

# Methoden der Skalierung

Mitschrift der Vorlesung  
von Dr. Christian Maag

im WS 07/08

Roland Pfister

Julius-Maximilians-Universität Würzburg

# Inhaltsverzeichnis

<b>0. Vorwort</b> .....	<b>4</b>
<b>1. Einführung</b> .....	<b>5</b>
1.1. Arten der Skalierung.....	5
1.2. Exkurs: Einige Skalen.....	5
1.2.1. Fahrenheit .....	5
1.2.2. Scoville-Skala .....	5
1.2.3. Mercalli-Skala .....	6
1.2.4. Beauford-Skala.....	6
1.3. Messtheorie .....	6
1.3.1. Probleme der Messtheorie.....	6
1.3.2. Skalenniveaus .....	6
1.3.3. Coomb'sche Datenklassifikation .....	7
<b>2. Signalentdeckungstheorie</b> .....	<b>8</b>
2.1. Grundbegriffe .....	8
2.2. Modell der SDT .....	9
2.2.1. Annahmen .....	9
2.2.2. Likelihood-Ratio.....	9
2.2.3. Parameter des Modells.....	9
2.3. Beispiel: Vigilanz.....	10
2.4. Sensitivität und Spezifität .....	11
2.5. ROC-Kurven.....	11
<b>3. Probabilistische Skalierung</b> .....	<b>13</b>
3.1. Urteilsverfahren.....	13
3.1.1. Direkte Methoden .....	13
3.1.2. Indirekte Verfahren .....	13
3.2. Law of Categorical Judgement .....	14
3.2.1. Voraussetzungen.....	14
3.2.2. Vorgehen .....	15
3.2.3. Interpretation .....	15
3.2.4. Zusammenfassung .....	16
3.3. Law of Comparative Judgement .....	16
3.3.1. Voraussetzungen.....	16
3.3.2. Vorgehen .....	16
3.3.3. Hinweise .....	17
3.3.4. Bewertung .....	17

<b>3.4. BTL-Modell (Bradley-Terry-Luce-Modell)</b> .....	<b>18</b>
3.4.1. Annahmen .....	18
3.4.2. Vorgehen .....	20
3.4.3. Probleme .....	22
<b>3.5. Vergleich von BTL und LCompJ</b> .....	<b>22</b>
<b>3.6. Übung: Probabilistische Skalierung</b> .....	<b>23</b>
3.6.1. Law of Categorical Judgement .....	23
3.6.2. Law of Comparative Judgement .....	24
3.6.3. BTL .....	24
3.6.4. LCJ- und BTL-Vorhersage .....	24
<b>4. Clusteranalyse</b> .....	<b>25</b>
<b>4.1. Fragestellung und Anwendungsbereiche</b> .....	<b>25</b>
<b>4.2. Vorgehen</b> .....	<b>25</b>
<b>4.3. Distanzmaße</b> .....	<b>25</b>
4.3.1. Nominalskalen .....	26
4.3.2. Metrische Skalen .....	28
<b>4.4. Gruppierungsverfahren: Fusionierungsalgorithmen</b> .....	<b>29</b>
4.4.1. Hierarchische Verfahren .....	30
4.4.2. Partitionierende Verfahren.....	33
<b>4.5. Zusammenfassung</b> .....	<b>33</b>
4.5.1. CA und FA .....	34
4.5.2. Vorschläge zur praktischen Anwendung .....	34
<b>4.6. Übung: Clusteranalyse</b> .....	<b>34</b>
4.6.1. Generelles .....	34
4.6.2. Output.....	35
<b>5. Conjoint-Analyse</b> .....	<b>38</b>
<b>5.1. Modell der Conjoint-Analyse</b> .....	<b>38</b>
<b>5.2. Definition der Urteilsobjekte</b> .....	<b>39</b>
5.2.1. Profilmethode .....	39
5.2.2. Trade-Off-Analyse .....	39
5.2.3. Symmetrische und asymmetrische Designs .....	40
5.2.4. Produktkarten .....	40
<b>5.3. Darbietung der Objekte</b> .....	<b>41</b>
<b>5.4. Berechnung</b> .....	<b>41</b>
5.4.1. Metrische Lösung: Kleinste Quadrate.....	42
5.4.2. Nichtmetrische Lösung: Monotone ANOVA.....	43
5.4.3. Aggregation von Nutzenwerten .....	43
<b>5.5. Interpretation der Teilnutzenwerte</b> .....	<b>44</b>
<b>5.6. Zusammenfassung</b> .....	<b>44</b>

<b>6. Faktorenanalyse</b>	<b>46</b>
<b>6.1. Schritt 1: Variablenauswahl und Korrelationsmatrix</b>	<b>46</b>
6.1.1. Variablenauswahl	46
6.1.2. Korrelationsmatrix	46
6.1.3. Anmerkung: Prüfung der Rohdaten	49
<b>6.2. Schritt 2: Extraktion der Faktoren</b>	<b>50</b>
6.2.1. Graphische Darstellung	50
<b>6.3. Schritt 3: Bestimmung der Kommunalitäten</b>	<b>51</b>
6.3.1. Hauptkomponentenanalyse	51
6.3.2. Hauptachsenanalyse	51
6.3.3. SPSS-Output	52
<b>6.4. Schritt 4: Zahl der Faktoren</b>	<b>52</b>
<b>6.5. Schritt 5: Faktoreninterpretation</b>	<b>53</b>
<b>6.6. Bestimmung der Faktorwerte</b>	<b>54</b>
<b>6.7. Zusammenfassung</b>	<b>54</b>
<b>6.8. Übung</b>	<b>55</b>
6.8.1. Anwendungsbeispiele	55
6.8.2. Übungsblatt	55
<b>7. Lineare Strukturmodelle</b>	<b>58</b>
<b>7.1. Beispiel</b>	<b>59</b>
<b>7.2. Kausalanalyse</b>	<b>60</b>
7.2.1. Kausale und nicht-kausale Komponenten	61
7.2.2. Beispiel	61
<b>7.3. LISREL-Ansatz</b>	<b>62</b>
7.3.1. Ablauf einer Analyse	63
7.3.2. Voraussetzungen	65
<b>7.4. Übung</b>	<b>66</b>
7.4.1. Software: AMOS 5.0/7.0	66
7.4.2. Interpretation	68
7.4.3. Anwendungsbeispiele	69
<b>8. Anmerkungen</b>	<b>71</b>
<b>8.1. Generell</b>	<b>71</b>
<b>8.2. Klausur</b>	<b>71</b>
<b>8.3. Fragen</b>	<b>71</b>

## 0. Vorwort

Dozenten: Dr. Christian Maag, Dipl.-Psych. Markus Schmitz

Termin: Mittwoch, 18:00 – 19:30, Kühle-Hörsaal

Klausur: Termin: 30.01.2008

Web: [www.ivzw.de](http://www.ivzw.de) bzw. [www.psychologie.uni-wuerzburg.de/methoden](http://www.psychologie.uni-wuerzburg.de/methoden)

Literatur: Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2006). Multivariate Analysemethoden. Berlin: Springer.

Anmerkung: eBooks – u. a. Bortz und Döring (2005) – gibt es unter [www.springerlink.com](http://www.springerlink.com) oder auf der Homepage der Unibibliothek. Bei Problemen mit den e-Books unter [viola.haemmer@bibliothek.uni-wuerzburg.de](mailto:viola.haemmer@bibliothek.uni-wuerzburg.de) nachfragen.

# 1. Einführung

## 1.1. Arten der Skalierung

Es werden 4 Arten bzw. Ziele der Skalierung unterschieden:

- Messung
- Anwendungsbereiche und Grenzen von Skalierungsmodellen
- Klassifikation und Organisation
- Skalierung als Urteilsmodell

Die primäre Zielsetzung der Skalierung ist dabei jedoch immer die Messung. Unter Messung versteht man die homomorphe Abbildung eines empirischen Relativs in ein numerisches. Das numerische Relativ wird als Skala bezeichnet.

## 1.2. Exkurs: Einige Skalen

### 1.2.1. Fahrenheit

Die Fahrenheit-Skala ist durch zwei Fixpunkte definiert: Die Temperatur der kältesten Winternacht des Entwicklungsjahres in Danzig wurde mit 0° F gleichgesetzt, die Körpertemperatur eines Menschen entspricht 100° F. Das Skalenintervall ist durch  $(F_1 - F_2)/96$  definiert.

### 1.2.2. Scoville-Skala

In der Scoville-Skala wird die Schärfe von Paprika in Scoville-Einheiten angegeben. Zu deren Bestimmung wurde Paprikapulver bis zur Unmerklichkeit verdünnt. Am schärfsten zeigten sich die Chili-Sorten Habanero und Scotch Bonnet mit einem Verdünnungsfaktor von 150000 – 300000.

Heute wird statt der Scoville-Skala der Schärfegrad (1-10) verwendet.

Schärfe-grad	Chili-Sorten	Scoville-Einheiten
10	Habanero, Scotch Bonnet	150000 - 300000
9	Tepin	75000 - 95000
8	Rocoto, Thai, Pequin	50000 - 75000
7	Aji-Sorten, de Arbol, Tabasco	35000 - 45000
6	Serrano, Piri-Piri	20000 - 30000
5	Jalapeño, ung. Kirschnaprika	10000 - 15000
4	Cascabel, Cayenne	6000 - 9000
3	Dutch Red, Peter Pepper, Cubanelle	3000 - 5000
2	Poblano, Guajillo, Pimento	2000 - 2500
1	Annaheim, New Mexican	500 - 1500
0	Gemüsepaprika	0

Neben objektiven Maßstäben können auch individuelle Reaktionen (etwa der Gesichtsausdruck) verwendet werden, um eine Skala zu entwickeln.

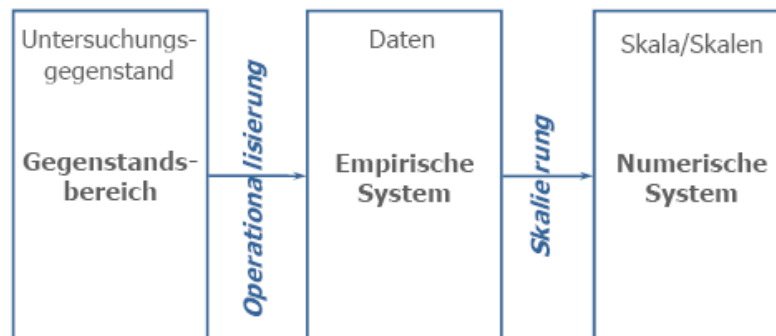
### 1.2.3. Mercalli-Skala

Die Verwendung von Verhalten bzw. Reaktionen von Menschen ist hier zur Bestimmung der Stärke von Erdbeben (I – XII) verwendet worden. Eine Umrechnung zur gebräuchlicheren Richter-Skala (1-8/9) ist leicht möglich.

### 1.2.4. Beauford-Skala

Die Beauford-Skala beschreibt die Windstärke.

## 1.3. Messtheorie



Messung wurde als homomorphe Abbildung empirischer in numerische Relativ beschreiben. Es können dabei zwei Aufgaben der Messung voneinander unterschieden werden:

- Aufdecken empirischer (psychischer) Strukturen
- Überprüfung, inwieweit diese empirische Strukturen in numerische Strukturen abgebildet werden können (Repräsentationsproblem)

### 1.3.1. Probleme der Messtheorie

- Repräsentationsproblem
- Eindeutigkeitsproblem
- Bedeutsamkeitsproblem

### 1.3.2. Skalenniveaus

Skalentyp	Zulässige Transformation	Invariant bleiben	Beispiel
<b>NOMINAL</b>	Jede eineindeutige Funktion	Eindeutigkeit der Messwerte	Nummerierung, Geschlecht
<b>ORDINAL</b>	Jede monoton steigende Funktion	Rangordnung der Messwerte	Härteskala, Schulnoten (?)
<b>INTERVALL</b>	Jede positiv lineare Funktion $y = bx + c$ mit $b > 0$	Verhältnisse der Intervalle der Messwerte	Temperatur, Nutzen
<b>VERHÄLTNIS</b>	Jede Ähnlichkeitsfunktion $y = bx$	Verhältnisse von Messwerten	Cm-g-sec-System
<b>ABSOLUT</b>	Identitätsfunktion $y = x$	Messwerte	Häufigkeiten, Wahrscheinlichkeiten

Anmerkung: Häufig wird die Bezeichnung Kardinalskala als Zusammenfassung von Intervall- und Verhältnisskala verwendet, etwa von Bortz.

### 1.3.3. Coomb'sche Datenklassifikation

Vergleich von	Dominanz von Paaren	Nähe von Paaren
Personen und Reizen	Einzelreize	Präferenz
Reizen	Reizvergleich	Ähnlichkeit

Dabei werden an dieser Stelle Präferenz, Reizvergleich und Ähnlichkeit behandelt.

Beispiele:

Vergleich von	Dominanz von Paaren	Nähe von Paaren
Personen und Reizen	Guttman-Skala Rasch-Modell	Coombs-Skala MDS Conjoint-Analyse
Reizen	Eindimensionale Skalierung BTL	MDS Clusteranalyse Faktorenanalyse



## 2. Signalentdeckungstheorie

Die SDT wurde ursprünglich als Anwendung der mathematischen Entscheidungstheorie auf Entdeckungsprobleme in verschiedenen Kontexten (Radar, Röntgenstrahlen) konzipiert. In der Psychologie wurde sie erstmals zu Beginn der 50er Jahre als psychophysisches Modell zur Kontrolle der Reaktionsneigung von Probanden bei Reizentdeckungs- und Reizdiskriminationsaufgaben (Schwellenbestimmung) verwendet.

Das Ziel bestand also in der Unterscheidung von **sensorischer Leistungsfähigkeit** und **Reaktionsneigung** (Strategien der Aufgabenbearbeitung, Urteilshaltungen).

### 2.1. Grundbegriffe

Man unterscheidet in einem Experiment zwischen Durchgängen, die **nur Rauschen (noise)**, **N**, und solchen, in denen zusätzlich zum Rauschen ein **Signal** dargeboten wird, **SN**. In jedem Durchgang soll reagiert werden, wenn ein Signal zu beobachten ist.

		Reaktion	
		nein	ja
Reiz	SN	<b>Verpasser</b>	<b>Treffer</b>
		falsch negativ (nein $\frac{1}{2}$ SN)	richtig positiv (ja $\frac{1}{2}$ SN)
	N	<b>Korrekte Zurückweisung</b>	<b>Falscher Alarm</b>
		richtig negativ (nein $\frac{1}{2}$ N)	falsch positiv (ja $\frac{1}{2}$ N)

SN = Signal und Rauschen  
 N = Rauschen

Per Konvention erfolgt eine Betrachtung der Treffer und der falschen Alarme. Für diese beiden Parameter werden 3 Kennwerte berechnet:

- $p$  = Fläche unter der Verteilungsfunktion der SNV
- $z$  = Zugehöriger Standardwert
- $y$  = Ordinatenwert von  $z$

Beispiel:

(Fast) perfekter Beobachter

	$p$	$z$	$y$		nein	ja
$p$ (Treffer)	0.98	2.05	0.049	SN	2	98
$p$ (falscher Alarm)	0.002	-2.88	0.006	N	998	2

Mutiger, sensibler Beobachter

	$p$	$z$	$y$		nein	ja
$p$ (Treffer)	0.98	2.05	0.05	SN	2	98
$p$ (falscher Alarm)	0.5	0	0.399	N	500	500

Das psychophysische Modell der SDT macht dabei nicht nur Annahmen über die Wirkung der Reize, sondern auch über das Urteilsverhalten der Probanden. Diese Annahmen werden im Modell der SDT zusammengefasst.

## 2.2. Modell der SDT

### 2.2.1. Annahmen

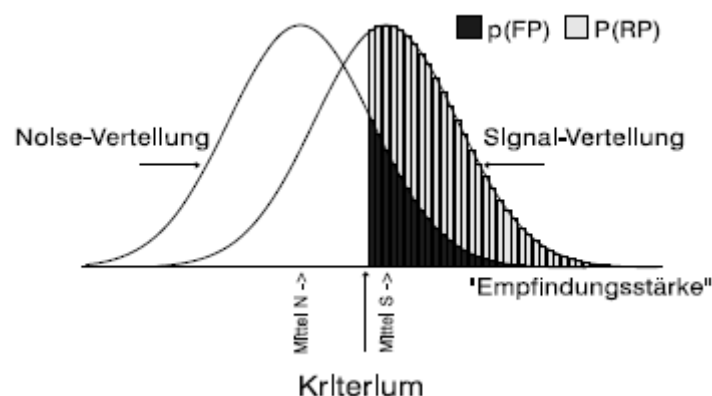
**Annahme 1:** Ein Signal löst eine Empfindungsstärke aus. Bei wiederholter Darbietung des Signals ist diese Empfindungsstärke nicht konstant, sondern normalverteilt nach  $N(\mu_{SN}; \sigma_{SN})$ .

**Annahme 2:** Der Beurteiler legt ein Reaktionskriterium fest, mit dem die Empfindungsstärke verglichen wird.

**Annahme 3:** Auch Rauschen löst Empfindungsstärken aus, die ebenfalls normalverteilt sind, aber mit einem anderen Mittelwert:  $\sim N(\mu_N; \sigma_N)$ . Die Kurven überlappen sich, d.h. der Beobachter kann nicht eindeutig zwischen SN und N unterscheiden.

### 2.2.2. Likelihood-Ratio

Aus dem Verhältnis der Ordinatenwerte der Dichtefunktionen  $f(x|N)$  und  $f(x|SN)$  lässt sich berechnen, mit welcher Wahrscheinlichkeit eine Empfindungsstärke aus der Darbietung von N bzw. SN resultiert.



Die schraffierte Fläche in der Signalverteilung entspricht dem Prozentsatz der Treffer, in der Noise-Verteilung entspricht sie dem Prozentsatz der falschen Alarme.

Das oben genannte Verhältnis wird auch als Likelihood-Ratio  $l(x)$  bezeichnet:

$$l(x) = \frac{f(x|SN)}{f(x|N)}$$

mit  $x$  = Ordinatenwert des Kriteriums in der jeweiligen Verteilungsfunktion.

### 2.2.3. Parameter des Modells

Die wichtigsten Parameter sind:

- Diskriminationsindex  $d'$
- Reaktionsneigung  $\beta$

#### 2.2.3.1. Diskriminationsindex $d'$

Der **Diskriminationsindex  $d'$**  steht für den sensorischen Aspekt der Signalentdeckung, also den Unterschied zwischen SN und N. Die Stärke des

Signals ist dabei immer relativ zum Rauschen zu sehen und insbesondere auch dann stärker, wenn die Streuung der N-Verteilung abnimmt.

$$d' = \frac{\mu_{SN} - \mu_N}{\sigma_N}$$

Es lassen sich zwei praktisch relevante Extremfälle unterscheiden:

- $d' = 0$ : Die Verteilungen liegen direkt aufeinander.
- $d' \gg 0$ : Fast keine Überlappung.

Immer wenn  $d' > 0$ , ist die Empfindungsstärke bei SN im Mittel stärker als bei N. Theoretisch beträgt der Wertebereich von  $d'$  dabei  $W_{d'} = ]-\infty; +\infty[$ , praktisch ist jedoch  $W_{d'} = ]0; +\infty[$ , weil S+N normalerweise mehr Empfindung auslösen sollte als N allein. In der psychologischen Realität treten  $d'$ -Werte bis ca. 4-5 auf.

### 2.2.3.2. Reaktionsneigung $\beta$

Die **Reaktionsneigung**  $\beta$  ist ein Parameter, der das Urteilsverhalten der Pbn ausdrückt. Er berechnet sich aus dem Verhältnis der Wahrscheinlichkeiten am Kriteriumspunkt:

$$\beta = \frac{f(x_{crit} | SN)}{f(x_{crit} | N)}$$

Für  $\beta > 1$  liegt ein „vorsichtiger“ Proband vor. Erst wenn die Wahrscheinlichkeit für SN deutlich größer ist als für N wird „ja“ gesagt. Für  $\beta < 1$  liegt ein „Ja-Sager“ vor, der auch dann antwortet, wenn die Wahrscheinlichkeit für SN kleiner ist als für N.

Der Wertebereich von  $\beta$  ist  $W_{\beta} = ]0; +\infty[$ .

### 2.2.3.3. Berechnung der Parameter

- $d' = z[p(\text{Treffer})] - z[p(\text{falscher Alarm})]$
- $\beta = \frac{y[p(\text{Treffer})]}{y[p(\text{falscher Alarm})]}$  mit  $y$  = Ordinatenwert der SNV

## 2.3. Beispiel: Vigilanz

In typischen Vigilanzuntersuchungen wird eine monotone, einfache Aufgabe über eine lange Zeitdauer bearbeitet; vorzugsweise mit schlafdeprivierten Probanden. Eine prototypische Aufgabe ist die Beobachtung des Zeigers einer Mackworth-Uhr, der normalerweise alle 0,5s auf die nächste Position springt. In je 15 Minuten des Experiments werden 10 Zeigersprünge übersprungen, worauf der Proband reagieren muss.

Es zeigt sich, dass die Leistung, gemessen anhand der Anzahl entdeckter Signale, mit der Zeit abnimmt. Es fragt sich nun, was sich verändert:

- Unterscheidbarkeit zwischen SN und N im Sinne eine Verminderung der Wahrnehmungsleistung (Vigilanzabfall) oder
- Urteilsverhalten des Probanden (kognitiver Effekt: konservativeres Urteilen)?

Eine Veränderung der Reaktionsneigung äußert sich dabei in einem Rückgang der Reaktionen, es wird also seltener und vorsichtiger reagiert. Eine Veränderung des Diskriminationsindex bedeutet, dass Treffer abnehmen, falsche Alarme jedoch zunehmen, das Signal also schlechter entdeckt wird.

Eine derartige Interpretation ist aber mit einigen Voraussetzungen und Einschränkungen verbunden:

- Bei extremen Unterschieden zwischen SN und N werden die Punktschätzungen für  $d'$  und  $\beta$  ungenau. Ideal ist  $p(\text{Treffer}) = 70\text{-}80\%$  und  $p(\text{Falscher Alarm}) > 5\%$ .
- Für eine gute Schätzung von  $d'$  und  $\beta$  ist eine ausreichende Anzahl an Durchgängen nötig. Richtwert: mindestens 50 SN und 50N.
- SN- und N-Verteilungen sollten normalverteilt sein.
- SN- und N-Verteilungen sollten die gleiche Varianz haben.
- Das Entscheidungskriterium sollte ein Wert des Likelihood-Ratios  $l(x)$  sein, also ein Kriterium der Empfindungsstärke.
- Beobachter sollten konsistent urteilen: Wenn Empfindungsstärke größer als Kriterium Reaktion (SN), ansonsten keine Reaktion (N).

## 2.4. Sensitivität und Spezifität

Die Begriffe Sensitivität SE und Spezifität SP werden verwendet, um die Güte eines Tests zu bewerten, einen bestimmten Zustand einer Person zu entdecken. Der Test kann eine positive (D+) oder negative (D-) Diagnose liefern, in Wahrheit kann der Pb krank (K+) oder gesund (K-) sein.

$$SE = p(RP)$$
$$SE = 1 - p(FP)$$

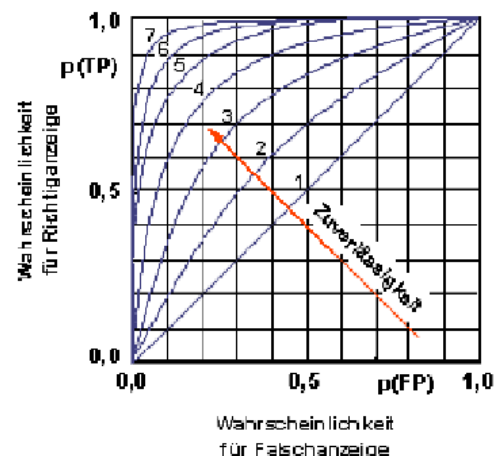
Zu beachten ist hierbei, dass normalerweise nicht zugleich hohe Sensitivität und hohe Spezifität zu realisieren sind.

## 2.5. ROC-Kurven

Die Ergebnisse eines Signalentdeckungs-Experiments lassen sich in einer ROC-Kurve (Receiver Operating Characteristic-Kurve) zusammenfassen, indem man die Wahrscheinlichkeit für einen falschen Alarm gegen die Wahrscheinlichkeit für einen Treffer abträgt. Eine Kurve beschreibt somit die Lage von Probanden mit demselben Wert für  $d'$ . Die Kurve für  $d' = 0$  (Winkelhalbierende) wird auch als Nullsensitivitätslinie bezeichnet.

Die gesamte Kurve erhält man durch systematisches Variieren der Pay-Off-Matrizen (Belohnung/Verluste → Reaktionsneigung) sowie der Signalwahrscheinlichkeit ( $d'$ ).

Signale von hoher Intensität oder sehr sensitive Versuchspersonen liegen im oberen linken Bereich und besitzen ein großes  $d'$ . Je größer der Wert von  $d'$  wird, desto weiter liegt die ROC-Kurve in der linken oberen Ecke.



Ein Entlangwandern an einer ROC-Kurve repräsentiert verschiedene Kriterien bei gleich bleibender Empfindlichkeit  $d'$ . Bei einem liberalen Kriterium liegen die Punkte im rechten oberen Bereich über der Hauptdiagonale, bei einem konservativen Kriterium liegen sie im linken unteren Bereich über der Hauptdiagonale (jeweils für  $d' > 0$ ). Die Hauptdiagonale selbst ( $d' = 0$ ) repräsentiert eine Diskriminationsleistung auf dem Zufallsniveau (Nullsensitivitätslinie).

Ein Maß für die Güte des Tests ist die Fläche unter der ROC-Kurve (AUC: Area Under Curve). Die Fläche kann Werte zwischen 0,5 und 1 annehmen, wobei ein höherer Wert die bessere Güte anzeigt.

[Anmerkung: ROC-Kurven mit SPSS. Die Kurve des Übungsblatts wurde mit folgenden Einstellungen generiert: Testvariable = test, State Variable = vorgruppe (Value = 0); Options: smaller test results indicate more positive tests. Whatever that means.]

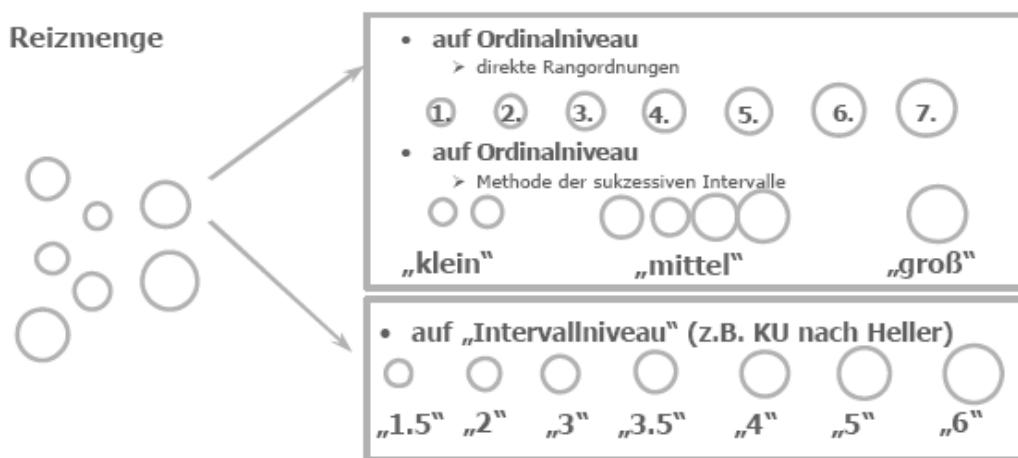
## 3. Probabilistische Skalierung<sup>1</sup>

Im Folgenden soll zunächst kurz auf direkte und indirekte Urteilsverfahren eingegangen werden. Anschließend werden zwei Fälle der indirekten Methoden erläutert: Law of Categorial Judgement und Law of Comparative Judgement. Im nächsten Abschnitt wird zudem das BTL-Modell betrachtet.

### 3.1. Urteilsverfahren

#### 3.1.1. Direkte Methoden

Bei direkten Methoden geben die Versuchspersonen sofort ein quantitatives Urteil ab (direkte Skalierung der Reaktion).



Hierbei existieren verschiedene Möglichkeiten:

- **Direkte Rangordnungen** (ordinal)
- **Methode der sukzessiven Intervalle** (ordinal): Der Proband ordnet die Reize der gegebenen Reizmenge in grobe Kategorien wie „klein“, „mittel“ und „groß“.
- **Kategorieurteil (KU) nach Heller** (Intervall): Hier wird neben der Kategorie zusätzlich die Tendenz angegeben:

Beispiel für das Kategorieurteil:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
sehr klein			klein			mittel			groß			sehr groß		

#### 3.1.2. Indirekte Verfahren

Bei indirekten Methoden indiziert die VP das Ausmaß ihrer Reaktion mittels bestimmter Antwortkategorien, z.B. im Paarvergleich. Ein numerischer Skalenwert ergibt sich erst durch die nachträgliche Anwendung bestimmter Messmodelle.



Kompletter Paarvergleich

<sup>1</sup> Die auf den Folien angegebene Literatur ist unvollständig bzw. schlecht ©. Besser mit den Folien lernen.

Alle indirekten Skalierungsverfahren bestehen also aus der Gewinnung von Skalenwerten des Reizes mittels psychometrischer Funktionen (z.B. wird aus Entdeckungs- oder Verwechslungshäufigkeiten über die SNV ein z-Wert ermittelt. Neben der kumulierten NV wird auch die logistische Funktion verwendet.

Die drei folgenden Berechnungsmethoden haben gemeinsam, dass sie:

- ein probabilistisches Empfindungsmodell unterstellen (Annahme von Messfehlern)
- das Datenniveau von Ordinalniveau auf höhere Niveaus verschieben.

Berechnungsmethode	Verschiebung des Datenniveaus
Law of Categorical Judgement	Ordinal → Intervallniveau
Law of Comparative Judgement	Ordinal → Intervallniveau
BTL-Modell	Ordinal → Verhältnisniveau

Das Law of Categorical Judgement geht dabei von Kategorialurteilen aus (Kat 1 > Kat 2 > Kat 3), während das Law of Comparative Judgement auf Paarvergleichen beruht.

## 3.2. Law of Categorical Judgement

### 3.2.1. Voraussetzungen

#### 3.2.1.1. Ausgangssituation

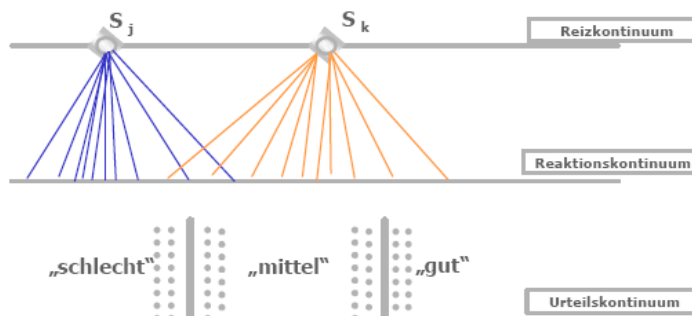
Die Reaktion der VP auf einen Reiz wird in Kategorien erfasst, z.B. „Wie gut ist Angela Merkel als Bundeskanzler geeignet?“

gar nicht	sehr schlecht	schlecht	mittel	gut	sehr gut
0	1	2	3	4	5

Nun sollen aus den Kategorialurteilen von mehreren Versuchspersonen Informationen über den Abstand der Objekte (intervallskaliert) ermittelt werden.

#### 3.2.1.2. Annahmen (Thurstone, 1927)

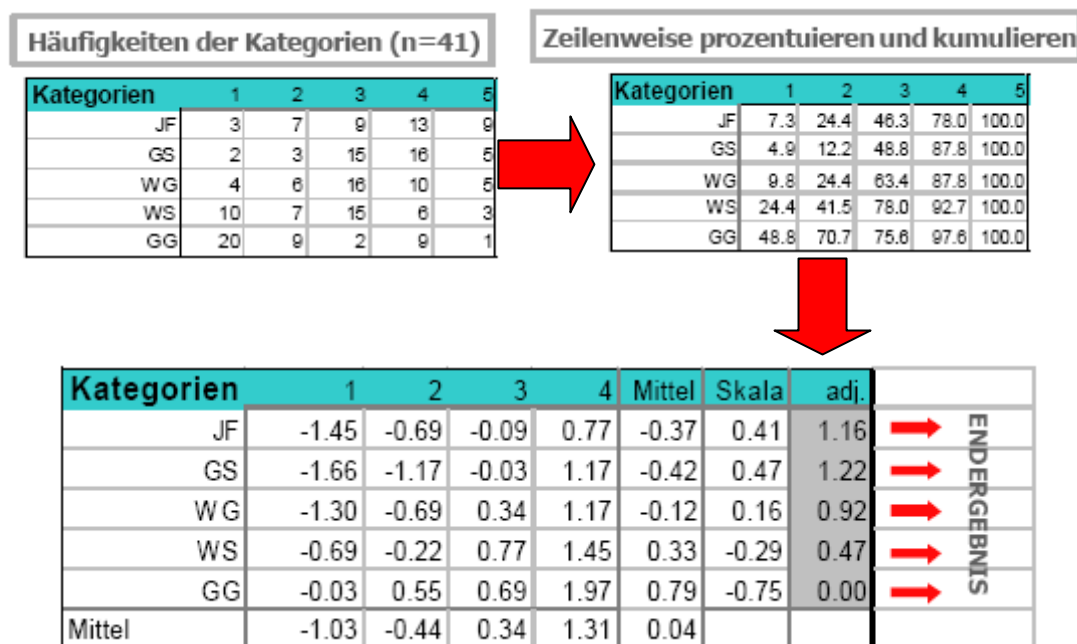
- Urteiler kann Merkmalskontinuum in Kategorien aufteilen
- Kategoriengrenzen schwanken um wahren Wert (normalverteilte Realisierung)
- Beurteilung einer Merkmalsausprägung schwankt ebenfalls zufällig (normalverteilte Realisierung)
- Kriterium als Urteilsgrenze



### 3.2.2. Vorgehen

- Absolute Häufigkeiten der Kategorialurteile in Tabelle zusammenstellen
- Umrechnung in kumulierte Prozentwerte (rel. H'keiten) für jedes Objekt
- z-Werte für W'keiten werden aus der SNV-Tabelle abgelesen;  $z(1) = \infty$  wird weggelassen<sup>2</sup>
- Zeilenmittel ausrechnen: Position der einzelnen Objekte
- Spaltenmittel ausrechnen: Position der oberen Kategoriengrenzen
- Gesamtmittel ausrechnen
- Abstände der Zeilenmittel vom Gesamtmittel ergibt die Skala
- Positivierung/Adjustierung der Skala durch Verschiebung um den kleinsten Wert
- Endergebnis gibt intervallskalierte Werte für jeden Reiz ab

Graphisch:



Die Kategoriengrenzen ergeben sich aus den Spaltenmitteln (Obergrenzen). Diese Kategoriengrenzen lassen sich selbstverständlich ebenfalls verschieben.

### 3.2.3. Interpretation

Wie bereits beschrieben resultieren intervallskalierte Werte, das Verhältnis von Abständen darf also interpretiert werden. Bsp.: Der Abstand zwischen Wolfgang Gerhard und Gregor Gysi ist doppelt so hoch wie der Abstand von Wolfgang Schäuble und Gregor Gysi.

Nicht möglich ist hingegen eine folgende Interpretation: Wolfgang Gerhard ist doppelt so geeignet wie Wolfgang Schäuble – die Skala lässt sich schließlich beliebig verschieben.

<sup>2</sup> Hier ergibt sich eine weitere (indirekte) Voraussetzung des Verfahrens: Die Objekte dürfen im Wahrnehmungsraum nicht zu weit entfernt sein und natürlich auch nicht aufeinander liegen. Ist dies der Fall, so landen alle in derselben Kategorie, sodass das Verfahren nicht mehr angewendet werden kann [ $z(1. \text{Kategorie}) = \infty$ ].



### 3.2.4. Zusammenfassung

Das Law of Categorical Judgement wird eingesetzt, um indirekte Beurteilungen von Reizen von Ordinalniveau auf Intervallniveau zu heben. Ausgangsdaten sind hierbei Häufigkeiten von Kategorienzuordnungen.

Durch die Normalverteilungsannahme können kumulierte Häufigkeiten in z-Werte (intervallskaliert) umgerechnet werden. Die Mittelung innerhalb der Kategoriengrenzen ergeben dabei die (ebenfalls als normalverteilt angenommen) mittleren Kategoriengrenzen, die Mittelung innerhalb eines Reizes hingegen ergibt die Position des Reizes auf einer intervallskalierten Skala.

Der Abstand der Zeilenmittelwerte vom Gesamtmittelwert ergibt die Position des Reizes bzgl. der Kategorien in der endgültigen Skala. Diese wird schließlich der Einfachheit halber adjustiert (positiviert).

## 3.3. Law of Comparative Judgement

Auch das Law of Comparative Judgement hat das Ziel, Informationen über den Abstand von Objekten zu ermitteln (intervallskaliert).

### 3.3.1. Voraussetzungen

#### 3.3.1.1. Ausgangssituation

Ausgangssituation ist die Präferenzmatrix eines Dominanzpaarvergleiches (Bevorzugungshäufigkeiten). Hierbei muss immer angegeben werden, wie die Tabelle zu lesen ist (z.B. „Zeile dominiert Spalte.“).<sup>3</sup>

Ein Problem sind Gleichurteile und Auslassungen. Sie können entweder durch Zufallsziehungen von  $>$  und  $<$  ersetzt oder völlig ignoriert werden, wobei bei der Berechnung von relativen Häufigkeiten als Anzahl der Paarvergleiche dann  $n(\text{Gesamt}) - n(\text{Gleichurteile})$  zu verwenden ist.

#### 3.3.1.2. Annahmen (Thurstone, 1927)

- Der Urteiler schätzt Merkmalsausprägungen mit Hilfe eines Diskriminationsprozesses ein. Wiederholte Beurteilungen fluktuieren dabei normalverteilt um einen wahren Wert.
- Das konkrete Urteil ist die Realisierung einer normalverteilten Zufallsvariable:  $p(x > y) = N(s(x) - s(y))$ . Differenzen zweier normalverteilter ZVn sind ebenfalls normalverteilt. Diese Differenz ist Richtung und Ausmaß der Unterschiedlichkeit zwischen den Reizen.
- Der Urteiler stuft Reiz 1 unterhalb von Reiz 2 ein, wenn die Reaktion auf die realisierte Merkmalsausprägung des Reizes 1 geringer ist als die Reaktion auf die realisierte Merkmalsausprägung des Reizes 2.
- Es hat also dasjenige Differenzurteil die höchste Wahrscheinlichkeit, das dem echten Wert der Differenz entspricht.

### 3.3.2. Vorgehen

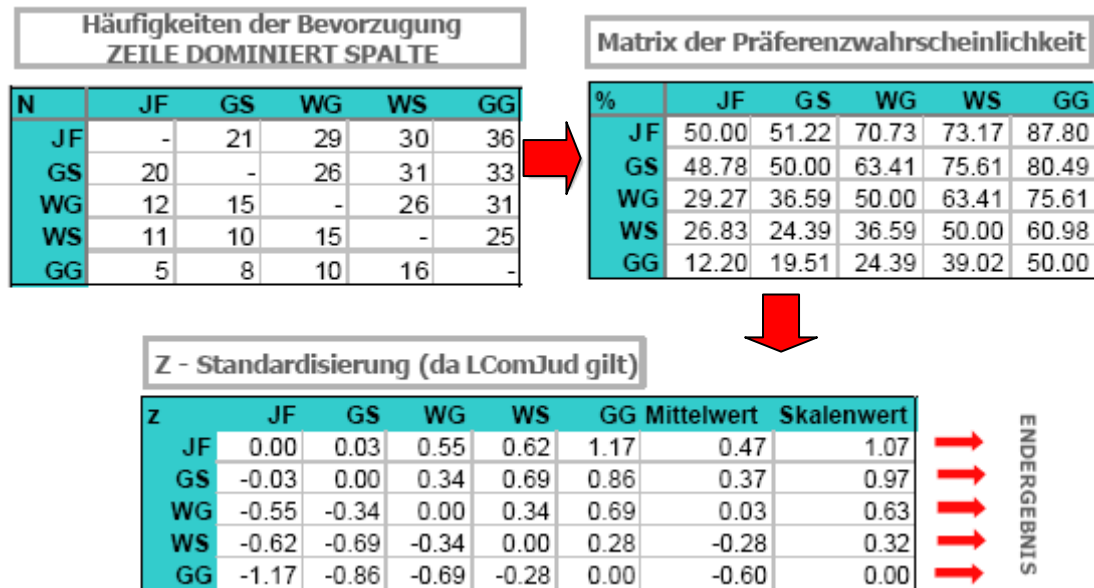
- Bevorzugungshäufigkeiten („Zeile dominiert Spalte“)
- Matrix der Präferenzwahrscheinlichkeiten

---

<sup>3</sup> Kontrollmöglichkeit: Die Bevorzugungshäufigkeiten müssen sich für jedes Paar zu N summieren.

- z-Standardisierung
- Zeilenmittelwerte
- Adjustierung

Graphisch:



Die Passung des Modells ist prüfbar, indem aus den Skalenwerten die Bevorzugungshäufigkeiten der Ausgangsmatrix vorhergesagt werden. Diese Vorhersage lässt sich mit den empirisch gefundenen Werten vergleichen.

### 3.3.3. Hinweise

Paarvergleichslisten sollten nach bestimmten Richtlinien generiert werden, um Urteilsverzerrungen zu vermeiden:

- Jeder Reiz sollte gleich oft als 1. und als 2. Reiz dargeboten werden.
- Kontrolle von Müdigkeitseffekten über umgekehrte Reihenfolge bei einer Hälfte der Probanden (Positionseffekte).
- Paare mit einem gleichen Reiz sollten möglichst weit voneinander entfernt angeordnet werden (Positionseffekte).
- „Korrekte“ Antworten sollten gleich oft an 1. oder 2. Stelle stehen (Seiteneffekte).
- Es sollte kein korrektes Antwortmuster und kein Muster unterschiedlich schwieriger Paarvergleiche geben.
- Falls eine Vp die Liste der Paarvergleiche mehrmals ausfüllt, sollte die Reihenfolge variiert werden, um Lerneffekte zu vermeiden.

### 3.3.4. Bewertung

#### 3.3.4.1. Vorteile

- Anhebung des Datenniveaus auf Intervallniveau
- Einfaches Urteil
- Soziale Erwünschtheit und Urteileffekte werden minimiert. Es muss z.B. immer eine Entscheidung getroffen werden und nicht „Es sind doch alle / keine toll.“

### 3.3.4.2. Probleme

- Verletzung der Normalverteilungsannahme weniger problematisch. Ein Bauerntest ist über Vorhersagewerte und nachgeschaltete  $X^2$ -Tests (for goodness of fit) möglich.
- Extremwahrscheinlichkeiten (s.o.); das Law of Comparative Judgement ist ebenfalls nur für Objekte anwendbar, die im psychologischen Wahrnehmungsraum nah beieinander liegen, da die Wahrscheinlichkeiten sonst 0 oder 1 betragen, bzw.  $z = -\infty$  oder  $+\infty$ .
- Exponentielle Anwachsen der notwendigen Paarvergleiche bei Erhöhung der Anzahl der Objekte  $\sim \binom{n}{2}$ .
- Zirkuläre Triaden (s. 3.4.3).

## 3.4. BTL-Modell (Bradley-Terry-Luce-Modell)

Das BTL-Modell beruht wie das Law of Comparative Judgement auf Paarvergleichsdaten (Dominanzpaarvergleich). Ausgangspunkt ist also die Matrix der Bevorzugungshäufigkeiten. Hieraus sollen aber nun Informationen über die Verhältnisse der Objekte zueinander ermittelt werden.

Gleichurteile sind auch hier problematisch, wobei sich wie beim Law of Comparative Judgement zwei Möglichkeiten bieten: Zufallsziehung von  $>/<$  oder ignorieren.

### 3.4.1. Annahmen

Bei dem Vergleich von zwei Reizen spielt eine Rolle,

- wie **groß** die beiden Reize sind (analog zu LCompJ) und
- wie **nah** sie bei einander liegen.

Die zweite Annahme ist eine Konsequenz des Weber'schen Gesetzes (ebenmerkliche Unterschiede). Die Bevorzugungswahrscheinlichkeit wird bei gleichem Abstand von der Größe der Reize bestimmt.<sup>4</sup>

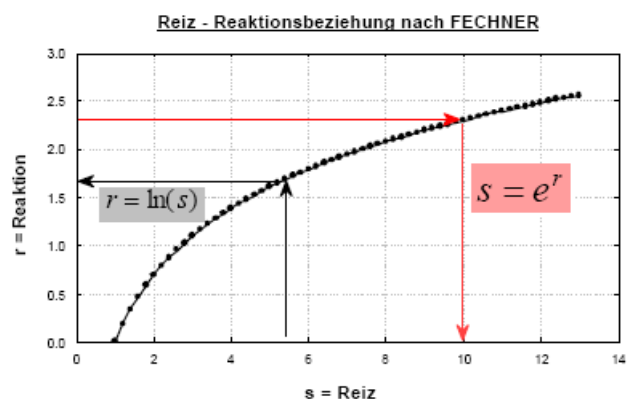
Nach dem BTL-Modell lässt sich die Bevorzugungswahrscheinlichkeit also als Funktion der Reizstärken ( $s$ ) darstellen:

$$p(x > y) = \frac{s_x}{s_x + s_y}$$

Da die Reizstärken jedoch nicht objektiv erfassbar (gerade die will man ja finden), muss die Gleichung über das Fechner'sche Gesetz umgeformt werden.

Mit  $s = e^r$  gilt:

$$p(x > y) = \frac{e^{r_x}}{e^{r_x} + e^{r_y}}$$



<sup>4</sup> Dies ist ein wichtiger Unterschied zum Law of Comparative Judgement, da dort die Bevorzugungswahrscheinlichkeit einzig und allein vom Abstand der Reize voneinander abhängt:  $p(x > y) = \sim N(s_x - s_y)$ .

Es kann somit eine Gleichung angegeben werden, welche die Bevorzugungswahrscheinlichkeit aus der Reaktion des Probanden bestimmt. Gebräuchlich ist dabei die Darstellung als logistische Funktion:

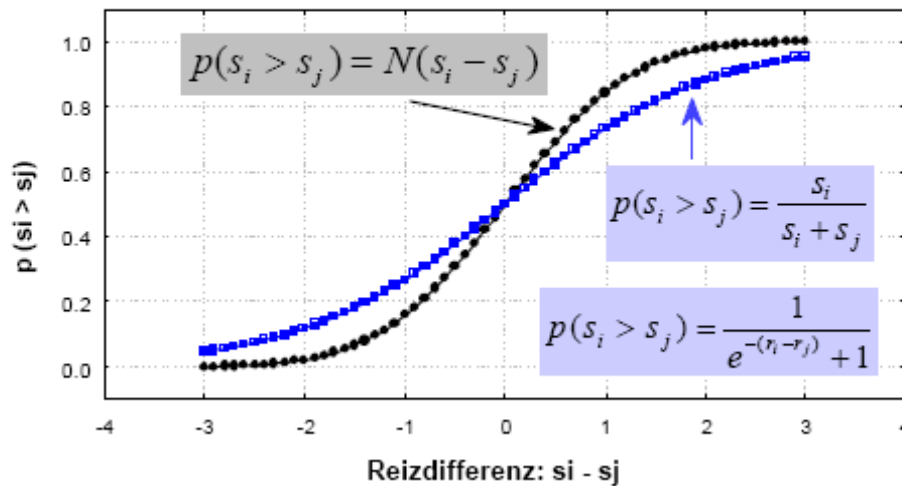
$$p(x > y) = \frac{1}{e^{-(r_x - r_y)} + 1}$$

[Herleitung  $\frac{e^{r_x}}{e^{r_x} + e^{r_y}} = \frac{\frac{e^{r_x}}{e^{r_x}}}{\frac{e^{r_x}}{e^{r_x}} + \frac{e^{r_y}}{e^{r_x}}} = \frac{1}{1 + \frac{e^{r_y}}{e^{r_x}}} = \frac{1}{1 + e^{r_y - r_x}} = \frac{1}{1 + e^{-(r_x - r_y)}} = \frac{1}{e^{-(r_x - r_y)} + 1}$  ] □

Die logistische Funktion beschreibt nun deutlich, dass die Wahrscheinlichkeit von Urteilen, die die Ordinalrelation zwischen x und y richtig abbilden, mit wachsendem Abstand zunächst rasch, dann immer langsamer zunimmt.

**Beziehung: Reizdifferenz - Reaktion**

LCJ und BTL



An der Grafik wird nochmals deutlich, dass die Reizdifferenz immer gleich in die Funktion des LCompJ eingeht, während sie beim BTL-Modell an der Reizgröße gewichtet wird.

Durch Umformung der logistischen Funktion (über einsetzen von  $s = e^r$ ) oder über die folgende einfache Umformung lässt sich nun ein Verhältnis von Bevorzugungs- und Gegenwahrscheinlichkeit angeben:

$$\frac{p(x > y)}{p(y > x)} = \frac{\frac{s_x}{s_x + s_y}}{\frac{s_y}{s_x + s_y}} = \frac{s_x}{s_y}$$

Die Annahmen des BTL-Modells lassen sich nun folgendermaßen zusammenfassen:

- $p(x > y) = \frac{s_x}{s_x + s_y}$  mit p = Wahrscheinlichkeit für die Bevorzugung von x und s = wahre Reizwerte, die wiederum normalverteilte Reaktionen hervorrufen.

- $p(x > y) = \frac{1}{e^{-(r_x - r_y)} + 1}$ : p hängt nur von dem Verhältnis der beiden Reize ab. Die Funktion geht gegen 1 für  $s_y \rightarrow 0$ , also je kleiner  $s_y$ , desto eher dominiert x. Sind beide Reize gleich groß ergibt sich auch eine Bevorzugungswahrscheinlichkeit von  $p = 0,5$ .  
 [Anmerkung zu Folie 12: Für Formel (2) gilt natürlich  $s_x \neq 0$ ].
- $p(x > y) + p(y > x) = 1$ : Es gibt nur zwei Alternativen. Ein „Unentschieden“ ist nicht möglich.
- $\frac{p(x > y)}{p(y > x)} \cdot \frac{p(y > z)}{p(z > y)} = \frac{p(x > z)}{p(z > x)}$  (Multiplikationsbedingung): Sind zwei Präferenzwahrscheinlichkeiten bekannt, lässt sich daraus die dritte berechnen. Das gilt jedoch nicht bei Interaktionen, Extremw'keiten und Mehrdimensionalität.  
 Bsp. für Mehrdimensionalität: Bayern ist besser als Nürnberg, verliert aber trotzdem, weil Nürnberg ein Angstgegner ist. Neben der Leistung kommt also eine weitere Dimension (z.B. Motivation) hinzu.

### 3.4.2. Vorgehen

- $p(x > y) = \frac{s_x}{s_x + s_y}$  und Gegenwahrscheinlichkeit  $p(y > x) = \frac{s_y}{s_x + s_y}$  auf Basis der Ausgangsmatrix der Bevorzugungshäufigkeiten berechnen.
- Division durch Gegenwahrscheinlichkeit  $\frac{p(x > y)}{p(y > x)}$  aufgrund der Annahme, dass das Verhältnis der wahren Werte über das Verhältnis der Wahrscheinlichkeiten  $p(x > y)$  und  $p(y > x)$  geschätzt werden kann.
- Bildung des durchschnittlichen Verhältnisses der Reize zum Gesamtmittelwert durch geometrisches Zeilenmittel:

$$\bar{s}_{(i)} = \sqrt[n]{\prod_{j=1}^n \frac{s(i)}{s(j)}} \quad \text{mit } i = \text{Anzahl der Zeilen, } j = \text{Anzahl der Spalten}$$

Durch Logarithmieren kann die Rechnung vereinfacht werden, da Multiplikationen so in Summen umgewandelt werden können [dieser Logarithmus hat natürlich nichts mit dem Logarithmus aus dem Fechner'schen Gesetz zu tun. Dort ist es ein empirisch gefundenes Gesetz, hier ein Rechentrick]:

$$\ln \bar{s}_{(i)} = \ln s(i) - \frac{\sum_{j=1}^n \ln s(j)}{n}$$

- Durch Rückgängigmachung der Logarithmierung erhält man das Verhältnis des Skalenwerts zum Gesamtmittelwert.

→ Verhältnisskala!

**Beispiel:**

Häufigkeiten der Bevorzugung						Matrix der Präferenzwahrscheinlichkeit					
	JF	GS	WG	WS	GG	%	JF	GS	WG	WS	GG
JF	-	21	29	30	36	JF	50.00	51.22	70.73	73.17	87.80
GS	20	-	26	31	33	GS	48.78	50.00	63.41	75.61	80.49
WG	12	15	-	26	31	WG	29.27	36.59	50.00	63.41	75.61
WS	11	10	15	-	25	WS	26.83	24.39	36.59	50.00	60.98
GG	5	8	10	16	-	GG	12.20	19.51	24.39	39.02	50.00

Verhältnisse aus Wahrscheinlichkeit und Gegenwahrscheinlichkeit

	JF	GS	WG	WS	GG
JF	1.00	1.05	2.42	2.73	7.20
GS	0.95	1.00	1.73	3.10	4.13
WG	0.41	0.58	1.00	1.73	3.10
WS	0.37	0.32	0.58	1.00	1.56
GG	0.14	0.24	0.49	0.64	1.00

Verhältnisse logarithmieren -> Zeilenmittel bilden -> Logarithmierung rückgängig

	JF	GS	WG	WS	GG	Zeilenmittel bilden (x)		ENDERGEBNIS
	JF	GS	WG	WS	GG	x	exp(x)	
JF	0.00	0.05	0.88	1.00	1.97	0.78	2.19	→
GS	-0.05	0.00	0.55	1.13	1.42	0.61	1.84	→
WG	-0.88	-0.55	0.00	0.55	1.13	0.05	1.05	→
WS	-1.00	-1.13	-0.55	0.00	0.45	-0.45	0.64	→
GG	-1.97	-1.42	-0.72	-0.45	0.00	-0.91	0.40	→

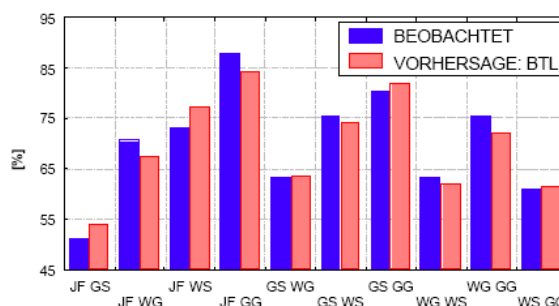
Das Endergebnis gibt das Verhältnis zum Gesamtmittelwert wieder. Werden nun Verhältnisse zwischen Reizen berechnet, kürzt sich der Gesamtmittelwert immer heraus (Konvention: Gesamtmittelwert = 1).

Nun sind Aussagen möglich wie: Joschka Fischer ist etwa doppelt so gut als Bundeskanzler geeignet wie Wolfgang Gerhard.

Die Passung des Modells lässt sich analog zum Law of Comparative Judgment berechnen, indem über die Formel

$$p(x > y) = \frac{s_x}{s_x + s_y}$$

durch einsetzen der Skalenwerte  $s_i$  Bevorzugungswahrscheinlichkeiten und Bevorzugungshäufigkeiten berechnen und diese mit den empirischen Werten vergleichen.



### 3.4.3. Probleme

- Extremwahrscheinlichkeiten: Nicht definierte Werte (z.B. Verhältnisse, bei denen eine Wahrscheinlichkeit 0 beträgt) und entsprechende Gegenwerte herausnehmen und für diese Zeilen mit verringertem n weiterrechnen.
- **Zirkuläre Triaden** (dieses Problem kann auch beim LCompJ auftreten; das Konzept ist für alle Paarvergleiche relevant). Zirkuläre Triaden sind eine Bezeichnung für eine bestimmte Art inkonsistenter Urteile, durch die ein logisches Paradoxon entsteht. Beispielsweise: JF > WS > WG > JF.

Mit zirkulären Triaden kann nach einem bestimmten Schema verfahren werden. Zunächst lässt man sich die Präferenzmatrix für die Versuchsperson ausgeben, bei der man eine zirkuläre Triade entdeckt hat.

	JF	GS	WG	WS	GG
JF	-	0	0	1	1
GS	1	-	1	1	1
WG	1	0	-	0	0
WS	0	0	1	-	1
GG	0	0	1	0	-
$D_i$	2	0	3	2	3
$D_i^2$	4	0	9	4	9

1: Zeile dominiert Spalte.  
 0: vice versa.  
 Hier:  
 $d = 2$   
 $d_{\max} = 5$   
 $T = 0.6$

Nun lässt sich über einige Hilfsgrößen die Teststatistik T für die relative Häufigkeit der konsistenten Triaden berechnen.

Berechnung der Anzahl zirkulärer Triaden:

$$d = \frac{k(k-1)(2k-1)}{12} - \frac{1}{2} \sum_i D_i^2 \quad \text{mit} \quad \begin{array}{l} k = \text{Anzahl der Reize und} \\ D_i = \text{Spaltensumme der Spalte } i \end{array}$$

Berechnung der maximalen Anzahl zirkulärer Triaden:

$$d_{\max} = \frac{k(k^2-4)}{24} \quad (\text{gerades } k) \quad \text{oder} \quad d_{\max} = \frac{k(k^2-1)}{24} \quad (\text{ungerades } k)$$

Teststatistik T als relative Häufigkeit der konsistenten Triaden:

$$T = 1 - \frac{d}{d_{\max}}$$

Faustregel: Für  $T > 0.85$  kann davon ausgegangen werden, dass der Beurteiler konsistent geurteilt hat. Bei nicht konsistenter Antwort kann der Proband ausgeschlossen werden. Haben viele Probanden  $T < 0.85$  so sollte das Vorgehen überdacht werden; evtl. sind die zu beurteilenden Reize beispielsweise nicht weit genug voneinander entfernt.

## 3.5. Vergleich von BTL und LCompJ

Die Frage, welches Modell besser zu verwenden ist, kann empirisch kaum beantwortet werden, da in der Regel die Unterschiede der beiden Modelle durch die Fehlervarianz überdeckt werden.

Kein Modell ist jedoch generell besser (s.u.),

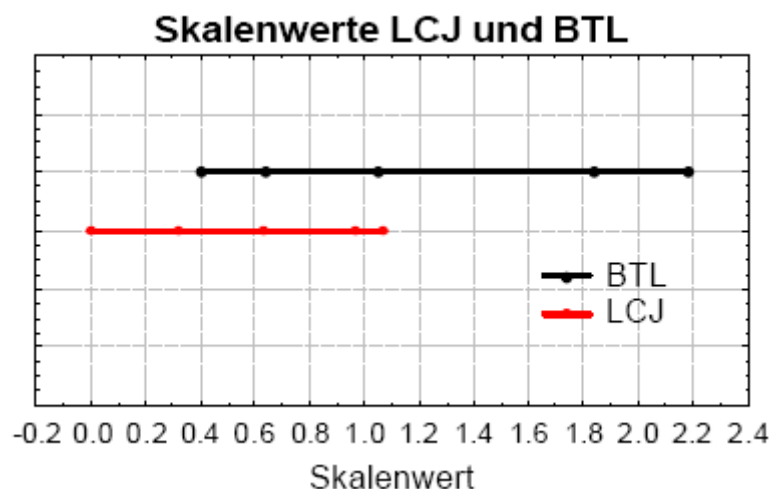
	Law of Comparative Judgement	BTL-Modell
<b>Formel</b>	$p(x > y) = N(s(x) - s(y))$	$p(x > y) = \frac{s_x}{s_x + s_y}$
	$z(1 > 2) = -1 \rightarrow p(1 > 2) = .159$	$p(1 > 2) = 1/(1+2) = .333$
	$z(5 > 6) = -1 \rightarrow p(5 > 6) = .159$	$p(5 > 6) = 5/(5+6) = .454$
<b>Leistungen</b>		
	Intervallskala	Verhältnisskala
	Aussagen über Unterschiede zwischen Abständen	Aussagen über Verhältnisse
	z-Werte	Reelle Zahlen
	Voraussetzung u.a. Transitivität von Differenzen	Voraussetzung u.a. Transitivität von Verhältnissen

Die Entscheidung, welches Modell zu verwenden ist, hängt von den Annahmen ab, die das Empfindungsmodell (Übertragung Reiz – Reaktion) betreffen:

- Lineare Übertragung: LCJ
- Logarithmische Übertragung: BTL

Zudem ist das BTL-Modell anfälliger für Extremwahrscheinlichkeiten, weil es mehr Annahmen macht.

Abschließend sollen die durch LCJ bzw. BTL erhaltenen Skalenwerte verglichen werden:



Hierbei fällt auf, dass das BTL-Modell nicht bei 0 anfängt, da der Skala ja ein absoluter Nullpunkt zugrunde liegt. Die Skala des LCJ hingegen lässt sich beliebig verschieben.

### 3.6. Übung: Probabilistische Skalierung

#### 3.6.1. Law of Categorical Judgement

Ausgangsdaten ist die VPn-Liste der Kategorienurteile. In Excel kann diese einfach in eine Häufigkeitstabelle überführt werden: `{=HÄUFIGKEIT(B11:B79;{1;2;3;4;5})}`.

Kontrolle: Spaltensummen (über Politiker) müssen n ergeben:  $ES1 + ES2 + ES3 + ES4 + ES5 = n$ .



Vorgehen:

- Häufigkeitstabelle: Kumulieren, Aufsummieren [ =F8+(G8\*100/N ]
- z-Werte zu Prozentwerten ermitteln [ = NORMINV() ]
- Zeilenmittel, Gesamtmittel
- Gesamtmittel – Zeilenmittel = Skalenwert
- Skala adjustieren

### 3.6.2. Law of Comparative Judgement

- Matrix der Bevorzugungshäufigkeiten
- Matrix der Bevorzugungswahrscheinlichkeiten mit  $p(x > y) = N(s(x) - s(y))$
- z-Werte [ = STANDNORMINV() ]
- Zeilenmittel, Gesamtmittel, Adjustierung

Interessant: Plötzlich zeigt sich ein großer Unterschied zwischen Claudia Roth und Kurt Beck, der durch das Law of Categorical Judgement nicht gefunden wurde. Das Prinzip des Dominanzpaarvergleichs zwingt zum Präferenzurteil!

### 3.6.3. BTL

- Matrix der Bevorzugungshäufigkeiten
- Matrix der Bevorzugungswahrscheinlichkeiten mit  $p(x > y) = \frac{s_x}{s_x + s_y}$
- Verhältnis der Präferenzwahrscheinlichkeiten
- Logarithmieren [ =LN() ], Zeilenmittel, Logarithmieren rückgängig machen [ =EXP() ] → Skalenwerte

Statt Logarithmierung kann man in Excel auch einfach das geometrische Mittel berechnen: =(C29\*D29\*E29\*F29\*G29)^(1/5)

### 3.6.4. LCJ- und BTL-Vorhersage

LCJ: Differenz der Skalenwerte → W'keit [ =NORMVERT() ] → W'keit \* n.

BTL: z.B. ES/(ES+KB) → p(ES > KB) → W'keit \* n.

Anmerkungen zur Präsentation:

- „Direkt“ in der Graphik zum Vergleich der Skalenwerte bedeutet, dass die Kategorie-Wahrscheinlichkeiten einfach durch Mittelung berechnet wurden (was natürlich nicht geht).
- Der Extremwert für Angela Merkel im BTL-Modell liegt vermutlich an der Verletzung der Annahmen, da  $AM > ES$  eine Extremwahrscheinlichkeit dargestellt hat.

## 4. Clusteranalyse

### 4.1. Fragestellung und Anwendungsbereiche

Die Clusteranalyse ist ein Verfahren zur Gruppierung bzw. Kategorisierung von Personen oder Objekten. **Es werden also Fälle und nicht Variablen kategorisiert (im Gegensatz zur Faktorenanalyse, s. 6.).**

Als Gruppe wird dabei eine Ansammlung von Objekten mit ähnlichen Eigenschaften verstanden. Gruppen von Personen werden also von Personen gebildet, die ähnlich reagieren, Gruppen von Variablen bestehen aus solchen, auf die ähnlich reagiert wird.

Es sollen heterogene Gesamtheiten von Objekten unter Berücksichtigung aller verfügbaren Eigenschaften in möglichst homogene Gruppen untergliedert werden. Dabei findet sich, wieder im Gegensatz zur Faktorenanalyse, kein Postulat der statistischen Unabhängigkeit. Ganz im Gegenteil: Die Gruppen sind abhängig.

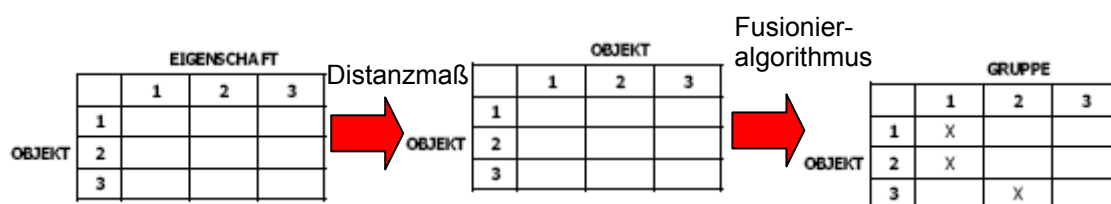
Beispiele:

- In welchen Städten sollen Testmärkte eingerichtet werden? → Suche nach ähnlichen Städten.
- Welche Arten von Organisationsstrukturen gibt es?
- Welche Persönlichkeitstypen gibt es?

### 4.2. Vorgehen

Ausgangspunkt bildet die Objekt x Eigenschaften (Variablen) – Matrix. Über ein Proximitätsmaß (Ähnlichkeits-/Distanzmaß; vgl. 4.3) wird daraus eine Objekt-Objekt-Matrix mit den Differenzen der Objekte zueinander errechnet. Die Diagonale dieser Matrix muss immer Null und die Matrix selbst symmetrisch sein.

Schließlich werden die Objekte nach und nach über einen Fusionieralgorithmus zu Gruppen zusammengefasst.



### 4.3. Distanzmaße

Es gibt eine Vielzahl von Proximitätsmaßen, sodass die Clusteranalyse sehr flexibel an die jeweiligen Gegebenheiten angepasst werden kann. Zentral ist hierbei jedoch die Unterscheidung von Proximitätsmaßen bei nominalskalierten vs. metrischen Daten.

Im Normalfall sollte eine Clusteranalyse immer nur auf einem Datenniveau stattfinden. Liegen gemischte Daten vor, so gibt es zwei mögliche Vorgehensweisen. Entweder werden die metrischen Daten künstlich kategorisiert, oder man rechnet eine zweistufige Clusteranalyse. Dabei werden im ersten Schritt nur die Nominaldaten analysiert und im zweiten Schritt die metrischen Daten hinzugenommen.

**Übersicht – Proximitätsmaße:**

Nominal-Skalen	Metrische Skalen
Tanimoto-Koeffizient	L1-Norm (City-Block)
RR-Koeffizient	L2-Norm (Euklidische Distanz)
M-Koeffizient	(quadrierte Euklidische Distanz)
Dice-Koeffizient	Q-Korrelationskoeffizient
Kulczynski-Koeffizient	Mahalanobis-Distanz

**4.3.1. Nominalskalen**

Proximitätskoeffizienten für Nominalskalen beruhen auf der Vierfeldertafel, in der kodiert ist, ob bei zwei Objekten ein Merkmal jeweils vorhanden bzw. nicht vorhanden ist.

		Objekt 2	
		Vorhanden	Nicht vorhanden
Objekt 1	Vorhanden	a = n(1;1)	c = n(1;0)
	Nicht vorhanden	b = n(0;1)	d = n(0;0)

Allgemeine Bestimmungsformel:

$$S_{ij} = \frac{a + \delta d}{a + \delta d + \lambda(b + c)}$$

Bsp.: Objekt 1 weist die Eigenschaft auf, Objekt 2 jedoch nicht.

Mit  $m = a + b + c + d$  und den Gewichtungsfaktoren  $\delta$  und  $\lambda$  ergeben sich folgende Koeffizienten:

Name des Koeffizienten	$\delta$	$\lambda$	Formel
Tanimoto	0	1	$\frac{a}{a + b + c}$
Simple-Matching (M)	1	1	$\frac{a + d}{m}$
Russel & Rao (RR)	-	-	$\frac{a}{m}$
Dice	0	0.5	$\frac{2a}{2a + b + c}$
Kulczynski	-	-	$\frac{a}{b + c}$

Der Gewichtungsfaktor  $\delta$  gibt an, ob das gemeinsame Nicht-Vorhandensein eines Merkmals zusätzlich zum gemeinsamen Vorhandensein berücksichtigt werden soll. Steckt darin eine Information (z.B. Symptom liegt nicht vor), sollte es auch berücksichtigt werden, ist die Information redundant (jemand der männlich ist, ist sicherlich auch nicht weiblich), sollte  $\delta = 0$  verwendet werden.

Die meisten Koeffizienten haben einen Wertebereich zwischen 0 und 1. Je größer der Koeffizient, desto größer die Ähnlichkeit. Will man von dem Ähnlichkeitsmaß zu einem Distanzmaß kommen, muss das Ähnlichkeitsmaß von 1 subtrahiert werden.

### 4.3.1.1. Beispiel

Beispiel: Es werden 10 Patienten hinsichtlich 7 Merkmalen verglichen. Für jedes Probandenpaar wird nun über alle Merkmale das  $\langle i,j \rangle$ -Tupel gebildet und die Tupel aufsummiert. Daraus resultiert die obige Vierfeldertafel.

ID	Melancholie	ängstlich	Schmerzen	Mimik	Dauer	zugänglich	Stimme
	ja/nein	ja/nein	ja/nein	wenig/viel	$\geq 5y / < 5y$	ja/nein	leise/normal
1	0	1	1	1	0	1	0
2	0	1	1	1	0	1	1
3	1	1	0	0	1	0	1
4	1	1	0	0	1	0	0
5	1	1	1	0	1	0	0
6	1	1	1	1	0	1	1
7	0	1	1	0	0	1	1
8	0	1	1	0	0	1	1
9	1	1	1	1	1	0	0
10	1	1	1	1	0	0	0

Für die Patienten 1 und 2 ergibt sich folgende Tafel:

		Pat 2	
		1	0
Pat 1	1	4	0
	0	1	2

Ähnlichkeit nach: RR-Koeffizient: 4/7

Tanimoto: 4/5 M-Koeffizient: 6/7

Anmerkungen – Symmetrie: Der Vergleich P1 mit P2 entspricht dem Vergleich von P2 mit P1.

### 4.3.1.2. Mehrkategoriale Merkmale

Mehrkategoriale Variablen lassen sich durch eine Umformung in das bekannte Schema integrieren. Jede Kategorie wird dabei als einzelne Variable gesehen, die entweder vorliegen kann oder nicht.

Dem Gewichtungskoeffizienten  $\delta$  kommt also bei mehrkategorialen Variablen eine besondere Bedeutung zu. Dieser sollte bei der Verwendung von umkodierten mehrkategorialen Variablen stets auf 0 gesetzt werden.

Als Beispiel soll die Variable „Dauer;  $>/< 5$  Jahre“ aus dem obigen Datensatz durch eine mehrkategoriale Variable ersetzt werden:

Dauer der Erkrankung	Stufe	Transformation in mehrere binäre Merkmale
0-2 Jahre	1	0001
2-4 Jahre	2	0010
4-6 Jahre	3	0100
$\geq 6$ Jahre	4	1000

Daraus ergibt sich folgende Ergänzung der Tabelle statt der Spalte „Dauer“:

$>/< 5$ J	0-2 J	2-4 J	4-6 J	$\geq 6$ J
0	→ 1	0	0	0
0	→ 0	1	0	0
1	→ 0	0	1	0
1	→ 0	0	0	1

„Da steckt eine enorme Redundanz drin.“

### 4.3.2. Metrische Skalen

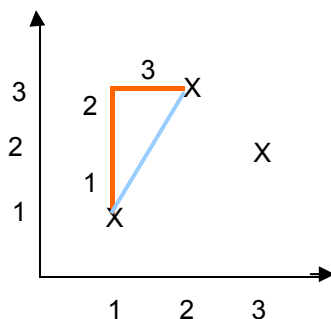
Bsp.: Befragung von Hausfrauen nach der Einschätzung verschiedener Emulsionsfette bzgl. Kaloriengehalt, Preis und Vitamingehalt auf einer 8-stufigen Skala (0-7). Frage: Welche Emulsionsfette sind sich hinsichtlich der gewählten Merkmale ähnlich?

In Frage kommen folgende Proximitätsmaße:

- L1-Norm (City-Block/Manhattan-Metrik)
- L2-Norm (Euklidische Distanz)
- Q-Korrelationskoeffizient
- Mahalanobis-Distanz

#### 4.3.2.1. L1- und L2-Norm

L1- und L2-Norm sollen kurz an einem Beispiel verdeutlicht werden:



	X	Y
A	1	1
B	2	3
C	3	2

L1-Norm:  $d = 3$

L2-Norm:  $d = \sqrt{2^2 + 1^2} = \sqrt{5}$

Die Euklidische Distanz misst also den direkten Abstand von zwei Objekten, während die City-Block-Metrik den Weg in Rechtecken nachfährt.

Für das Emulsionsfett-Beispiel ergibt sich aus der Rohdatendatenmatrix (Objekte x Eigenschaften) für den Abstand von Rama und Hama auf allen Variablen:

- L1-Norm (City-Block-Metrik):

$$d = |1-2| + |2-3| + |1-3| = 1+1+2 = 4$$

- L2-Norm (Euklidische Distanz):

$$d = \sqrt{(1-2)^2 + (2-3)^2 + (1-3)^2} = \sqrt{6} = 2.45$$

- Quadrierte Euklidische Distanz:

$$d = (1-2)^2 + (2-3)^2 + (1-3)^2 = 1+1+4 = 6$$

Diese Ähnlichkeitskoeffizienten werden nun für alle Objektpaare über alle Variablen berechnet. Die Fusionieralgorithmen beruhen dabei meist auf der euklidischen Distanz. Die City-Block-Metrik kommt hingegen nur bei der Bestimmung von örtlichen Distanzen zum Einsatz.

In der Veranstaltung wird ausschließlich mit der (quadrierten) Euklidischen Distanz gerechnet.

#### 4.3.2.2. Q-Korrelationskoeffizient

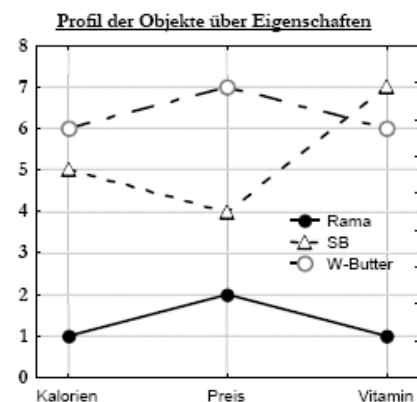
Die beiden Metriken errechnen aus der ObjektxEigenschaften-Matrix eine ObjektxObjekt-Matrix, die die Distanzen der Objekte zueinander enthält.

Wenn jedoch eher die Ähnlichkeit der Profile (= Struktur) analysiert, ist der Q-Korrelationskoeffizient angebracht. Dieser gibt die Ähnlichkeit zweier Profile unabhängig von deren Ausgangslage/Niveau an.

In der nebenstehenden Grafik könnte bspw. interessieren, ob sich SB und Weihnachtsbutter vom Niveau her ähneln (L1-/L2-Norm) oder ob das Profil von Weihnachtsbutter und Rama vergleichbar ist.

Wird der Q-Korrelationskoeffizient verwendet gehen also Informationen über die Unterschiede der Objekte verloren.

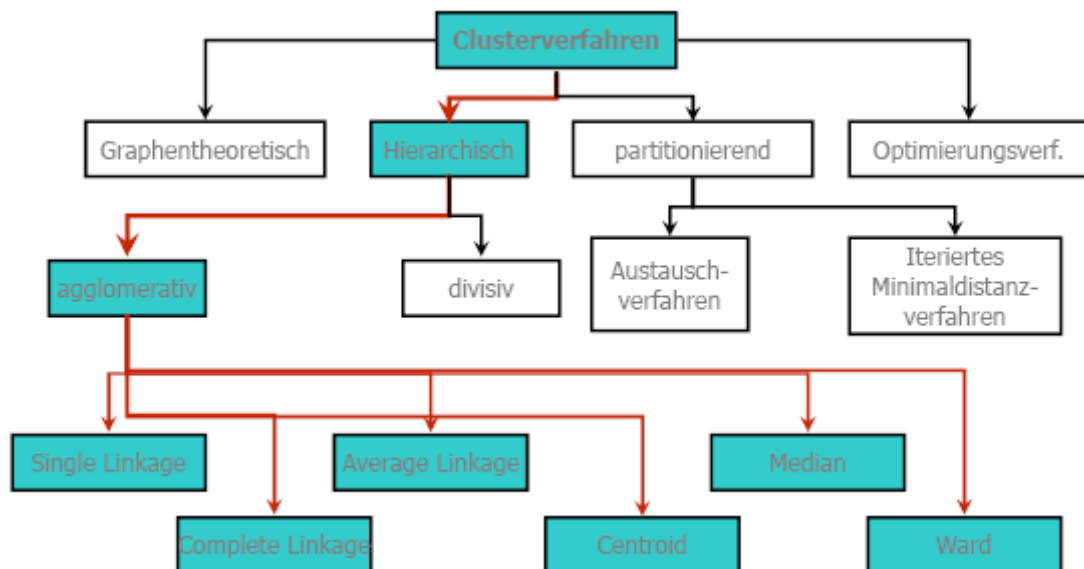
Ein Beispiel für eine solche Untersuchung wäre der Vergleich der 12-Monats-Kennzahlen von einem großen und einem kleinen Unternehmen.



## 4.4. Gruppierungsverfahren: Fusionierungsalgorithmen

Nachdem die ObjektxEigenschaften-Matrix über ein Proximitätsmaß in eine ObjektObjekt-Matrix überführt wurde, können Fusionierungsalgorithmen verwendet werden, um die Objekte zu Gruppen zusammenzufassen.

Auch hier steht eine Vielzahl verschiedener Verfahren zur Auswahl:



In der Veranstaltung werden nur hierarchische agglomerative Verfahren sowie partitionierende Verfahren behandelt. Der Grundgedanke dieser Verfahren ist es, die Distanz innerhalb einer Gruppe zu minimieren (Ähnlichkeit maximieren) und die Distanz zwischen Gruppen zu maximieren (Ähnlichkeit minimieren).

Agglomerative hierarchische Verfahren gehen dabei zunächst davon aus, dass jedes Objekt sein eigenes Cluster darstellt (keine Gruppierung) und fassen benachbarte Objekte dann schrittweise zusammen. Divisive hierarchische Verfahren gehen vom anderen Extrem aus und beginnen mit einem Cluster, das nach und nach unterteilt wird.

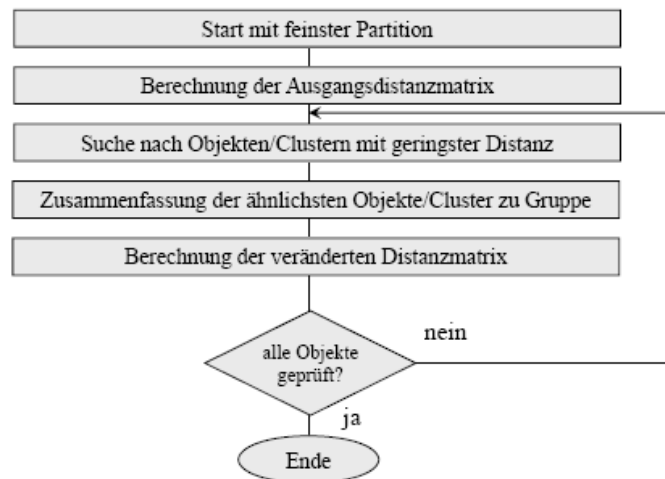
Partitionierende Verfahren hingegen gehen von einer beliebigen bestehenden Gruppierung der Objekte aus. Danach werden Objekte zwischen den Gruppen so lange ausgetauscht, bis die Gruppen möglichst homogen sind und sich möglichst große Abstände zwischen den Gruppen einstellen.

Die einzelnen Schritte bei hierarchischen Verfahren sind also irreversibel (Gruppenzusammenfassungen werden nicht wieder aufgehoben). Bei partitionierenden Verfahren kann ein Objekt auch mehrmals hin und her geschoben werden.

### 4.4.1. Hierarchische Verfahren

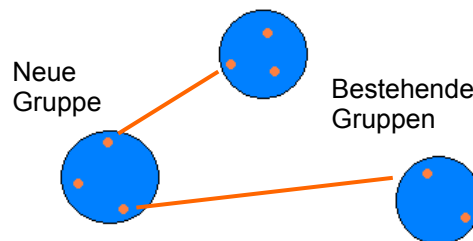
Der Standard sind hierarchische agglomerative Verfahren:

- Anfangs ist jedes Objekt ein Cluster.
- Distanzmatrix: Zusammenfassung der beiden ähnlichsten Objekte zu einem Cluster.
- Neuberechnung der (reduzierten) Distanzmatrix: Die Gruppe stellt ein einziges Objekt dar.
- Zusammenfassung der nun ähnlichsten Objekte.

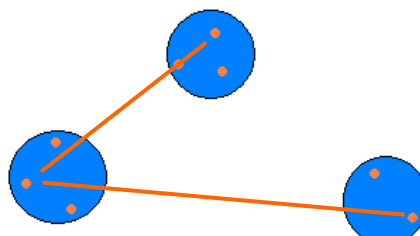


Problematisch ist die Berechnung des Abstandes von Objekten zu Gruppen bzw. zwischen Gruppen. Hierfür gibt es mehrere Verfahren:

- **Single Linkage SL** (auch Nächster Nachbar bzw. Nearest Neighbour): Es wird der geringste Abstand verwendet.



- **Complete Linkage CL** (Entferntester Nachbar; Furthest Neighbour): Es wird die maximale Distanz verwendet.



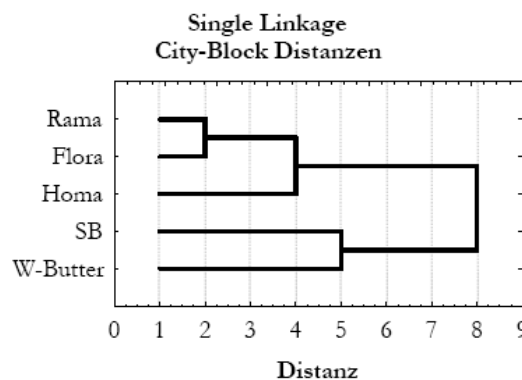
- **Average Linkage AL:** Es wird der mittlere Abstand verwendet.
- **WARD:** Die Streuung (Fehlerquadratsumme) der Abstände wird minimiert.
- **Zentroid:** Abstände werden zum Flächenmittel der Gruppe berechnet.
- **Median**

Hierarchische Verfahren gehen dabei immer soweit, bis nur noch ein Cluster existiert, also alle Objekte zu einer einzigen Gruppe zusammengefasst sind. Da dies vielleicht nicht allzu sinnvoll ist, muss manuell ein Stopp-Kriterium definiert werden.

Die Resultate einer hierarchischen Cluster-Analyse werden in einer Grafik ausgegeben. Hierbei stehen zwei Alternativen zur Verfügung:

- **Eiszapfendiagramm**
- **Dendrogramm**

Das Dendrogramm liest sich v.l.n.r.:



Im ersten Schritt fusionieren im Beispiel Rama und Flora bei einer Distanz von  $d = 2$  (hier im SL). Im nächsten Schritt kommt Homa hinzu (3-Objekt-Cluster). Schließlich bilden SB und W eine Gruppe. Die beiden Gruppen sind deutlich voneinander getrennt und fusionieren erst bei  $d = 8$ .

Im Folgenden sollen Rechenbeispiele für die vier gängigsten Verfahren vorgestellt werden. Diese lassen sich noch problemlos von Hand rechnen, was bei vielen Variablen nicht mehr möglich ist. In SPSS/Statistica lässt sich die Anzahl der Cluster vorgeben, bei denen der Algorithmus abbrechen soll.

#### 4.4.1.1. Single Linkage SL

Die Distanz zwischen zwei Clustern im Single-Linkage-Verfahren (Nächster Nachbar) wird bestimmt, indem der geringste Abstand zwischen zwei Objekten der beiden Cluster berechnet wird.

Für die Fusionierung von SB und W-Butter des obigen Beispiels lässt sich für das SL-Verfahren folgende Distanzmatrix angeben:

	F/R/H	SB	W-Butter
SB	8	0	
W-Butter	11	5	0

$11 > 8 > 5$ : SB und W-Butter fusionieren.

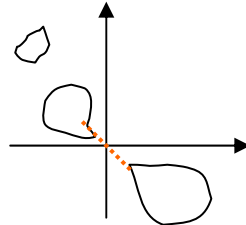
SPSS würde danach unsinnigerweise die Distanzmatrix für den letzten Schritt ebenfalls bestimmen, um alle Objekte in einer Gruppe zusammenzufassen:

	F/R/H
SB/W	8



Über Single Linkage werden bevorzugt große Gruppen (Ketten) gebildet und große Distanzen betont. Daher eignet sich das Verfahren v.a. um Ausreißer zu finden.

Die Kettenbildung kann allerdings auch dazu führen, dass Objekte geclustert werden, die eigentlich nicht zusammengehören.



Nachdem Ausreißer gefunden wurden sollte also auf andere Verfahren zurückgegriffen werden.

#### 4.4.1.2. Complete Linkage CL

Genauso wie SL, nur dass immer der maximale Abstand verwendet wird.

#### 4.4.1.3. Average Linkage AL

Nach dem AL-Verfahren lassen sich folgende Distanzmatrizen berechnen:

	F/R/H	SB	W-Butter
SB	9,5	0	
W-Butter	12,5	5	0

12,5 > 9,5 > 5: SB und W-Butter fusionieren.



	F/R/H
SB/W	$\frac{9,5+12,5}{2} = 11$

In SPSS wird Average Linkage nur als Linkage bezeichnet.

#### 4.4.1.4. WARD

Die Ward-Methode hat einen etwas anderen Ansatz. Hier werden keine Distanzen berechnet, sondern die Veränderung der Fehlerquadratsumme betrachtet, die entsteht, wenn ein Objekt in eine Gruppe integriert wird.

Zu Beginn der Clusteranalyse hat jedes Cluster eine Fehlerquadratsumme von 0. Die Objekt-Objekt-Matrix besteht hier aus hypothetischen Fehlerquadratsummen für alle möglichen Paarungen. Es wird dasjenige Paar fusioniert, das einen möglichst geringen Anstieg an Fehlerquadratsumme mit sich bringt.

Eine beispielhafte Gruppe wären Rama (1|2|1) und Flora (3|2|1). Der Centroid dieser Gruppe ergibt sich aus den jeweiligen Mittelwerten zu (2|2|1). Der Anstieg der Fehlerquadratsumme ergibt sich aus der Summe aller quadrierten Abweichungen. Es lässt sich dabei zeigen, dass der Anstieg der Fehlerquadratsumme gleich der halben quadrierten Euklidischen Distanz ist (logisch; s. 4.7).

Vorteil: Es werden homogene und gleichmäßige („sinnvolle“) Gruppen gebildet. Das Ward-Verfahren ist daher die Methode der Wahl unter den hierarchischen Verfahren.

Das Verfahren findet jedoch keine Ausreißer, sodass vorher eine CA mit dem SL-Verfahren gerechnet werden sollte.

#### 4.4.1.5. Wann welches Verfahren?

Single Linkage findet Ausreißer, sodass diese eliminiert werden können. Das Verfahren bildet große Gruppen (Ketten).

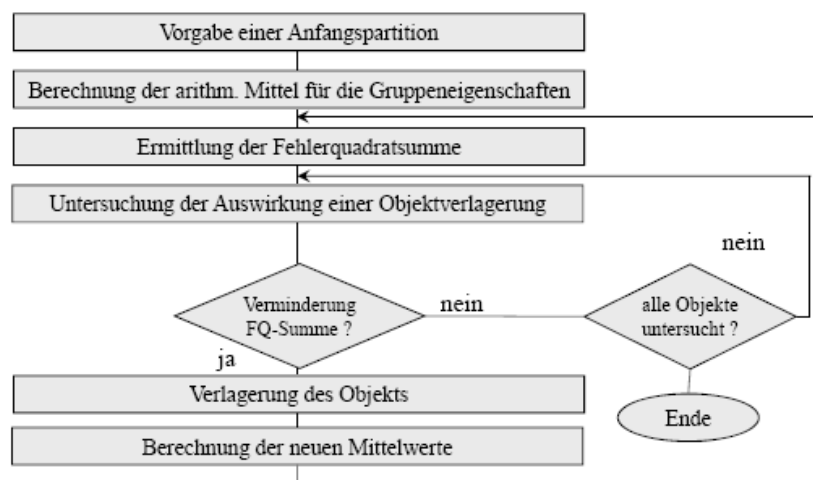
Average Linkage: Bildet viele kleine Gruppen mit wenigen Elementen.

WARD-Verfahren: Bildet ungefähr gleich große Gruppen, findet meist „sehr gute“ Gruppen und ordnet die Elemente meist „richtig“ zu.

### 4.4.2. Partitionierende Verfahren

Hierarchische Verfahren geben einen Überblick darüber, welche Objekte sinnvollerweise zu Gruppen zusammengefasst werden können. Partitionierende Verfahren können diese Lösungen überprüfen und evtl. optimieren.

Hierzu werden einzelne Objekte zwischen bestehenden Gruppen (Anzahl der Cluster vorgegeben!) verschoben und der dadurch entstehende Fehler betrachtet. Wenn die Fehlerquadratsumme bei keiner Verschiebung mehr sinkt, so ist die optimale Lösung gefunden. Es handelt sich also um extrem rechenintensive Verfahren.



Partitionierende Verfahren sind jedoch auch sehr angenehm, da ihre Schritte nicht endgültig sind. Objekte können also mehrmals verschoben werden.

## 4.5. Zusammenfassung

Die Clusteranalyse bietet viele Möglichkeiten an die Daten heranzugehen. Es werden also auch wenige Voraussetzungen an die Daten gestellt.

Sowohl für Nominaldaten als auch metrische Daten gibt es eine Vielzahl von Proximitätsmaßen. Liegen sowohl metrische als auch Nominaldaten vor, so gibt es verschiedene Herangehensweisen (kategorisieren vs. schrittweise).

Auf Basis der ObjektxObjekt-Matrix können wiederum verschiedene Fusionieralgorithmen verwendet werden.

Das Ergebnis der CA ist aber in jedem Fall eine Gruppierung bzw. Kategorisierung der untersuchten Objekte. Diese wird in Eiszapfendiagrammen bzw. Dendrogrammen dargestellt.

Es gibt dabei natürlich immer einen Schritt weniger als Objekte.

### 4.5.1. CA und FA

Durch die Anwendung der CA entstehen abhängige Gruppen (im Gegensatz zu den unabhängigen Gruppen der FA). Auch legt die CA kein Modell über das Zustandekommen der Daten zugrunde.

Die einzige Voraussetzung der CA ist die Einheitlichkeit der verwendeten Skalen.

### 4.5.2. Vorschläge zur praktischen Anwendung

Als Distanzmaß sollte die Euklidische Distanz verwendet werden. Die Daten sollten damit in drei Schritten analysiert werden:

- Fusionierung mit Single Linkage: Eliminieren von Ausreißern
- Finden von sinnvollen Clustern mit dem WARD-Verfahren
- Optimierung der Lösung mit einem partitionierenden Verfahren

## 4.6. Übung: Clusteranalyse

### 4.6.1. Generelles

Die (hierarchische) Clusteranalyse findet sich unter *Analysieren\Klassifizieren*. Gegenstand der Clusteranalyse sind Fälle (Cases) und nicht Variablen.

Unter *Statistik* (Statistics) lässt sich die Zuordnungs-Matrix (Agglomeration Schedule) auswählen. Diese gibt aus, welches Objekt bei einer vorgegeben Anzahl der Cluster in welchem Cluster zu finden ist. Auch lässt sich hier die *Distanzmatrix* (Proximity matrix) oder auch Näherungsmatrix auswählen, die der besprochenen Objekt-Objekt-Matrix entspricht. SPSS gibt Distanzmatrizen (Unähnlichkeitsmatrizen) aus. Es existieren jedoch auch Ähnlichkeitsmatrizen. Welche Art jeweils verwendet wurde findet sich links unter der Tabelle.

Das wichtigste Auswahlfeld ist *Methode*. Hier lassen sich die verschiedenen Berechnungsverfahren einstellen:

- Linkage steht hier für das Verfahren Average Linkage AL.
- Die Quadrierte Euklidische Distanz betont große Distanzen (Ausreißer).
- Minkowski-Metriken: Ur-Metriken
- Toll: Die Werte lassen sich auch hier Standardisieren. Im Zweifelsfall sollte diese Option immer gewählt werden. Bevor man die Standardisierung durchführen lässt ist aber auch ein kurzer Blick auf die Rohdaten angebracht, um die Plausibilität einer Normalverteilung abschätzen zu können.

Unter *Speichern* ist es möglich, die Clusterzugehörigkeit von Items als neue Variable zu kodieren.

## 4.6.2. Output

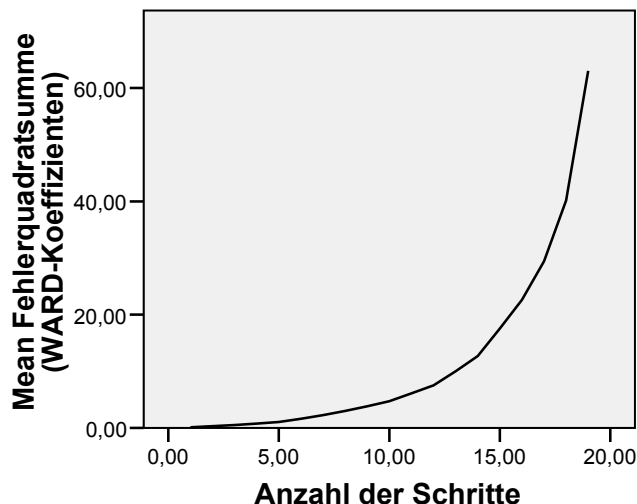
### 4.6.2.1. Zuordnungsübersicht (Agglomeration Schedule)

In den Zeilen werden die einzelnen Schritte der Clusteranalyse dargestellt. Die ersten beiden Spalten erhalten Informationen darüber, welche Objekte bzw. Cluster im entsprechenden Schritt miteinander fusioniert wurden. Ein Cluster wird dabei immer durch das Objekt mit der niedrigsten Nummer repräsentiert.

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	6	7	,210	0	0	4
2	2	11	,388	0	0	7
...	...	...	...	...	...	...
18	1	8	9,412	17	0	19
19	1	5	16,134	18	16	0

In der Spalte *Koeffizienten* finden sich die Distanzmaße für das jeweils verwendete Verfahren. Wurde die WARD-Methode eingesetzt, so entspricht der Koeffizient dem jeweiligen Anstieg der Fehlerquadratsumme für diese Anzahl an Clustern.

Ein Sprung in den Koeffizienten zeigt, dass die fusionierten Gruppen extrem weit voneinander entfernt waren (Abbruchkriterium). Dafür empfiehlt es sich auch, die Koeffizienten gegen die Anzahl der Schritte zu plotten. Hierzu markiert man die Tabelle mit einem Doppelklick und kopiert die Spalte Koeffizienten. Diese setzt man als neue Variable in SPSS ein. Daneben wird die Anzahl der Schritte eingegeben. Diese beiden Variablen lassen sich nun in einem Liniendiagramm gegeneinander abtragen (*Chart Builder* unter *Graphics*).



Für diese Grafik lässt sich dabei ein Ellbogenkriterium definieren, um die Anzahl der zu extrahierenden Cluster zu bestimmen:

Lässt sich ein deutlicher Knick erkennen (hier schwer möglich), so erfolgt in diesem Schritt ein großer Zuwachs an Heterogenität. Also sollte bei diesem Schritt abgebrochen werden.

In den Spalten zum *Ersten Vorkommen* ist protokolliert, in welchem Schritt das jeweilige Objekt bzw. Cluster erstmalig fusioniert ist. 0 steht für „Dieses Objekt war noch nie Teil eines Clusters“, 2 steht für „Dieses Cluster wurde im Schritt Nr. 2 gebildet“. Steht in beiden Spalten eine 0 ( | 0 | 0 | ), besteht das gebildete Cluster aus komplett neuen Objekten.

Eine einzelne 0 in einem der letzten Schritte ( | X | 0 | ) spricht für einen Ausreißer.

Unter der Spalte *Nächster Schritt* (Next Stage) wird angegeben, in welchem Schritt das gebildete Cluster erneut fusioniert. 0 steht hierbei für die endgültige Lösung (es wird nicht mehr fusioniert).

#### 4.6.2.2. Dendrogramm

Das Dendrogramm lässt in SPSS keine Rückschlüsse auf die Distanzen der Objekte zu, weil die verwendete Skala immer auf 25 normiert wird.

#### 4.6.2.3. Eiszapfendiagramm (Icicle)

Im Eiszapfendiagramm werden die einzelnen Objekte in Spalten angetragen. Zwischen zwei Objekte ist jeweils eine freie Spalte. In den Zeilen steht die Anzahl der Cluster. In der ersten Zeile ist also der letzte Schritt der Clusteranalyse abgebildet, in dem alle Objekte in einem einzigen Cluster zusammengeführt sind, während in der letzten Zeile noch keine Clustering durchgeführt wurde und somit jedes Objekt sein eigenes Cluster bildet.

Vertical Icicle

Number of clusters	16 :P or sc he		5: Co rv et- te		8: Fo rd			2: Au di		14 :Ol ds		1: Ac ur a
1	X	X	X	X	X	...	...	X	X	X	X	X
2	X	X	X		X	...	...	X	X	X	X	X
3	X	X	X		X	...	...	X	X	X	X	X
4	X	X	X		X	...	...	X		X	X	X
...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...
16	X		X		X	...	...	X		X		X
17	X		X		X	...	...	X		X		X
18	X		X		X	...	...	X		X		X
19	X		X		X	...	...	X		X		X

Wird in der Spalte zwischen zwei Objekten ein X gesetzt, so sind diese beiden Objekte miteinander in einem Cluster. Ist die Spalte frei, so sind liegen verschiedene Cluster vor.

#### 4.6.2.4. Heterogene Skalenniveaus

Liegen Variablen in metrischen und nominalen Skalen vor, so lässt sich die hierarchische Clusteranalyse nicht mehr ohne weiteres durchführen. Am geschicktesten ist es hier zu fragen, ob wirklich alle Variablen benötigt werden, oder ob evtl. einige oder gar alle nominalskalierte Variablen entfernt werden können.

Im Idealfall lassen sich also alle dichotomen Variablen entfernen und die hCA nur mit standardisierten metrischen Variablen durchführen.

Wird mit Nominaldaten gerechnet müssen bei polytomen Variablen Umkodierungen vorgenommen werden. Dies geht unter *Transformieren\Umkodieren in andere Variable (Recode into different Variables)*. Für eine Variable mit 3 Kategorien müssen also drei neue Variablen definiert werden:

Var_1	Var_2	Var_3	
1 → 1	1 → 0	1 → 0	
2 → 0	2 → 1	2 → 0	→ Zuweisen (Change)
3 → 0	3 → 0	3 → 1	

#### 4.6.2.5. Anmerkungen

Häufig zeigen sich bei verschiedenen Clusterlösungen sog. **Einzelobjektverteilungen**. Wenn statt 3 Clustern 4 Cluster verwendet werden, wird das vierte Cluster nur von einem einzigen Objekt gebildet.

Ein guter Ansatz zur Entscheidung zwischen verschiedenen Clusterlösungen ist, dass die Varianz innerhalb von Clustern geringer sein sollte als die Gesamtvarianz. SPSS kann hierfür auch den Bruch  $\text{Var}(\text{Cluster})/\text{Var}(\text{ges})$  berechnen. Dieser lässt sich wiederum mit verschiedenen Optimierungsverfahren weiterverwenden.

Die einfachere Methode ist jedoch der Anstieg der Fehlerquadratsumme zwischen verschiedenen Clusterlösungen.

### 4.7. Anhang

Beweis: Der Zuwachs an Fehlerquadratsumme entspricht der halben quadrierten euklidischen Distanz (L2). Gegeben seien 2 Objekte, die auf zwei Variablen eingestuft wurden.

Variable	1	2
Objekt 1	1	2
Objekt 2	5	6

Der Zuwachs an Fehlerquadratsumme entspricht mit  $x_i$  = Ausprägung einer Variable bei Objekt i:

$$\left(\frac{x_1 + x_2}{2} - x_1\right)^2 + \left(\frac{x_1 + x_2}{2} - x_2\right)^2,$$

die quadrierte euklidische Distanz berechnet sich für diese Variable als

$$(x_1 - x_2)^2.$$

Der Beweis ist durch eine einfache Umformung möglich:

$$\begin{aligned} &\left(\frac{x_1 + x_2}{2} - x_1\right)^2 + \left(\frac{x_1 + x_2}{2} - x_2\right)^2 = \\ &= \frac{(x_1 + x_2)^2}{4} - x_1x_2 - x_1^2 + x_1^2 + \frac{(x_1 + x_2)^2}{4} - x_1x_2 - x_2^2 + x_2^2 = \\ &= \frac{(x_1 + x_2)^2}{2} - 2x_1x_2 = \\ &= \frac{x_1^2 + 2x_1x_2 + x_2^2}{2} - \frac{4x_1x_2}{2} = \\ &= \frac{x_1^2 - 2x_1x_2 + x_2^2}{2} = \frac{1}{2}(x_1 - x_2)^2 \quad \square \end{aligned}$$

## 5. Conjoint-Analyse

Idee: Der Nutzen eines Gesamtobjekts ergibt sich aus der Verknüpfung der Nutzenwerte seiner Komponenten. Die Beiträge dieser Einzelkomponenten zum Gesamtnutzen sollen über die Conjoint-Analyse bestimmt werden, wobei die Probanden nur das Gesamtobjekt bewerten.

Die Conjoint-Analyse dient hierbei nicht nur der statistischen Auswertung, sondern bereitet auch die Datenerhebung vor. Die Conjoint-Analyse ist in diesem Zusammenhang als dekompositionelles Verfahren zu sehen<sup>5</sup>. Die einzelnen Objekte werden vor der Datenerhebung in Produktkarten (ganzes Objekt mit seinen jeweiligen Eigenschaften) organisiert.

Grundsätzliches Vorgehen: Die Objekte werden zunächst in eine Rangreihe gebracht – entweder durch direkte Rangordnung oder aber auch durch einen Paarvergleich. Die Rangreihe bildet den Ausgangspunkt zur Ableitung von Teilnutzenwerten für die einzelnen Eigenschaftsausprägungen. **Aus den ordinalen Gesamtnutzenwerten werden also metrische Teilnutzenwerte abgeleitet**, die durch Addition metrische Gesamtnutzenwerte bilden können.

### 5.1. Modell der Conjoint-Analyse

Die Conjoint-Analyse geht von einem additiven Modell aus: Der Gesamtnutzen  $y$  eines Objekts ergibt sich aus der Summe der Teilnutzenwerte  $\beta$  seiner Merkmale.

$$y = \beta_1 \cdot A + \beta_2 \cdot B$$

Wichtig: Es wird allerdings davon ausgegangen, dass Probanden ein Globalurteil fällen, wenn sie ein Objekt bewerten, ohne dieses in seine Eigenschaften zu zerlegen.

Es lassen sich daher einige Richtlinien zur **Auswahl der Eigenschaften** und deren Ausprägung aufstellen:

- Keine Ausschlusskriterien (K.O.-Kriterien): Es sollten keine Merkmale variiert werden, deren Vorhandensein alle anderen dominiert – bspw. ein Telefon ohne Hörer.
- Kompensatorische Beziehung der Eigenschaftsausprägungen: „Wenn das Telefon billig ist, ist mir die Farbe auch nicht so wichtig...“ oder „Für einige nette Features zahle ich gerne auch ein bisschen mehr.“
- Annahme eines einstufigen Entscheidungsprozesses (s.o.)
- Unabhängigkeit der Eigenschaften
- Begrenztheit der Anzahl der Eigenschaften und ihrer Ausprägungen. Werden zu viele Merkmale/Ausprägungen verwendet ist die Untersuchung aus praktischen Gründen nicht mehr durchführbar.
- Beeinflussbarkeit: Die Eigenschaften müssen (im Sinne einer UV) variiert werden können, z.B. in der Produktgestaltung.
- Realisierbarkeit: Technische Durchführbarkeit für den Hersteller.
- Relevanz: Ausgewählte Eigenschaften sollten eine Bedeutung für den Gesamtnutzen haben.

---

<sup>5</sup> Kompositionelle Verfahren: Erhebung der Nutzenwerte für einzelne Eigenschaften (Merkmal für Merkmal) und anschließende Verknüpfung der Einzelmerkmale zu einem Gesamturteil.  
Dekompositionelle Verfahren: Erhebung von Gesamtnutzenurteilen für Objekte (Präferenzurteile) zur Bestimmung der Nutzenwerte der Einzelkomponenten.

Neben der Definition der Eigenschaften muss in einem 2. Schritt für jede Objekteigenschaft ein geeignetes **Präferenzmodell** gewählt werden. Unter einem Präferenzmodell versteht man die Art des Zusammenhangs zwischen den Rangdaten und Ausprägungen einer Eigenschaft (z.B. linear, quadratische Zusammenhänge mit Ideal- oder Antiidealpunkt).

Dies kann u.U. problematisch sein, da scheinbar offensichtliche Zusammenhänge (je höher der Preis, desto uninteressanter ist das Objekt) häufig durch andere Prozesse überlagert werden. Ein sehr niedrigerer Preis suggeriert bspw. eine minderwertige Qualität eines Produkts.

Bisher wurden also zwei Schritte behandelt:

- [Definition der Eigenschaften](#)
- [Wahl eines geeigneten Präferenzmodells für jede Eigenschaft](#)

Die Wahl des Präferenzmodells beeinflusst wiederum die Wahl der Eigenschaftsausprägungen, die in der Untersuchung verwendet werden sollen (dritter Schritt). Alle drei Kriterien wirken sich deutlich auf die Validität der Ergebnisse aus.

## **5.2. Definition der Urteilsobjekte**

Es werden zwei Methoden behandelt: Profilmethode und Trade-Off-Analyse.

### **5.2.1. Profilmethode**

Die [Profilmethode](#) geht davon aus, dass ein Objekt aus der Kombination je einer Ausprägung aller Eigenschaften besteht. Dabei werden alle möglichen Objekte berücksichtigt.

Sollen 3 Eigenschaften ABC, A mit 3, B und C mit 2 Ausprägungen, verglichen werden, ergeben sich nach der Profilmethode  $2 \times 2 \times 3 = 12$  Objekte.

Aus dem Plan aller möglichen Eigenschaftskombinationen werden nun die besten Repräsentanten extrahiert, um den Erhebungsaufwand zu verringern. Diese werden jedem Probanden vorgegeben.

Die Profilmethode wird dabei als realitätsnäher als die Trade-Off-Analyse angesehen, da nur hier ganze Produkte und nicht isolierte Eigenschaften miteinander verglichen werden. Dabei ist jedoch unwahrscheinlich, dass wirklich alle Eigenschaften vom Probanden realisiert werden – gerade bei vielen Faktoren ist es wahrscheinlicher, dass nur die wichtigsten überhaupt beachtet werden.

### **5.2.2. Trade-Off-Analyse**

Die [Trade-Off-Analyse](#) wird auch als „[2-factors-at-a-time-evaluation](#)“ bezeichnet. Zur Bildung der Objekte werden jeweils nur zwei Eigenschaften (Faktoren) herangezogen. Für jedes Eigenschaftspaar wird eine sog. Trade-Off-Matrix erstellt, die die Kombinationen der Ausprägungen der beiden Eigenschaften enthält.

N Eigenschaften führen also zu  $\binom{n}{2}$  Trade-Off-Matrizen:



	B1	B2
A1		
A2		
A3		

	C1	C2
A1		
A2		
A3		

	C1	C2
B1		
B2		

Jede Zelle der Matrix entspricht einem zu beurteilenden Objekt. Dadurch können pro Durchgang weniger Objekte verwendet werden, was die Probanden entlastet. Allerdings wird das Objekt hierbei wieder künstlich zergliedert. Auch steigt die Zahl der zu beurteilenden Objekte bei der Profilmethode deutlich schneller an als bei der Trade-Off-Analyse.

Allerdings kann es vorkommen, dass andere Faktoren das eigentlich einfache Urteil beeinflussen. Bspw. könnte der die Ausprägung „hohe Fahrleistung“ in einem Durchgang aus antizipierten Kostengründen vermieden werden, obwohl die Kosten nicht zu beurteilen sind.

### 5.2.3. Symmetrische und asymmetrische Designs

Bei **symmetrischen Designs** haben alle Eigenschaften die gleiche Anzahl von Ausprägungen. Im vollständigen Design für 5 Eigenschaften mit je 3 Ausprägungen wären  $3^5 = 243$  Objekte in eine Rangreihe zu bringen.

Es wird daher nach einer Teilmenge von Objekten gesucht, die das vollständige Design möglichst gut repräsentiert (**reduziertes Design**). In der Regel trifft man hier eine systematische Auswahl.

Spezialfall: Lateinisches Quadrat (siehe F19).

Für alle symmetrischen Designs existieren Tafeln für reduzierte Pläne, die auch in SPSS integriert sind. Diese helfen bei der Auswahl der Objekte.

Liegt ein **asymmetrisches Design** vor (nicht alle Eigenschaften haben die gleiche Anzahl von Ausprägungen), werden die fehlenden Zellen mit Dummies aufgefüllt: Statt der nicht-existent Ausprägung 3 wird für binäre Eigenschaften je zu 50% 1 oder 2 gesetzt. Dadurch lassen sich die Tafeln für symmetrische Designs verwenden.

### 5.2.4. Produktkarten

Die Objekte des reduzierten Designs werden in Produktkarten<sup>6</sup> organisiert. Neben den eigentlichen Objekten der Conjoint-Analyse können zusätzliche Objekte in Form von Holdout- und Simulationskarten eingeführt werden.

**Holdout-Karten** gehen nicht mit die Conjoint-Analyse ein, werden aber trotzdem bewertet. Sie dienen zur Validierung der gefundenen Teilnutzenwerte.

Besonders interessierende Objekte könnten auf **Simulationskarten** dargestellt werden. Sie gehen nicht in die Conjoint-Analyse ein und werden von den Probanden auch nicht bewertet; es wird jedoch ein Gesamtnutzen für diese Objekte berechnet. Dies ist besonders für die Marktforschung relevant.

---

<sup>6</sup> Die in der Vorlesung gezeigten Karten entstammen dem Programm „Alaska“ von Hogrefe, mit dem die Interviews vorbereitet und schöner durchgeführt werden können.

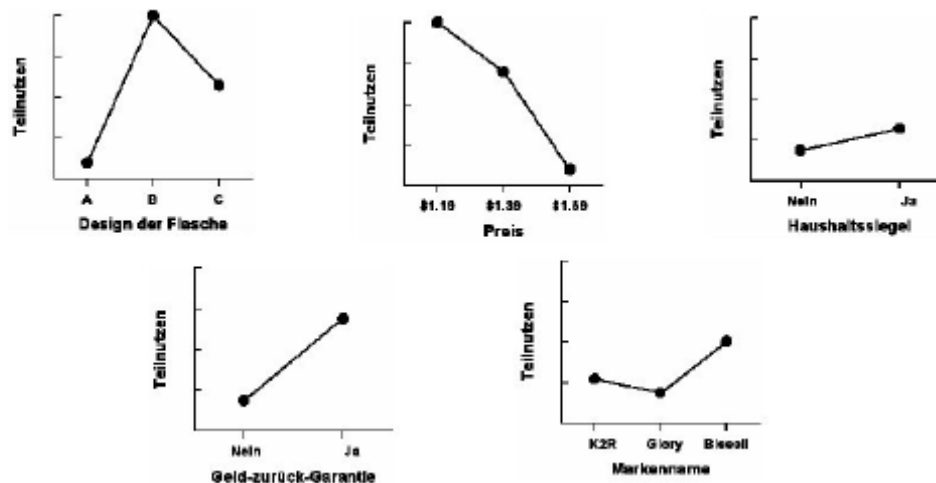
## **5.3. Darbietung der Objekte**

Die Darbietung der Urteilsobjekte ist abhängig vom Produkt bzw. den zu untersuchenden Eigenschaften. Möglich sind bspw. Abbildungen, Klangbeispiele, verbale Beschreibungen oder Geschmackstests.

Die Art der Skalierung ist weitgehend frei, solange eine Rangreihe der Präferenzen des Probanden (Nutzenvorstellungen) resultiert: direkte Rangordnung oder schrittweise rangordnen (Grobeinteilung → Rangordnung innerhalb der Gruppen). Ebenfalls denkbar sind Paarvergleiche. Wichtig ist für die Conjoint-Analyse, dass definiert ist, ob die beste oder schlechteste Kombination den Rang 1 erhält.

Beispiel: Untersuchung von Green und Wind (1975) zu Präferenzen bzgl. 18 Fleckenentfernern, die zur Rangordnung vorgelegt wurden. Dieses Beispiel findet sich in allen Lehrbüchern zur Conjoint-Analyse.

Die 18 Fleckenentferner-Kombinationen wurden aus 108 möglichen Objekten ermittelt, die aus der Kombination von 5 Eigenschaften resultieren. Dargestellt sind die Ergebnisse für den Teilnutzen der 5 Faktoren:



Hier sind sowohl Niveau (Kurvenhöhe) als auch Profil (Verlauf) relevant. Für das Flaschendesign (1. Faktor) zeigt sich beispielsweise eine hohe Differenz zwischen A und B, wobei B einen extrem hohen Teilnutzen bringt. Der Faktor ist also äußerst wichtig und seine Abstufungen sehr relevant.

Das Haushaltssiegel (3. Faktor) zeigt ein sehr niedriges Niveau und einen flachen Verlauf. Scheinbar ist dieser Faktor also irrelevant.

Weitere Interpretationen finden sich auf F28.

Insgesamt gilt: Das Niveau gibt an, wie viel die Eigenschaft zum Gesamtnutzen beiträgt, das Profil enthält Informationen darüber, was durch eine Änderung der Ausprägungen erreicht werden kann.

## **5.4. Berechnung**

Die Bestimmung und Verrechnung der einzelnen Teilnutzenwerte ist auf Ordinal- und Intervallniveau möglich. Die Nutzenwerteschätzung ist ebenfalls mit Alaska möglich.

### 5.4.1. Metrische Lösung: Kleinste Quadrate

Für eine metrische Lösung wird das Prinzip der KQ-Schätzer verwendet. Annahme: Die Abstände zwischen den vergebenen Rangwerten sind äquidistant, d.h. den empirischen Rangwerten wird nicht mehr Ordinal- sondern Intervallniveau unterstellt. Das Modell wird ergänzt durch den Durchschnittsrang  $\mu$  (**Basisnutzen**) der vergebenen Rangwerte.

$$y = \mu + \beta_1 A + \beta_2 B$$

Die metrische Lösung wird nun in 3 Schritten errechnet:

- 1.) Ermitteln des Durchschnittsrangs  $\mu = \frac{\sum(Ränge)}{Anzahl(Ränge)}$
- 2.) Ermitteln der Teilnutzenwerte für jede Eigenschaftsausprägung: Differenz zu  $\mu$
- 3.) Die Teilnutzenwerte  $\beta$  sind KQ-Schätzungen, also eine Minimierung der Summe der quadrierten Abweichungen zwischen empirischen und geschätzten Nutzenwerten.

Beispiel für eine Analyse mit zwei Merkmalen (A mit 4 Ausprägungen, B mit 3 Ausprägungen):

		B			Zeilenmittel	Teilnutzen
		1	2	3		
A	1	2	1	3	2	-4,5
	2	4	5	6	5	-1,5
	3	7	11	10	9,33	2,83
	4	12	8	9	9,67	3,17
Spaltenmittel		6,25	6,25	7	$\mu = 6.5$	
Teilnutzen		-0,25	-0,25	0,5		

Aus der Tabelle geht hervor, dass Eigenschaft A nicht nur einen insgesamt höheren Teilnutzen erzielen kann, sondern dass die unterschiedlichen Ausprägungen von A auch einen stärkeren Effekt haben als die Ausprägungen von B. Für die einzelnen Objekte ergibt sich folgende Tabelle:

Objekt		p	Basisnutzen	Teilnutzen A	Teilnutzen B	y	(p-y)	(p-y) <sup>2</sup>
I	A1 B1	2	6.5	-4.5	-0.25	1.75	0.25	0.06
II	A1 B2	1	6.5	-4.5	-0.25	1.75	-0.75	0.56
III	A1 B3	3	6.5	-4.5	0.5	2.5	0.5	0.25
IV	A2 B1	4	6.5	-1.5	-0.25	4.75	-0.75	0.56
V	A2 B2	5	6.5	-1.5	-0.25	4.75	0.25	0.06
VI	A2 B3	6	6.5	-1.5	0.5	5.5	0.5	0.25
VII	A3 B1	7	6.5	2.83	-0.25	9.08	-2.08	4.34
VIII	A3 B2	11	6.5	2.83	-0.25	9.08	1.92	3.67
IX	A3 B3	10	6.5	2.83	0.5	9.83	0.17	0.03
X	A4 B1	12	6.5	3.17	-0.25	9.42	2.58	6.67
XI	A4 B2	8	6.5	3.17	-0.25	9.42	-1.42	2.01
XII	A4 B3	9	6.5	3.17	0.5	10.17	-1.17	1.36
S		78				78	0	19.83

mit  $y$  = Errechneter Gesamtnutzen (Basisnutzen +  $\Sigma$ (Teilnutzen)) und  $p$  = Tatsächlicher subjektiver Nutzen.  $\Sigma(p-y)^2$  sollte also minimal sein, da dies für eine optimale Schätzung spricht. Entsprechende Programme geben sofort die optimale Lösung aus.

### 5.4.2. Nichtmetrische Lösung: Monotone ANOVA

Die nichtmetrische Lösung basiert auf einer monotonen Varianzanalyse. Annahme: Die vergebenen Rangwerte sind ordinalskaliert. Zur Schätzung der Teilnutzenwerte wird daher das Verfahren der **monotonen Varianzanalyse** nach Kruskal (1965) verwendet. Hierbei handelt es sich um ein iteratives Verfahren, das mit hohem Rechenaufwand verbunden ist.

Dieses Verfahren ist in etwa vergleichbar mit den partitionierenden Verfahren, die im Kontext der Clusteranalyse kurz behandelt wurden (siehe 4.4.2).

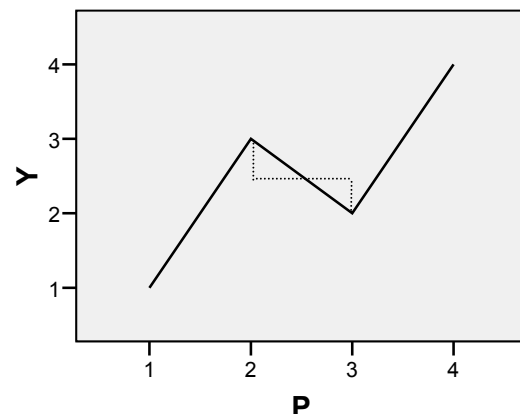
Der Iterationsprozess startet mit einer metrischen Lösung. Anschließend werden mit einem Zufallsgenerator unterschiedliche Ausgangslösungen generiert, wobei das sog. **STRESS-Maß** optimiert werden soll:

$$\underset{f_M}{\text{Min}} \underset{\beta}{\text{Min}} \text{ STRESS} = \sqrt{\frac{\sum_{k=1}^K (z_k - y_k)^2}{\sum_{k=1}^K (y_k - \bar{y})^2}}$$

Die Bedeutung des STRESS-Maßes lässt sich folgendermaßen illustrieren:

Abgebildet ist das Ergebnis einer Conjoint-Analyse. Der zweite tatsächliche Rang trifft mit dem dritten geschätzten Rang zusammen. Die Kurve sollte allerdings bei einer adäquaten Schätzung niemals fallen (keine negative Steigung).

Um die negative Steigung zu eliminieren, muss der Wert für  $p = 2$  abgesenkt, der Wert für  $p = 3$  angehoben werden (Mittelwertsbildung). Wie viel angehoben bzw. abgesenkt werden muss drückt sich im STRESS-Maß aus.



Problem: Das iterative Verfahren kann – abhängig von der Ausgangslösung – gegen suboptimale Lösungen (lokale Minima) konvergieren.

### 5.4.3. Aggregation von Nutzenwerten

Die Conjoint-Analyse ermittelt zunächst Schätzungen individueller Nutzenwerte. Da aber meist die Nutzenstruktur einer Gruppe von Personen interessiert, müssen die individuellen Nutzenwerte aggregiert werden.

Hierzu gibt es zwei Möglichkeiten:

- **Individualanalysen und anschließende Aggregation der Teilnutzenwerte**
- **Gemeinsame Conjoint-Analyse für mehrere Beurteiler**

Voraussetzung für die Individualanalyse ist die Normierung der individuellen Teilnutzenwerte. Der niedrigste Teilnutzen sollte 0 betragen. Bei der normalen Berechnung (siehe Beispiel 5.4.1) können abhängig vom Urteilsverhalten unterschiedliche minimale Teilnutzenwerte entstehen.

Die Normierung sollte also über die Subtraktion der niedrigsten Wertes erfolgen:  $X_{i\_normiert} = X_i - \min(\text{Teilnutzen})$ .

Dann werden die Ergebnisse aller Beurteiler gleichzeitig zur Schätzung der Teilnutzenwerte herangezogen. Die Aggregation der Nutzenwerte über mehrere Personen setzt eine hinreichende Homogenität der individuellen Teilnutzenwerte voraus (überprüfbar mit Hilfe der Clusteranalyse).

Bei Heterogenität: Durchführung von „Individualanalysen“ in jeder Gruppe (segmentspezifische Analysen).

Die Berechnung von Individualanalysen lässt sich also auf folgende Schritte zusammenfassen:

- 1.) Normierung auf den gleichen Nullpunkt.
- 2.) Die Eigenschaftsausprägung mit dem kleinsten Teilnutzenwert wird auf 0 gesetzt durch Berechnung der Differenzen Teilnutzenwerte –  $\min(\text{Teilnutzenwerte})$ .

	Teilnutzenwert	min (Teilnutzen)	transformierter Teilnutzen
A1	-4.5	-4.5	0
A2	-1.5	-4.5	3
A3	2.83	-4.5	7.33
A4	3.17	-4.5	7.67
B1	-0.25	-0.25	0
B2	-0.25	-0.25	0
B3	0.5	-0.25	0.75

## **5.5. Interpretation der Teilnutzenwerte**

Die Höhe der Teilnutzenwerte sagt etwas aus über die Bedeutsamkeit einer Eigenschaftsausprägung für den Gesamtnutzenwert eines Objekts. Sie sagt jedoch nichts über die relative Wichtigkeit einer Eigenschaft zur Präferenzveränderung aus.

Haben alle Teilnutzenwerte der Eigenschaftsausprägungen einer Eigenschaft hohe Werte, so ist diese Eigenschaft also nicht wichtiger als andere Eigenschaften hinsichtlich der Präferenzveränderung (Veränderung des Gesamtnutzenwerts).

Die Spannweite (Range) hingegen sagt aus, ob es sich lohnt, eine Eigenschaft zu variieren.

## **5.6. Zusammenfassung**

Die Conjoint-Analyse ist ein altes Verfahren und findet sich in älteren Darstellungen auch unter dem deutschen Namen „Verbundene Messung“. Dieser Name drückt aus, dass Teilnutzenwerte für einzelne Eigenschaften durch die Messung verbunden werden.

Das Vorgehen lässt sich in folgende Schritte gliedern:

- Bestimmung von Eigenschaften, Präferenzmodell und Ausprägungen
- Festlegung eines Erhebungsdesigns
- Bewertung der Stimuli
- Schätzung der Nutzenwerte
- Aggregation der Nutzenwerte
- Interpretation

Die Conjoint-Analyse ist nicht in SPSS integriert, daher findet sich auch keine Übung dazu. Ein alternatives Programm ist Alaska (Hogrefe).

Anwendungsgebiete sind unter anderem Käuferpräferenzen bei Ackerschleppern, günstige und ungünstige Gelegenheiten zu kriminellen Handlungen, Ermittlung von Präferenzen verschiedener Anspruchsgruppen für die Landschaft in einem Naturschutzgebiet oder die Ermittlung zuschauerorientierter Leistungen im Fußball.

## **6. Faktorenanalyse**

Die Skalierung beschäftigt sich mit dem messen von Objekten bzw. dem messbar machen. Die Faktorenanalyse zählt zum Gebiet der Skalierung, weil mit ihr ermittelt werden kann, wie viele Beschreibungsdimensionen benötigt werden, um eine Menge von Objekten zu beschreiben.

Ziel der Faktorenanalyse ist also die Entdeckung von wenigen (untereinander unabhängigen) Beschreibungsvariablen. Dies kann aus verschiedenen Gründen interessieren: Hypothesengenerierung, Überprüfung der Dimensionalität komplexer Merkmale, Entdeckung untereinander unabhängiger Beschreibungsdimensionen oder auch Datenreduktion.

Die Datenreduktion ist vor allem dann sinnvoll, wenn die gefundenen Dimensionen als Eingangsgrößen für nachfolgende Analysen (z.B. multiple Regression) verwendet werden. Eine multiple Regression mit 70 Items als Prädiktoren ist bspw. wenig sinnvoll; werden dagegen 5 extrahierte Dimensionen verwendet, lässt sich eine multiple Regression durchführen.

Der Ablauf einer Faktorenanalyse gliedert sich immer in die folgenden Schritte:

- 1.) Variablenauswahl und Errechnung der Korrelationsmatrix
- 2.) Extraktion der Faktoren
- 3.) Bestimmung der Kommunalitäten
- 4.) Bestimmung der Zahl der Faktoren
- 5.) Interpretation der Faktoren
- 6.) Bestimmung der Faktorwerte

### **6.1. Schritt 1: Variablenauswahl und Korrelationsmatrix**

#### **6.1.1. Variablenauswahl**

Wichtig ist die Zuverlässigkeit der Ausgangsdaten, da von ihr die Güte der gesamten Faktorenanalyse abhängig ist. Hier ist vor allem wichtig, dass die erhobenen Merkmale bzw. Variablen für den Untersuchungsgegenstand relevant sind und methodisch korrekt erhoben wurden (also: gute Instruktion, übersichtliches Layout...).

Neben der Relevanz der erhobenen Merkmale ist die Unabhängigkeit der Beurteilung eine zwingende Voraussetzung. Zudem müssen mindestens intervallskalierte Variable vorliegen. Die Anzahl der Fälle muss größer als die Anzahl der verwendeten Variablen sein – besser sogar mindestens 3x so groß.

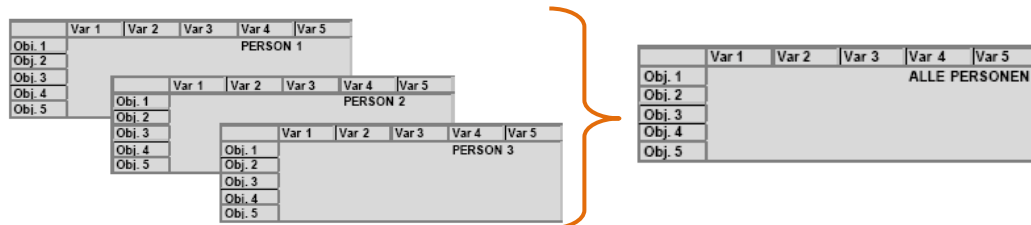
Zusätzlich ist eine möglichst homogene Befragungsstichprobe von Vorteil, da hierdurch hohe Korrelationen zwischen den Variablen entstehen.

#### **6.1.2. Korrelationsmatrix**

Sind die Daten erhoben, wird die reduzierte Objekte-x-Variablen-Matrix in eine Variablen-x-Variablen-Matrix (Korrelationsmatrix) transformiert. Es existieren hierbei einige Prüfkriterien, um zu testen, ob sich die Korrelationsmatrix für eine FA eignet.

### 6.1.2.1. Reduzierte Objekt-x-Variablen-Matrix

Die reduzierte Objekt-x-Variablen-Matrix entsteht aus der Mittelung der Ergebnisse verschiedener Personen über die gesamte Objekt-x-Variablen-Matrix. Gemittelt wird für jede Eigenschaft jedes Objekts über alle Personen.



Die Ausgangsmatrix für eine Faktorenanalyse ist (in SPSS) also immer eine Matrix, in der die beurteilten Objekte in den Zeilen stehen und die Eigenschaften (Variablen) auf denen sie beurteilt wurden in den Spalten. Die Eigenschaften sind dabei das zu gruppierende Merkmal.

Anmerkung: Wird die Persönlichkeit von Probanden mit einem Fragebogen erfasst, so sind die Probanden die Objekte und füllen somit die Zeilen. Eine Reduktion der Objekt-x-Variablen-Matrix ist also nicht notwendig:

	Var1	Var2	Var3
Pb1			
Pb2			
Pb3			

### 6.1.2.2. Güte der Korrelationsmatrix

In SPSS lässt sich nun für die (reduzierte) Objekt-x-Variablen-Matrix eine Korrelationsmatrix (Variablen-x-Variablen-Matrix) berechnen.

Korrelationsmatrix<sup>a</sup>

		Ungesättigte Fettsäuren	Kaloriengehalt	Vitamingehalt	Haltbarkeit	Preis
Korrelation	UngesättigteFettsäuren	1.000	.712	.981	.109	.044
	Kaloriengehalt	.712	1.000	.704	.138	.067
	Vitamingehalt	.981	.704	1.000	.078	.024
	Haltbarkeit	.109	.138	.078	1.000	.983
	Preis	.044	.067	.024	.983	1.000
Signifikanz (1-seitig)	UngesättigteFettsäuren		.058	.001	.419	.467
	Kaloriengehalt	.058		.059	.397	.450
	Vitamingehalt	.001	.059		.441	.482
	Haltbarkeit	.419	.397	.441		.000
	Preis	.467	.450	.482	.000	

a. Determinante = .001

Für diese Korrelationsmatrix existieren wie erwähnt Prüfkriterien bzgl. ihrer Eignung für eine Faktorenanalyse.

- **Signifikanz der Korrelationskoeffizienten.**

Gibt an, mit welcher Wahrscheinlichkeit (1-p) Korrelationen unterschiedlich von 0 sind. Es gibt jedoch keine Hausregel, wie viele Korrelationen prozentual signifikant sein müssen...es müssen aber irgendwelche signifikanten dabei sein ☺.

- **Inverse der Korrelationsmatrix.**



Die Inverse einer Matrix ist diejenige Matrix, mit der man eine Matrix multiplizieren muss, damit die entsprechende Einheitsmatrix entsteht.

$$\text{Einheitsmatrix (3x3):} \quad \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}$$

Die Werte außerhalb der Diagonalen der Inversen der Korrelationsmatrix sollten möglichst nahe bei 0 liegen (die Inverse sollte also der Einheitsmatrix ähneln). Jedoch ist auch hier kein numerisches Kriterium vorhanden.

- **Bartlett-Test / Test of sphericity.**

Der Bartlett-Test ist das erste wichtige Kriterium, da er als Signifikanztest auch ein numerisches Prüfkriterium besitzt. Er sollte daher auf jeden Fall beachtet werden.  $H_0$ : Die Variablen sind unkorreliert; die Korrelationsmatrix weicht nur zufällig von der Einheitsmatrix ab. Voraussetzungen: Normalverteilung der Variablen in der Stichprobe (z.B. über Kolmogorov-Smirnov-Anpassungstest prüfbar).

- **Anti-Image-Kovarianzmatrix.**

Ausgangspunkt dieses Kriteriums ist die Teilung der Varianz einer Variablen in ein Image und ein Anti-Image. Das Image ist derjenige Teil der Varianz einer Variablen, der durch die anderen Variablen aufgeklärt werden kann, das Anti-Image ist derjenige Teil, der unabhängig von den anderen Variablen ist.

Das Anti-Image sollte also möglichst gering sein – die Werte außerhalb der Diagonalen der Anti-Image-Kovariationsmatrix (Kovariationsmatrix zwischen zwei Anti-Images) sollten also nahe 0 liegen. Konvention: Wenn mehr als 25% der Nicht-diagonal-Elemente  $\neq 0$  – praktisch: Betrag  $> 0.09$  – dann ist die Matrix für eine FA ungeeignet.

- **Anti-Image-Korrelationsmatrix und Kaiser-Meyer-Olkin-Kriterium (KMO).**

Das KMO ist eine Maßzahl, die auf Basis der Anti-Image-Korrelations-Matrix berechnet wird. Synonym: „measure of sampling adequacy“ (MSA). Es stellt das beste Gütekriterium dar.

Das KMO/MSA gibt an, in welchem Umfang die Ausgangsvariablen zusammengehören. Es hat einen Wertebereich zwischen 0 und 1. Eine Besonderheit ist, dass das KMO sowohl für die gesamte Korrelationsmatrix als auch für einzelne Variablen berechnet werden kann.

Interpretation des KMO:

< 0.5	untragbar
= 0.5	kläglich
= 0.6	mittelmäßig
= 0.7	ziemlich gut
= 0.8	verdienstvoll
= 0.9	erstaunlich

Es gibt also eine Reihe von Gütekriterien: Signifikanz der Korrelationskoeffizienten, Bartlett-Test, Inverse der Korrelationsmatrix, Anti-Image-Kovarianzmatrix sowie Anti-Image-Korrelationsmatrix und KMO/MSA.

Es sollte dabei immer mehr als ein Prüfkriterium betrachtet werden, da die Güte der Korrelationsmatrix das Ergebnis der FA beeinflusst. In der Realität werden die Kriterien fast nie erreicht. Es sollte aber zumindest eine Intervallskalierung der Merkmale sichergestellt sein.

Empfehlung: Es sollten zumindest 3 Kriterien verwendet werden: Bartlett-Test (sollte signifikant sein), KMO > 0.5 für alle Variablen sowie die Gesamttabelle sowie die Anti-Image-Kovarianzmatrix.

KMO- und Bartlett-Test

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		.578
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat df	17.371 10
	Signifikanz nach Bartlett	.067

Anti-Image-Matrizen

		Ungesättigte Fettsäuren	Kaloriengehalt	Vitamingehalt	Haltbarkeit	Preis
Anti-Image-Kovarianz	Ungesättigte Fettsäuren	.089	-.019	-.065	-.010	.010
	Kaloriengehalt	-.019	.459	-.027	-.026	.025
	Vitamingehalt	-.065	-.027	.071	.009	-.008
	Haltbarkeit	-.010	-.026	.009	.026	-.026
	Preis	.010	.025	-.008	-.026	.027
Anti-Image-Korrelation	Ungesättigte Fettsäuren	.597 <sup>a</sup>	-.108	-.926	-.237	.223
	Kaloriengehalt	-.108	.878 <sup>a</sup>	.160	-.239	.227
	Vitamingehalt	-.926	-.160	.598 <sup>a</sup>	.207	-.193
	Haltbarkeit	-.237	-.239	.207	.471 <sup>a</sup>	-.988
	Preis	.223	.227	-.193	-.988	.467 <sup>a</sup>

a. Maß der Stichprobeneignung

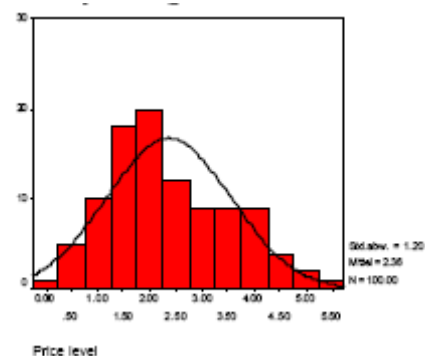
Anmerkung 1: Das KMO für Einzelvariable finde sich in der Diagonale der Anti-Image-Korrelationsmatrix.

Anmerkung 2: Im Beispiel ist das KMO für Einzelvariablen nicht immer > 0.5. Die Korrelationsmatrix ist also nicht so ideal – hier kann man als guter Psychologe allerdings anfangen zu argumentieren, warum man sie trotzdem nimmt 😊.

### 6.1.3. Anmerkung: Prüfung der Rohdaten

Neben der Korrelationsmatrix sollten auch die Ausgangsdaten selbst auf ihre Eignung für eine FA geprüft werden. Die Variablen in der Stichprobe sollten dabei normalverteilt oder mindestens halbwegs gleichartig verteilt sein<sup>7</sup>.

Hier ist eine graphische Darstellung mithilfe eines Histogramms hilfreich. Die Normalverteilungsannahme kann über einen Anpassungstest geprüft werden.



<sup>7</sup> Der Bartlett-Test benötigt eine Normalverteilung, was sich über einen Kolmogorov-Smirnov-Anpassungstest prüfen lässt, der natürlich nicht signifikant werden sollte. Dieser Test findet sich in SPSS unter *Analysieren/Nonparametrisch/1...* Die FA selbst benötigt jedoch keine Normalverteilung.

## 6.2. Schritt 2: Extraktion der Faktoren

Nach dem Fundamentalsatz von Thurstone gilt, dass sich die beobachteten Werte der Personen in den Variablen als Linearkombination mehrerer nicht beobachtbarer Variablen beschreiben lassen. Die empirische Korrelationsmatrix ist durch das Faktorenmuster reproduzierbar.

Mit  $k$  Objekten,  $i$  Merkmalen und  $q$  Faktoren gilt:

$$X_{ik} = a_{i1} p_{1k} + a_{i2} p_{2k} + \dots + a_{iq} p_{qk} \quad (+ \text{Rest})$$

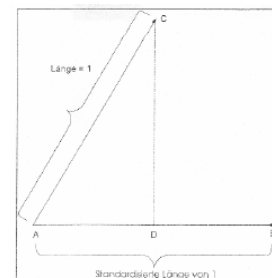
Die FA unterstellt also ein additives Modell für das Zustandekommen der Messwerte. Unter der Faktorladung  $a_{iq}$  versteht man die Korrelation zwischen Faktor und Variable. Der Faktorwert  $p_{qk}$  bezeichnet die Ausprägung eines Faktors bei einem Objekt.

### 6.2.1. Graphische Darstellung

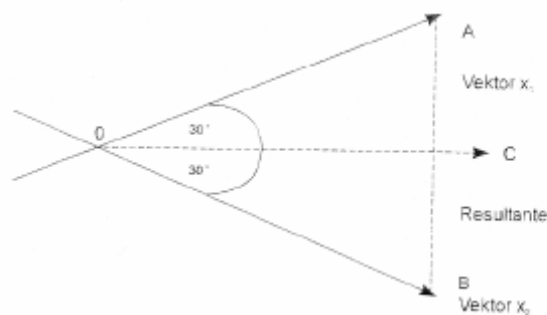
Graphisch lassen sich Variablen als Punkte im  $n$ -dimensionalen Versuchspersonenraum darstellen. Der Vektor zwischen Ursprung und Variablenpunkt wird als Variablenvektor bezeichnet.

Da die standardisierte Länge eines Variablenvektors 1 beträgt, lässt sich eine Korrelation im graphischen Fall als  $\cos(\alpha)$  darstellen mit  $\alpha$  = Winkel zwischen zwei Vektoren.

Zur Abbildung: Da  $AC = AB = 1$  entspricht der Kosinus (= Ankathete / Hypotenuse) des Winkels zwischen  $AC$  und  $AB$  der Strecke  $AD$ <sup>8</sup>.



Die Extraktion des ersten Faktors verläuft nun durch Bildung des Summenvektors aller Variablenvektoren. „Man bringe an einem Kasten Bier Vektoreisenle der standardisierten Länge 1 an und ziehe an allen mit gleicher Kraft.“ Der erste Faktor entsteht also als Linearkombination (Resultante) der Vektoren.



Im orthogonalen Fall wird der zweite Faktor rechtwinklig zum ersten Faktor durch den Ursprung des Variablenbündels gelegt. Die nachfolgenden Faktoren werden ebenfalls orthogonal zu den jeweils vorhandenen Faktoren erstellt.

Die so gefundene Lösung muss nicht optimal sein, da die extrahierten Faktoren zur einfacheren Interpretation später rotiert werden können.

<sup>8</sup> Anmerkung zu F21: Mit  $\sphericalangle X_1 X_2 + \sphericalangle X_2 X_3 = \sphericalangle X_1 X_3$  lassen sich die drei Vektoren im zweidimensionalen Raum darstellen. Das muss natürlich nicht immer so sein.

## **6.3. Schritt 3: Bestimmung der Kommunalitäten**

Die Kommunalität gibt den Umfang der Varianzaufklärung an, den die Faktoren gemeinsam für eine Ausgangsvariable leisten (= durch alle Faktoren aufgeklärter Anteil der Varianz einer Variablen).

Die Kommunalitäten liefern unter anderem Hinweise darauf, welche Variablen aus der FA ausgeschlossen werden sollen.

Das Kommunalitätenproblem der FA besteht darin, die Anfangskommunalität zu schätzen. Dies ist meist ein subjektives Vorab-Urteil des Anwenders. Die Bestimmung der Anfangskommunalitäten vor der Faktorenextraktion ist dabei unmittelbar an die Methode der Faktorenermittlung gekoppelt. Die Bestimmung der Kommunalitäten hängt also eng mit der Extraktion der Faktoren zusammen.

Zur Bestimmung der Kommunalitäten stehen zahlreiche Verfahren zur Verfügung, von denen jedoch nur zwei näher betrachtet werden sollen:

- **Hauptkomponentenanalyse**
- **Hauptachsenanalyse**
- Image-Analyse
- Alpha-Faktorenanalyse
- Ungewichtete Least-Square
- Verallgemeinerte Least-Square
- ML

Die verschiedenen Extraktionsmethoden führen i.d.R. zu ähnlichen Ergebnissen. Welche der verschiedenen Methoden verwendet werden sollte ist an inhaltliche Überlegungen geknüpft („Was ist das Ziel meiner Analyse?“).

### **6.3.1. Hauptkomponentenanalyse**

Die Anfangskommunalität wird auf 1 gesetzt, d.h. der Anwender geht davon aus, dass die Varianz der Ausgangsvariablen vollständig durch die Extraktion der Faktoren erklärt werden kann. Die Kommunalität von 1 wird dann vollständig reproduziert, wenn ebenso viele Faktoren wie Variablen extrahiert werden.

Das Ziel der Hauptkomponentenanalyse ist dabei die **Reproduktion der Datenstruktur** und keine kausale Interpretation der Faktoren.

Interpretation: „Wie lassen sich die auf einem Faktor hoch ladenden Variablen durch einen Sammelbegriff zusammenfassen?“

### **6.3.2. Hauptachsenanalyse**

Die Anfangskommunalitäten werden auf einen Wert  $< 1$  gesetzt. Der Anwender geht also davon aus, dass die Varianz der Ausgangsvariablen aus der Kommunalität *und* Einzelrestvarianz (Spezifität) besteht.

Der verwendete Schätzwert entspricht meist (u. a. auch in SPSS) dem quadrierten multiplen Regressionskoeffizienten  $R^2$ .

Das Ziel der Hauptachsenanalyse ist dabei eine **kausale Interpretation** der Faktoren.

Interpretation: „Wie lässt sich die Ursache bezeichnen, die für die hohen Ladungen der Variablen auf diesem Faktor verantwortlich ist?“ bzw. „Wie lässt sich eine der Variablenstruktur zugrunde liegende Ursache bezeichnen?“

### 6.3.3. SPSS-Output

Kommunalitäten

	Anfänglich	Extraktion
UngesättigteFettsäuren	.931	.968
Kaloriengehalt	.541	.526
Vitamingehalt	.929	.953
Haltbarkeit	.974	.991
Preis	.973	.981

Extraktionsmethode: Hauptachsen-Faktorenanalyse

Reproduzierte Korrelationen

	Ungesättigte Fettsäuren	Kaloriengehalt	Vitamingehalt	Haltbarkeit	Preis
Reproduzierte Korrelation UngesättigteFettsäuren	.968 <sup>a</sup>	.712	.960	.110	.042
Kaloriengehalt	.712	.526 <sup>b</sup>	.705	.127	.077
Vitamingehalt	.960	.705	.953 <sup>b</sup>	.085	.017
Haltbarkeit	.110	.127	.085	.991 <sup>b</sup>	.983
Preis	.042	.077	.017	.983	.981 <sup>b</sup>
Residuum <sup>a</sup> UngesättigteFettsäuren	.000	.000	.001	-.001	.001
Kaloriengehalt	.000	.000	-.001	.011	-.011
Vitamingehalt	.001	-.001	.000	-.006	.006
Haltbarkeit	-.001	.011	-.006	.954E-005	9.54E-005
Preis	.001	-.011	.006	9.54E-005	9.54E-005

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

<sup>a</sup>. Residuen werden zwischen beobachteten und reproduzierten Korrelationen berechnet. Es liegen 0 (.0%) nicht redundante Residuen mit absoluten Werten größer 0,05 vor.

<sup>b</sup>. Reproduzierte Kommunalitäten

Die Diagonale der Reproduzierten Korrelationen entspricht den reproduzierten Kommunalitäten.

## 6.4. Schritt 4: Zahl der Faktoren

Die Faktorenextraktion liefert zunächst so viele Faktoren wie Variablen. Das Ziel ist es nun, möglichst wenige Faktoren (Dimensionen) zu finden, die die Zusammenhänge der Variablen möglichst gut erklären. Das Kriterium ist hierbei immer der Anteil der erklärten Varianz an der Gesamtvarianz. Die Gesamtvarianz ist immer gleich der Anzahl der Variablen, weil diese jeweils eine standardisierte Varianz von 1 aufweisen.

Hierfür können 3 Kennwerte herangezogen werden:

- **Eigenwert:** Kumulierte Varianzaufklärung eines Faktors  $\sum_i a_{iq}^2$  („Wie gut repräsentiert der Faktor die Variablen?“).
- **Kommunalität:** Kumulierte aufgeklärte Varianz einer Variablen  $\sum_q a_{iq}^2$  („Wie gut wird diese Variable von allen Faktoren repräsentiert?“).
- **Faktorladung:** Korrelation zwischen Variable und Faktor.

Unter **Faktorenproblem** versteht man nun die Frage nach der Anzahl der zu extrahierenden Faktoren. Grundsätzlich sollten selbstverständlich weniger Faktoren als Variablen extrahiert werden. Allerdings bedeutet eine Verringerung der Anzahl der Faktoren auch immer einen größeren Anteil nicht aufgeklärter Varianz.

Die Zahl der Faktoren lässt sich daher entweder durch Vorgabe anhand theoretischer Überlegungen bestimmen oder aber durch die sukzessive Extraktion der Faktoren. Letztlich bleibt die Wahl jedoch immer subjektiv.

Für die Methode der sukzessiven Extraktion lassen sich aber zumindest zwei Heuristiken angeben:

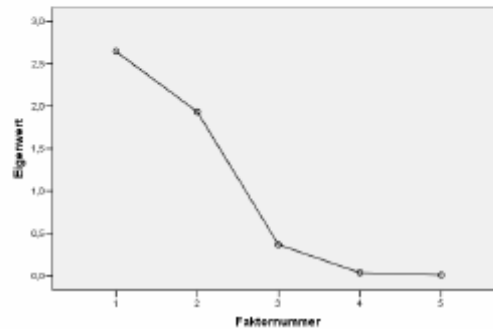
- **Kaiser-Kriterium:** Alle Faktoren mit Eigenwerten  $> 1$  (alle Faktoren die mehr Varianz aufklären als eine einzelne Variable).
- **Scree-Kriterium (Geröll-Kriterium):** Im **Scree-Plot** der Eigenwerte wird als letzter Faktor derjenige extrahiert, der links vom Knick ist.

SPSS-Output:

Faktor	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2.645	52.903	52.903	2.507	50.138	50.138	2.447	48.946	48.946
2	1.934	38.678	91.581	1.911	38.225	88.363	1.971	39.417	88.363
3	.369	7.374	98.955						
4	.039	.786	99.741						
5	.013	.259	100.000						

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

Screeplot



Warum sind die anfänglichen Eigenwerte größer als die Eigenwerte in der zweiten Spalte (Summen von quadrierten Faktorladungen)? Das liegt daran, dass die linke Spalte *immer* als Hauptkomponentenanalyse gerechnet wird (Anfangskommunalität = 1), unabhängig davon, was in SPSS als Methode eingestellt ist.

## **6.5. Schritt 5: Faktoreninterpretation**

Nach der Faktorenextraktion müssen die zunächst abstrakten Faktoren interpretiert werden. Dies geschieht auf Basis der auf ihnen hoch ladenden Variablen.

Zur einfacheren Interpretation können die Faktoren im Raum gedreht werden, um das Kriterium der Einfachstruktur nach Thurstone zu erfüllen. Hier stellt sich jedoch das **Rotationsproblem**: Nach welchen Kriterien soll rotiert werden?

Auch hier gibt es verschiedene Verfahren, die jedoch nicht näher behandelt werden:

- **VARIMAX**
- **QUARTIMAX**
- EQUAMAX
- BIQUARTIMAX
- Schiefwinklige Rotationen

Per Konvention gelten alle Ladungen  $> 0.5$  nach der Rotation als groß und dürfen daher interpretiert werden. Allerdings ist auch die Interpretation ein äußerst subjektiver Schritt.

Wenn eine Variable keine Ladung  $> 0.5$  besitzt oder eine Variable auf mehreren Faktoren ähnlich hohe Ladungen besitzt, sollte diese Variable eliminiert werden. Im Anschluss muss die FA neu berechnet werden.

Anmerkung: Ladungen werden nach der Rotation neu berechnet, Eigenwerte haben jedoch nur vor der Rotation eine Bedeutung und dürfen daher auch nur davor berechnet werden (zumindest nach Hommers).

Für ein Beispiel zur Rotation und Interpretation sie F36 und F37.

Mit der Interpretation der Faktoren endet die eigentliche Faktorenanalyse. Es ist jedoch noch möglich, die Faktorwerte einzelner Objekte zu bestimmen, um diese auf den gefundenen Dimensionen anzuordnen.



## 6.8. Übung

### 6.8.1. Anwendungsbeispiele

Anwendungsbeispiel I: Emulsionsfette. Siehe Folien.

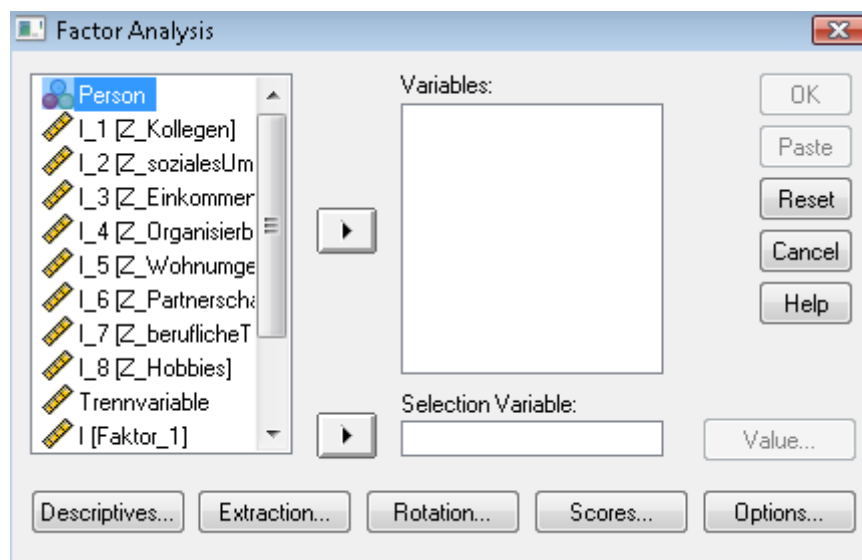
Anwendungsbeispiel II (siehe Folien) zeigt eine alternative Verwendung der Faktorwerte: Neben der Angabe der Lage einer Person im Faktorenraum (Mapping) können die Faktorwerte auch verwendet werden um Beziehungen zwischen Variablen herzustellen.

Anwendungsbeispiel III: Siehe Folien. Wichtig: Hier geht es nicht um Kausalinterpretationen, sondern um Oberbegriffe für die Datenstruktur. Daher wird die Hauptkomponentenanalyse verwendet.

### 6.8.2. Übungsblatt

#### 6.8.2.1. Einstellungen

Die Faktorenanalyse findet sich unter *Analysieren\Dimensionsreduktion*. Im Auswahlfeld für die einzubeziehenden Variablen werden String-Variablen gar nicht erst angezeigt. Nur numerische Variablen können einbezogen werden und daher können auch nur diese ausgewählt werden.



**Deskriptive Statistiken:** Schön sind die deskriptiven Statistiken für Einzelvariablen. Ansonsten sind alle anderen Punkte ebenfalls relevant, lediglich die Inverse der Korrelationsmatrix (aufgrund des fehlenden Kriteriums hierfür) sowie die Determinante („Keine Ahnung was die macht.“) sind nicht unbedingt zu verwenden.

Wird die Anti-Image-Kovariationsmatrix ausgegeben, kann man in dieser zusätzlich die KMO-Werte für Einzelvariable einsehen. Sehr praktisch.

**Extraktion:** Hier lässt sich die zu verwendende Methode zur Bestimmung der Kommunalitäten einstellen (Hauptkomponenten- vs. Hauptachsenanalyse, s. 6.3). Auch sollte der Scree-Plot mit ausgegeben werden.

**Rotation:** Klar.



**Faktorwerte (Scores):** Regression ist die Standardmethode.

**Optionen:** Die Option „Sortiert nach Größe“ sollte zur besseren Lesbarkeit der Ladungsmatrizen aktiviert werden. Auch die Option „Niedrige Ladungen unterdrücken“ (z.B.  $< .4$ ) sorgt für mehr Übersichtlichkeit.

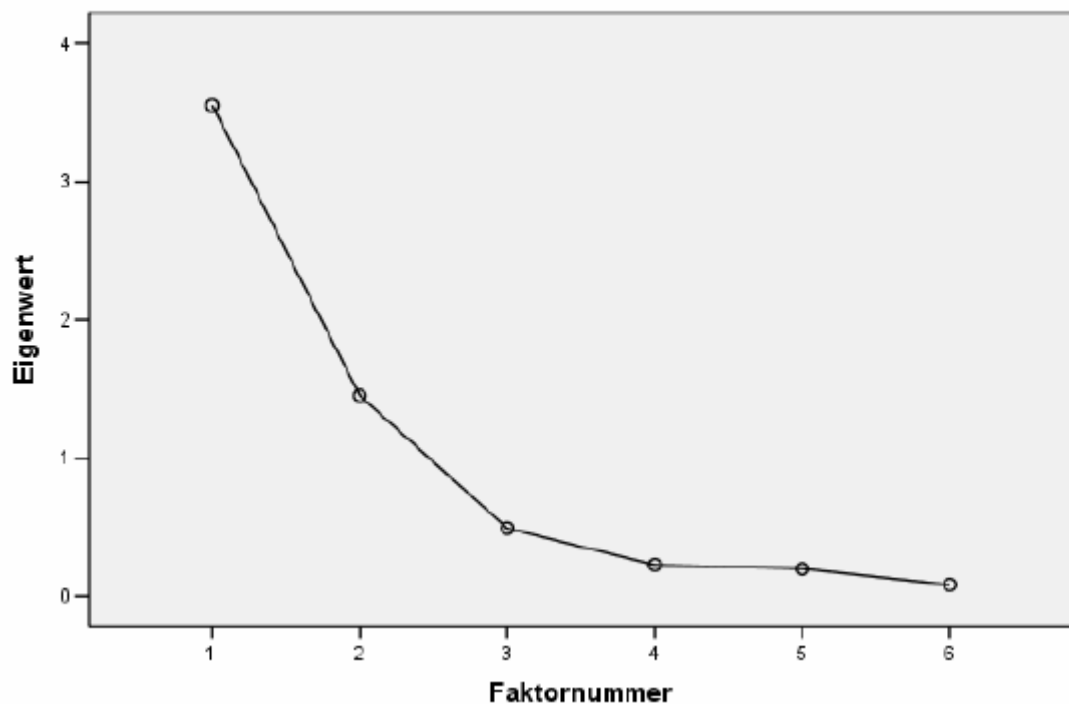
### 6.8.2.2. Güte der Korrelationsmatrix

Bei der Betrachtung der **Anti-Image-Kovariationsmatrix** sollte natürlich nur eine der symmetrischen Hälften betrachtet werden, wenn die Anzahl der von Null verschiedenen ( $> 0.09$ ) Nicht-diagonal-Elemente ermittelt wird.

**Scree-Plot:** Der Knick beginnt dort „wo es flach wird“. Dort ist kein großer Unterschied mehr zwischen den Eigenwerten aufeinanderfolgender Faktoren zu finden. Diese sind sich demnach ähnlich und nicht mehr so interessant.

Im Beispiel beginnt der Knick bspw. bei Faktor Nr. 3. Es würden also zwei Faktoren extrahiert werden.

Screeplot



### 6.8.2.3. Ergebnisdarstellung

Tabelle: „**Total Variance Explained**“. Die anfänglichen Eigenwerte werden immer durch eine Hauptkomponentenanalyse ermittelt. Das Ergebnis der Hauptachsenanalyse steht in der mittleren Spalte. Dabei entspricht die Unterspalte „Kumulierte %“ der Varianzaufklärung in der durchgeführten Analyse.

In der rechten Spalte stehen die rotierten „Eigenwerte“, welche sich durch die Rotation ändern [Anmerkung: Laut Hommers haben Eigenwerte nach der Rotation keine Bedeutung mehr!]. Die gesamte Varianzaufklärung bleibt natürlich gleich.

Auch interessant ist die Tabelle der **reproduzierten Kommunalitäten** für Einzelvariable. Hier sieht man, welche Variable gut durch die FA erfasst wird und welche nur unzureichend erklärt werden kann. Diese Tabelle wird bei vielen Variablen jedoch sehr unübersichtlich.

Toll: Die reproduzierten Kommunalitäten lassen sich aus den Faktorladungen berechnen. Dafür werden die Ladungen der einzelnen Variablen auf einem Faktor quadriert und aufsummiert ( $\sum a^2$ ).

#### 6.8.2.4. Interpretation

Zu **Beginn** sollte immer erst genannt werden, welcher Anteil der Varianz durch die Faktorenanalyse aufgeklärt werden kann.

Die Interpretation der **Faktoren** bezieht sich immer auf die rotierte Ladungsmatrix. Dabei sollten alle Ladungen  $>.5$  interpretiert werden. Weißt eine Variable auf mehreren Faktoren entsprechende Ladungen auf, so muss sie natürlich für die Interpretation beider Faktoren berücksichtigt werden. Werden 5 Faktoren gefunden, so spricht man auch von einer „5-Faktor-Struktur“ der Daten.

Die Faktorwerte lassen sich ebenfalls gut interpretieren, da sie auf  $\mu = 0$  und  $\sigma \approx 1$  normiert sind. Dafür gibt es in SPSS unterschiedliche Methoden, wobei manche  $\sigma = 1$  gewährleisten und andere nur  $\sigma \approx 1$  ausgeben.

#### 6.8.2.5. Graphische Darstellung

Es geht auch mit dem einfachen Streudiagramm unter den veralteten Methoden. Hier muss die Fallbeschriftung sowohl in die entsprechende Zelle gezogen werden, als auch unter Optionen zusätzlich aktiviert werden.

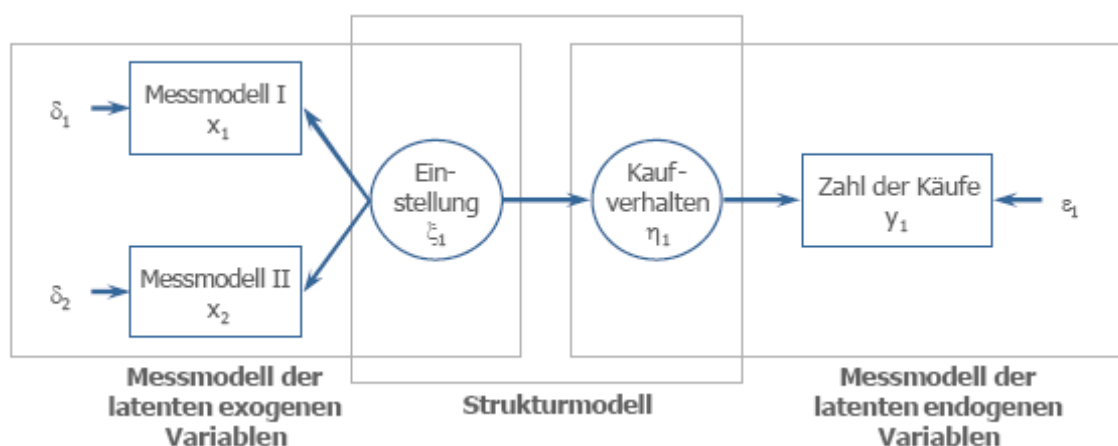
## 7. Lineare Strukturmodelle

Andere Bezeichnungen der linearen Strukturmodelle sind Kovarianzstrukturanalyse, LISREL oder Ähnliches. Der prominenteste Ansatz ist das lineare Modell LISREL (LInear Structural RELationship) von Jöreskog und Sörbom. Die Modelle sind insbesondere in der pädagogischen Psychologie äußerst beliebt und finden sich in zahlreichen Publikationen (SEM = structural equation model).

Ziel der linearen Strukturmodelle ist die Kausalanalyse, also die Untersuchung kausaler Abhängigkeiten, zwischen nicht experimentell erfassten Variablen. Hierbei geht die explizite Struktur von UVn und AVn verloren: Abhängige Variablen können selbst wiederum unabhängige Variablen für weitere Variablen sein.

Das Verfahren hat einen konfirmatorischen, also hypothesenprüfenden Charakter. Dabei werden auch Zusammenhänge zwischen latenten Variablen untersucht.

Ein LISREL-Modell besteht immer aus Struktur- und Messmodell. Das Strukturmodell beschreibt die Abhängigkeitsstruktur der latenten Variablen, während die Messmodelle den Zusammenhang der latenten Variablen mit ihren jeweiligen Indikatoren spezifizieren.



**Bezeichnungen:** Durch Kreise symbolisierte Variablen sind latente Variablen (griech. Buchstaben), rechteckig markierte Variablen sind deren manifeste Indikatoren (lateinische Buchstaben). Latente Variablen, von denen nur Pfeile wegzeigen, werden als exogene Variablen bezeichnet. Zeigt ein Pfeil auf eine latente Variable, so wird diese als endogene Variable bezeichnet. Die Indikatoren der exogenen Seite werden mit lateinischem  $x_i$  bezeichnet, die Indikatoren auf endogener Seite mit  $y_i$ . Indikatoren selbst sind natürlich immer endogene Variablen.

Auf allen Ebenen werden spezifische Fehlerterme angenommen ( $\delta$  für Indikatoren der latenten exogenen Variablen,  $\zeta$  (ZETA) für Messfehler der latenten Variablen und  $\varepsilon$  für Messfehler der Indikatoren der latenten endogenen Variablen). Auch die Messfehler sind exogene Variablen.

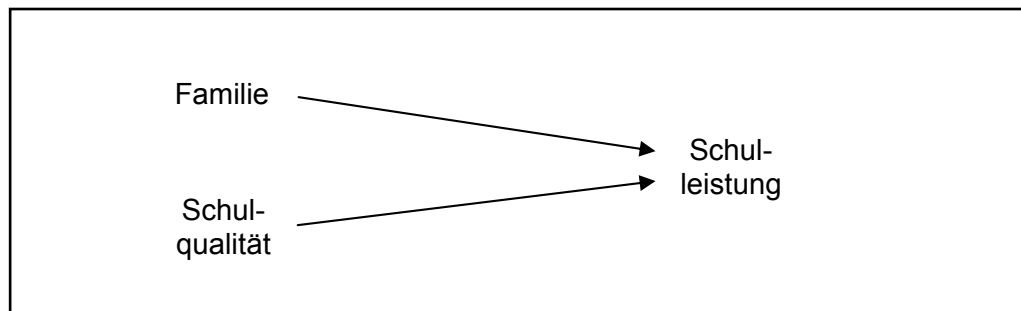
Alle einseitigen Pfeile stehen für eine kausale Beeinflussung, bidirektionale Verbindungen für einen korrelativen Zusammenhang. Ein Korrelationspfeil macht eine Variable dabei nicht endogen.

Ausgangspunkt der Berechnung ist die Korrelationsmatrix der Indikatorvariablen. Aus dieser Matrix sollen Beziehungen zwischen latenten und ihren Indikatorvariablen (vgl. Faktorenanalyse) und v.a. zwischen den latenten exogenen und endogenen Variablen untereinander abgeleitet werden.

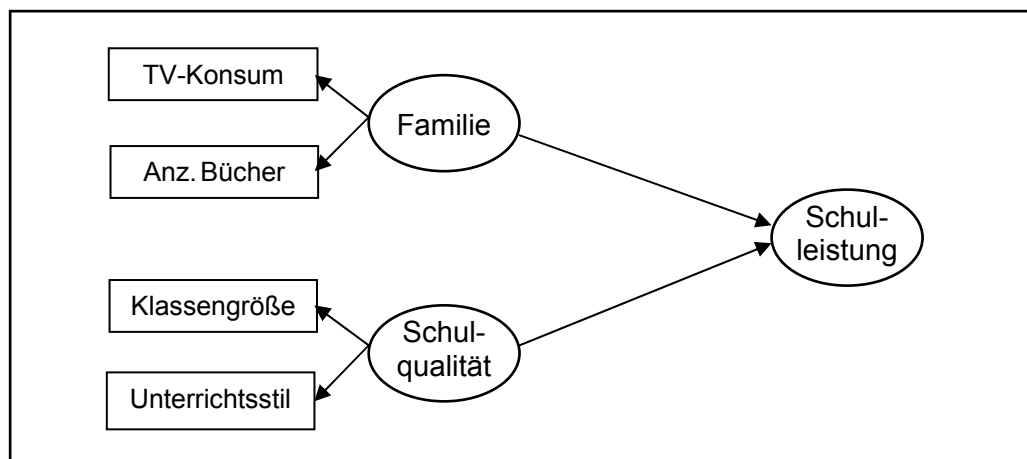
Über ein LISREL-Modell können jedoch lediglich unpassende Modelle abgelehnt werden. Die Nicht-Ablehnung eines Modells ist jedoch nicht gleichbedeutend mit seiner Richtigkeit. Im Sinne der Daten gibt es viele „richtige“ Modelle, sodass ein im Voraus theoretisch fundiertes Hypothesensystem vorhanden sein muss, um eine konfirmatorische LISREL-Analyse zu beginnen.

## 7.1. Beispiel

Grundsätzlich soll überprüft werden, wie sich Familie und Schulqualität auf die Schulleistung auswirken:

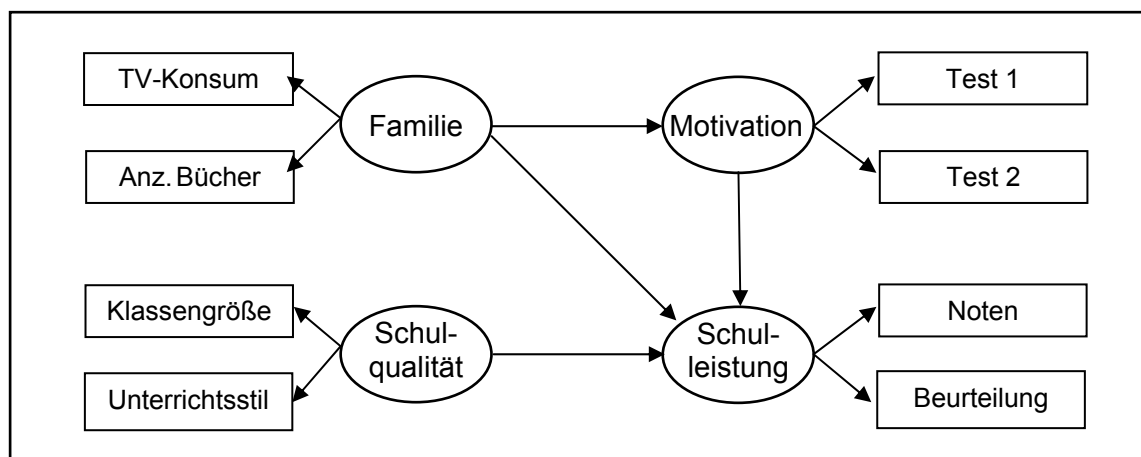


Familie und Schulqualität sind jedoch nicht direkt messbar (latent), sodass man sich geeignete Indikatoren überlegen muss.



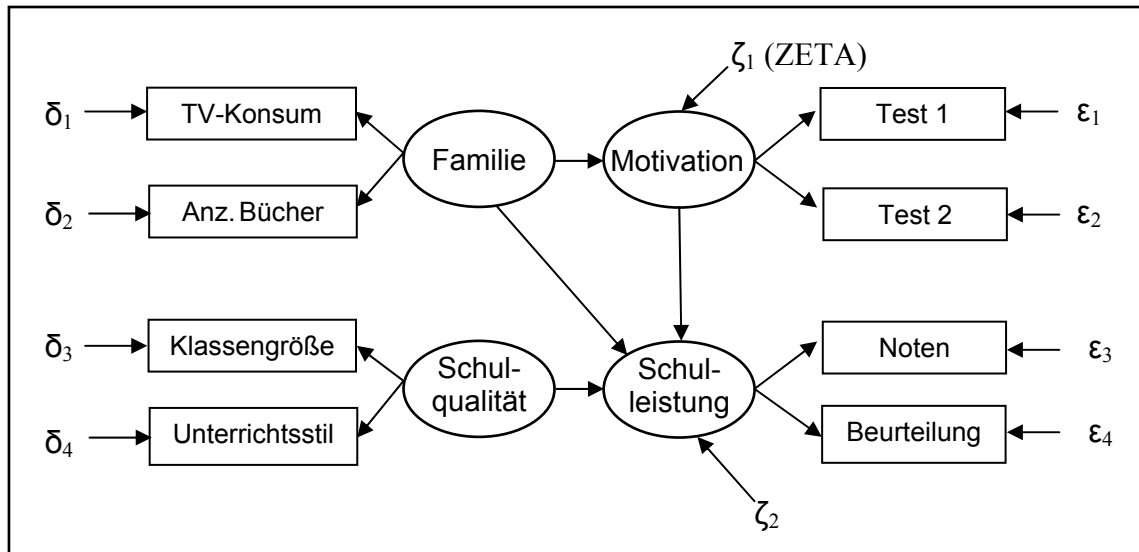
Der Pfeil der latenten Variablen zu den Indikatorvariablen kann auch als „operationalisiert durch“ gedeutet werden.

Auch die Schulleistung muss auf eine geeignete Art und Weise erfasst werden. Zusätzlich sollten Moderatorvariablen berücksichtigt werden.



Hier wird noch einmal deutlich, dass die Variablen nicht explizit als UVn oder AVn klassifiziert werden können. Die endogene Variable Motivation wird durch das familiäre Umfeld beeinflusst (AV), wirkt sich aber gleichzeitig selbst auf die Schulleistung aus (UV).

Die latenten Variablen erklären jedoch nicht die gesamte Varianz der Indikatoren. Von daher müssen spezifische Fehlerterme angenommen werden.



Die Fehlerterme werden also je nach Bedeutung der entsprechenden Variablen unterschiedlich benannt.

## 7.2. Kausalanalyse

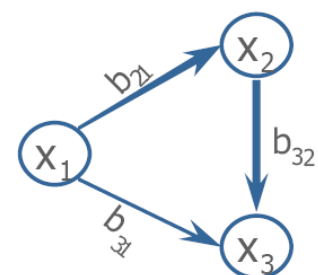
Die Kausalanalyse stellt den ersten Grundstein der linearen Strukturgleichungsmodelle dar. Hier wird versucht, aus der Korrelationsmatrix manifester Variablen eine kausale Struktur abzuleiten (Pfadanalyse).

Dabei lässt eine Korrelation zwischen x und y mehrere Interpretationen zu:

- Kausal:  $x \rightarrow y$
- Kausal:  $y \rightarrow x$
- Drittvariable:  $z \rightarrow (x, y)$

Die Überprüfung kausaler Zusammenhänge erfolgt hierbei über die Regressionsanalyse. Ausgangspunkt ist ein hypothetisches Pfaddiagramm, welches meist grafisch veranschaulicht wird. Darin werden die vermuteten Wirkungsrichtungen spezifiziert.

Hier können kausale Wirkungen insofern untersucht werden, als dass partielle Korrelationen der interessierenden Variablen berechnet werden. Drittvariablen liegen vor, wenn  $r_{zx} \cdot r_{zy} = r_{xy}$ . Ist dies nicht der Fall, so kann von einer kausalen Wirkrichtung ausgegangen werden.



Anmerkung: Die Pfadkoeffizienten folgen einer standardisierten Nomenklatur.  $b_{21}$  bedeutet, dass dieser Pfeil auf die Variable 2 zeigt und von der Variable 1 kommt.

Aus dem Pfaddiagramm lassen sich Strukturgleichungen ableiten. Im Beispiel:

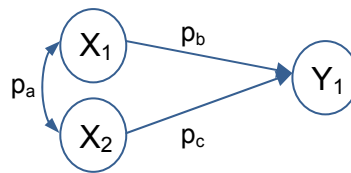
$$x_2 = b_1 + b_{21} x_1 \qquad x_3 = b_2 + b_{31} x_1 + b_{32} x_2$$

Bei standardisierten Pfadkoeffizienten fällt der konstante Term weg. Sie werden statt b mit p gekennzeichnet:

$$z_2 = p_{21} z_1 \qquad z_3 = p_{31} z_1 + p_{32} z_2$$

Korrelationen zwischen Variablen lassen sich als verbundene Pfade in einem Pfaddiagramm darstellen:

	X <sub>1</sub>	X <sub>2</sub>	Y <sub>1</sub>
X <sub>1</sub>	1.0		
X <sub>2</sub>	0.5	1.0	
Y <sub>1</sub>	0.6	0.7	1.0



$$r_{X_1 X_2} = p_a$$

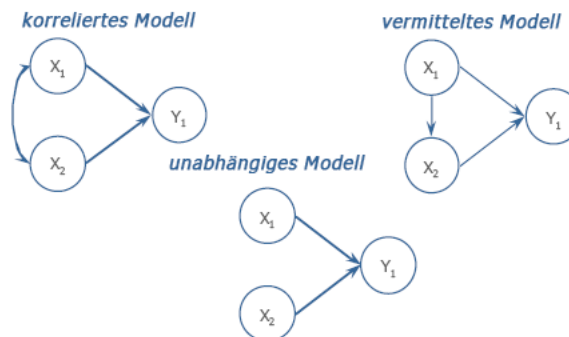
$$r_{X_1 Y_1} = p_b + p_a p_c$$

$$r_{X_2 Y_1} = p_c + p_a p_b$$

Dazu werden alle direkten und indirekten Pfade verwendet. Das dargestellte Gleichungssystem ist lösbar und kann also im Rahmen der Kausalanalyse verwendet werden.

In diesem Fall ergibt sich  $p_c = 0.53$ ;  $(53\%)^2 = 28.1\%$  der Varianz der Variablen Y werden also durch die entsprechenden Variable X<sub>2</sub> direkt erklärt.

Dabei sind natürlich auch andere Pfaddiagramme denkbar – jeweils abhängig von den entsprechenden theoretischen Vorüberlegungen:



### 7.2.1. Kausale und nicht-kausale Komponenten

Mit Hilfe der errechneten Pfadkoeffizienten lassen sich die empirischen Korrelationen zwischen den Variablen in kausale und nicht-kausale Komponenten zerlegen.

Der kausale Anteil entspricht demjenigen Anteil der Korrelation, der über die Pfadkoeffizienten erklärbar ist. Dabei wird zwischen direkten und indirekten kausalen Effekten unterschieden. Der nicht erklärbare Anteil wird als nicht-kausale Komponente bezeichnet.

[Anmerkung F12: Die nicht-kausale Komponente in  $r_{z_3 z_2}$  ist  $z_1$ .]

### 7.2.2. Beispiel

Das Beispiel zeigt ein hypothetisches Pfaddiagramm. Es lassen sich folgende 4 Wege von X1 zu Y finden:

- X1 → Y
- X1 → X2 → Y
- X1 → X2 → X3 → Y
- X1 → X3 → Y

→ 1 direkter, 3 indirekte Wege.

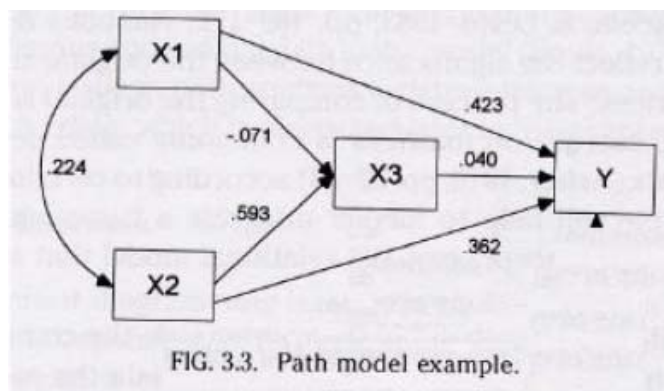


FIG. 3.3. Path model example.

Die Passung des Modells kann über die unten stehende Tabelle berechnet werden. Unter der Diagonale sind dabei die Pfadkoeffizienten angegeben, über der Diagonale stehen die ursprünglichen Korrelationen.

**TABLE 3.1**  
 Original and Reproduced Correlations in the Path Model

Variable	Y	X1	X2	X3	
Y	1.000	.507	.480	.275	Original correlations
X1	.423	1.000	.224	.062	
X2	.362	.224	1.000	.577	
X3	.040	-.070	.593	1.000	

$r_{12} = p_{12} = .224$   
(C)

$r_{13} = p_{31} + p_{32}p_{21} = -.071 + (.593)(.224) = .062$   
(DE) (IE)

$r_{23} = p_{32} + p_{31}p_{21} = .593 + (-.071)(.224) = .577$   
(DE) (IE)

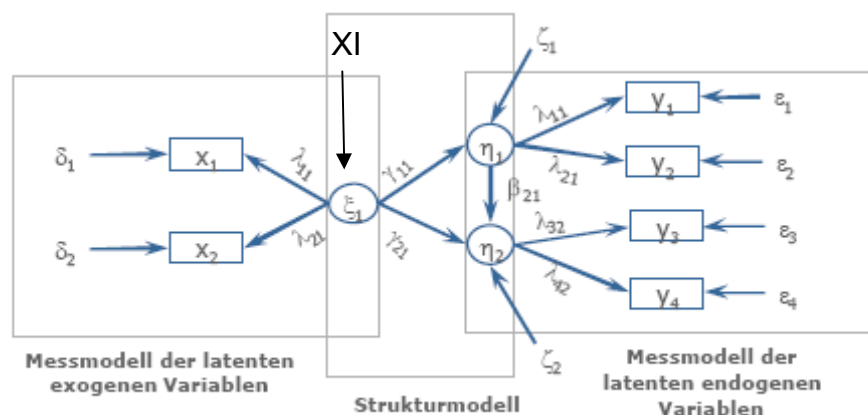
### 7.3. LISREL-Ansatz

Der LISREL-Ansatz stellt nun eine Erweiterung der Kausalanalyse auf latente Variablen dar. Bestandteile sind **Strukturmodell** der latenten Variablen, die daraus ableitbaren **Strukturgleichungen** (vgl. Kausalanalyse) sowie **Messmodelle** der latenten exogenen und endogenen Variablen (siehe F15).

Im LISREL-Ansatz der Kausalanalyse werden also die Beziehungen zwischen latenten Variablen in einem Strukturmodell abgebildet, das dem regressionsanalytischen Denkansatz entspricht. Die latenten Variablen werden dabei durch direkt beobachtbare Indikatorvariablen operationalisiert (faktorenanalytisches Denkmodell), wobei für endogene und exogene Variable getrennte Messmodelle aufgestellt werden.

Es werden also zwei Faktorenmodelle sowie ein Regressionsmodell integriert, wodurch theoretisch postulierte Beziehungen zwischen latenten Variablen überprüft werden können.

Das vollständige LISREL-Modell mit einer exogenen und zwei endogenen Variablen (mit je zwei Indikatorvariablen) lässt sich also wie folgt darstellen:



Im Folgenden soll nun der Ablauf einer Analyse vorgestellt werden. Desweiteren werden Voraussetzungen des LISREL-Ansatzes betrachtet.

### 7.3.1. Ablauf einer Analyse

Eine Analyse lässt sich in 6 aufeinanderfolgende Schritte untergliedern. (1) Zunächst wird eine Hypothese über relevante Variablen und deren Zusammenhänge erstellt. (2) Anschließend erfolgt eine Veranschaulichung dieser Überlegungen in einem Pfaddiagramm, wobei die Hypothesen (3) in mathematische Gleichungen überführt werden. (4) Nun muss geprüft werden, ob das Gleichungssystem lösbar ist (Identifikation der Modellstruktur).

Ist das Gleichungssystem lösbar, so können (5) Input-Matrix und Schätzmethode ausgewählt werden wodurch die Schätzung der einzelnen Modellparameter möglich wird. In diesem Schritt resultiert ein spezifiziertes Modell.

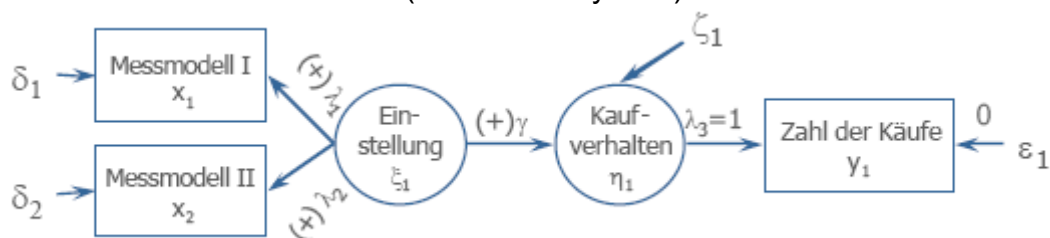
Schließlich kann dieses Modell (6) bzgl. seiner Passung an die empirischen Daten geprüft werden.

#### 7.3.1.1. Schritt 1: Hypothesenbildung

Aufgrund theoretischer Überlegungen wird festgelegt, welche Variablen in das Modell eingehen und wie die Beziehungen zwischen diesen Variablen aussehen sollen. Auch werden hier die Indikatoren der latenten Variablen festgelegt.

#### 7.3.1.2. Schritt 2: Pfaddiagramm

Die postulierten Zusammenhänge werden grafisch in einem Pfaddiagramm veranschaulicht. Hierbei sind immer Strukturmodell (regressionsanalytisch) und die beiden Messmodelle (faktorenanalytisch) enthalten.



#### 7.3.1.3. Schritt 3: Gleichungssystem

Nun erfolgt die Spezifikation der Modellstruktur. Für das obige Beispiel ergibt sich:

- Strukturmodell:  $\eta_{k1} = \gamma \cdot \xi_{k1} + \zeta_1 \cdot \zeta_{k1}$
- Messmodell der latenten exogenen Variablen:
 
$$x_{k1} = \lambda_1 \cdot \xi_{k1} + \delta_1 \cdot \delta_{k1}^9$$

$$x_{k2} = \lambda_2 \cdot \xi_{k1} + \delta_2 \cdot \delta_{k2}$$
- Messmodell der latenten endogenen Variable:  $y_{k1} = \lambda_3 \cdot \eta_{k1} + \varepsilon_1 \cdot \varepsilon_{k1}$

Die Hypothesen werden also in mathematische Gleichungen überführt.

<sup>9</sup> Der Fehler besteht sowohl aus der absoluten Fehlergröße ( $\delta_i$ ) sowie dem Pfad zur Indikatorvariablen ( $\delta_{ki}$ ). Das ist aber Kniefieselei.



#### 7.3.1.4. Schritt 4: Identifikation der Modellstruktur

Nach der Spezifikation muss geprüft werden, ob das Gleichungssystem lösbar ist. Dafür muss die Zahl der Gleichungen mindestens der Zahl der zu schätzenden Parameter entsprechen.

Die Zahl der Gleichungen entspricht dabei der Zahl der Korrelationskoeffizienten und berechnet sich daher mit  $n$  = Zahl der Indikatorvariablen durch:

$$n(\text{Gleichungen}) = \frac{n \cdot (n+1)}{2}.$$

Anmerkung: Die Zahl der Gleichungen hat nichts mit den Gleichungen in Schritt 3 zu tun.

Die Zahl der zu schätzenden Parameter kann aus dem Pfaddiagramm abgelesen werden, wobei immer alle Varianzen der exogenen Variablen (inkl. Fehler!) sowie alle nicht spezifizierten Pfadgewichte geschätzt werden müssen (siehe auch 7.4.1.1). Die Freiheitsgrade des Modells ergeben sich aus der Differenz der beiden Größen:

$$df = n(\text{Gleichungen}) - n(\text{Zu spezifizierende Parameter})$$

Für das Beispiel:

- $n(\text{Gleichungen}) = 3 \cdot (3+1) / 2 = 6$
- $n(\text{Zu schätzende Parameter}) = 9$  (Pfade:  $\lambda_1, \lambda_2, \lambda_3, \gamma$ , Varianzen:  $\varepsilon_1, \zeta_1, \xi_1, \delta_1, \delta_2$ )
- Es müssen also 3 Parameter festgelegt bzw. restringiert werden (s. u.), z.B. soll das Kaufverhalten ohne Messfehler erhoben werden ( $\lambda_3 = 1, \varepsilon_1 = 0, \xi_1 = 1$ ).

Ist die Zahl der Gleichungen gleich der Zahl der zu schätzenden Parameter, so spricht man von **just-identified**. Ist die Zahl der Gleichungen größer bzw. kleiner als die Zahl der zu schätzenden Parameter, so ist das Modell **over-** bzw. **under-identified**.

Dabei sollten over-identified Modelle am häufigsten vorkommen. Ist das Modell under-identified, so ist es nicht lösbar. Die Analyse muss jedoch nicht abgebrochen werden, da in einem solchen Fall bestimmte Variablen festgelegt oder restringiert werden können. Restringiert bedeutet in diesem Zusammenhang, dass ein Parameter als ebenso groß definiert wird wie ein anderer. Die Güte des Modells kann (über einen  $X^2$ -Wert) nur bestimmt werden, wenn das Modell over-identified ist.

Häufig werden die Varianzen der exogenen Variablen und/oder die Pfadgewichte zu Indikatorvariablen gleich 1 gesetzt (siehe 7.4.1.1).

#### 7.3.1.5. Schritt 5: Parameterschätzung

Ist das Gleichungssystem lösbar kann eine Input-Matrix sowie eine entsprechende Schätzmethode (z.B. ML) ausgewählt werden.

Aus der empirischen Korrelationsmatrix  $R$  wird dann eine modelltheoretische Korrelationsmatrix  $R'$  durch eine Schätzung der entsprechenden Parameter errechnet. Das Ziel ist eine möglichst gut Anpassung von  $R'$  an  $R$ , also  $(R-R') \rightarrow 0$ .

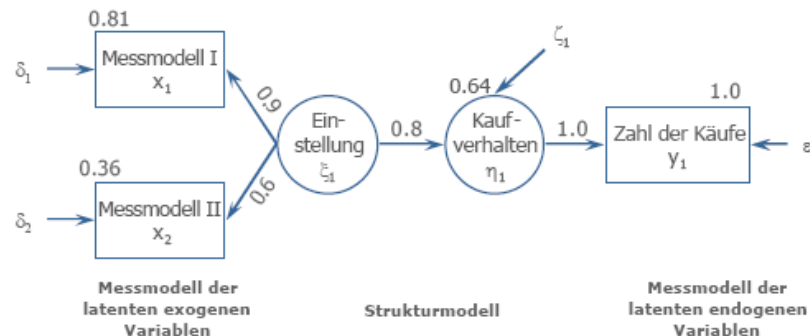
Das Resultat der Parameterschätzung ist die Ausgabe eines Pfaddiagramms mit den entsprechenden Gewichten (Pfadkoeffizienten).

### 7.3.1.6. Beurteilung und Interpretation der Schätzergebnisse

Streng genommen wird die Modellstruktur erst in diesem Schritt als Pfaddiagramm dargestellt.

Stimmen die postulierten Vorzeichen der Parameter mit den Vorzeichen der errechneten Parameter überein werden die Hypothesen nicht abgelehnt.

In der Praxis wird  $R'$  iterativ geschätzt, sodass  $(R-R') \rightarrow \min$ . In diesen Fällen können statistische Gütekriterien für die Modellschätzung errechnet werden. Dies ist jedoch nur möglich, wenn das Modell over-identified ist.



## 7.3.2. Voraussetzungen

Es lassen sich inhaltliche, stichprobenbezogene und statistische Anforderungen unterscheiden.

### 7.3.2.1. Inhaltliche Anforderungen

Es müssen gesicherte Theorien über die Zusammenhänge zwischen den Variablen existieren. Es sollten möglichst viele Informationen (Variablen) in die Analyse eingehen.

### 7.3.2.2. Stichprobe

Zur Ermittlung der Korrelationsmatrix sollten mindestens 100-150 Personen untersucht werden. Besser sind Stichprobengrößen von  $n > 200$ . Ein alternatives Kriterium sind ca. 10 Pbn pro gemessener Variable.

Zudem sollten mindestens 5 Pbn pro zu schätzendem Parameter untersucht worden sein.

### 7.3.2.3. Statistische Kriterien

- Die beobachteten Variablen sollten einer Multi-Normalverteilung folgen (abhängig vom Schätzverfahren). Es sollten mindestens 4-5 Indikatoren pro latenter Variable verwendet werden.
- Die Messmodelle entsprechen dem Grundmodell der Faktorenanalyse.
- Im Strukturmodell dürfen die Residuen nicht mit den exogenen latenten Variablen korrelieren; die Erwartungswerte der Residuen im Strukturmodell sind gleich 0.
- Die Messfehler korrelieren mit keiner anderen Variable.
- Linearität und Additivität der Konstrukte und Messhypothesen
- Invertierbarkeit der modelltheoretischen Kovarianz-Matrix (d. h. Identifizierbarkeit des Modells)

## 7.4. Übung

### 7.4.1. Software: AMOS 5.0/7.0

LISREL-Analysen sind mit SPSS nicht durchführbar. Daher wird eine Erweiterung, etwa die Software AMOS (Analysis of Moment Structures) benötigt. Dies ist ein spezielles Werkzeug zur Analyse von Strukturgleichungsmodellen mit latenten Variablen. Die Software wird von SPSS vertrieben, ist aber ein selbstständiges Windows-Programm.

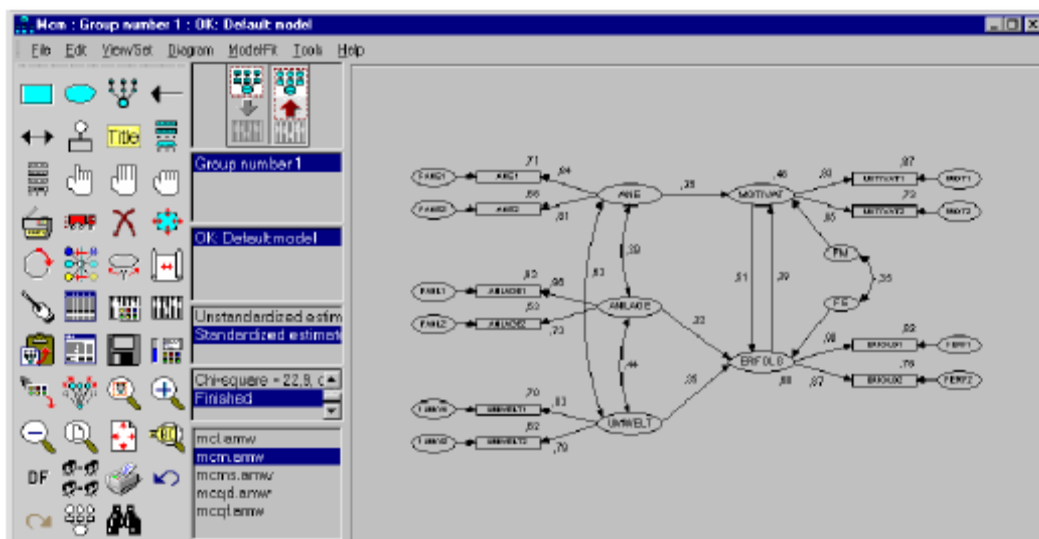
Eine Studentenversion von Amos (v 5.0) ist unter [www.amosdevelopment.com](http://www.amosdevelopment.com) zum Download verfügbar. Die Version ist voll funktionsfähig, hat aber die Einschränkung, dass maximal 8 Variablen betrachtet werden können. In üblichen Anwendungsfällen sollen jedoch weit mehr Variablen berücksichtigt werden.

Besonders einfach ist die Modellspezifikation in AMOS, da diese durch den Aufbau eines entsprechenden Pfaddiagramms erfolgt. Zusätzlich ist die Makrosprache Amos Basic integriert. Grundlegende Dateitypen des Programms sind:

- Projektdatei .amw
- Sicherungskopien der Projektdatei .bk1, .bk2. Die entsprechenden Sicherungen werden automatisch erstellt.
- Amos-Output: .AmosOutput

Amos-Pfaddiagramme können in die Zwischenablage kopiert und so einfach in andere Anwendungen übertragen werden. Es ist allerdings auch ein Output ins pdf-Format möglich.





Für die meisten Analysen benötigt Amos keine Rohdaten sondern lediglich die Stichprobenmomente, also Mittelwerte, Varianzen sowie Kovarianzen bzw. Korrelationen.





In die Zeichnungsfläche rechts kann das Modell eingetragen werden. Dabei wird zwischen Input-Modell und Output-Modell unterschieden (mittlere Spalte oben). Latente Variablen werden mit Kreisen symbolisiert, manifeste Variablen mit Rechtecken.

Auf den Folien 29 und 30 werden die einzelnen Symbole der Werkzeugpalette kurz erklärt.

Wichtige uneindeutige Werkzeuge sind:

-  : Müssen gezogen werden.
-  : Duplicate Objects. Toll, um gleich große Variablen zu bauen.
-  : Hiermit können Fehlerterme zu Indikatorvariablen oder aber auch zu endogenen latenten Variablen hinzugefügt werden. Der Pfadkoeffizient wird dann automatisch auf 1 gesetzt.
-  : Mit dem Zauberstab können die Pfeile ausgerichtet werden.

### 7.4.1.1. Daten einlesen und Modell spezifizieren

Wenn das Modell in Amos spezifiziert wurde können die Daten eingelesen werden. Das geht über das Werkzeug *Select Data Files* (). Es wird eine SPSS-Datei (.sav) eingelesen, die die entsprechende Struktur aufweisen sollte. Anschließend lässt man sich die enthaltenen Variablen anzeigen () und zieht sie auf die entsprechenden Kästchen für die Indikatorvariablen.

Die restlichen Objekte werden beliebig benannt. Unter Objekteigenschaften lassen sich zudem Parameter bzw. Varianzen spezifizieren (Zahlenwert eintragen) oder restringieren (Name *für die Varianz* eintragen; gleiche Namen werden gleich geschätzt). Hier sollten bspw. die Pfadkoeffizienten von Fehlervariablen immer gleich 1 gesetzt werden.

Auch empfiehlt sich die Spezifikation der Varianz der exogenen Variablen (ebenfalls 1), bzw. die Spezifikation eines Pfadgewichts jeder latenten Variablen zu einem ihrer Indikatoren. Hier sollte jeweils derjenige Indikator verwendet werden, der die latente Variable vermutlich bestmöglich repräsentiert. Diese Spezifikation ist keine wirkliche Restriktion des jeweiligen Parameters, sondern lediglich ein technischer Handgriff. Die endgültigen Varianzen der standardisierten Variablen können also von diesem Wert abweichen.

**Anmerkung 1:** „Warum müssen gerade die Varianzen der exogenen Variablen geschätzt werden?“. Klar, diese sind ja auch verantwortlich für die Varianz der Indikatoren – auch auf der endogenen Seite (hier indirekt). Die Pfade von Fehlern zu Indikatoren werden nicht explizit geschätzt, weil sie per Definitionem gleich 1 gesetzt werden.

**Anmerkung 2:** Die Struktur der eingelesenen Daten muss exakten Regeln folgen (vgl. F 33). Wichtig sind die Unterstriche bei `rowtype_` und `varname_`. Die Zeilen 4+ in der Spalte `rowtype_` geben an, ob es sich um Korrelationen oder Kovarianzen handelt.


	rowtype	varname_	y1	x1	x2	va
1	STDDEV		1.00	1.00	1.00	
2	MEAN		.00	.00	.00	
3	N		234.00	234.00	234.00	
4	CORR	y1	1.00	.72	.48	
5	CORR	x1	.72	1.00	.54	
6	CORR	x2	.48	.54	1.00	
7						

### 7.4.1.2. Berechnung

Nun müssen die Analyseeigenschaften (Analysis properties) festgelegt werden. Hier empfiehlt sich das ML-Schätzverfahren. Danach kann die Schätzung durchgeführt werden (Calculate estimates).

An optionalen Ausgaben sollten „Minimization history“, „Standardized estimates“ und „Squared multiple correlations“ ausgewählt werden.

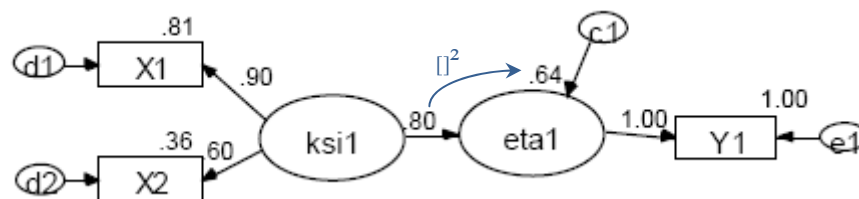
### 7.4.1.3. Ergebnis

Im Output-Fenster ist nun das Ausgabepfaddiagramm zu sehen (standardized estimates). Zur besseren Übersicht können die Parameter verschoben werden (  ). Dabei ist die Zahl über dem Kästchen der Indikatorvariablen deren Varianz.

Im Amos-Viewer (View text) lassen sich zusätzlich die Kennwerte der Schätzung als Textoutput betrachten. Dies ist für die Bewertung und Interpretation von Bedeutung.

## 7.4.2. Interpretation

Zunächst kann das (standardisierte) **Modell (Ausgabepfaddiagramm)** als Ergebnis der Schätzung interpretiert werden.



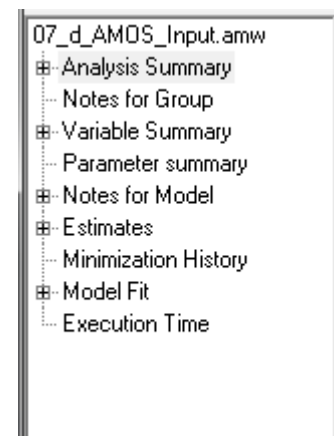
Wichtig ist hier vor allem, dass der quadrierte Pfadkoeffizient gleich der erklärten Varianz ist, bspw.  $0.9^2 = 0.81$  im Fall des Indikators  $X_1$ , der nur von der exogenen Variable  $\xi_1$  erklärt wird. Diese Berechnung entspricht der Kommunalität bei der Faktorenanalyse (Pfadkoeffizient = Faktorladung).

Was natürlich eigentlich interessant ist, sind die Pfadkoeffizienten und aufklärten Varianzen im Strukturmodell. Hier gilt ebenfalls die Regel:  $\beta^2 =$  erklärte Varianz. Im Beispiel werden also 64% der Varianz der latenten endogenen Variablen  $\eta_1$  durch die exogene Variable  $\xi_1$  erklärt.

Neben dem Ausgabepfaddiagramm kann auch der **Textoutput** von Amos zur Interpretation herangezogen werden.

Während Analysis Summary und Notes f. Group noch weniger interessant sind, verbirgt sich unter Variable Summary eine Umschreibung der Informationen, die im Ausgabepfaddiagramm enthalten sind, wobei nach verschiedenen Kriterien aufgeschlüsselt wird (Anzahl endogener, exogener, latenter, manifester Variablen).

Die ersten wirklich relevanten Informationen sind unter **Notes for Model** zu finden. Distinct sample Moments steht dabei für die Anzahl der Gleichungen (= Anzahl der Korrelationskoeffizienten). Ist das Modell over-identified sind hier auch Informationen zur Güte des Modells zu finden (in Form eines  $X^2$ -Tests).



Unter **Estimates** werden die Schätzungen für Varianzen und Pfadgewichte ausgegeben. Interpretiert werden nur die standardisierten Werte (Tabelle 2), wobei diese auch im Ausgabepfaddiagramm zu finden sind. Hier können jedoch aus der oberen Tabelle Informationen über deren Signifikanz mit zur Interpretation herangezogen werden.

**Minimization History** beschreibt den Ablauf der Analyse und die Anzahl der notwendigen Iterationen.

Am interessantesten ist jedoch die Ausgabe **Model Fit**, in der verschiedene Gütemaße zu finden sind, sofern es sich um ein over-identified model handelt. Die einzelnen Maße sollen kurz erklärt werden:

- **X<sup>2</sup>**: Der X<sup>2</sup>-Wert sollte nicht signifikant sein, da er ein Wert für die Abweichung von empirischer und geschätzter Zusammenhangsmatrix ist.
- **CMIN/DF**: Für gültige Modelle ist ein Quotient nahe 1 zu erwarten. Die Obergrenze für den X<sup>2</sup>/df-Bruch schwankt in der Literatur zwischen 2 und 5.
- **RMSEA**: Maß für die nicht aufgeklärte Varianz (Root Mean Square Error). Dieser Wert sollte also möglichst gering sein, wobei ein RMSEA < 0.05 als gut bezeichnet wird. Ein RMSEA zwischen .05 und .08 ist akzeptabel, alles andere ist inakzeptabel. Nach Möglichkeit sollte auch der Maximalwert (HI) kleiner als 0.05 sein.

Der Wert **PCLOSE** gibt zusätzlich die Wahrscheinlichkeit aus, mit der der RMSEA in Wirklichkeit < 0.05 ist. Dieser Wert sollte größer als 0.95 sein.

- **GFI/AGFI**: Der (Adjusted) Goodness-of-Fit-Index ist ein Maß für die erklärte Ausgangsvarianz. Der AGFI stellt zusätzlich die Anzahl der verwendeten Parameter in Rechnung. Beide sollten er > .90 oder besser > .95 sein.
- **NFI/CFI**: Normed bzw. Comparative Fit Index. Sollte > .90 sein.

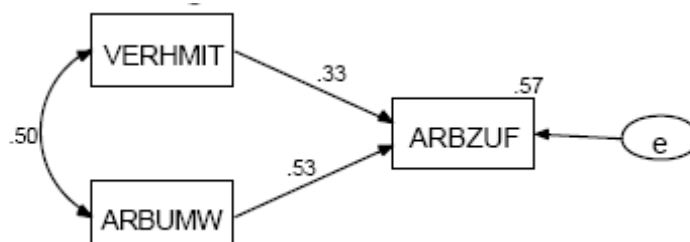
Für alle Maße gilt: *Default Model* bezeichnet das spezifizierte Modell.

### 7.4.3. Anwendungsbeispiele

#### 7.4.3.1. Folienbeispiel: Arbeitszufriedenheit

In diesem Beispiel soll demonstriert werden, dass Amos auch Kausalanalysen für eine Menge beobachteter Variablen durchführen kann. Hierfür wurde das unter 7.2 verwendete Modell (vgl. F 10) mit Amos bearbeitet.

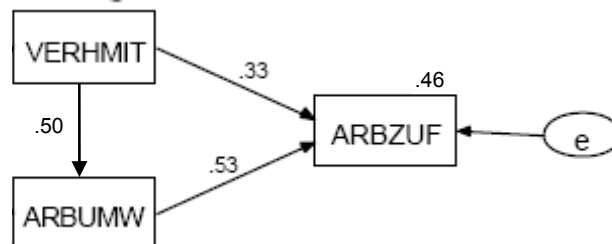
Die Schätzung entspricht der händisch errechneten. Wie im ersten Folienbeispiel ist das Modell just-identified (0 df), sodass kein X<sup>2</sup>-Test zur Bewertung der Modellanpassung möglich ist.



**Wichtig:** Wenn die aufgeklärten Varianzanteile für verbundene Pfade (indirekte Effekte) geschätzt werden, so werden die Pfadkoeffizienten nicht quadriert.

Für das Beispiel gilt für den indirekten Effekt VERHMIT → ARBUMW → ARBZUF:  $0.5 * (0.53^2) = 0.14$ .

Sähe das Modell hingegen eine kausale Verbindung der beiden exogenen Variablen vor, so würde das Pfadgewicht quadriert werden:



Jetzt berechnet sich der indirekte Effekt VERHMIT → ARBUMW → ARBZUF zu:  $(0.5^2) * (0.53^2) = 0.07$ .

Unabhängig davon berechnen sich kausale Effekte immer durch das Quadrieren der jeweiligen Pfadgewichte. Die aufgeklärte Varianz der Variable ARBZUF lässt sich im kausalen Beispiel also in 3 Komponenten zerlegen:

$$K1: \text{VERHMIT} \rightarrow \text{ARBZUF} = .33^2$$

$$K2: \text{ARBUMW} \rightarrow \text{ARBZUF} = .53^2$$

$$K3: \text{VERHMIT} \rightarrow \text{ARBUMW} \rightarrow \text{ARBZUF} = .50^2 * .53^2$$

Hier werden also nur 46% der Varianz aufgeklärt.

#### 7.4.3.2. Anwendungsbeispiel I

Es sollen die Bedingungen von beruflichen Ambitionen bei Jugendlichen untersucht werden; auch hier werden nur manifeste Variablen verwendet.

Das Modell hat 2 df (over-identified), sodass eine Prüfung über einen  $\chi^2$ -Test möglich wird. Hier sind alle Kriterien unter Model Fit sehr gut – der einzige Wehrmutstropfen ist vielleicht der etwas zu hohe  $RMSEA_{max}$ . Insgesamt passt das Modell jedoch sehr gut.

#### 7.4.3.3. Anwendungsbeispiel II

In diesem Beispiel werden mehr Variablen verwendet, als mit der Studentenversion von Amos gehandhabt werden können. Auch hier bestätigen die Gütekriterien eine gute Passung.

Anmerkung: \*\*\* bei Signifikanzniveaus steht für hochsignifikante Ergebnisse.

## 8. Anmerkungen

### 8.1. Generell

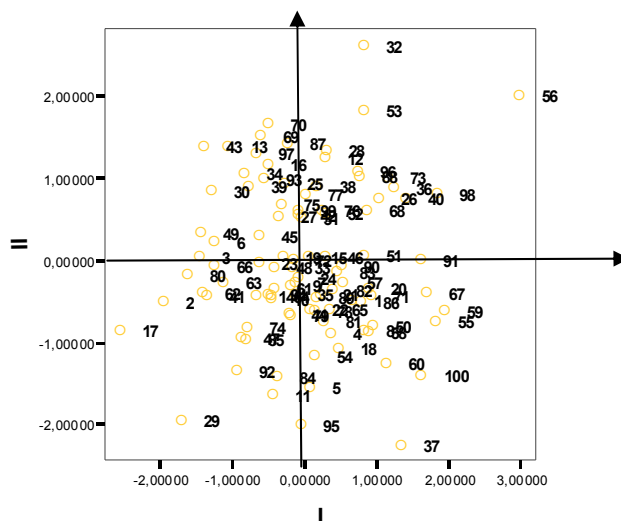
- Der Backhaus ist ein tolles Lehrbuch. Die Methodenleute wollen vom Skript-Lernen weg und definieren daher für jede Veranstaltung ein Lehrbuch.
- Wenn man unsicher ist, ob man wirklich metrische Daten hat hilft eine z-Standardisierung.

### 8.2. Klausur

- Namen der Proximitätsmaße der Clusteranalyse kennen: 4.3.
- LISREL: Wichtig ist vor allem der Grundgedanke (Ableiten kausaler Zusammenhänge zw. latenten Variablen) sowie der Ablauf einer Analyse

### 8.3. Fragen

- Wie kriegt man mit SPSS ganz normale Koordinatensysteme hin? Bzw.: Wie kriege ich Linien in einen Scatterplot? → Scheinbar wirklich nur über einen zusätzlichen Funktionenplot in Statistica oder R ( $x = 0$  und  $y = 0$ ) oder nachträgliches reinpinseln mit Paint. **Die Lösung:** Im Output doppelklicken! Da geht das.



- Warum zum Henker ist =STANDNORMINV(0.5) in Excel nicht genau 0? In Excel XP und Excel 2007 ergeben sich auch noch unterschiedliche Werte.
- LISREL: Wie komme ich auf .27 auf F45. Excel-File.