der für die vorliegende Zielsetzung relevanten Lebensmittelgruppen, dargestellt. Zudem werden nur wenige besonders aussagekräftige Hauptkomponenten herangezogen, die einen möglichst großen Anteil der Varianz erklären (Schulze 2002 S 51). In der vorliegenden Auswertung wurden bei beiden Geschlechtern jeweils 13 Hauptkomponenten extrahiert, von denen auf Basis des üblicherweise verwendeten Kaiser-Kriterium, bei den Frauen vier und bei den Männern fünf Hauptkomponenten ausgewählt wurden. Mit diesen Hauptkomponenten werden 43,5 % bzw. 52,5 % der Varianz des Lebensmittelverzehrs erklärt. Durch diese Auswahl einer bestimmten Anzahl an Hauptkomponenten bleibt, wie von Moosbrugger und Kelava (2007 S 309f) beschrieben, ein Teil der Gesamtvarianz unberücksichtigt. So wird beispielsweise von Hearty und Gibney (2009 S 604) beschrieben, dass durch die in ihrer Studie angewendete Hauptkomponentenanalyse nicht alle Ernährungsmuster, die in den Daten vorliegen, aufgezeigt werden. Zu diesem Schluss kamen sie durch einen Vergleich der Ergebnisse mit denen einer Clusteranalyse, die mit dem gleichen Ziel und dem gleichen Datensatz angewendet wurde. Über beide Methoden wurden zwar vergleichbare Muster gefunden, durch die Clusteranalyse konnten allerdings mehr Muster identifiziert werden. Für die vorliegende Auswertung bedeutet dies, dass ein Teil der Information des Lebensmittelverzehrs durch die Anwendung der Hauptkomponentenanalyse unberücksichtigt bleibt.

Über die Clusteranalyse werden in den gebildeten Clustern Personen mit einem ähnlichen Lebensmittelverzehr gruppiert (Bamia et al. 2005 S 103, Hu 2002 S 4). Die Darstellung des Lebensmittelverzehrs erfolgt durch die Mittelwerte der Verzehrsmengen aller 13 einbezogenen Lebensmittelgruppen der Personen, die einem Cluster zugeordnet wurden.

Durch den Entscheidungsbaum werden wie bei der Clusteranalyse Personen, die sich ähnlich ernähren, gruppiert. Allerdings werden die Gruppen nicht auf Basis von Gemeinsamkeiten, sondern auf Basis von Unterschieden im Lebensmittelverzehr bezogen auf die beispielhaft gewählte Differenzierungsvariable Bio-Kauf, gebildet. Dabei wird bei jedem Knoten des Entscheidungsbaums die Lebensmittelgruppe ermittelt, aufgrund derer sich Frauen und Männer bezogen auf den Bio-Kauf am besten differenzieren lassen. Ist eine Lebensmittelgruppe zur Differenzierung der Personengruppe nicht relevant, wird sie nicht verwendet. Damit wurden zwar in der vorliegenden Arbeit bei der Berechnung des Entscheidungsbaums alle 13 einbezogenen Lebensmittelgruppen mit den entsprechenden Verzehrsmengen berücksichtigt, allerdings, bedingt durch das Berechnungsverfahren, nur

die für die Differenzierung relevanten Lebensmittelgruppen Fleisch, Gemüse, Getreide und bei den Frauen zusätzlich Obst und Fisch dargestellt.

Für das Kriterium möglichst alle Lebensmittelgruppen und deren verzehrten Lebensmittelmengen darzustellen, bedeutet dies, dass in der vorliegenden Auswertung lediglich durch die Clusteranalyse das Kriterium erfüllt wird. Durch den HEI-NVS II werden nur 10 Lebensmittelgruppen, für die es quantifizierbare Empfehlung gibt, direkt über die Indexwerte und indirekt über die Indexsumme dargestellt. Das Kriterium möglichst alle Lebensmittelgruppen und deren verzehrten Lebensmittelmengen darzustellen, wird über den HEI-NVS II weitgehend erfüllt. Bei der Hauptkomponentenanalyse ist die Darstellung aller Lebensmittelgruppen mit den verzehrten Lebensmittelmengen zwar grundsätzlich möglich, durch das ausschließliche Darstellen der charakteristischen Lebensmittelgruppen in einer Hauptkomponenten bleibt ein Teil der Information des Lebensmittelverzehrts allerdings unberücksichtigt. Auch durch den Entscheidungsbaum wird das Kriterium nicht erfüllt, da bei den Frauen nur drei und bei den Männern nur fünf der 13 Lebensmittelgruppen für die Differenzierung relevant waren und über den Baum dargestellt werden.

#### **Kriterium: Identifizierung von Personengruppen mit ähnlichem Lebensmittelverzehr**

Die Indexsumme entsteht additiv und liefert, im Gegensatz zu Ergebnissen explorativer Methoden wie der Clusteranalyse, keine Informationen darüber, durch welche möglicherweise charakteristische Kombinationen an verzehrten Lebensmitteln sich eine Personengruppe auszeichnet. Damit ist es auch nicht möglich, Personengruppen mit einem ähnlichen Lebensmittelverzehr zu identifizieren. Es können ausschließlich Personengruppen, die sich stärker oder weniger stark nach den Empfehlungen der DGE ernähren, durch Aufteilung auf Basis einer hohen und niedrigen Indexsumme gruppiert werden. Auch die Indexwerte einzelner Lebensmittelgruppen tragen nicht zur Erfüllung dieses Kriteriums bei, da in der vorliegenden Auswertung Personengruppen auf Basis möglichst vieler Lebensmittelgruppen identifiziert werden sollen.

Bei der Hauptkomponentenanalyse werden, im Gegensatz zur Clusteranalyse oder zum Entscheidungsbaum, keine Personen mit einem ähnlichen Lebensmittelverzehr gruppiert, sondern Linearkombinationen der Ausgangsdaten, die Hauptkomponenten, extrahiert. Damit stehen, wie durch Hearty und Gibney (2009) beschrieben, durch die Hauptkomponenten- und Clusteranalyse unterschiedliche Ergebnisse zur Verfügung. Bei der Hauptkomponentenanalyse wird der Lebensmittelverzehr einer Person, wie durch Schulze und Hoffmann (2006 S 861) beschrieben, durch den Faktorwert ausgedrückt, der Auskunft über die Übereinstimmung zu jeder einzelnen Hauptkomponente liefert. Damit lässt sich der Lebensmittelverzehr einer Person, im Gegensatz zur Clusteranalyse, nicht durch ein einziges Muster, sondern nur durch die Kombination aller extrahierten Hauptkomponenten beschreiben

(Newby und Tucker 2004 S 197). In der vorliegenden Auswertung wurde auf Basis der Faktorwerte grafisch die Zuordnung der Studienteilnehmenden der NVS II zu einer der beiden aussagekräftigsten ersten beiden Hauptkomponenten geprüft. Sowohl bei den Frauen als auch bei den Männern konnten nur ein kleiner Personenanteil spezifisch einer der ersten beiden Hauptkomponenten zugeordnet werden. Das bedeutet, dass über die Hauptkomponentenanalyse keine Personengruppen mit einem ähnlichen Lebensmittelverzehr identifiziert werden können.

Im Gegensatz zum Index und zur Hauptkomponentenanalyse können über die Clusteranalyse (Bamia et al. 2005 S 103, Hu 2002 S 4) und den Entscheidungsbaum Personen mit einem ähnlichen Lebensmittelverzehr gruppiert werden. Theoretisch weisen die den einzelnen Clustern und Blättern zugeordneten Personen einen ähnlichen Lebensmittelverzehr auf. Allerdings wurden durch die vorliegende Clusteranalyse in der Gruppe der Männer sehr unterschiedlich große Cluster gebildet. Es ist fraglich, inwiefern ein Cluster mit 90 % und 1,6 % der Männer zur Identifizierung von Personengruppen mit einem ähnlichen Lebensmittelverzehr beitragen kann. Über die Clusteranalyse werden grundsätzlich alle Personen über Cluster gruppiert. Allerdings kann nicht ausgeschlossen werden, dass Objekte nur zufällig durch das Verfahren einem Cluster zugeordnet werden und möglicherweise nicht in das Cluster passen. In der Gruppe der Frauen wurden zwei ähnlich große Cluster gebildet. Durch den Entscheidungsbaum konnten Personen mit einem ähnlichen Lebensmittelverzehr gruppiert werden. Allerdings können hierfür nur die Lebensmittelgruppen, die zur Differenzierung der in einem Blatt gruppierten Personen beigetragen haben, zur Beschreibung des ähnlichen Lebensmittelverzehrs herangezogen werden.

Das Kriterium, Personen mit ähnlichem Lebensmittelverzehr zu identifizieren, kann grundsätzlich durch die Ergebnisse der Clusteranalyse, des Entscheidungsbaums und durch die Ergebnisse der Hauptkomponentenanalyse, nicht aber durch die Ergebnisse des HEI-NVS II erfüllt werden. Allerdings weisen in der vorliegenden Arbeit die Ergebnisse der Clusteranalyse nur in der Gruppe der Frauen und nicht in der Gruppe der Männer auf eine sinnvolle Identifizierung von Personengruppen mit einem ähnlichen Lebensmittelverzehr hin. Bei dem Entscheidungsbaum konnten Personen mit einem ähnlichen Lebensmittelverzehr über die Blätter identifiziert werden. Bei der Hauptkomponentenanalyse konnten weder bei den Frauen noch bei den Männern Personen mit einem ähnlichem Lebensmittelverzehr identifiziert werden.

**Kriterium: Differenzierung von Bio-Käufer/innen und Nicht-Bio-Käufer/innen**

Neben der Prüfung der Methoden für ihre Eignung Personengruppen mit einem ähnlichen Lebensmittelverzehr zu identifizieren, wurde geprüft, ob sich die Bio-Käufer/innen und Nicht-Bio-Käufer/innen über die Ergebnisse der einzelnen Methoden differenzieren lassen.

Bei dem HEI-NVS II kann die Indexsumme individuell für jede Personen berechnet werden. Bei den Indexwerten und -summen konnten deutliche Unterschiede zwischen den Bio-Käufer/innen und Nicht-Bio-Käufer/innen bezogen auf den Lebensmittelverzehr aufgezeigt werden. Damit ist es möglich, über den HEI-NVS II Bio-Käufer/innen und Nicht-Bio-Käufer/innen zu differenzieren. Bei der Hauptkomponenten- und Clusteranalyse wurde nach Anwendung der Methoden und damit indirekt geprüft, ob eine Differenzierung möglich ist. Die Ergebnisse beider Analysen weisen auf keine gute Differenzierung der Bio-Käufer/innen und Nicht-Bio-Käufer/innen hin. Bei der Hauptkomponentenanalyse ist keine spezifische Zuordnung der Bio-Käufer/innen und Nicht-Bio-Käufer/innen zu einer der ersten beiden Hauptkomponenten, die am meisten der Varianz des Lebensmittelverzehrs erklären, möglich. Dass durch die Ergebnisse der Hauptkomponentenanalyse grundsätzlich Personengruppen differenziert werden können, wurde anhand anderer Variablen der gleichen Studiengruppen aufgezeigt.

In der Clusteranalyse konnten bei den Frauen keine Unterschiede im Anteil der Bio-Käufer/innen und Nicht-Bio-Käufer/innen zwischen den beiden Clustern festgestellt werden. Bei den Männern, bei denen zwar Unterschiede festgestellt wurden, sind die gebildeten Cluster aufgrund der großen Gruppengrößenunterschiede zur Differenzierung allerdings nicht geeignet. Ein Cluster, das 90 % der Männer enthält, kann nicht zur Differenzierung einer weiteren Studiengruppe beitragen.

Im Gegensatz zum HEI-NVS II, der Hauptkomponenten- und Clusteranalyse ist es Ziel des Entscheidungsbaums, eine Differenzierung auf Basis einer Klassifizierungsvariable vorzunehmen. Durch den Entscheidungsbaum wurden die Bio-Käufer/innen und Nicht-Bio-Käufer/innen anhand ihrer Lebensmittelverzehrmengen voneinander getrennt. Allerdings liefert auch diese Methode keine eindeutige Differenzierung. Die hierfür herangezogenen Korrekt- und Falsch-Klassifikationsraten sind sehr gering. Um zu prüfen, ob sich der Entscheidungsbaum, der in der Literatur in diesem Bereich bisher nicht beschrieben wurde, grundsätzlich zum Differenzieren von Personengruppen auf Basis des Lebensmittelverzehrs eignet, wurden neben der Variabel Bio-Kauf weitere Variablen zur Differenzierung herangezogen. Diese weiteren Beispiele auf Basis der Daten der NVS II zeigen, dass sich der Entscheidungsbaum durchaus eignen kann, Personengruppen auf Basis des Lebensmittelverzehrs zu differenzieren.

Für das Kriterium der Differenzierung bedeutet dies, dass in der vorliegenden Auswertung lediglich der HEI-NVS II geeignet ist, die Gruppe der Bio-Käufer/innen und Nicht-Bio-Käufer/innen erfolgreich zu differenzieren. Allerdings wird bei der Hauptkomponentenanalyse und dem Entscheidungsbaum anhand anderer Differenzierungsvariablen mit der gleichen Studiengruppen aufgezeigt, dass sich diese Methoden grundsätzlich zum Differenzieren von Personengruppen eignen. Bei der Clusteranalyse zeigt die Literatur, dass die Methode ebenfalls zum Differenzieren geeignet sein kann.

### **Kriterium: Charakterisierung der identifizierten und differenzierten Personengruppen**

Werden durch die angewendeten Methoden Personengruppen identifiziert und differenziert, ist es von Bedeutung, diese mit weiteren ernährungs- und gesundheitsrelevanten Variablen zu charakterisieren.

Die über die Indexsumme des HEI-NVS II voneinander differenzierten Bio-Käufer/innen und Nicht-Bio-Käufer/innen konnten erfolgreich mit zahlreichen weiteren Variablen charakterisiert werden. Es zeigt sich, dass Bio-Käufer/innen mit einer höheren Indexsumme und damit mit ihrer Ernährung eher an den Empfehlungen liegen, gleichzeitig ein gesundheitsbewussteres Verhalten aufweisen.

Wie von Newby und Tucker (2004 S 197) beschrieben, sind die Ergebnisse der Hauptkomponentenanalyse im Vergleich zu anderen Methoden wie der Clusteranalyse prädestiniert dafür, weitere Analysen auf Basis der extrahierten Hauptkomponenten anzuschließen. Hierfür werden in der Literatur beispielsweise Personengruppen mit einem hohen versus niedrigen Faktorwert für eine Hauptkomponente durch die Verknüpfung mit weiteren Variablen verglichen, um die Ursache-Wirkungs-Zusammenhänge zwischen der Ernährung und Erkrankungen einer Risikogruppe weiter zu untersuchen (Beispiele: Edefonti et al. 2010, de Stefani et al. 2008a). In der vorliegenden Auswertung konnten durch die Anwendung der Hauptkomponentenanalyse weder Personengruppen mit einem ähnlichem Lebensmittelverzehr identifiziert noch Bio-Käufer/innen und Nicht-Bio-Käufer/innen differenziert werden, die weiter charakterisiert werden können. Dennoch wurden beispielhaft die Frauen und Männer mit einem hohen versus niedrigen Faktorwert für die ersten beiden Hauptkomponenten anhand weiterer Variablen charakterisiert. Die Ergebnisse der deskriptiven Auswertung und des HEI-NVS II konnten durch diese Ergebnisse allerdings nicht bestätigt werden. Es konnte kein Zusammenhang zwischen einer günstigeren Ernährung und einem gesundheitsbewussteren Verhalten festgestellt werden.

Ebenso wie die Hauptkomponentenanalyse wird die Clusteranalyse häufig eingesetzt, um den Ursache-Wirkungs-Zusammenhang zwischen der Ernährung und Bioparametern oder Krankheiten zu untersuchen (Beispiele: Nyholm et al. 2012, de Kort et al. 2011, Pedone et al. 2011, Millen et al. 2005). Im Gegensatz zur Hauptkomponentenanalyse, bei welcher häufig

untersucht wird, ob sich Personen, die sich eher nach einer Hauptkomponente ernähren, ein höheres oder niedrigeres Erkrankungsrisiko haben, wird bei der Clusteranalyse untersucht, ob Personen, die sich ähnlich ernähren, ein höheres oder niedrigeres Erkrankungsrisiko haben. Die anschließende Beschreibung mit weiteren Variablen, wie körperlicher Aktivität und Raucherstatus, dient wie bei der Hauptkomponentenanalyse bei diesen Studien der Charakterisierung möglicher Risikogruppen. Bei der in der vorliegenden Auswertung gebildeten Clustern konnten in der Gruppe der Frauen bei den beschreibenden Variablen allerdings nur wenige Unterschiede im ernährungs- und gesundheitsrelevanten Verhalten ermittelt werden. Bei den Männern konnten zwar mehr Unterschiede festgestellt werden, die allerdings aufgrund der großen Gruppenunterschiede der Cluster nicht zur Charakterisierung beitragen können.

Wie bei dem HEI-NVS II, der Hauptkomponenten- und Clusteranalyse wurden die Ergebnisse des Entscheidungsbaums genutzt, um Personengruppen weiter zu charakterisieren. Hierfür wurden beispielhaft bei den Frauen und Männern jeweils die Personengruppen der beiden Blätter weiter charakterisiert, denen am meisten Bio-Käufer/innen und Nicht-Bio-Käufer/innen zugeordnet wurden. Die Ergebnisse sind mit den Ergebnissen des HEI-NVS II vergleichbar. Sowohl die Ergebnisse des Entscheidungsbaums als auch die des HEI-NVS II zeigen Zusammenhänge zwischen einer ungünstigeren Lebensmittelauswahl und einem ungünstigeren Gesundheitsverhalten. Bei der Charakterisierung der beiden Blätter der Männer konnten deutlich weniger Zusammenhänge festgestellt werden, was möglicherweise auf die ungleiche Verteilung der Bio-Käufer zurückzuführen ist.

Da in der vorliegenden Arbeit nur über die Clusteranalyse und den Entscheidungsbaum Personengruppen auf Basis ihres ähnlichen Lebensmittelverzehrs identifiziert und nur über den HEI-NVS II Bio-Käufer/innen und Nicht-Bio-Käufer/innen differenziert werden können, war es auch nur für diese Personengruppen möglich, eine für die vorliegende Arbeit relevante Charakterisierung vorzunehmen. Durch andere Differenzierungsvariablen oder Literaturbeispiele wird aufgezeigt, dass alle vier Methoden geeignet sind, identifizierte und differenzierte Personengruppen weiter zu charakterisieren.

### **Weitere Aspekte zur Anwendung der Methoden**

Entscheidend für die Anwendung einer der vier Methoden sind auch die vorausgesetzten statistischen Kenntnisse und Vergleichbarkeit der Ergebnisse mit denen anderer Studien.

Bei der statistischen Anwendung lässt sich festhalten, dass der Index im Vergleich zu den anderen drei Methoden schnell, unkompliziert und ohne große statistische Vorkenntnisse anwendbar ist und die Ergebnisse leicht interpretierbar sind. Im Gegensatz zum Index müssen bei der Anwendung der Hauptkomponentenanalyse, Clusteranalyse und des Entscheidungsbaums, zahlreiche für das Ergebnis relevante Entscheidungen getroffen

werden, für die entsprechende statistische Expertise notwendig ist. Auch die Interpretation der Ergebnisse der Hauptkomponentenanalyse ist nicht ohne entsprechende statistische Kenntnisse möglich. Wie von Newby und Tucker (2004 S 197) beschrieben, sind die Ergebnisse der Clusteranalyse einfacher zu interpretieren als die Ergebnisse der Hauptkomponentenanalyse. Das Gleiche lässt sich für die Ergebnisse des Entscheidungsbaums festhalten. Die Ergebnisse des Entscheidungsbaums sind leicht nachvollziehbar aufgebaut.

Ein weiterer Vorteil von Indizes, wie dem HEI-NVS II gegenüber den anderen drei angewendeten Methoden ist, dass diese Methode durch den zugrunde gelegten hypothesenorientierten Ansatz insofern nicht studiengruppenspezifisch ist, als dass ein Index für weitere Personengruppen angewendet werden kann (Miller et al. 2010 S 1272). Wird der gleiche Index bei unterschiedlichen Studiengruppen angewendet, können die Ergebnisse, unter Voraussetzung der Anwendung einer gleichen Ernährungserhebungsmethode, z. B. einem Diet-History-Interview, miteinander verglichen werden. Dagegen sind die Ergebnisse der Hauptkomponenten- und Clusteranalyse spezifisch für eine Studiengruppe (Bamia et al. 2005 S 409, Kant 2004 S 630, Schulze et al. 2003a S 111). Das bedeutet, dass die gebildeten Cluster und extrahierten Hauptkomponenten nicht auf andere Studiengruppen übertragbar sind und nicht miteinander verglichen werden können. Dies liegt, wie von Kant (2004 S 630) beschrieben, an den unterschiedlichen Variablen, die bei der Analyse einbezogen werden, sowie an den unterschiedlichen Berechnungsmöglichkeiten bei der Anwendung einer Methode. Das Gleiche trifft auf die Ergebnisse des Entscheidungsbaums zu. Auch die gebildeten Blätter sind spezifisch für eine Studiengruppe und können nicht mit Blättern anderer Bäume aus anderen Studien verglichen werden. Wird das Ziel verfolgt, die Ergebnisse mit denen anderer Studien zu vergleichen, kann ausschließlich der Index angewendet werden.

## **5.2 Limitierungen der Arbeit**

Bei allen vier angewendeten Methoden können unterschiedliche Vorgehensweisen gewählt werden, die maßgeblich Einfluss auf die Ergebnisse haben. Zwar wurden für die vorliegenden Auswertungen mit unterschiedlichen Experten das statistische Vorgehen bei der Anwendung der verschiedenen Methoden abgestimmt, um das für das Untersuchungsziel jeweils optimale Verfahren zu finden. Dennoch können, wie auch von Newby und Tucker (2004 S 198) beschrieben, nicht alle getroffenen Entscheidungen objektiv erfolgen. Auch Kant (2004 S 629) macht deutlich, dass bei jeder Methode zahlreiche subjektive Entscheidungen getroffen werden, welche die Ergebnisse maßgeblich beeinflussen.

So müssen bei der Entwicklung eines Indexes zahlreiche Entscheidungen, wie die Auswahl der zu berücksichtigenden Lebensmittelgruppen, Übersetzung der Empfehlungen in ein

Punktesystem und Bildung der Indexsumme, getroffen werden. Auch bei der Hauptkomponentenanalyse sind neben der Wahl des geeigneten Verfahrens der Faktorenanalyse (Bamia et al. 2005 S 111), die Wahl der einbezogenen Variablen, die Festlegung des Schwellenwertes, ab dem eine Faktorladung noch als hoch angesehen wird (Heidemann 2006 S 34), sowie die Wahl der Anzahl an extrahierten Hauptkomponenten (Bamia et al. 2005 S 111) vom jeweiligen Untersucher abhängig. Bei der Clusteranalyse nehmen die Wahl des geeigneten Verfahrens (Bamia et al. 2005 S 111), die Wahl der einbezogenen Variablen sowie der Festlegung der Clusteranzahl Einfluss auf die Ergebnisse (Schendera 2004 S 540f). Bei der Anwendung des Entscheidungsbaums fließen neben der Auswahl der zur Klassifizierung verwendeten Lebensmittelgruppen, zahlreiche Optimierungsmöglichkeiten wie das Beschneiden von Ästen sowie das Festlegen einer Mindestanzahl an Objekten in einem Blatt eine gewisse Subjektivität mit ein. All diese Entscheidungen beeinflussen das Ergebnis dieser Methoden. Daher müssen die Anwendung der Methoden und die gewonnenen Ergebnisse unter diesen Aspekten stets kritisch betrachtet werden. Für die vorliegende Auswertung bedeutet dies, dass all diese Entscheidungen das Ergebnis beeinflussen können. Da in der vorliegenden Auswertung nicht die inhaltlichen Ergebnisse im Vordergrund standen, sondern die Prüfung der Eignung der Methoden, wurde nicht geprüft, in welcher Form die hier vorliegenden Ergebnisse durch diese Entscheidungen beeinflusst wurden.

### **5.3 Fazit**

In der aktuellen Literatur werden bislang lediglich Vergleiche einzelner Methoden, nicht aber Vergleiche aller vier angewendeten Methoden beschrieben. Zudem gibt es keine Veröffentlichungen, in denen die Anwendung des Entscheidungsbaums zur umfassenden Darstellung des Lebensmittelverzehrs beschrieben wird. Die vorliegende Arbeit füllt diese Lücke.

Die Prüfung der angewendeten Methoden ergab, dass keine Methode gleichzeitig alle hierfür definierten Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs anhand der Daten der NVS II erfüllt. Dennoch haben alle Methoden gemeinsam, dass sie den Lebensmittelverzehr umfassender darstellen können, als dies über rein deskriptive Vergleiche einzelner Lebensmittelgruppen möglich ist. Daneben haben alle vier Methoden unterschiedliche Vor- und Nachteile, die bei der umfassenden Darstellung von Bedeutung sind und bei der Wahl einer geeigneten Methode berücksichtigt werden müssen. Beispielsweise können durch den HEI-NVS II im Gegensatz zur Hauptkomponenten- und Clusteranalyse sowie dem Entscheidungsbaum jede beliebige Personengruppe wie die Bio-Käufer/innen und Nicht-Bio-Käufer/innen differenziert werden und Aussagen über die Qualität der Ernährung, bezogen auf den Vergleich mit den Empfehlungen der DGE,

getroffen werden. Allerdings konnten in der vorliegenden Auswertung nur 10 der 13 relevanten Lebensmittelgruppen einbezogen werden. Durch die Hauptkomponentenanalyse wurden am wenigsten Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs erfüllt. Dafür ist diese Methode prädestiniert, um Ursache-Wirkungs-Zusammenhänge zu untersuchen, was in der vorliegenden Auswertung allerdings nicht von Bedeutung war. Aufgrund der vorliegenden Ergebnisse ist es fraglich, ob die Hauptkomponentenanalyse nicht nur bei der vorliegenden Auswertung, sondern grundsätzlich einen wesentlichen Beitrag zur umfassenden Darstellung des Lebensmittelverzehrs, bezogen auf die definierten Kriterien, leisten kann. Durch die Clusteranalyse können theoretisch am besten Personengruppen mit einem ähnlichen Lebensmittelverzehr identifiziert werden, was in der vorliegenden Auswertung allerdings nicht gelungen ist. Der Entscheidungsbaum, der in diesem Bereich in der Literatur bisher nicht beschrieben wurde, bietet, auch wenn er nicht zur Differenzierung der Bio-Käufer/innen und Nicht-Bio-Käufer/innen beitragen konnte, Potential, Personengruppen auf Basis des Lebensmittelverzehrs zu differenzieren und weiter zu untersuchen. Zudem liefert diese Methode Erkenntnisse über die Wichtigkeit von Lebensmittelgruppen die zur Differenzierung von Personengruppen beitragen, was durch keine der anderen Methoden möglich ist.

Dies zeigt, dass durch jede einzelne Methode Erkenntnisse über die Studiengruppen und in Bezug auf die Forschungsfrage gewonnen werden können. Als Konsequenz sollten die Methoden als komplementär angesehen werden, um von den unterschiedlichen Stärken der Methoden zu profitieren. In Abhängigkeit der Forschungsfrage sollte daher in Erwägung gezogen werden, mehr als eine Methode anzuwenden, um die daraus gewonnenen Erkenntnisse zu kombinieren. Beispielsweise ist es möglich, die Ergebnisse der Clusteranalyse oder des Entscheidungsbaums mit den Ergebnissen des HEI-NVS II zu kombinieren. Hierdurch kann der charakteristische Lebensmittelverzehr identifizierter und differenzierter Personengruppen über die im HEI-NVS II zugrunde gelegten Empfehlungen zusätzlich bewertet werden.

## 6 Zusammenfassung

In der vorliegenden Arbeit werden ein Index, eine Hauptkomponentenanalyse, eine Clusteranalyse und ein Entscheidungsbaum auf ihre Eignung zur umfassenden Darstellung des Lebensmittelverzehrs mit Daten der NVS II geprüft. In der aktuellen Literatur werden lediglich Vergleiche einzelner Methoden, nicht aber Vergleiche aller vier Methoden beschrieben. Zudem gibt es keine Veröffentlichungen, in denen die Anwendung des Entscheidungsbaums zur umfassenden Darstellung des Lebensmittelverzehrs beschrieben wird.

Für die Prüfung der Methoden werden die fünf Kriterien (1) Berücksichtigung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Mengen bei Anwendung der Methode (2) Darstellung möglichst aller Lebensmittelgruppen und der dazugehörigen verzehrten Mengen durch die Ergebnisse (3) Identifizierung von Personengruppen mit ähnlichem Lebensmittelverzehr (4) Differenzierung von Bio-Käufer/innen und Nicht-Bio-Käufer/innen und (5) Charakterisierung der identifizierten und differenzierten Personengruppen herangezogen.

Mit dem entwickelten Index (HEI-NVS II) wird der Lebensmittelverzehr mit Empfehlungen der DGE verglichen und über ein Punktesystem bewertet. Als Ergebnis stehen die Indexwerte und die Indexsumme zur Verfügung. Bei der Hauptkomponentenanalyse erfolgt die Darstellung des Lebensmittelverzehrs durch Hauptkomponenten, die durch Korrelationen im Lebensmittelverzehr gebildet werden. Durch jede Hauptkomponente wird eine charakteristische Kombination an verzehrten Lebensmitteln ausgedrückt. Jeder Person wird für jede Hauptkomponente ein Faktorwert berechnet. Der Faktorwert gibt Auskunft darüber, wie gut der Lebensmittelverzehr einer Person durch eine Hauptkomponente wiedergespiegelt wird. Bei der Clusteranalyse werden homogene Personengruppen mit ähnlichem Lebensmittelverzehr gebildet. Die Darstellung des Lebensmittelverzehrs erfolgt durch die Mittelwerte der Verzehrsmengen der in einem Cluster zugeordneten Personen. Bei einem Entscheidungsbaum werden Personen in Abhängigkeit einer Klassifizierungsvariable aufgrund ihrer Unterschiede im Lebensmittelverzehr getrennt. Die Darstellung des Lebensmittelverzehrs erfolgt durch die Blätter, in denen Personen mit einem ähnlichem Lebensmittelverzehr gruppiert werden. Die Korrekt- und Falschklassifikationsraten geben Auskunft darüber, wie zuverlässig die Personen auf Basis einer Klassifizierungsvariablen getrennt werden.

Das Kriterium, möglichst alle Lebensmittelgruppen und die dazugehörigen verzehrten Lebensmittelmengen zu berücksichtigen, wird durch die drei explorativen Methoden Hauptkomponentenanalyse, Clusteranalyse und Entscheidungsbaum vollständig und durch den HEI-NVS II fast vollständig erfüllt.

Das Kriterium möglichst alle Lebensmittelgruppen und deren verzehrten Lebensmittelmengen über die Ergebnisse darzustellen, wird lediglich durch die Clusteranalyse erfüllt. Durch den HEI-NVS II werden nur Lebensmittelgruppen, für die es quantifizierbare Empfehlung gibt, dargestellt. Bei der Hauptkomponentenanalyse werden nur die charakteristischen Lebensmittelgruppen einer Hauptkomponente dargestellt. Auch durch den Entscheidungsbaum wird das Kriterium nicht erfüllt, da nur die Lebensmittelgruppen, die für die Differenzierung relevant waren, über den Baum dargestellt werden.

Das Kriterium, Personengruppen mit ähnlichem Lebensmittelverzehr zu identifizieren, wird durch die Ergebnisse der Clusteranalyse und des Entscheidungsbaums, nicht aber durch die Ergebnisse des HEI-NVS II und der Hauptkomponentenanalyse erfüllt. Allerdings weisen die Ergebnisse der Clusteranalyse in der Gruppe der Männer auf keine sinnvolle Identifizierung von Personengruppen mit einem ähnlichen Lebensmittelverzehr hin.

Das Kriterium, eine Differenzierung von Bio-Käufer/innen und Nicht-Bio-Käufer/innen vorzunehmen, wurde lediglich durch den HEI-NVS II erfüllt. Bei der Hauptkomponentenanalyse und dem Entscheidungsbaum wurde anhand anderer Differenzierungsvariablen mit der gleichen Studiengruppen aufgezeigt, dass sich diese Methoden grundsätzlich zum Differenzieren von Personengruppen eignen. Wie in der Literatur gezeigt, kann auch die Clusteranalyse zum Differenzieren geeignet sein.

Das Kriterium der Charakterisierung der identifizierten und differenzierten Personengruppen konnten nur teilweise geprüft werden, da nur teilweise Personengruppen identifiziert und differenziert werden konnten. Grundsätzlich können alle vier Methoden geeignet sein, Personengruppen weiter zu charakterisieren.

Die Prüfung der vier Methoden zeigt, dass keine Methode alle Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs anhand der Daten der NVS II erfüllt. Alle vier Methoden haben unterschiedliche Vor- und Nachteile, die bei der umfassenden Darstellung des Lebensmittelverzehrs von Bedeutung sind und bei der Wahl einer geeigneten Methode berücksichtigt werden müssen. Beispielsweise kann durch den HEI-NVS II im Gegensatz zu den anderen drei Methoden jede beliebige Personengruppe wie Bio-Käufer/innen und Nicht-Bio-Käufer/innen differenziert werden. Des Weiteren können mit dem HEI-NVS II Aussagen über die Qualität der Ernährung, bezogen auf den Vergleich mit den Empfehlungen der DGE, getroffen werden. Durch die Hauptkomponentenanalyse wurden am wenigsten Kriterien zur umfassenden Darstellung des Lebensmittelverzehrs erfüllt. Stattdessen ist diese Methode prädestiniert, Ursache-Wirkungs-Zusammenhänge zu untersuchen. Die Clusteranalyse ist theoretisch am besten geeignet, um Personengruppen mit ähnlichem Lebensmittelverzehr zu identifizieren. Allerdings ist dies in der vorliegenden Auswertung nicht gelungen. Der Entscheidungsbaum bietet Potential, Personengruppen auf Basis des Lebensmittelverzehrs

zu identifizieren und zu differenzieren, auch wenn die Methode in der vorliegenden Arbeit nicht zwischen Bio-Käufer/innen und Nicht-Bio-Käufer/innen differenzieren konnte. Zudem liefert der Entscheidungsbaum Erkenntnisse über die Wichtigkeit von Lebensmittelgruppen, die zur Differenzierung von Personengruppen beitragen, was durch keine der anderen Methoden möglich ist. Der Entscheidungsbaum, welcher in der Literatur in diesem Forschungsbereich bisher nicht beschrieben wurde, weist Potential zur umfassenden Darstellung des Lebensmittelverzehr auf.

Diese Ergebnisse zeigen, dass durch jede einzelne Methode Erkenntnisse über die Studiengruppen in Bezug auf die Forschungsfrage gewonnen werden können. Die Methoden haben unterschiedliche Stärken und sollten als sich ergänzend gesehen werden. In Abhängigkeit der Forschungsfrage sollte daher in Erwägung gezogen werden, mehr als eine Methode anzuwenden, um die daraus gewonnenen Erkenntnisse zu kombinieren.

## 7 Summary

In presented dissertation an index, a principal component analysis, a cluster analysis and a decision tree are tested for their applicability to perform a comprehensive description of food consumption using data of the NVS II. In current literature, comparison of only single but not all four methods are described. Furthermore, there is no publication using a decision tree for a comprehensive description of food consumption.

For testing these methods, five criteria are used: (1) consideration of preferably all food groups and their corresponding consumption by using the method (2) presentation of preferably all food groups and their corresponding consumption in the results (3) identification of subgroups with similar food consumption, (4) differentiation of buyers and non-buyers of organic food and (5) characterization of the identified and differentiated subgroups.

With the developed index (HEI-NVS II) food consumption is compared to the dietary guidelines of the DGE and rated by using a point-based system. This results in an index score and an index sum. Applying principal component analysis food consumption is expressed as main components, which are created by correlations of food consumption. Each principal component expresses a characteristic combination of food consumption. For each person, a factor score is calculated for each principal component. The factor score gives information on how well a person's food consumption is mirrored by a specific principal component. With cluster analysis homogeneous subgroups with similar food consumption are classified. The description of food consumption in each cluster is presented by the mean food consumption of the persons allocated to a cluster. With the method of the decision tree, subgroups are separated depending on a classification variable using differences in food consumption. Food consumption is described through leafs grouping subjects with similar food consumption. The ratios of correct and false classification provide information about how reliable persons can be separated by using this classification variable.

The criterion to consider of preferably all food groups and their corresponding consumption is fulfilled by the three explorative methods, principal component analysis, cluster analysis and decision tree and nearly fulfilled by the HEI-NVS II.

The criterion to present in the results of preferably all food groups, and their corresponding consumption is only fulfilled by cluster analysis. In the HEI-NVS II, only food groups which are quantified by dietary guidelines can be described. With the principal component analysis only characteristic food groups of the respective principal components are described. Also with the decision tree this criterion is not fulfilled since only food groups which are relevant for differentiation are described.

The criterion to identify subgroups with similar food consumption is fulfilled by the results of the cluster analysis and the decision tree but not by the results of the HEI-NVS II and the principal component analysis. However, results of cluster analysis for men do not indicate a reasonable identification of subgroups with similar food consumption.

The criterion to differentiate between buyers and non-buyers of organic food was solely fulfilled by the HEI-NVS II. However, as demonstrated by using other differentiating variables on the same study group principal component analysis and decision tree are generally suitable to differentiate between subgroups. As it is shown in literature, cluster analysis can also be used to achieve differentiation.

The criterion to characterize the identified and differentiated subgroups could only partially be tested, because identification and differentiation was only partially possible. Generally, all four methods can be used to characterize these subgroups.

The assessment of the four methods shows, that none of the methods fulfills all criteria for a comprehensive description of food consumption based on data of the NVS II. All four methods have different strengths and limitations which are relevant for the comprehensive description of food consumption and which have to be considered for the selection of a suitable method. For example, in contrast to the other three methods the HEI-NVS II allows to differentiate subgroups like buyers and non-buyers of organic food. Furthermore, the HEI-NVS II provides information about quality of diet compared to dietary guidelines of the DGE. Principal component analysis fulfilled the fewest criteria for a comprehensive description of food consumption. Instead, this method is predestined to analyse cause-effect-relationships. In theory, cluster analysis is most suitable to identify subgroups with similar food consumption. However, in this study this could not be demonstrated. The decision tree shows potential to identify and differentiate subgroups on the basis of food consumption even though this method in the presented analysis was not able to differentiate between buyers and non-buyers of organic food. Additionally, the decision tree provides insights into the relevance of food groups which contributes to the differentiation between subgroups, which no other method provides. The decision tree, which was never mentioned in literature in this field of research, has potential for a comprehensive description of food consumption.

The results show, that each method provides insights about the study population in regard to the specific research question. The methods exhibit different strengths and should be seen as additional. Depending on the research question it should be taken into consideration to apply more than one method to combine the gained insights.

## 8 Literatur

- Anderson AL, Harris TB, Tyllavsky FA, Perry SE, Houston DK, Hue TF, Strotmeyer ES, Sahyoun NR, Hlth ABCS: Dietary patterns and survival of older adults. *J Am Diet Assoc* 111 (1), 84–91, 2011
- Angelopoulos P, Kourlaba G, Kondaki K, Fragiadakis GA, Manios Y: Assessing children's diet quality in Crete based on Healthy Eating Index: the Children Study. *Eur J Clin Nutr* 63 (8), 964–969, 2009
- Arkkola T, Uusitalo U, Kronberg-Kippilä C, Männistö S, Virtanen M, Kenward MG, Veijola R, Knip M, Ovaskainen ML, Virtanen SM: Seven distinct dietary patterns identified among pregnant Finnish women - Associations with nutrient intake and sociodemographic factors. *Public Health Nutr* 11 (2), 176–182, 2008
- Azadbakht L, Esmailzadeh A: Dietary patterns and attention deficit hyperactivity disorder among Iranian children. *Nutrition* 28 (3), 242–249, 2012
- Backhaus K, Erichson B, Plinke W, Weiber R: Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Springer, Heidelberg, 12. Auflage, 2008
- Bamia C, Orfanos P, Ferrari P, Overvad K, Hundborg HH, Tjønneland A, Olsen A, Kesse E, Boutron-Ruault MC, Clavel-Chapelon F, Nagel G, Boffetta P, Boeing H, Hoffmann K, Trichopoulos D, Baibas N, Psaltopoulou T, Norat T, Slimani N, Palli D, Krogh V, Panico S, Tumino R, Sacerdote C, Bueno-de-Mesquita HB, Ocke MC, Peeters PH, van Rossum CT, Quiros JR, Sanchez MJ, Navarro C, Barricarte A, Dorronsoro M, Berglund G, Wirfalt E, Hallmans G, Johansson I, Bingham S, Khaw KT, Spencer EA, Roddam AW, Riboli E, Trichopoulos A: Dietary patterns among older Europeans: the EPIC-Elderly study. *Brit J Nutr* 94 (1), 100–113, 2005
- Bau A-M, Gothe R, Borde T: Ernährungsverhalten von 3- bis 6-jährigen Kindern verschiedener Ethnien – Ergebnisse einer Kitastudie in Berlin. *Ernährungs-Umschau* 50 (6), 214–218, 2003
- Bazelmans C, De Henauw S, Matthys C, Dramaix M, Kornitzer M, De Backer G, Leveque A: Healthy food and nutrient index and all cause mortality. *European journal of epidemiology* 21 (2), 145–152, 2006
- Beierle C, Kern-Isberner G: Methoden wissenschaftlicher Systeme. Grundlagen, Algorithmen, Anwendung. Fried. Vieweg & Sohn, Wiesbaden, 3., erweiterte Auflage, 2006
- Bertuccio P, Edefonti V, Bravi F, Ferraroni M, Pelucchi C, Negri E, Decarli A, La Vecchia C: Nutrient dietary patterns and gastric cancer risk in Italy. *Cancer Epidem Biomar* 18 (11), 2882–2886, 2009
- Biesalski HK, Grimm P: Taschenatlas der Ernährung. Thieme, Stuttgart, 4. Auflage, 2004
- Bingham CM, Jallinoja P, Lahti-Koski M, Absetz P, Paturi M, Pihlajamäki H, Sahi T, Uutela A: Quality of diet and food choices of Finnish young men: a sociodemographic and health behaviour approach. *Public Health Nutr* 13 (6A), 980–986, 2010
- Bodenstedt A, Oltersdorf U, Boeing H, Hendrichs A, Behrens U: Erfassung und Deutung des menschlichen Ernährungsverhaltens „Ernährungsmodell-Studie in Giessen“ (EMSIG). Forschungsbericht, Giessen, 1983. Internet: [http://www.ernaehrungsdenkwerkstatt.de/fileadmin/user\\_upload/EDWText/TextElemente/Publikationen/071\\_EMSIG\\_Berichtmenschlichen\\_Ernaehrungsverhaltens.pdf](http://www.ernaehrungsdenkwerkstatt.de/fileadmin/user_upload/EDWText/TextElemente/Publikationen/071_EMSIG_Berichtmenschlichen_Ernaehrungsverhaltens.pdf)
- Boggs DA, Palmer JR, Spiegelman D, Stampfer MJ, Adams-Campbell LL, Rosenberg L: Dietary patterns and 14-y weight gain in African American women. *Am J Clin Nutr* 94 (1), 86–94, 2011
- Bortz J, Döring N: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Springer-Verlag, Berlin, 4. Auflage, 2009
- Bortz J, Schuster H: Statistik: Für Human- und Sozialwissenschaftler. Springer, Berlin, Heidelberg 7. Auflage, 2010
- Bosetti C, Gallus S, Trichopoulos A, Talamini R, Franceschi S, Negri E, La Vecchia C: Influence of the Mediterranean diet on the risk of cancers of the upper aerodigestive tract. *Cancer epidemiol biomarkers Prev* 12 (10), 1091–1094, 2003

- Boynton A, Neuhauser ML, Sorensen B, McTiernan A, Ulrich CM: Predictors of diet quality among overweight and obese postmenopausal women. *J Am Diet Assoc* 108 (1), 125–130, 2008
- Bradshaw PT, Siega-Riz AM, Campbell M, Weissler MC, Funkhouser WK, Olshan AF: Associations Between Dietary Patterns and Head and Neck Cancer The Carolina Head and Neck Cancer Epidemiology Study. *Am J Epidemiol* 175 (12), 1225–1233, 2012
- Breiman L, Friedmann JH, Olshen RA, Stone CJ: Classification and regression trees. Wadsworth international group, Belmont, California, 1984
- Brennan SF, Cantwell MM, Cardwell CR, Velentzis LS, Woodside JV: Dietary patterns and breast cancer risk: a systematic review and meta-analysis. *Am J Clin Nutr* 91 (5), 1294–1302, 2010
- Brombach C, Wagner U, Eisinger-Watzl M, Heyer A: Die Nationale Verzehrsstudie II. Ziel: Aktuelle und belastbare Primärdaten für die Ernährungsberichterstattung des Bundes generieren. *Ernährungs-Umschau* 53 (6), 4–9, 2006
- Burke BS, Beal VA, Kirkwood SB, Stuart HC: The influence of nutrition during pregnancy upon the condition of the infant at birth. *Jour Nutrition* 26 (6), 569–583, 1943
- Burke BS, Stuart HC: A method of diet analysis. Application in research and pediatric practice. *The Journal of Pediatrics* 12 (4), 493–503, 1938
- Butler LM, Wu AH, Wang R, Koh W-P, Yuan J-M, Yu MC: A vegetable-fruit-soy dietary pattern protects against breast cancer among postmenopausal Singapore Chinese women. *Am J Clin Nutr* 91 (4), 1013–1019, 2010
- Charreire H, Kesse-Guyot E, Bertrais S, Simon C, Chaix B, Weber C, Touvier M, Galan P, Hercberg S, Oppert J-M: Associations between dietary patterns, physical activity (leisure-time and occupational) and television viewing in middle-aged French adults. *Brit J Nutr* 105 (6), 902–910, 2011
- Christof K: Formale Segmentierungsverfahren In: Pepels W (Hrsg.): Marktsegmentierung. Erfolgsnischen finden und besetzen. Symposion Publishing GmbH, Düsseldorf, 2. aktualisierte Auflage, 109–140, 2007
- Costacou T, Bamia C, Ferrari P, Riboli E, Trichopoulos D, Trichopoulou A: Tracing the Mediterranean diet through principal components and cluster analyses in the Greek population. *Eur J Clin Nutr* 57 (11), 1378–1385, 2003
- Cutler GJ, Flood A, Hannan P, Neumark-Sztainer D: Multiple sociodemographic and socioenvironmental characteristics are correlated with major patterns of dietary intake in adolescents. *J Am Diet Assoc* 111 (2), 230–240, 2011
- Daniel CR, Prabhakaran D, Kapur K, Graubard BI, Devasenapathy N, Ramakrishnan L, George PS, Shetty H, Ferrucci LM, Yurgalevitch S, Chatterjee N, Reddy KS, Rastogi T, Gupta PC, Mathew A, Sinha R: A cross-sectional investigation of regional patterns of diet and cardio-metabolic risk in India. *Nutrition Journal* 10, 1–13, 2011
- de Kort CAR, Nieuwenhuijsen MJ, Mendez MA: Relationship between maternal dietary patterns and hypospadias. *Paediatric and Perinatal Epidemiology* 25 (3), 255–264, 2011
- de Stefani E, Boffetta P, Fagundes RB, Deneo-Pellegrini H, Ronco AL, Acosta G, Mendilaharsu M: Nutrient patterns and risk of squamous cell carcinoma of the esophagus: a factor analysis in Uruguay. *Anticancer Research* 28 (4C), 2499–2506, 2008a
- de Stefani E, Boffetta P, Ronco AL, Deneo-Pellegrini H, Acosta G, Gutierrez LP, Mendilaharsu M: Nutrient patterns and risk of lung cancer: a factor analysis in Uruguayan men. *Lung Cancer* 61 (3), 283–291, 2008b
- Deichsel G, Trampisch HJ: Biometrie. Clusteranalyse und Diskriminanzanalyse. Gustav Fischer Verlag, Stuttgart, 1985
- del Mar Bibiloni M, Martinez E, Llull R, Pons A, Tur JA: Western and Mediterranean dietary patterns among Balearic Islands' adolescents: socio-economic and lifestyle determinants. *Public Health Nutr* 15 (4), 683–692, 2012

- Devlin UM, McNulty BA, Nugent AP, Gibney MJ: The use of cluster analysis to derive dietary patterns: methodological considerations, reproducibility, validity and the effect of energy mis-reporting. *The Proceedings of the Nutrition Society*, 1–11, 2012
- DGE (Deutsche Gesellschaft für Ernährung): Die dreidimensionale Lebensmittelpyramide. DGE-special 2/2005 2005. Internet: <http://www.dge.de/modules.php?name=News&file=article&sid=481>
- DGE (Deutsche Gesellschaft für Ernährung e.V.): Gemüse und Obst in der Prävention ausgewählter chronischer Krankheiten. Stellungnahme der Deutschen Gesellschaft für Ernährung e.V., Bonn, 2012. Internet: <http://www.dge.de/pdf/ws/DGE-Stellungnahme-Gemuese-Obst-2012.pdf>
- DGE, AID (Deutsche Gesellschaft für Ernährung e.V., AID Infodienst Ernährung Landwirtschaft Verbraucherschutz e.V.): Vollwertig essen und trinken nach den 10 Regeln der DGE. DCM Druck Center, Bonn, 25., überarbeitete Auflage, 2013
- DGE, ÖGE, SGE, SVE (Deutsche Gesellschaft für Ernährung, Österreichische Gesellschaft für Ernährung, Schweizerische Gesellschaft für Ernährungsforschung, Schweizerische Vereinigung für Ernährung) (Hrsg.): Referenzwerte für die Nährstoffzufuhr. Umschau Verlag, Neustadt an der Weinstrasse, 1. Auflage, 5., korrigierter Nachdruck, 2013
- Drescher LS, Thiele S, Mensink GBM: A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *Journal of nutrition* 137 (3), 647–651, 2007
- Drewnowski A, Fiddler EC, Dauchet L, Galan P, Hercberg S: Diet quality measures and cardiovascular risk factors in France: applying the Healthy Eating Index to the SU.VI.MAX study. *Journal of the American College of Nutrition* 28 (1), 22–29, 2009
- Drewnowski A, Henderson SA, Driscoll A, Rolls BJ: The Dietary variety score: assessing diet quality in healthy young and older adults. *J Am Diet Assoc* 97 (3), 266–271, 1997
- Drewnowski A, Henderson SA, Shore AB, Fischler C, Preziosi P, Hercberg S: Diet quality and dietary diversity in France: implications for the French paradox. *J Am Diet Assoc* 96 (7), 663–669, 1996
- Edefonti V, Bravi F, Garavello W, La Vecchia C, Parpinel M, Franceschi S, Dal Maso L, Bosetti C, Boffetta P, Ferraroni M, Decarli A: Nutrient-based dietary patterns and laryngeal cancer: Evidence from an exploratory factor analysis. *Cancer Epidem Biomar* 19 (1), 18–27, 2010
- Edefonti V, Decarli A, La Vecchia C, Bosetti C, Randi G, Franceschi S, Dal Maso L, Ferraroni M: Nutrient dietary patterns and the risk of breast and ovarian cancers. *International Journal of Cancer* 122 (3), 609–613, 2008
- Edefonti V, Randi G, La Vecchia C, Ferraroni M, Decarli A: Dietary patterns and breast cancer: a review with focus on methodological issues. *Nutr Rev* 67 (6), 297–314, 2009
- Eisenführ F, Weber M, Langer T: Rationales Entscheiden. Springer, Heidelberg, 5. Auflage, 2010
- Eisinger-Watzl M, Hoffmann I: Beschreibung der Untersuchungsgruppe. In: Hoffmann I, Spiller A (Hrsg.): Auswertung der Daten der Nationalen Verzehrsstudie II (NVS II): eine integrierte verhaltens- und lebensstilbasierte Analyse des Bio-Konsums. 11–50, 2010
- Eisinger-Watzl M, Wittig F, Heuer T, Hoffmann I: Nationale Verzehrsstudie II (NVS II). In: Hoffmann I, Spiller A (Hrsg.): Auswertung der Daten der Nationalen Verzehrsstudie II (NVS II): eine integrierte verhaltens- und lebensstilbasierte Analyse des Bio-Konsums. 5–10, 2010
- Erber E, Hopping BN, Grandinetti A, Park S-Y, Kolonel LN, Maskarinec G: Dietary patterns and risk for diabetes The multiethnic cohort. *Diabetes care* 33 (3), 532–538, 2010
- Ester M, Sander J: Knowledge discovery in databases. Techniken und Anwendung. Springer, Berlin, 2000
- Flury B, Riedwyl H: Angewandte multivariate Statistik. Computergestützte Analyse mehrdimensionaler Daten. Gustav Fischer Verlag, Stuttgart, 1983

- Ford ES, Bergmann MM, Kroger J, Schienkiewitz A, Weikert C, Boeing H:* Healthy living Is the best revenge findings from the European Prospective Investigation Into Cancer and Nutrition-Potsdam Study. *Arch Intern Med* 169 (15), 1355–1362, 2009
- Ford ES, Mokdad AH, Liu S:* Healthy Eating Index and C-reactive protein concentration: findings from the National Health and Nutrition Examination Survey III, 1988-1994. *Eur J Clin Nutr* 59 (2), 278–283, 2005
- Fung TT, Hu FB, Barbieri RL, Willett WC, Hankinson SE:* Dietary patterns, the Alternate Healthy Eating Index and plasma sex hormone concentrations in postmenopausal women. *International journal of cancer* 121 (4), 803–809, 2007
- Fung TT, Willett WC, Stampfer MJ, Manson JE, Hu FB:* Dietary patterns and the risk of coronary heart disease in women. *Arch Intern Med* 161 (15), 1857–1862, 2001
- Ganesan S, Raman R, Kulothungan V, Sharma T:* Influence of dietary-fibre intake on diabetes and diabetic retinopathy: Sankara Nethralaya-Diabetic Retinopathy Epidemiology and Molecular Genetic Study (report 26). *Clin Experiment Ophthalmol* 40 (3), 288–294, 2012
- Greenwood DC, Cade JE, Draper A, Barrett JH, Calvert C, Greenhalgh A:* Seven unique food consumption patterns identified among women in the UK Women's Cohort Study. *Eur J Clin Nutr* 54 (4), 314–320, 2000
- Gu Y, Scarmeas N:* Dietary patterns in Alzheimer's disease and cognitive aging. *Current Alzheimer research* 8 (5), 510–519, 2011
- Guallar-Castillon P, Rodriguez-Artalejo F, Tormo MJ, Sanchez MJ, Rodriguez L, Quiros JR, Navarro C, Molina E, Martinez C, Marin P, Lopez-Garcia E, Larranaga N, Huerta JM, Dorronsoro M, Chirlaque MD, Buckland G, Barricarte A, Banegas JR, Arriola L, Ardanaz E, Gonzalez CA, Moreno-Iribas C:* Major dietary patterns and risk of coronary heart disease in middle-aged persons from a Mediterranean country: the EPIC-Spain cohort study. *Nutrition, Metabolism and Cardiovascular Diseases* 22 (3), 192–199, 2012
- Guenther PM, Reedy J, Krebs-Smith SM:* Development of the Healthy Eating Index-2005. *J Am Diet Assoc* 108 (11), 1896–1901, 2008
- Guenther PM, Reedy J, Krebs-Smith SM, Reeve BB, Basiotis PP:* Development and evaluation of the Healthy Eating Index-2005: Technical Report. Center for Nutrition Policy and Promotion, U.S. Department of Agriculture, Washington, D.C., 2007. Internet: <http://www.cnpp.usda.gov/Publications/HEI/HEI-2005/HEI-2005TechnicalReport.pdf>
- Guo H, Niu K, Monma H, Kobayashi Y, Guan L, Sato M, Minamishima D, Nagatomi R:* Association of Japanese dietary pattern with serum adiponectin concentration in Japanese adult men. *Nutrition, Metabolism and Cardiovascular Diseases* 22 (3), 277–284, 2012
- Guo X, Warden BA, Paeratakul S, Bray GA:* Healthy Eating Index and obesity. *Eur J Clin Nutr* 58 (12), 1580–1586, 2004
- Hamer M, Mishra GD:* Dietary patterns and cardiovascular risk markers in the UK Low Income Diet and Nutrition Survey. *Nutrition Metabolism and Cardiovascular Diseases* 20 (7), 491–497, 2010
- Hare-Bruun H, Togo P, Andersen LB, Heitmann BL:* Adult Food Intake Patterns Are Related to Adult and Childhood Socioeconomic Status. *Journal of Nutrition* 141 (5), 928–934, 2011
- Hartmann C, Siegrist M, van der Horst K:* Snack frequency: associations with healthy and unhealthy food choices. *Public Health Nutr*, 1–10, 2012
- Hearty AP, Gibney MJ:* Analysis of meal patterns with the use of supervised data mining techniques--artificial neural networks and decision trees. *Am J Clin Nutr* 88 (6), 1632–1642, 2008
- Hearty AP, Gibney MJ:* Comparison of cluster and principal component analysis techniques to derive dietary patterns in Irish adults. *Br J Nutr* 101 (4), 598–608, 2009

- Heidemann C*: Identifizierung und Charakterisierung eines mit dem Risiko für Typ-2-Diabetes assoziierten Ernährungsmusters in der EPIC-Potsdam-Studie, Berlin, Technische Universität Berlin, 2006
- Heidemann C, Hoffmann K, Spranger J, Klipstein-Grobusch K, Mohlig M, Pfeiffer AF, Boeing H*: A dietary pattern protective against type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC) - Potsdam Study cohort. *Diabetologia* 48 (6), 1126–1134, 2005
- Hische M, Luis-Dominguez O, Pfeiffer AF, Schwarz PE, Selbig J, Spranger J*: Decision trees as a simple-to-use and reliable tool to identify individuals with impaired glucose metabolism or type 2 diabetes mellitus. *Eur J Endocrinol* 163 (4), 565–571, 2010
- Hoffmann I*: Transcending reductionism in nutrition research. *Am J Clin Nutr* 78 (3 Suppl), 514S–516S, 2003
- Hoffmann I, Schneider K, Leitzmann C (Hrsg.)*: Ernährungsökologie. Komplexen Herausforderungen integrativ begegnen. Oekom, München, 2011
- Hoffmann I, Spiller A*: Auswertung der Daten der Nationalen Verzehrsstudie II (NVS II): eine integrierte verhaltens- und lebensstilbasierte Analyse des Bio-Konsums. 2010a. Internet: <http://orgprints.org/18055>
- Hoffmann I, Spiller A*: Ziele und Aufgabenstellung des Projektes. In: Hoffmann I, Spiller A (Hrsg.): Auswertung der Daten der Nationalen Verzehrsstudie II (NVS II): eine integrierte verhaltens- und lebensstilbasierte Analyse des Bio-Konsums. 1–4, 2010b
- Hoffmann I, Wittig F*: Ernährungsforschung: Reduktionismus und integrative Ansätze. In: Hoffmann I, Schneider K, Leitzmann C (Hrsg.): Ernährungsökologie. Komplexen Herausforderungen integrativ begegnen. oekom, München, 112–120, 2011
- Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U, Boeing H*: Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *Am J Epidemiol* 159 (10), 935–944, 2004
- Hong S, Song Y, Lee KH, Lee HS, Lee M, Jee SH, Joung H*: A fruit and dairy dietary pattern is associated with a reduced risk of metabolic syndrome. *Metabolism-Clinical and Experimental* 61 (6), 883–890, 2012
- Hornik K, Buchta C, Zeileis A*: Open-source machine learning: R meets Weka. *Computational statistics* 24 (2), 225–232, 2009
- Hu FB*: Dietary pattern analysis: a new direction in nutritional epidemiology. *Current opinion in lipidology* 13 (1), 3–9, 2002
- Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, Willett WC*: Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *The New England journal of medicine* 345 (11), 790–797, 2001
- Hunt E, Marin J, Stone PJ*: Experiments in induction. Academic Press, New York, 1966
- Hurley KM, Oberlander SE, Merry BC, Wroblewski MM, Klassen AC, Black MM*: The healthy eating index and youth healthy eating index are unique, nonredundant measures of diet quality among low-income, African American adolescents. *The journal of nutrition* 139 (2), 359–364, 2009
- Jacobs DR, Jr., Gross MD, Tapsell LC*: Food synergy: an operational concept for understanding nutrition. *Am J Clin Nutr* 89 (5), 1543S–1548S, 2009
- Jacobs DR, Jr., Steffen LM*: Nutrients, foods, and dietary patterns as exposures in research: a framework for food synergy. *Am J Clin Nutr* 78 (3 Suppl), 508S–513S, 2003
- Jacques PF, Tucker KL*: Are dietary patterns useful for understanding the role of diet in chronic disease? *Am J Clin Nutr* 73 (1), 1–2, 2001
- Jakovljevic J, Touillaud MS, Bondy ML, Singletary SE, Pillow PC, Chang S*: Dietary intake of selected fatty acids, cholesterol and carotenoids and estrogen receptor status in premenopausal breast cancer patients. *Breast Cancer Research and Treatment* 75 (1), 5–14, 2002
- James FC, Mcculloch CE*: Multivariate-Analysis in ecology and Systematics - Panacea or Pandora Box. *Annu Rev Ecol Syst* 21, 129-166, 1990

- Janssen I, Katzmarzyk PT, Srinivasan SR, Chen W, Malina RM, Bouchard C, Berenson GS:* Combined influence of body mass index and waist circumference on coronary artery disease risk factors among children and adolescents. *Pediatrics* 115 (6), 1623–1630, 2005
- Jordan I, Hebestreit A, Swai B, Krawinkel MB:* Dietary patterns and breast cancer risk among women in northern Tanzania: a case-control study. *Eur J Nutr* 52 (3), 905–915, 2013
- Kant A, Graubard B, Schatzkin A:* Dietary patterns predict mortality in a national cohort: the National Health Interview Surveys, 1987 and 1992. *J Nutr* 134 (7), 1793–1799, 2004
- Kant AK:* Indexes of overall diet quality: a review. *J Am Diet Assoc* 96 (8), 785–791, 1996
- Kant AK:* Dietary patterns and health outcomes. *J Am Diet Assoc* (104), 615–635, 2004
- Kant AK, Schatzkin A, Graubard BI, Schairer C:* A prospective study of diet quality and mortality in women. *JAMA* 283 (16), 2109–2115, 2000
- Kant AK, Schatzkin A, Harris TB, Ziegler RG, Block G:* Dietary diversity and subsequent mortality in the First National Health and Nutrition Examination Survey Epidemiologic Follow-up Study. *Am J Clin Nutr* 57 (3), 434–440, 1993
- Kasper H:* Ernährungsmethodik und Diätetik. Urban & Fischer, München, 2000
- Kennedy ET, Ohls J, Carlson S, Fleming K:* The Healthy Eating Index: design and applications. *J Am Diet Assoc* 95 (10), 1103–1108, 1995
- Kessler W:* Multivariate Datenanalyse: für die Pharma-, Bio- und Prozessanalytik: Für Die Pharma, Bio Und Prozessanalytik Wiley-VCH Verlag GmbH & Co. KGaA, 2006
- Kessler W:* Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik. Ein Lehrbuch. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2007
- Kim J, Jo I, Joung H:* A rice-based traditional dietary pattern is associated with obesity in Korean adults. *J Acad Nutr Diet* 112 (2), 246–253, 2012
- Kim S, Haines PS, Siega-Riz AM, Popkin BM:* The Diet Quality Index-International (DQI-I) provides an effective tool for cross-national comparison of diet quality as illustrated by China and the United States. *The journal of nutrition* 133 (11), 3476–3484, 2003
- Kleiser C, Mensink GB, Scheidt-Nave C, Kurth BM:* HuSKY: A healthy nutrition score based on food intake of children and adolescents in Germany. *The British journal of nutrition* 102 (4), 610–618, 2009
- Klir GJ:* Facets of systems science. Kluwer Academic/Plenum Publisher, New York, Volume 15, 2. ed., 2001
- Knol LL, Haughton B, Fitzhugh EC:* Dietary patterns of young, low-income US children. *J Am Diet Assoc* 105 (11), 1765–1773, 2005
- Kokai G:* Erfolge und Probleme evolutionärer Algorithmen, induktiver logischer Programmierung und ihrer Kombination, Technische Fakultät der Universität Erlangen-Nürnberg, 2002
- Krems C, Bauch A, Götz A, Heuer T, Hild A, Möseneder J, Brombach C:* Methoden der Nationalen Verzehrsstudie II. *Ernährungs-Umschau* 53 (6), 44–50, 2006
- Krems C, Brombach C:* Wege zu einem Healthy Nutrition Lifestyle Index auf Grundlage der Zweiten Nationalen Verzehrsstudie. *Proceedings of the German Nutrition Society* (10), 2007
- Kruzich LA, Marquis GS, Wilson CM, Stephensen CB:* HIV-infected US youth are at high risk of obesity and poor diet quality: a challenge for improving short- and long-term health outcomes. *J Am Diet Assoc* 104 (10), 1554–1560, 2004
- Lee JE, Kim J-H, Son SJ, Ahn Y, Lee J, Park C, Lee L, Erickson KL, Jung I-K:* Dietary pattern classifications with nutrient intake and health-risk factors in Korean men. *Nutrition* 27 (1), 26–33, 2011
- Leitzmann C:* "Healthy Eating Index" and "Optimal Diet Membership" Nutritional Evaluation. *Ernährungs Umschau* 59 (7), 380–385, 2012
- Lockheart MS, Steffen LM, Rebnord HM, Fimreite RL, Ringstad J, Thelle DS, Pedersen JJ, Jacobs DR, Jr.:* Dietary patterns, food groups and myocardial infarction: a case-control study. *Brit J Nutr* 98 (2), 380–387, 2007

- Ma Y, Li W, Olendzki BC, Pagoto SL, Merriam PA, Chiriboga DE, Griffith JA, Bodenlos J, Wang Y, Ockene IS: Dietary quality 1 year after diagnosis of coronary heart disease. *J Am Diet Assoc* 108 (2), 240–246, 2008
- Mahner M, Bunge M: *Philosophische Grundlagen der Biologie*. Springer, Berlin, 2000
- Mair J, Smidt J, Lechleitner P, Dienstl F, Puschendorf B: A decision tree for the early diagnosis of acute myocardial infarction in nontraumatic chest pain patients at hospital admission. *Chest* 108 (6), 1502–1509, 1995
- Malik VS, Fung TT, van Dam RM, Rimm EB, Rosner B, Hu FB: Dietary patterns during adolescence and risk of type 2 diabetes in middle-aged women. *Diabetes care* 35 (1), 12–18, 2012
- Manios Y, Kourlaba G, Kondaki K, Grammatikaki E, Birbilis M, Oikonomou E, Roma-Giannikou E: Diet quality of preschoolers in Greece based on the Healthy Eating Index: the GENESIS study. *J Am Diet Assoc* 109 (4), 616–623, 2009
- Martinez E, Llull R, del Mar Bibiloni M, Pons A, Tur JA: Adherence to the Mediterranean dietary pattern among Balearic Islands adolescents. *Brit J Nutr* 103 (11), 1657–1664, 2010
- Maruyama K, Iso H, Date C, Kikuchi S, Watanabe Y, Wada Y, Inaba Y, Tamakoshi A: Dietary patterns and risk of cardiovascular deaths among middle-aged Japanese: JACC Study. *Nutrition, metabolism, and cardiovascular diseases : NMCD* 1–9, 2012
- McCullough ML, Feskanich D, Stampfer MJ, Giovannucci EL, Rimm EB, Hu FB, Spiegelman D, Hunter DJ, Colditz GA, Willett WC: Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *Am J Clin Nutr* 76 (6), 1261–1271, 2002
- McGowan CA, McAuliffe FM: Maternal dietary patterns and associated nutrient intakes during each trimester of pregnancy. *Public Health Nutr*, 1–11, 2012
- Messina M, Lampe JW, Birt DF, Appel LJ, Pivonka E, Berry B, Jacobs DR, Jr.: Reductionism and the narrowing nutrition perspective: time for reevaluation and emphasis on food synergy. *J Am Diet Assoc* 101 (12), 1416–1419, 2001
- Michie D (Hrsg.): *Expert systems in the micro electronic age*. University Press, Edinburgh, 1979
- Millen BE, Quatromoni PA, Pencina M, Kimokoti R, Nam BH, Cobain S, Kozak W, Appugliese DP, Ordovas J, D'Agostino RB: Unique dietary patterns and chronic disease risk profiles of adult men: the Framingham nutrition studies. *J Am Diet Assoc* 105 (11), 1723–1734, 2005
- Miller PE, Lazarus P, Lesko SM, Muscat JE, Harper G, Cross AJ, Sinha R, Ryczak K, Escobar G, Mauger DT, Hartman TJ: Diet index-based and empirically derived dietary patterns are associated with colorectal cancer risk. *Journal of Nutrition* 140 (7), 1267–1273, 2010
- Miller WL, Crabtree BF, Evans DK: Exploratory study of the relationship between hypertension and diet diversity among Saba Islanders. *Public Health Reports* 107 (4), 426–432, 1992
- Mirmiran P, Azadbakht L, Azizi F: Dietary diversity within food groups: An indicator of specific nutrient adequacy in Tehranian women. *Journal of the American College of Nutrition* 25 (4), 354–361, 2006
- Mirmiran P, Azadbakht L, Esmailzadeh A, Azizi F: Dietary diversity score in adolescents - a good indicator of the nutritional adequacy of diets: Tehran lipid and glucose study. *Asia Pacific Journal of Clinical Nutrition* 13 (1), 56–60, 2004
- Moosbrugger H, Kelava A: *Testtheorie und Fragebogenkonstruktion*. Springer, Heidelberg, 2007
- Moursi MM, Arimond M, Dewey KG, Treche S, Ruel MT, Delpeuch F: Dietary diversity is a good predictor of the micronutrient density of the diet of 6- to 23-month-old children in Madagascar. *Journal of Nutrition* 138 (12), 2448–2453, 2008

- MRI (Max Rubner-Institut, Bundesforschungsinstitut für Ernährung und Lebensmittel): Nationale Verzehrsstudie II. Ergebnisbericht, Teil 1. Die bundesweite Befragung zur Ernährung von Jugendlichen und Erwachsenen. 2008a. Internet: [http://www.mri.bund.de/fileadmin/Institute/EV/NVS\\_II\\_Abschlussbericht\\_Teil\\_1\\_mit\\_Ergaenzungsbericht.pdf](http://www.mri.bund.de/fileadmin/Institute/EV/NVS_II_Abschlussbericht_Teil_1_mit_Ergaenzungsbericht.pdf)
- MRI (Max Rubner-Institut, Bundesforschungsinstitut für Ernährung und Lebensmittel): Nationale Verzehrsstudie II. Ergebnisbericht, Teil 2. Die bundesweite Befragung zur Ernährung von Jugendlichen und Erwachsenen. 2008b. Internet: [http://www.mri.bund.de/fileadmin/Institute/EV/NVSII\\_Abschlussbericht\\_Teil\\_2.pdf](http://www.mri.bund.de/fileadmin/Institute/EV/NVSII_Abschlussbericht_Teil_2.pdf)
- MRI (Max Rubner-Institut, Bundesforschungsanstalt für Ernährung und Lebensmittel): Der Bundeslebensmittelschlüssel (BLS) Version 3.01. 2010. Internet: <http://www.blisdb.de>
- Mullie P, Clarys P, Hulens M, Vansant G: Dietary patterns and socioeconomic position. *Eur J Clin Nutr* 64 (3), 231–238, 2010
- Mursu J, Steffen LM, Meyer KA, Duprez D, Jacobs DR, Jr.: Diet quality indexes and mortality in postmenopausal women: the Iowa Women's Health Study. *Am J Clin Nutr* 98 (2), 444–453, 2013
- Naska A, Fouskakis D, Oikonomou E, Almeida MD, Berg MA, Gedrich K, Moreiras O, Nelson M, Trygg K, Turrini A, Remaut AM, Volatier JL, Trichopoulou A: Dietary patterns and their socio-demographic determinants in 10 European countries: data from the DAFNE databank. *Eur J Clin Nutr* 60 (2), 181–190, 2006
- Newby P, Muller D, Tucker K: Associations of empirically derived eating patterns with plasma lipid biomarkers: a comparison of factor and cluster analysis methods. *Am J Clin Nutr* 80 (3), 759–767, 2004
- Newby PK, Tucker KL: Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* 62 (5), 177–203, 2004
- Noethlings U, Murphy SP, Wilkens LR, Boeing H, Schulze MB, Bueno-de-Mesquita HB, Michaud DS, Roddam A, Rohrmann S, Tjonneland A, Clavel-Chapelon F, Trichopoulou A, Sieri S, Rodriguez L, Ye W, Jenab M, Kolonel LN: A food pattern that is predictive of flavonol intake and risk of pancreatic cancer. *Am J Clin Nutr* 88 (6), 1653–1662, 2008
- Northstone K, Emmett PM: Dietary patterns of men in ALSPAC: associations with socio-demographic and lifestyle characteristics, nutrient intake and comparison with women's dietary patterns. *Eur J Clin Nutr* 64 (9), 978–986, 2010
- Nyholm M, Lissner L, Hornell A, Johansson I, Hallmans G, Weinehall L, Winkvist A: Exploring dietary patterns, obesity and sources of bias: the Vasterbotten Intervention Programme (VIP). *Public Health Nutr*, 1–8, 2012
- Okubo H, Miyake Y, Sasaki S, Murakami K, Tanaka K, Fukushima W, Kiyohara C, Tsuboi Y, Yamada T, Oeda T, Shimada H, Kawamura N, Sakae N, Fukuyama H, Hirota Y, Nagai M, Fukuoka Kinki Parkinsons Dis S: Dietary patterns and risk of Parkinson's disease: a case-control study in Japan. *European Journal of Neurology* 19 (5), 681–688, 2012
- Okubo H, Miyake Y, Sasaki S, Tanaka K, Murakami K, Hirota Y, Osaka M, Child Hlth S: Nutritional adequacy of three dietary patterns defined by cluster analysis in 997 pregnant Japanese women: the Osaka Maternal and Child Health Study. *Public Health Nutr* 14 (4), 611–621, 2011
- Olinto MTA, Willett WC, Gigante DP, Victora CG: Sociodemographic and lifestyle characteristics in relation to dietary patterns among young Brazilian adults. *Public Health Nutr* 14 (1), 150–159, 2011
- Oltersdorf US: Ernährungsepidemiologie. Ulmer, Stuttgart, 1995
- Osler M, Heitmann BL, Gerdes LU, Jorgensen LM, Schroll M: Dietary patterns and mortality in Danish men and women: a prospective observational study. *Brit J Nutr* 85 (2), 219–225, 2001
- Ovaskainen ML, Nevalainen J, Uusitalo L, Tuokkola JJ, Arkkola T, Kronberg-Kippilä C, Veijola R, Knip M, Virtanen SM: Some similarities in dietary clusters of pre-school children and their mothers. *Brit J Nutr* 102 (3), 443–452, 2009

- Palli D, Russo A, Decarli A*: Dietary patterns, nutrient intake and gastric cancer in a high-risk area of Italy. *Cancer Cause Control* 12 (2), 163–172, 2001
- Pan Y, Pratt CA*: Metabolic syndrome and its association with diet and physical activity in US adolescents. *J Am Diet Assoc* 108 (2), 276–286, 2008
- Park HS, Dailey R, Lemus D*: The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research* 28 (4), 562–577, 2002
- Patterson RE, Haines PS, Popkin BM*: Diet quality index: capturing a multidimensional behavior. *J Am Diet Assoc* 94 (1), 57–64, 1994a
- Patterson RE, Haines PS, Popkin BM*: Health lifestyle patterns of U.S. adults. *Prev Med* 23 (4), 453–460, 1994b
- Pavlopoulos SA, Stasis AC, Loukis EN*: A decision tree - based method for the differential diagnosis of Aortic Stenosis from Mitral Regurgitation using heart sounds. *Biomedical engineering online* 3 (1), 21, 2004
- Pedone C, Napoli N, Pozzilli P, Rossi FF, Lauretani F, Bandinelli S, Ferrucci L, Antonelli-Incalzi R*: Dietary pattern and bone density changes in elderly women: a longitudinal study. *Journal of the American College of Nutrition* 30 (2), 149–154, 2011
- Pepels W*: Marktsegmentierung. Erfolgsnischen finden und besetzen. Symposium Publishing GmbH, Düsseldorf, 2. aktualisierte Auflage, 2007a
- Pepels W*: Segmentierungsdeterminanten im Käuferverhalten. In: *Pepels W (Hrsg.): Marktsegmentierung. Erfolgsnischen finden und besetzen. Symposium Publishing GmbH, Düsseldorf, 2. aktualisierte Auflage, 75–108, 2007b*
- Petersohn H*: Data Mining: Verfahren, Prozesse, Anwendungsarchitektur. Oldenbourg, München, 2005
- Previdelli AN, Lipi M, Castro MA, Marchioni DM*: Dietary quality and associated factors among factory workers in the metropolitan region of Sao Paulo, Brazil. *J Am Diet Assoc* 110 (5), 786–790, 2010
- Quinlan JR*: Discovering rules by induction from large collections of examples. In: *Michie D (Hrsg.): Expert systems in the micro electronic age. University Press, Edinburgh, 168–201, 1979*
- Quinlan JR*: Induction of decision trees. *Mach Learn* 1 (1), 81–106, 1986
- Quinlan JR*: C4.5: programs for machine learning. Kaufmann, 1993
- R Core Team*: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., 2014. Internet: <http://www.R-project.org/>
- Randi G, Edefonti V, Ferraroni M, La Vecchia C, Decarli A*: Dietary patterns and the risk of colorectal cancer and adenomas. *Nutrition Reviews* 68 (7), 389–408, 2010
- Reedy J, Wirfalt E, Flood A, Mitrou PN, Krebs-Smith SM, Kipnis V, Midthune D, Leitzmann M, Hollenbeck A, Schatzkin A, Subar AF*: Comparing 3 dietary pattern methods-cluster analysis, factor analysis, and index analysis-with colorectal cancer risk. *Am J Epidemiol* 171 (4), 479–487, 2010
- Rezazadeh A, Rashidkhani B, Omidvar N*: Association of major dietary patterns with socioeconomic and lifestyle factors of adult women living in Tehran, Iran. *Nutrition* 26 (3), 337–341, 2010
- Richter A, Heidemann C, Schulze MB, Roosen J, Thiele S, Mensink GBM*: Dietary patterns of adolescents in Germany - Associations with nutrient intake and other health related lifestyle characteristics. *BMC Pediatrics* 12 (35), 1–14, 2012
- Rifas-Shiman SL, Rich-Edwards JW, Kleinman KP, Oken E, Gillman MW*: Dietary quality during pregnancy varies by maternal characteristics in Project Viva: a US cohort. *J Am Diet Assoc* 109 (6), 1004–1011, 2009
- Rodriguez-Ramirez S, Mundo-Rosas V, Garcia-Guerra A, Shamah-Levy T*: Dietary patterns are associated with overweight and obesity in Mexican school-age children. *Archivos latinoamericanos de nutricion* 61 (3), 270–278, 2011
- Rokach L, Maimon O*: Top-down induction of decision trees classifiers - A survey. *IEEE T Syst Man Cy C* 35 (4), 476–487, 2005

- Royston P*: Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing* 2 (3), 117–119, 1992
- Runkler TA*: *Data Mining. Methoden und Algorithmen intelligenter Datenanalyse*. Vieweg + Teubner, Wiesbaden, 1. Auflage, 2010
- SAS: The cluster procedure. In: SAS (Hrsg.): *SAS/STAT 9.2 Users` s Guide*. SAS Institute, North Carolina, 2nd, 1230-1316, 2009a
- SAS: *SAS/STAT 9.2 Users` s Guide*. SAS Institute, North Carolina, 2nd, 2009b
- SAS Institute Inc*: Cary, North Carolina, USA 2012
- Schauder P, Ollenschläger G*: *Ernährungsmedizin. Prävention und Therapie*. Urban & Fischer, München, 2. Auflage, 2003
- Schendera CFG*: *Datenmanagement und Datenanalyse mit dem SAS-System. Lehr- und Handbuch der Statistik*. Oldenbourg Wissenschaftsverlag GmbH, München, 2004
- Schendera CFG*: *Clusteranalyse mit SPSS: Mit Faktorenanalyse*. Oldenbourg Wissenschaftsverlag GmbH, München, 2010
- Schneider K, Hoffmann I*: Das Konzept der Ernährungsökologie: Herausforderung annehmen. In: Hoffmann I, Schneider K, Leitzmann C (Hrsg.): *Ernährungsökologie. Komplexen Herausforderungen integrativ begegnen*. Oekom, München, 38–45, 2011
- Schneider R (Hrsg.)*: *Vom Umgang mit Zahlen und Daten*. Umschau Zeitschriftenverlag, Frankfurt am Main, 1997
- Schnell R, Hill P, Esser E*: *Methoden der empirischen Sozialforschung*. Oldenbourg Wissenschaftsverlag GmbH, München, 8. Auflage, 2008
- Schulze M*: *Ernährungsmuster als neuer Ansatz zum Verständnis der Rolle der Ernährung in der Ätiologie der essentiellen arteriellen Hypertonie*. Dissertation, Berlin, 2002. Internet: [http://opus.kobv.de/tuberlin/volltexte/2002/412/pdf/schulze\\_matthias.pdf](http://opus.kobv.de/tuberlin/volltexte/2002/412/pdf/schulze_matthias.pdf)
- Schulze MB, Hoffmann K*: Methodological approaches to study dietary patterns in relation to risk of coronary heart disease and stroke. *Brit J Nutr* 95 (5), 860–869, 2006
- Schulze MB, Hoffmann K, Kroke A, Boeing H*: Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. *Brit J Nutr* 85 (3), 363–373, 2001
- Schulze MB, Hoffmann K, Kroke A, Boeing H*: An approach to construct simplified measures of dietary patterns from exploratory factor analysis. *Brit J Nutr* 89 (3), 409–419, 2003a
- Schulze MB, Hoffmann K, Kroke A, Boeing H*: Risk of hypertension among women in the EPIC-Potsdam Study: Comparison of relative risk estimates for exploratory and hypothesis-oriented dietary patterns. *Am J Epidemiol* 158 (4), 365–373, 2003b
- Schwerin HS, Stanton JL, Riley AM, Jr., Schaefer AE, Leveille GA, Elliott JG, Warwick KM, Brett BE*: Food eating patterns and health: a reexamination of the Ten-State and HANES I surveys. *Am J Clin Nutr* 34 (4), 568–580, 1981
- Schwerin HS, Stanton JL, Smith JL, Riley AM, Jr., Brett BE*: Food, eating habits, and health: a further examination of the relationship between food eating patterns and nutritional health. *Am J Clin Nutr* 35 (5 Suppl), 1319–1325, 1982
- Sell K, Gedrich K, Fischer B, Döring A*: Trends im Ernährungsverhalten in der Region Augsburg. *Ernährungs-Umschau* 50 (6), 208–213, 2003
- Seymour JD, Calle EE, Flagg EW, Coates RJ, Ford ES, Thun MJ*: Diet Quality Index as a predictor of short-term mortality in the American Cancer Society Cancer Prevention Study II Nutrition Cohort. *Am J Epidemiol* 157 (11), 980–988, 2003
- Shah BS, Freeland-Graves JH, Cahill JM, Lu H, Graves GR*: Diet quality as measured by the healthy eating index and the association with lipid profile in low-income women in early postpartum. *J Am Diet Assoc* 110 (2), 274–279, 2010
- Shaheen SO, Jameson KA, Syddall HE, Sayer AA, Dennison EM, Cooper C, Robinson SM, Hertfordshire Cohort Study G*: The relationship of dietary patterns with adult lung function and COPD. *European respiratory journal* 36 (2), 277–284, 2010
- Shatenstein B, Nadon S, Godin C, Ferland G*: Diet quality of Montreal-area adults needs improvement: estimates from a self-administered food frequency questionnaire furnishing a dietary indicator score. *J Am Diet Assoc* 105 (8), 1251–1260, 2005

- Slattery ML, Edwards SL, Boucher KM, Anderson K, Caan BJ*: Lifestyle and colon cancer: an assessment of factors associated with risk *Am J Epidemiol* 150 (8), 869–877, 1999
- Statistisches Bundesamt*: Bevölkerung in Privathaushalten im Alter von 18–80 Jahren (Jahresdurchschnitt). Sonderauswertung aus dem Mikrozensus 2006. Wiesbaden, 2010
- Stookey JD, Wang Y, Ge K, Lin H, Popkin BM*: Measuring diet quality in china: the INFH-UNC-CH diet quality index. *Eur J Clin Nutr* 54 (11), 811–821, 2000
- Su Y, Shen J, Qian H, Ma H, Ji J, Ma L, Zhang W, Meng L, Li Z, Wu J, Jin G, Zhang J, Shou C*: Diagnosis of gastric cancer using decision tree classification of mass spectral data. *Cancer science* 98 (1), 37–43, 2007
- Tande DL, Magel R, Strand BN*: Healthy Eating Index and abdominal obesity. *Public Health Nutr* 13 (2), 208–214, 2010
- Torjusen H, Lieblein G, Naes T, Haugen M, Meltzer HM, Brantsaeter AL*: Food patterns and dietary quality associated with organic food consumption during pregnancy; data from a large cohort of pregnant women in Norway. *BMC Public Health* 12, 612, 2012
- Trichopoulou A, Costacou T, Bamia C, Trichopoulos D*: Adherence to a Mediterranean diet and survival in a Greek population. *New Engl J Med* 348 (26), 2599–2608, 2003
- Trichopoulou A, Kourisblazos A, Wahlqvist ML, Gnardellis C, Lagiou P, Polychronopoulos E, Vassilakou T, Lipworth L, Trichopoulos D*: Diet and overall survival in elderly people. *Brit Med J* 311 (7018), 1457–1460, 1995
- Tsien CL, Fraser HS, Long WJ, Kennedy RL*: Using classification tree and logistic regression methods to diagnose myocardial infarction. *Studies in health technology and informatics* 52 Pt 1, 493–497, 1998
- Turner-McGrievy GM, Barnard ND, Cohen J, Jenkins DJ, Gloede L, Green AA*: Changes in nutrient intake and dietary quality among participants with type 2 diabetes following a low-fat vegan diet or a conventional diabetes diet for 22 weeks. *J Am Diet Assoc* 108 (10), 1636–1645, 2008
- Tutz G*: Die Analyse kategorialer Daten. Oldenbourg, München, 2000
- USDA (United States Department of Agriculture, Center for Nutrition Policy and Promotion)*: Healthy Eating Index. CNPP-1. Washington DC, 1995. Internet: <http://www.cnpp.usda.gov/publications/hei/HEI89-90report.pdf>
- van Dam RM*: New approaches to the study of dietary patterns. *The British journal of nutrition* 93 (5), 573–574, 2005
- Vitolins MZ, Toozee JA, Golden SL, Arcury TA, Bell RA, Davis C, Devellis RF, Quandt SA*: Older adults in the rural South are not meeting healthful eating guidelines. *J Am Diet Assoc* 107 (2), 265–272, 2007
- von Normann K*: Münsteraner Ernährungsmuster- und Lebensstilstudie bei Schulkindern „MEALS“ 2006. IÖB-Diskussionspapier 5/07 Münster, 2007. Internet: [http://www.wiwi.uni-muenster.de/ioeb/downloads/forschen/paper/IOEB\\_DP\\_05\\_2007.PDF](http://www.wiwi.uni-muenster.de/ioeb/downloads/forschen/paper/IOEB_DP_05_2007.PDF)
- von Rimscha M*: Algorithmen kompakt und verständlich. Lösungsstrategien am Computer. Vieweg + Teubner, Wiesbaden, 2. Auflage, 2010
- von Rüsten A, Illner A-K, Boeing H, Flothkötter M*: Die Bewertung der Lebensmittelaufnahme mittels eines ‚Healthy Eating Index‘ (HEI-EPIC). *Ernährungs-Umschau* 8 (9), 450–456, 2009
- von Rüsten A, Illner AK, Buijsse B, Heidemann C, Boeing H*: Adherence to recommendations of the German food pyramid and risk of chronic diseases: results from the EPIC-Potsdam study. *Eur J Clin Nutr*, 2010
- Waijers PMCM, Feskens EJM, Ocke MC*: A critical review of predefined diet quality scores. *Brit J Nutr* 97 (2), 219–231, 2007
- WCRF, AICR (World Cancer Research Fund, American Institute for Cancer Research)*: Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington DC, 2007. Internet: <http://www.dietandcancerreport.org>

- WHO (World Health Organization):* Obesity. Preventing and managing the global epidemic. WHO Technical Report Series 894, Genf, 2000. Internet: [http://www.who.int/nutrition/publications/obesity/WHO\\_TRS\\_894/en/](http://www.who.int/nutrition/publications/obesity/WHO_TRS_894/en/)
- WHO (World Health Organization):* Diet, nutrition and the prevention of chronic diseases. Report of a Joint WHO/FAO Expert Consultation, Geneva, 2003. Internet: [http://whqlibdoc.who.int/trs/WHO\\_trs\\_916.pdf](http://whqlibdoc.who.int/trs/WHO_trs_916.pdf)
- Winkler G, Döring A:* Kurzmethoden zur Charakterisierung des Ernährungsmusters: Einsatz und Auswertung eines Food-Frequency-Fragebogens. *Ernährungs-Umschau* 42 (8), 289–291, 1995
- Yakub M, Iqbal MP, Iqbal R:* Dietary patterns are associated with hyperhomocysteinemia in an urban Pakistani population. *Journal of Nutrition* 140 (7), 1261–1266, 2010

## 9 Anhang

### 9.1 Lebensmittelgruppen

Tab. 9.1: Einteilung der Lebensmittel zu den einzelnen Lebensmittelgruppen

Lebensmittelgruppen	Lebensmittel
<b>Obst</b>	rohes und gegartes Obst, Obsterzeugnisse wie Kompott und Trockenobst
<b>Gemüse</b>	rohes und gegartes Gemüse, Pilze, Salat, Hülsenfrüchte
<b>Getreide</b>	Getreide/-erzeugnisse wie Brot, Backwaren, Kekse, Cerealien, Reis, Knödel, Teigwaren
<b>Kartoffeln</b>	Kartoffeln u. a. Knollenfrüchte
<b>Milch</b>	Trinkmilch und Milcherzeugnisse wie Käse, Joghurt, Quark
<b>Fisch</b>	Fisch/-erzeugnisse und Krusten-/Schalentiere
<b>Fleisch</b>	Fleisch/-erzeugnisse und Wurstwaren
<b>Eier</b>	Eier/-erzeugnisse
<b>Alkohol</b>	Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke
<b>Streichfett</b>	Butter, Margarine
<b>Getränke</b>	alkoholfreie Getränke wie Wasser, Kaffee, Tee, Fruchtsäfte und -nektare, Gemüsesäfte, Limonaden, Fruchtsaftgetränke
<b>Süßwaren</b>	Süßigkeiten wie Schokolade sowie Speiseeis, süße Aufstriche, Süßungsmittel, Cremes und Süßspeisen
<b>Nüsse und Samen</b>	Nüsse und Samen sowie geröstete und gesalzene Nüsse

## 9.2 Faktorenanalyse

### Ergebnisse zur dritten und vierten Hauptkomponenten der Frauen

Tab. 9.2: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der dritten Hauptkomponente

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev)	3. Hauptkomponente (MW±Std dev)		Signifikanzen <sup>1</sup>	Faktorladung der 3. Hauptkomponente	Interpretation der 3. Hauptkomponente
		1. Quartile	4. Quartile			
Obst	292±233	233±176	368±310	<0,001	<b>0,20</b>	hoher Obstverzehr
Gemüse	251±150	264±158	257±167	<0,05		
Getreide	239±99	194±79	299±116	<0,001	<b>0,36</b>	hoher Getreideverzehr
Kartoffeln	72±43	86±48	61±39	<0,001		
Nüsse und Samen	4±11	3±8	5±15	n.s.		
Milch	239±211	169±143	347±294	<0,001	<b>0,30</b>	hoher Milchverzehr
Eier	12±13	19±18	8±9	<0,001	<b>-0,30</b>	niedriger Eiverzehr
Fisch	23±22	33±31	15±15	<0,001	<b>-0,29</b>	niedriger Fischverzehr
Fleisch	75±48	97±57	58±43	<0,001	<b>-0,28</b>	niedriger Fleischverzehr
Streichfett	21±17	18±14	25±21	<0,001		
Getränke	2335±889	2172±855	2575±962	<0,001		
Alkohol <sup>2</sup>	6±9	12±13	2±3	<0,001	<b>-0,51</b>	niedriger Alkoholverzehr
Süßwaren	61±56	39±35	95±79	<0,001	<b>0,38</b>	hoher Süßwarenverzehr

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

**Tab. 9.3: Charakterisierung der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der dritten Hauptkomponente**

		Gesamtgruppe (%)	3. Hauptkomponente 1. Quartile (%)	4. Quartile (%)	Signifikanzen <sup>1</sup>
<b>Bio-Kauf</b>	ja	50,3	46,2	55,9	<0,001
	nein	49,7	53,9	44,1	<0,001
<b>Altersgruppe</b>	18–34 Jahre	20,7	15,9	26,4	<0,001
	35–50 Jahre	35,3	33,7	37,9	<0,01
	51–64 Jahre	24,2	31,1	19,7	<0,001
	65–80 Jahre	19,8	19,3	16,0	<0,05
<b>Soziale Schicht<sup>2</sup></b>	untere Schicht	21,0	20,7	19,2	n.s.
	mittlere Schicht	29,9	28,4	29,1	n.s.
	obere Schicht	49,1	50,8	51,8	n.s.
<b>BMI</b>	Untergewicht	1,6	1,2	2,6	<0,01
	Normalgewicht	46,0	43,0	51,4	<0,001
	Übergewicht	27,9	29,0	26,2	n.s.
	Adipositas	18,9	20,8	14,8	<0,001
	keine Angabe	5,6	6,1	5,0	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 88 cm	65,8	63,1	70,8	<0,001
	über 88 cm	29,4	31,7	24,6	<0,001
	keine Angabe	4,8	5,2	4,7	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüft- umfang)</b>	bis 0,85	74,2	70,8	77,4	<0,001
	über 0,85	21,0	24,0	17,9	<0,001
	keine Angabe	4,8	5,2	4,7	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	1968±589			<0,001
<b>Ernährungskennnisse</b>	gut	40,9	36,1	46,6	<0,001
	mittel	33,9	34,9	32,4	n.s.
	schlecht	24,2	27,8	20,2	<0,001
	keine Angabe	1,0	1,2	0,8	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	58,9	54,4	64,4	<0,001
	nein	40,6	45,1	35,2	<0,001
	keine Angabe	0,6	0,5	0,3	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	21,7	26,4	18,7	<0,001
	nein	78,0	73,3	80,7	<0,001
	keine Angabe	0,3	0,3	0,6	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	76,2	75,6	78,3	n.s.
	mittel	20,2	20,4	18,4	n.s.
	schlecht	3,5	4,0	3,0	n.s.
	keine Angabe	0,1	0,1	0,2	--- <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	78,8	77,9	78,8	n.s.
	schlecht	20,7	21,5	20,9	n.s.
	keine Angabe	0,5	0,6	0,3	--- <sup>2</sup>
<b>HEI-NVS II</b>	arithm. MW±SD	84,8±9,8	80,7	87,3	<0,001

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und dem HEI-NVS II über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

**Tab. 9.4: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der vierten Hauptkomponente**

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev)	4. Hauptkomponente (MW±Std dev)		Signifikanzen <sup>1</sup>	Faktorladung der 4. Hauptkomponente	Interpretation der 4. Hauptkomponente
		1. Quartile	4. Quartile			
Obst	292±233	378±312	246±190	<0,001		
Gemüse	251±150	265±166	252±151	<0,05		
Getreide	239±99	195±80	281±121	<0,001	<b>0,27</b>	hoher Getreideverzehr
Kartoffeln	72±43	106±47	51±34	<0,001	<b>-0,42</b>	niedriger Kartoffelverzehr
Nüsse und Samen	4±11	1±2	10±20	<0,001	<b>0,58</b>	hoher Nüsse- und Samenverzehr
Milch	239±211	256±261	233±192	n.s.		
Eier	12±13	12±13	13±13	n.s.		
Fisch	23±22	31±29	19±19	<0,001		
Fleisch	75±48	79±52	72±49	<0,001		
Streichfett	21±17	20±16	22±20	<0,05		
Getränke	2335±889	2076±819	2594±949	<0,001		
Alkohol <sup>2</sup>	6±9	2±3	13±13	<0,001	<b>0,55</b>	hoher Alkoholverzehr
Süßwaren	61±56	53±47	64±67	<0,001		

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

**Tab. 9.5: Charakterisierung der Gesamtgruppe sowie der Frauen der ersten und vierten Quartile der vierten Hauptkomponente**

		Gesamtgruppe (%)	4. Hauptkomponente		Signifikanzen <sup>1</sup>
			1. Quartile (%)	4. Quartile (%)	
<b>Bio-Kauf</b>	ja	50,3	45,3	57,3	<0,001
	nein	49,7	54,7	42,7	<0,001
<b>Altersgruppe</b>	18–34 Jahre	20,7	14,7	23,1	<0,001
	35–50 Jahre	35,3	25,1	42,0	<0,001
	51–64 Jahre	24,2	25,8	24,9	n.s.
	65–80 Jahre	19,8	34,4	10,1	<0,001
<b>Soziale Schicht<sup>2</sup></b>	untere Schicht	21,0	30,9	12,7	<0,001
	mittlere Schicht	29,9	35,0	24,6	<0,001
	obere Schicht	49,1	34,1	62,7	<0,001
<b>BMI</b>	Untergewicht	1,6	1,0	2,1	<0,01
	Normalgewicht	46,0	36,8	55,4	<0,001
	Übergewicht	27,9	30,5	24,9	<0,001
	Adipositas	18,9	26,2	12,2	<0,001
	keine Angabe	5,6	5,5	5,5	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 88 cm	65,8	55,8	75,1	<0,001
	über 88 cm	29,4	39,3	20,3	<0,001
	keine Angabe	4,8	5,0	4,6	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüft-umfang)</b>	bis 0,85	74,2	64,7	81,2	<0,001
	über 0,85	21,0	30,3	14,2	<0,001
	keine Angabe	4,8	5,0	4,6	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	1968±589	1869±570	2170±627	<0,001
<b>Ernährungskenntnisse</b>	gut	40,9	33,3	44,5	<0,001
	mittel	33,9	31,8	36,5	<0,01
	schlecht	24,2	33,1	18,3	<0,001
	keine Angabe	1,0	1,8	0,7	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	58,9	52,2	66,7	<0,001
	nein	40,6	47,0	33,1	<0,001
	keine Angabe	0,6	0,8	0,2	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	21,7	17,5	26,6	<0,001
	nein	78,0	82,3	73,1	<0,001
	keine Angabe	0,3	0,2	0,4	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	76,2	67,3	81,4	<0,001
	mittel	20,2	26,7	16,4	<0,001
	schlecht	3,5	6,0	2,1	<0,001
	keine Angabe	0,1	0,1	0,2	--- <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	78,8	77,2	78,3	n.s.
	schlecht	20,7	22,2	21,3	n.s.
	keine Angabe	0,5	0,6	0,5	--- <sup>2</sup>
<b>HEI-NVS II</b>	arithm. MW±SD	84,8±9,8	87±9	82±11	<0,001

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und dem HEI-NVS II über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

## Ergebnisse zur dritten, vierten und fünften Hauptkomponente der Männer

**Tab. 9.6: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der dritten Hauptkomponente**

Lebensmittel- gruppen	Gesamtgruppe (MW±Std dev)	3. Hauptkomponente (MW±Std dev)		Signifikanzen <sup>1</sup>	Faktorladung der 3. Haupt- komponente	Interpretation der 3. Hauptkomponente
		1. Quartile	4. Quartile			
<b>Obst</b>	249±223	283±257	211±188	<0,001		
<b>Gemüse</b>	239±143	204±127	279±165	<0,001		
<b>Getreide</b>	303±127	369±147	261±114	<0,001	<b>-0,28</b>	niedriger Getreideverzehr
<b>Kartoffeln</b>	92±57	74±48	114±69	<0,001	<b>0,22</b>	hoher Kartoffelverzehr
<b>Nüsse und Samen</b>	5±13	3±9	6±16	<0,001		
<b>Milch</b>	249±258	392±366	166±172	<0,001	<b>-0,29</b>	niedriger Milchverzehr
<b>Eier</b>	16±18	13±15	21±24	<0,001		
<b>Fisch</b>	29±29	19±18	42±39	<0,001	<b>0,27</b>	hoher Fischverzehr
<b>Fleisch</b>	133±83	108±65	171±106	<0,001	<b>0,27</b>	hoher Fleischverzehr
<b>Streichfett</b>	30±27	36±33	29±25	<0,001		
<b>Getränke</b>	2332±981	2616±1094	2085±967	<0,001		
<b>Alkohol<sup>2</sup></b>	16±18	6±7	35±25	<0,001	<b>0,62</b>	hoher Alkoholverzehr
<b>Süßwaren</b>	68±63	111±84	44±41	<0,001	<b>-0,37</b>	niedriger Süßwarenverzehr

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

**Tab. 9.7: Charakterisierung der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der dritten Hauptkomponente**

		Gesamtgruppe (%)	3. Hauptkomponente		Signifikanzen <sup>1</sup>
			1. Quartile (%)	4. Quartile (%)	
<b>Bio-Kauf</b>	ja	38,6	42,6	35,5	<0,001
	nein	61,5	57,4	64,5	<0,001
<b>Altersgruppe</b>	18–34 Jahre	19,7	23,7	14,8	<0,001
	35–50 Jahre	32,0	38,9	28,3	<0,001
	51–64 Jahre	25,3	20,3	30,1	<0,001
	65–80 Jahre	23,0	17,1	26,8	<0,001
<b>Soziale Schicht<sup>3</sup></b>	untere Schicht	17,6	16,1	19,1	<0,05
	mittlere Schicht	28,4	30,1	27,6	n.s.
	obere Schicht	54,1	53,9	53,3	n.s.
<b>BMI</b>	Untergewicht	0,7	0,9	0,7	n.s.
	Normalgewicht	30,0	38,9	24,9	<0,001
	Übergewicht	45,0	42,2	45,6	n.s.
	Adipositas	20,1	14,0	24,0	<0,001
	keine Angabe	4,3	4,0	4,8	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 102 cm	68,6	76,3	62,9	<0,001
	über 102 cm	28,0	20,2	33,1	<0,001
	keine Angabe	3,4	3,5	4,0	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüft-umfang)</b>	bis 1,0	81,2	87,3	75,5	<0,001
	über 1,0	15,4	9,3	20,5	<0,001
	keine Angabe	3,4	3,5	4,0	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	2569±850	2846±922	2665±911	<0,001
<b>Ernährungskennnisse</b>	gut	17,1	18,8	13,4	<0,001
	mittel	38,2	40,8	35,2	<0,01
	schlecht	43,8	39,5	50,9	<0,001
	keine Angabe	0,9	0,9	0,5	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	56,2	60,3	53,1	<0,001
	nein	43,4	39,1	46,6	<0,001
	keine Angabe	0,5	0,6	0,3	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	28,1	25,3	33,9	<0,001
	nein	71,6	74,5	65,8	<0,001
	keine Angabe	0,2	0,3	0,3	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	75,2	77,4	74,6	n.s.
	mittel	20,7	18,7	21,5	<0,05
	schlecht	3,9	3,8	3,8	n.s.
	keine Angabe	0,1	0,2	0,1	--- <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	83,6	82,4	82,8	n.s.
	schlecht	16,0	17,3	17,0	n.s.
	keine Angabe	0,4	0,3	0,1	--- <sup>2</sup>
<b>HEI-NVS II</b>	arithm. MW±SD	80,2±10,7	81,9±10,6	75,4±10,6	< 0,001

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und dem HEI-NVS II über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

**Tab. 9.8: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der vierten Hauptkomponente**

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev)	4. Hauptkomponente (MW±Std dev)		Signifikanzen <sup>1</sup>	Faktorladung der 4. Hauptkomponente	Interpretation der 4. Hauptkomponente
		1. Quartile	4. Quartile			
Obst	249±223	202±194	304±265	<0,001		
Gemüse	239±143	252±154	234±143	<0,001		
Getreide	303±127	318±138	301±128	<0,01		
Kartoffeln	92±57	84±56	105±61	<0,001		
Nüsse und Samen	5±13	2±6	10±23	<0,001	<b>0,27</b>	hoher Nüsse- und Samenverzehr
Milch	249±258	197±187	332±352	<0,001	<b>0,20</b>	hoher Milchverzehr
Eier	16±18	18±21	16±18	n.s.		
Fisch	29±29	23±24	39±37	<0,001	<b>0,22</b>	hoher Fischverzehr
Fleisch	133±83	173±108	109±65	<0,001	<b>-0,32</b>	niedriger Fleischverzehr
Streichfett	30±27	20±19	45±35	<0,001	<b>0,36</b>	hoher Streichfettverzehr
Getränke	2332±981	3375±932	1539±644	<0,001	<b>-0,73</b>	niedriger Getränkeverzehr
Alkohol <sup>2</sup>	16±18	16±18	17±20	n.s.		
Süßwaren	68±63	60±59	82±74	<0,001		

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

**Tab. 9.9: Charakterisierung der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der vierten Hauptkomponente**

		Gesamtgruppe (%)	4. Hauptkomponente		Signifikanzen <sup>1</sup>
			1. Quartile (%)	4. Quartile (%)	
<b>Bio-Kauf</b>	ja	38,6	35,1	39,3	<0,05
	nein	61,5	64,9	60,7	<0,05
<b>Altersgruppe</b>	18–34 Jahre	19,7	31,3	11,6	<0,001
	35–50 Jahre	32,0	41,3	23,2	<0,001
	51–64 Jahre	25,3	19,2	27,5	<0,001
	65–80 Jahre	23,0	8,3	37,9	<0,001
<b>Soziale Schicht<sup>3</sup></b>	untere Schicht	17,6	18,0	18,5	n.s.
	mittlere Schicht	28,4	28,1	29,4	n.s.
	obere Schicht	54,1	53,9	52,1	n.s.
<b>BMI</b>	Untergewicht	0,7	0,7	0,6	n.s.
	Normalgewicht	30,0	29,6	31,3	n.s.
	Übergewicht	45,0	42,5	45,0	n.s.
	Adipositas	20,1	23,2	18,5	<0,01
	keine Angabe	4,3	4,0	4,7	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 102 cm	68,6	69,5	67,7	n.s.
	über 102 cm	28,0	27,2	28,6	n.s.
	keine Angabe	3,4	3,3	3,7	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüft- umfang)</b>	bis 1,0	81,2	84,0	79,2	<0,001
	über 1,0	15,4	12,7	17,1	<0,001
	keine Angabe	3,4	3,3	3,7	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	2569±850	2618±940	2750±884	<0,001
<b>Ernährungskenntnisse</b>	gut	17,1	16,6	17,1	n.s.
	mittel	38,2	42,9	32,5	<0,001
	schlecht	43,8	40,0	49,4	<0,001
	keine Angabe	0,9	0,6	1,1	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	56,2	60,6	52,4	<0,001
	nein	43,4	39,0	46,9	<0,001
	keine Angabe	0,5	0,4	0,7	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	28,1	37,3	23,7	<0,001
	nein	71,6	62,6	76,2	<0,001
	keine Angabe	0,2	0,1	0,1	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	75,2	75,3	73,4	n.s.
	mittel	20,7	20,7	21,8	n.s.
	schlecht	3,9	3,8	4,6	n.s.
	keine Angabe	0,1	0,2	0,2	--- <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	83,6	81,0	84,6	<0,01
	schlecht	16,0	18,6	15,2	<0,05
	keine Angabe	0,4	0,4	0,1	--- <sup>2</sup>
<b>HEI-NVS II</b>	arithm. MW±SD	80,2±10,7	78,7±10,7	79,8±11,0	<0,01

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und dem HEI-NVS II über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden

**Tab. 9.10: Lebensmittelverzehr (g/Tag) der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der fünften Hauptkomponente**

Lebensmittelgruppen	Gesamtgruppe (MW±Std dev)	5. Hauptkomponente (MW±Std dev)		Signifikanz <sup>1</sup>	Faktorladung der 5. Hauptkomponente	Interpretation der 5. Hauptkomponente
		1. Quartile	4. Quartile			
Obst	249±223	328±276	217±196	<0,001		
Gemüse	239±143	285±165	215±132	<0,001		
Getreide	303±127	353±143	289±124	<0,001		
Kartoffeln	92±57	105±63	86±57	<0,001		
Nüsse und Samen	5±13	1±4	12±23	<0,001	<b>0,43</b>	hoher Nüsse- und Samenverzehr
Milch	249±258	133±105	476±375	<0,001	<b>0,59</b>	hoher Milchverzehr
Eier	16±18	9±9	28±28	<0,001	<b>0,48</b>	hoher Eiverzehr
Fisch	29±29	38±37	26±26	<0,001		
Fleisch	133±83	136±88	139±95	n.s.		
Streichfett	30±27	47±34	22±21	<0,001	<b>-0,33</b>	hoher Streichfettverzehr
Getränke	2332±981	2244±940	2403±1049	<0,001		
Alkohol <sup>2</sup>	16±18	13±15	18±22	<0,001		
Süßwaren	68±63	81±70	69±67	<0,001		

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über den Rangsummentest Mann-Whitney U-Test).

<sup>2</sup>Alkohol (Ethanol) als Inhaltsstoff alkoholischer Getränke

**Tab. 9.11: Charakterisierung der Gesamtgruppe sowie der Männer der ersten und vierten Quartile der fünften Hauptkomponente**

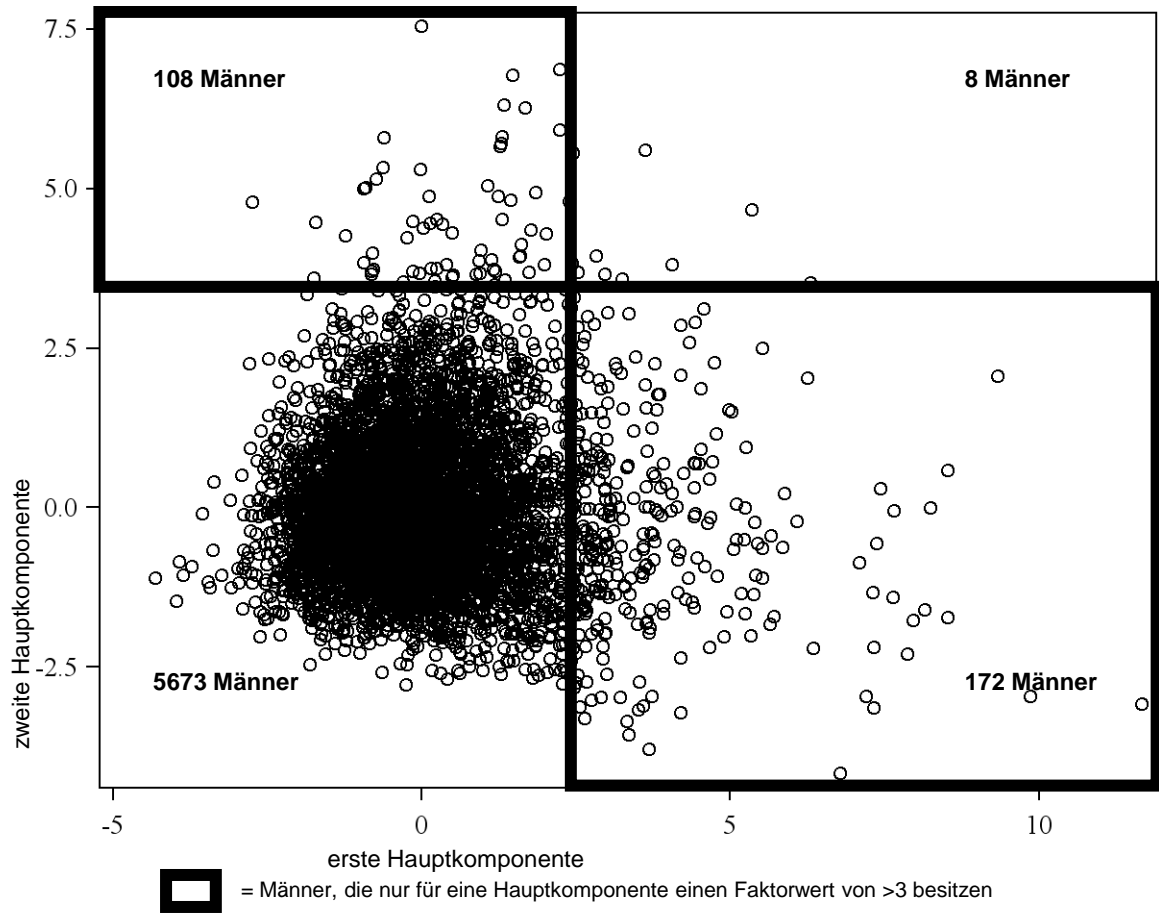
		Gesamtgruppe (%)	5. Hauptkomponente		Signifikanzen <sup>1</sup>
			1. Quartile (%)	4. Quartile (%)	
<b>Bio-Kauf</b>	ja	38,6	39,9	36,4	<0,05
	nein	61,5	60,1	63,6	<0,05
<b>Altersgruppe</b>	18–34 Jahre	19,7	13,1	27,0	<0,001
	35–50 Jahre	32,0	27,9	33,4	<0,01
	51–64 Jahre	25,3	29,0	23,6	<0,001
	65–80 Jahre	23,0	30,0	16,0	<0,001
<b>Soziale Schicht<sup>3</sup></b>	untere Schicht	17,6	18,1	16,7	n.s.
	mittlere Schicht	28,4	30,3	27,3	n.s.
	obere Schicht	54,1	51,6	56,0	<0,05
<b>BMI</b>	Untergewicht	0,7	0,5	0,9	n.s.
	Normalgewicht	30,0	29,3	34,0	<0,01
	Übergewicht	45,0	46,4	43,7	n.s.
	Adipositas	20,1	19,7	16,9	<0,05
	keine Angabe	4,3	4,1	4,6	--- <sup>2</sup>
<b>Taillenumfang</b>	bis 102 cm	68,6	67,9	73,5	<0,001
	über 102 cm	28,0	28,9	23,0	<0,001
	keine Angabe	3,4	3,3	3,6	--- <sup>2</sup>
<b>WHR (Taillen- zu Hüft-umfang)</b>	bis 1,0	81,2	81,1	84,2	<0,05
	über 1,0	15,4	15,6	12,2	<0,01
	keine Angabe	3,4	3,3	3,6	--- <sup>2</sup>
<b>Energiezufuhr</b>	Kcal/Tag (MW)	2569±850	2757±840	2763±955	n.s.
<b>Ernährungskenntnisse</b>	gut	17,1	15,8	18,6	<0,05
	mittel	38,2	35,4	39,3	<0,05
	schlecht	43,8	47,5	41,3	<0,001
	keine Angabe	0,9	1,3	0,9	--- <sup>2</sup>
<b>Sportlich aktiv</b>	ja	56,2	54,0	59,9	<0,01
	nein	43,4	45,3	39,7	<0,01
	keine Angabe	0,5	0,7	0,4	--- <sup>2</sup>
<b>Raucherstatus</b>	ja	28,1	23,4	35,0	<0,001
	nein	71,6	76,2	64,8	<0,001
	keine Angabe	0,2	0,4	0,1	--- <sup>2</sup>
<b>Allgemeiner Gesundheitszustand<sup>4</sup></b>	gut	75,2	74,7	78,4	<0,05
	mittel	20,7	21,9	17,9	<0,01
	schlecht	3,9	3,3	3,6	n.s.
	keine Angabe	0,1	0,1	0,2	--- <sup>2</sup>
<b>Schlafqualität<sup>4</sup></b>	gut	83,6	85,0	83,4	n.s.
	schlecht	16,0	14,6	16,5	n.s.
	keine Angabe	0,4	0,3	0,2	--- <sup>2</sup>
<b>HEI-NVS II</b>	arithm. MW±SD	80,2±10,7	81,7±10,8	78,3±10,8	<0,001

<sup>1</sup>Das Signifikanzniveau p gibt an, ob ein statistischer Unterschied zwischen der ersten und vierten Quartile vorliegt (getestet über: Chi-Quadrat-Test bzw. bei der Kalorienaufnahme und dem HEI-NVS II über den Rangsummentest Mann-Whitney U-Test).

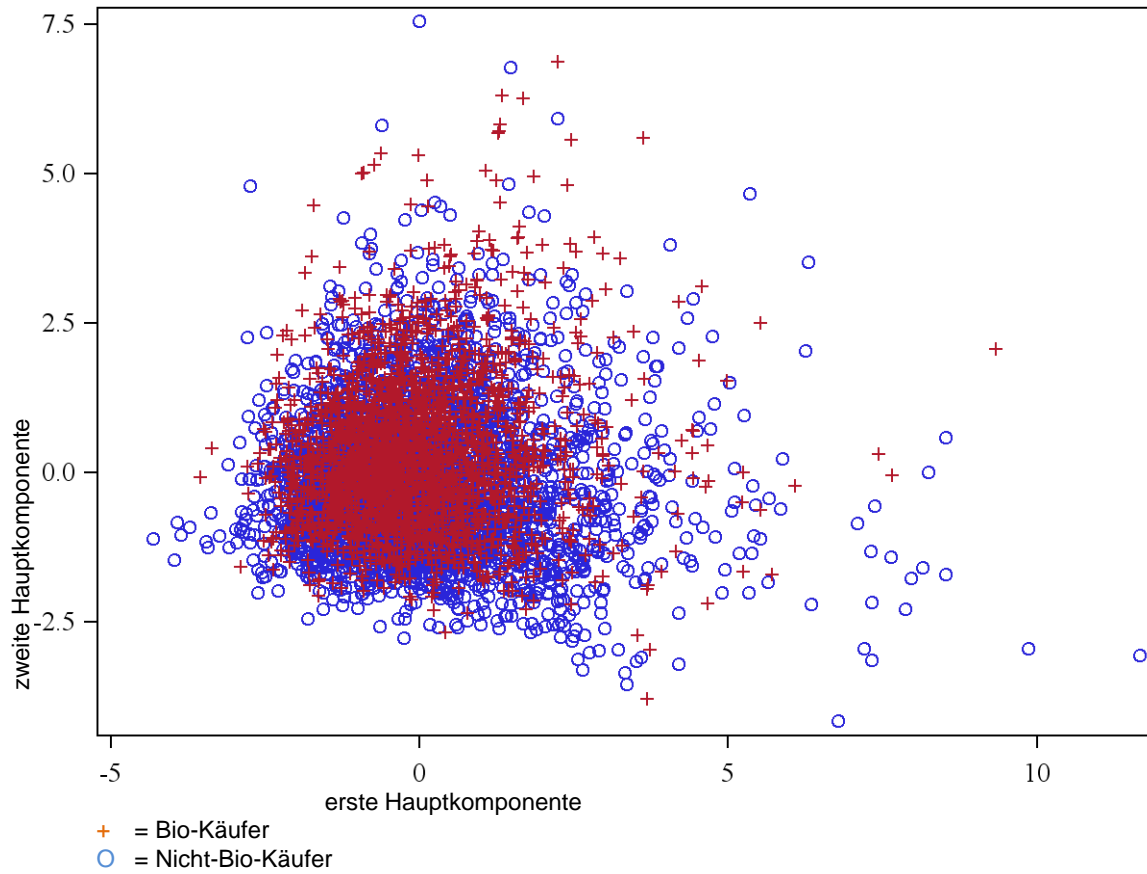
<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>4</sup>Selbsteinschätzung der Teilnehmenden



**Abb. 9.1** Zuordnung der Männer zu den ersten beiden Hauptkomponenten auf Basis der Faktorwerte



**Abb. 9.2** Differenzierung der Bio-Käufer und Nicht-Bio-Käufer über die ersten beiden Hauptkomponenten auf Basis der Faktorwerte

### 9.3 Clusteranalyse

Tab. 9.12: Kriterien zur Festlegung der Clusteranzahl (Frauen)

Anzahl der Cluster	CCC	Pseudo F-Wert	Pseudo T-Wert
1	0,00	.	314
2	-59	314	267
3	-75	301	250
4	-83	287	251
5	-86	280	209
6	-88	272	321
7	-89	268	168
8	-86	266	175
9	-82	264	168
10	-81	256	124
11	-80	249	105
12	-77	244	126
13	-74	240	143
14	-69	234	176
15	-68	227	99,2
16	-64	221	62,8
17	-60	216	89,4
18	-57	211	37,1
19	-55	206	74,8
20	-52	202	102
21	-49	198	79,0
22	-47	193	62,8
23	-45	190	53,3
24	-43	186	49,6
25	-40	183	80,1
26	-38	180	49,2
27	-36	177	113

Anzahl der Cluster	CCC	Pseudo F-Wert	Pseudo T-Wert
28	-34	174	29,4
29	-32	172	85,1
30	-29	169	31,4
31	-28	167	68,2
32	-26	164	49,3
33	-25	162	31,6
34	-23	160	23,4
35	-22	157	82,3
36	-21	155	48,8
37	-20	153	24,4
38	-19	151	28,0
39	-19	149	49,9
40	-18	146	45,2
41	-17	144	41,8
42	-17	142	70,7
43	-16	141	45,4
44	-15	139	16,9
45	-15	137	44,7
46	-14	135	46,7
47	-13	134	31,1
48	-13	132	31,8
49	-12	131	43,7
50	-11	129	16,0
51	-10	128	43,3
52	-9,8	127	18,9
53	-9,0	125	34,7
54	-8,4	124	14,1

Anzahl der Cluster	CCC	Pseudo F-Wert	Pseudo T-Wert
55	-7,8	123	23,1
56	-7,2	122	22,1
57	-6,6	121	34,3
58	-6,1	119	38,6
59	-5,5	118	22,3
60	-5,0	117	16,5
61	-4,5	116	43,4
62	-4,0	115	12,7
63	-3,5	114	26,7
64	-3,1	113	19,4
65	-2,6	112	27,4
66	-2,1	111	21,7
67	-1,7	110	11,9
68	-1,2	109	30,0
69	-,84	108	31,4
70	-,43	107	15,5
71	-,02	107	17,3
72	0,41	106	14,9
73	0,84	105	29,8
74	1,29	104	20,9
75	1,74	103	27,8
76	2,21	103	25,1

**Tab. 9.13: Kriterien zur Festlegung der Clusteranzahl (Männer)**

Anzahl der Cluster	CCC	Pseudo F-Wert	Pseudo T-Wert
1	0,00	.	290
2	-54	290	298
3	-63	293	262
4	-68	285	270
5	-68	279	228
6	-69	269	216
7	-70	260	135
8	-69	251	191
9	-69	242	192
10	-68	234	70,1
11	-67	227	115
12	-64	221	81,6
13	-58	217	142
14	-52	213	98,1
15	-49	208	67,3
16	-48	202	135
17	-45	197	94,6
18	-42	192	91,2
19	-39	188	49,7
20	-37	184	101
21	-35	179	107
22	-34	175	48,8
23	-32	171	30,3
24	-31	167	70,2
25	-30	164	46,3
26	-28	160	17,7
27	-27	157	23,1
28	-26	154	67,2

Anzahl der Cluster	CCC	Pseudo F-Wert	Pseudo T-Wert
29	-26	152	56,1
30	-24	149	52,9
31	-23	147	17,3
32	-22	144	22,1
33	-21	142	41,7
34	-20	140	27,4
35	-20	137	53,0
36	-19	135	52,6
37	-18	133	24,6
38	-17	131	34,0
39	-17	129	21,7
40	-16	127	21,9
41	-16	126	25,0
42	-15	124	7,7
43	-14	122	28,9
44	-14	121	20,7
45	-13	119	57,1
46	-13	118	41,6
47	-12	116	6,4
48	-12	115	30,4
49	-11	114	32,8
50	-10	112	23,1
51	-9,8	111	20,7
52	-9,2	110	26,0
53	-8,8	109	11,7
54	-8,5	107	37,4
55	-8,1	106	12,8
56	-7,7	105	11,4

Anzahl der Cluster	CCC	Pseudo F-Wert	Pseudo T-Wert
57	-7,3	104	28,3
58	-6,8	103	30,8
59	-6,4	102	14,2
60	-5,9	101	15,1
61	-5,4	100	20,4
62	-5,0	99,3	27,0
63	-4,5	98,4	35,6
64	-4,1	97,5	25,2
65	-3,7	96,7	34,6
66	-3,3	95,9	11,7
67	-2,9	95,0	8,4
68	-2,4	94,3	21,6
69	-2,0	93,5	13,1
70	-1,5	92,8	15,1
71	-1,1	92,1	38,4
72	-,69	91,4	14,3
73	-,27	90,7	9,5
74	0,17	90,0	28,3
75	0,60	89,3	19,5
76	1,03	88,7	28,9
77	1,48	88,1	11,4
78	1,92	87,5	19,4
79	2,36	86,9	34,5

## 9.4 Entscheidungsbaum

Tab. 9.14: Charakterisierung der Gesamtgruppe sowie der Frauen der einzelnen Blätter des Entscheidungsbaums - Teil 1

Gesamtgruppe (%)			Blätter				
			1. Blatt (%)	2. Blatt (%)	3. Blatt (%)	4. Blatt (%)	5. Blatt (%)
<b>Bio-Kauf</b>	ja	50,3	21,7	54,7	68,3	82,7	36,8
	nein	49,7	78,3	45,3	31,7	17,3	63,2
<b>Altersgruppe</b>	18–34 Jahre	20,7	41,3	26,0	18,0	32,4	33,9
	35–50 Jahre	35,3	28,3	23,3	30,0	41,6	37,3
	51–64 Jahre	24,2	15,2	21,5	28,4	16,2	15,6
	65–80 Jahre	19,8	15,2	29,2	23,7	9,8	13,2
<b>Soziale Schicht<sup>1</sup></b>	untere Schicht	21,0	41,3	25,1	17,7	11,6	24,1
	mittlere Schicht	29,9	41,3	32,3	29,5	23,7	30,7
	obere Schicht	49,1	17,4	42,6	52,8	64,7	45,2
<b>BMI</b>	Untergewicht	1,6	0,0	2,2	1,6	5,2	1,8
	Normalgewicht	46,0	54,4	49,3	51,8	71,7	48,4
	Übergewicht	27,9	28,3	24,2	26,8	13,9	26,2
	Adipositas	18,9	10,9	16,6	13,7	5,2	17,4
	keine Angabe <sup>2</sup>	5,6	6,5	7,6	6,2	4,1	6,3
<b>Taillenumfang</b>	bis 88 cm	65,8	71,7	67,7	72,5	85,6	68,0
	über 88 cm	29,4	23,9	26,9	23,0	9,3	26,1
	keine Angabe <sup>3</sup>	4,8	4,4	5,4	4,6	5,2	6,0
<b>WHR (Taillen- zu Hüftumfang)</b>	bis 0,85	74,2	80,4	69,1	76,6	87,3	76,5
	über 0,85	21,0	15,2	25,6	18,8	7,5	17,6
	keine Angabe <sup>2</sup>	4,8	4,4	5,4	4,6	5,2	6,0
<b>Energiezufuhr</b>	Kcal/Tag (MW (95. KI des MW))	1968 (1954–1981)	1286 (1146–1426)	1520 (1458–1581)	1702 (1671–1733)	2370 (2286–2455)	1966 (1931–2000)
<b>Ernährungs- kenntnisse</b>	gut	40,9	26,1	30,5	45,2	54,3	35,1
	mittel	33,9	32,6	35,4	35,5	31,2	36,3
	schlecht	24,2	41,3	32,7	18,5	14,5	27,5
	keine Angabe <sup>2</sup>	1,0	0,0	1,4	0,8	0,0	1,1
<b>Sportlich aktiv</b>	ja	58,9	32,6	63,7	69,0	76,3	54,9
	nein	40,6	67,4	35,9	29,4	23,7	44,7
	keine Angabe <sup>2</sup>	0,6	0,0	0,5	1,6	0,0	0,5
<b>Raucherstatus</b>	ja	21,7	45,7	20,2	17,7	12,1	29,7
	nein	78,0	54,4	79,4	82,1	86,7	70,1
	keine Angabe <sup>2</sup>	0,3	0,0	0,5	0,2	1,2	0,2
<b>Allgemeiner Gesundheits- zustand<sup>3</sup></b>	gut	76,2	60,9	70,9	77,3	85,0	75,9
	mittel	20,2	28,3	23,3	18,6	9,8	20,9
	schlecht	3,5	10,9	5,8	3,7	4,6	3,1
	keine Angabe <sup>2</sup>	0,1	0,0	0,0	0,5	0,6	0,1
<b>Schlafqualität<sup>3</sup></b>	gut	78,8	76,1	72,6	79,5	78,6	77,4
	schlecht	20,7	23,9	26,5	20,0	20,8	22,3
	keine Angabe <sup>2</sup>	0,5	0,0	0,9	0,6	0,6	0,3
<b>HEI-NVS II</b>	MW (95 KI des MW)	84,8(84,6–85,1)	65,0(62,6–67,3)	78,8(77,7–79,9)	88,0(87,4–88,6)	89,4(88,1–90,8)	77,0(76,5–77,5)

<sup>1</sup> Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>2</sup> Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen.

<sup>3</sup> Selbsteinschätzung der Teilnehmenden

**Tab. 9.15: Charakterisierung der Gesamtgruppe sowie der Frauen der einzelnen Blätter des Entscheidungsbaums - Teil 2**

			Blätter			
			6. Blatt (%)	7. Blatt (%)	8. Blatt (%)	9. Blatt (%)
<b>Bio-Kauf</b>	ja	50,3	33,7	45,5	57,1	59,3
	nein	49,7	66,3	54,5	43,0	40,7
<b>Altersgruppe</b>	18–34 Jahre	20,7	25,8	12,7	17,4	15,0
	35–50 Jahre	35,3	42,6	29,0	41,5	37,9
	51–64 Jahre	24,2	17,2	30,6	22,7	28,6
	65–80 Jahre	19,8	14,4	27,7	18,4	18,5
<b>Soziale Schicht<sup>1</sup></b>	untere Schicht	21,0	22,9	23,7	17,6	17,7
	mittlere Schicht	29,9	31,2	31,4	27,4	28,1
	obere Schicht	49,1	45,9	44,9	55,0	54,3
<b>BMI</b>	Untergewicht	1,6	2,4	1,1	1,4	1,2
	Normalgewicht	46,0	44,8	38,5	50,7	43,3
	Übergewicht	27,9	25,8	32,4	25,7	29,7
	Adipositas	18,9	19,6	23,1	17,6	20,9
	keine Angabe <sup>2</sup>	5,6	7,3	4,9	4,6	4,9
<b>Taillenumfang</b>	bis 88 cm	65,8	64,1	58,6	67,0	65,3
	über 88 cm	29,4	29,1	37,4	29,2	30,8
	keine Angabe <sup>3</sup>	4,8	6,8	4,1	3,8	3,9
<b>WHR (Taillen- zu Hüftumfang)</b>	bis 0,85	74,2	74,1	68,9	77,2	74,3
	über 0,85	21,0	19,1	27,1	19,0	21,8
	keine Angabe <sup>2</sup>	4,8	6,8	4,1	3,8	3,9
<b>Energiezufuhr</b>	Kcal/Tag (MW (95. KI des MW))	1968 (1954–1981)	1911 (1871–1952)	1857 (1835–1879)	2362 (2321–2402)	2148 (2117–2180)
<b>Ernährungs-kenntnisse</b>	gut	40,9	37,0	36,9	49,3	46,6
	mittel	33,9	37,4	32,6	32,5	30,9
	schlecht	24,2	25,1	29,0	16,8	21,7
	keine Angabe <sup>2</sup>	1,0	0,5	1,5	1,4	1,0
<b>Sportlich aktiv</b>	ja	58,9	48,7	54,4	59,3	64,8
	nein	40,6	51,2	45,2	40,4	34,6
	keine Angabe <sup>2</sup>	0,6	0,1	0,4	0,3	0,7
<b>Raucherstatus</b>	ja	21,7	30,2	18,0	14,9	19,8
	nein	78,0	69,1	82,0	84,6	79,8
	keine Angabe <sup>2</sup>	0,3	0,7	0,0	0,5	0,4
<b>Allgemeiner Gesundheitszustand<sup>3</sup></b>	gut	76,2	72,0	74,5	78,1	79,2
	mittel	20,2	24,1	21,7	19,2	17,7
	schlecht	3,5	3,9	3,7	2,5	3,1
	keine Angabe <sup>2</sup>	0,1	0,0	0,1	0,2	0,1
<b>Schlafqualität<sup>3</sup></b>	gut	78,8	76,7	79,2	80,5	80,9
	schlecht	20,7	22,3	20,4	19,2	18,8
	keine Angabe <sup>2</sup>	0,5	1,0	0,4	0,3	0,3
<b>HEI-NVS II</b>	MW (95 KI des MW)	84,8(84,6–85,1)	76,7(76,2–77,1)	86,8(86,4–87,1)	88,8(88,3–89,2)	91,6(91,2–92,0)

<sup>1</sup> Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

<sup>2</sup> Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede wurde hier nicht durchgeführt.

<sup>3</sup> Selbsteinschätzung der Teilnehmenden

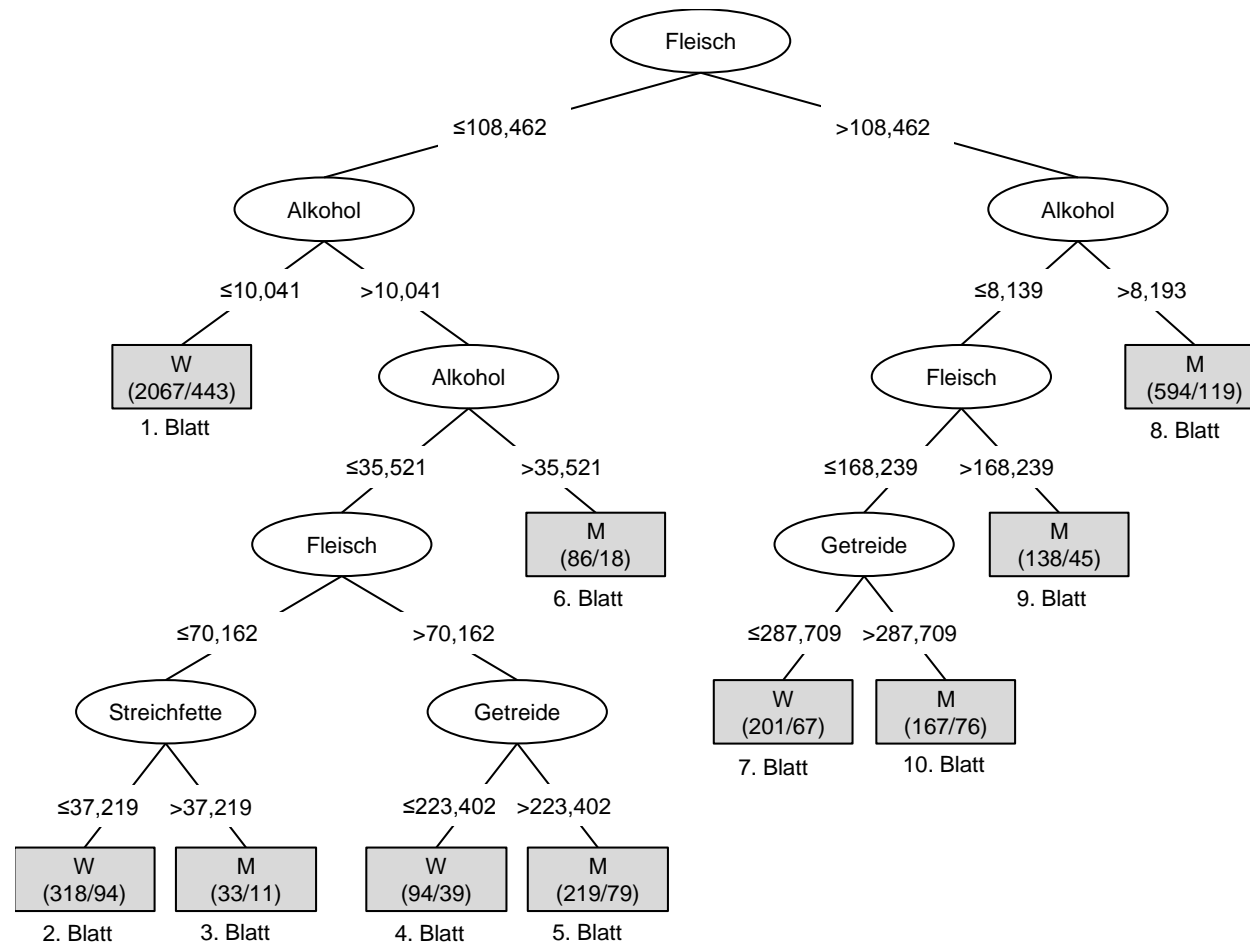
**Tab. 9.16: Charakterisierung der Gesamtgruppe sowie der Männer der einzelnen Blätter des Entscheidungsbaums**

		Gesamtgruppe (%)	Blätter				
			1. Blatt (%)	2. Blatt (%)	3. Blatt (%)	4. Blatt (%)	5. Blatt (%)
<b>Bio-Kauf</b>	ja	38,6	31,1	37,1	55,7	54,7	31,1
	nein	61,5	68,9	62,9	44,3	45,3	68,9
<b>Altersgruppe</b>	18–34 Jahre	19,7	26,0	9,4	10,4	14,4	25,8
	35–50 Jahre	32,0	27,5	24,9	28,1	33,6	34,7
	51–64 Jahre	25,3	23,5	28,5	26,1	26,4	23,7
	65–80 Jahre	23,0	23,0	37,3	35,5	25,6	15,8
<b>Soziale Schicht<sup>1</sup></b>	obere Schicht	54,1	49,5	51,3	57,1	62,0	51,6
	mittlere Schicht	28,4	30,4	30,3	28,6	24,7	29,1
	untere Schicht	17,6	20,1	18,4	14,3	13,3	19,3
<b>BMI</b>	Untergewicht	0,7	1,7	0,4	0,0	0,4	0,8
	Normalgewicht	30,0	31,4	20,9	18,2	35,6	31,2
	Übergewicht	45,0	43,9	50,1	48,3	44,4	43,4
	Adipositas	20,1	18,4	21,9	26,1	15,5	21,4
	keine Angabe <sup>2</sup>	4,3	4,7	6,8	7,4	4,1	3,2
<b>Taillenumfang</b>	bis 102 cm	68,6	69,4	62,0	62,6	73,8	68,8
	über 102 cm	28,0	26,2	33,4	32,5	23,2	28,4
	keine Angabe <sup>2</sup>	3,4	4,4	4,6	4,9	3,0	2,9
<b>WHR (Taillen- zu Hüftumfang)</b>	bis 1,0	81,2	80,9	75,2	79,3	84,7	81,8
	über 1,0	15,4	14,7	20,2	15,8	12,4	15,3
	keine Angabe <sup>2</sup>	3,4	4,4	4,6	4,9	3,0	2,9
<b>Energiezufuhr</b>	Kcal/Tag (MW (95. KI des MW))	2569 (2548–2591)	2092 (2021–2162)	1920 (1889–1951)	2095 (2018–2173)	2544 (2511–2576)	2914 (2881–2946)
<b>Ernährungs-kenntnisse</b>	gut	17,1	12,3	15,3	21,2	21,8	15,8
	mittel	38,2	37,3	33,6	35,0	39,9	39,4
	schlecht	43,8	48,8	49,7	41,9	37,6	44,2
	keine Angabe <sup>2</sup>	0,9	1,7	1,5	2,0	0,7	0,6
<b>Sportlich aktiv</b>	ja	56,2	49,8	50,0	61,1	64,0	55,1
	nein	43,4	50,0	49,1	37,9	35,5	44,6
	keine Angabe <sup>2</sup>	0,5	0,3	0,9	1,0	0,6	0,3
<b>Raucherstatus</b>	ja	28,1	33,3	24,8	22,7	18,9	33,4
	nein	71,6	66,4	75,2	76,8	80,8	66,3
	keine Angabe <sup>2</sup>	0,2	0,3	0,0	0,5	0,3	0,3
<b>Allgemeiner Gesundheitszustand<sup>3</sup></b>	gut	75,2	71,1	72,6	70,5	78,3	75,6
	mittel	20,7	22,1	21,2	22,2	18,5	21,3
	schlecht	3,9	6,9	6,1	6,9	3,1	2,9
	keine Angabe <sup>2</sup>	0,1	0,0	0,1	0,5	0,1	0,1
<b>Schlafqualität<sup>3</sup></b>	gut	83,6	82,1	83,1	78,8	86,6	82,9
	schlecht	16,0	17,7	16,5	20,7	12,8	16,8
	keine Angabe <sup>2</sup>	0,4	0,3	0,4	0,5	0,6	0,3
<b>HEI-NVS II</b>	MW (95 KI des MW)	80,3(80,0–80,5)	73,9(72,9–74,9)	81,4(80,9–82,0)	89,9(88,7–91,2)	86,4(86,0–86,9)	77,1(76,7–77,5)

<sup>1</sup>Soziale Schicht: Obere Schicht = Oberschicht + obere Mittelschicht, Mittelschicht = Mittelschicht, Untere Schicht = untere Mittelschicht + Unterschicht

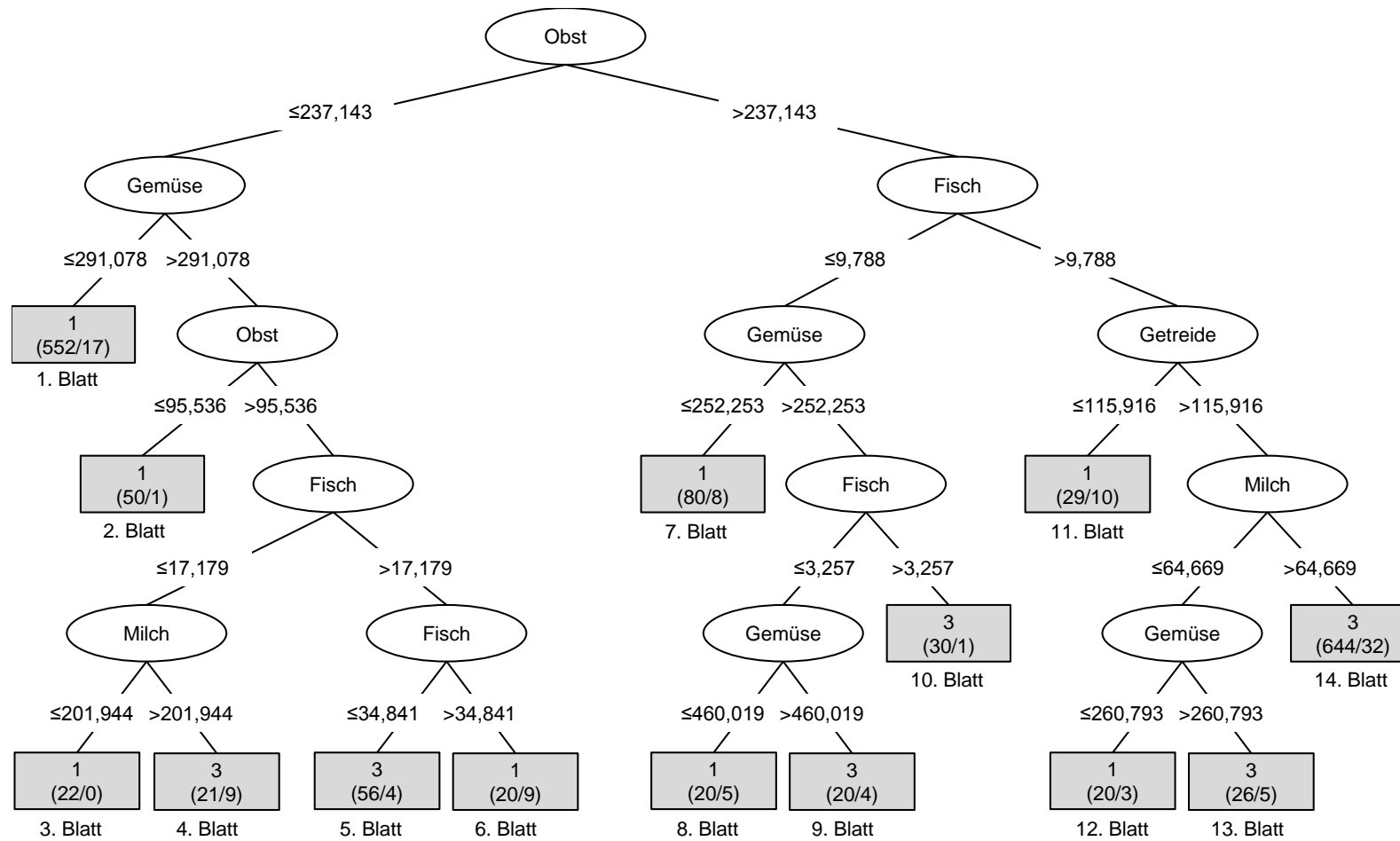
<sup>2</sup>Nicht alle Teilnehmenden haben alle Fragen beantwortet bzw. haben an anthropometrischen Messungen teilgenommen. Ein Test auf signifikante Unterschiede bei diesem Personenanteil wurde nicht durchgeführt.

<sup>3</sup>Selbsteinschätzung der Teilnehmenden



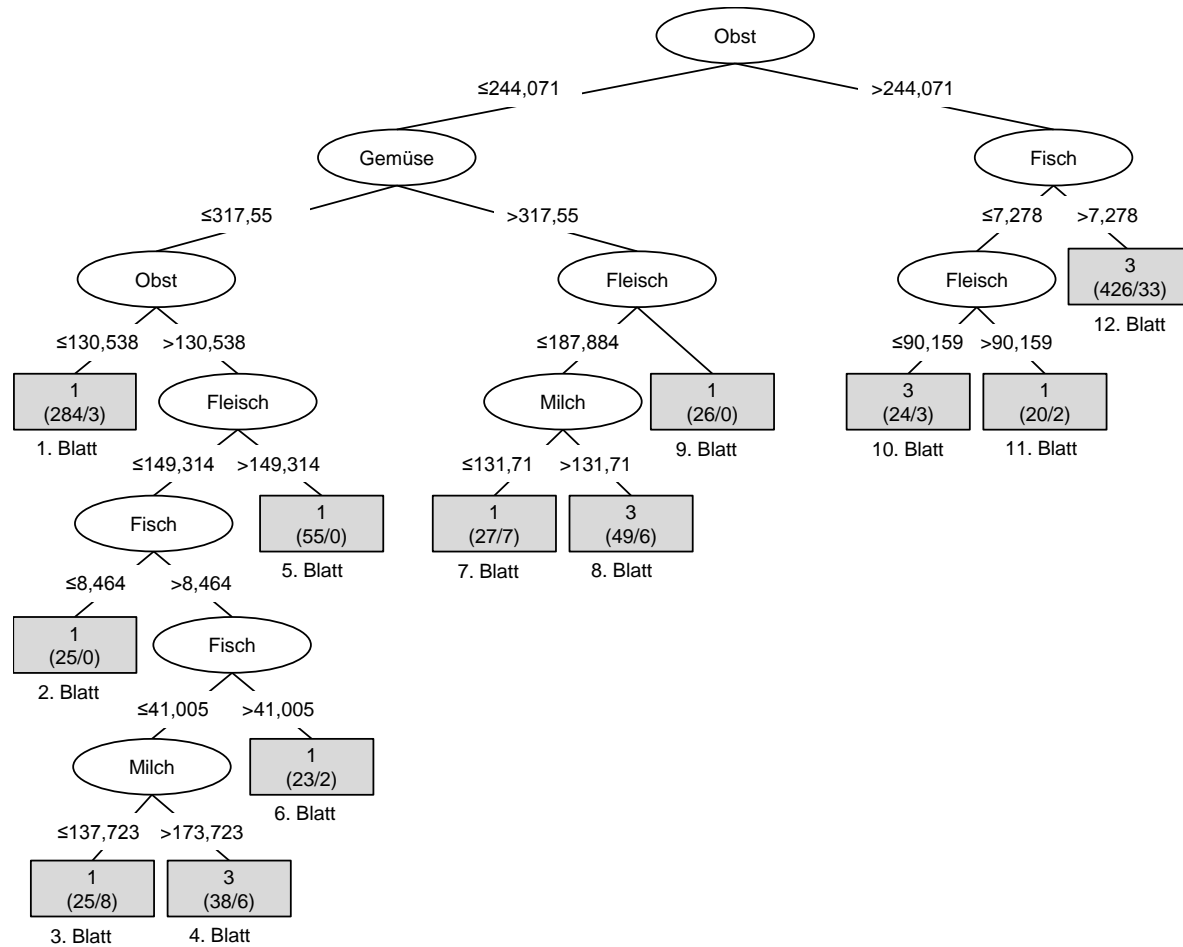
Die Lebensmittelverzehrsmengen auf dem Pfad sind als kleiner/gleich und größer als Angaben in Gramm pro Tag angegeben.  
 Die Angaben W und M weisen auf ein durch Frauen (W) oder Männer (M)-dominantes Blatt hin.  
 Die Zahlen in Klammer geben die Anzahl der Frauen und Männer an, wobei immer der größere Anteil zuerst genannt wird.

**Abb. 9.3: Entscheidungsbaum der Bio-Käufer/innen, klassifiziert nach Frauen und Männern**



Die Lebensmittelverzehrsmengen auf dem Pfad sind als kleiner/gleich und größer als Angaben in Gramm pro Tag angegeben.  
 Die Zahlen 1 und 3 weisen auf ein Blatt hin, dass mehr Frauen enthält, die die Empfehlungen weniger erreichen (1) oder gut erreichen (3)  
 Die Zahlen in Klammer geben die Anzahl der Frauen der beiden Gruppen an, wobei immer der größere Anteil zuerst genannt wird.

**Abb. 9.4:** Entscheidungsbaum der Bio-Käuferinnen, klassifiziert nach „Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“, gemessen an der HEI-NVS II-Indexsumme



Die Lebensmittelverzehrsmengen auf dem Pfad sind als kleiner/gleich und größer als Angaben in Gramm pro Tag angegeben.  
 Die Zahlen 1 und 3 weisen auf ein Blatt hin, dass mehr Männer enthält, die die Empfehlungen weniger erreichen (1) oder gut erreichen (3)  
 Die Zahlen in Klammer geben die Anzahl der Männer der beiden Gruppen an, wobei immer der größere Anteil zuerst genannt wird.

**Abb. 9.5: Entscheidungsbaum der Bio-Käufer, klassifiziert nach „Empfehlungen gut erreicht“ versus „Empfehlungen weniger erreicht“, gemessen an der HEI-NVS II-Indexsumme**

### **Erklärung**

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.