

Henning Stolze

# MULTIVARIATE METHODEN FÜR DIE UMFRAGEANALYSE

Skriptum zur Vorlesung



**Universität Lüneburg**

Fachbereich Wirtschafts- und Sozialwissenschaften

Forschungsinstitut Freie Berufe

Professur 'Statistik und Freie Berufe'

<http://ffb.uni-lueneburg.de>



# **Multivariate Methoden für die Umfrageanalyse**

Skriptum zur Vorlesung

Henning Stolze

1. Auflage: Sommersemester 2004

Universität Lüneburg

Fachbereich Wirtschafts- und Sozialwissenschaften

Forschungsinstitut Freie Berufe

Copyright © 2004

# INHALT

<b>1</b>	<b>Einführung.....</b>	<b>4</b>
1.1	Einordnung des Themas in den Ablauf einer Umfrage.....	4
1.2	Übersicht über das Thema .....	5
1.2.1	Unterscheidungsmerkmale.....	5
1.2.2	Skalenniveaus.....	6
1.2.3	Regressionsanalyse.....	7
1.2.4	Faktorenanalyse.....	7
1.2.5	Clusteranalyse .....	8
1.2.6	Neuronale Netze.....	8
1.3	Exkurs: Gestaltung von Fragebögen .....	9
<b>2</b>	<b>Regressionsanalyse .....</b>	<b>10</b>
2.1	Schätzung der Regressionsfunktion.....	11
2.2	Prüfung des Regressionsmodells .....	13
2.2.1	Prüfung der Regressionsfunktion .....	13
2.2.2	Prüfung der Koeffizienten .....	16
2.2.3	Prüfung der Grundannahmen für das Modell.....	16
<b>3</b>	<b>Logistische Regression.....</b>	<b>20</b>
3.1	Binomiale Logistische Regression.....	20
3.1.1	Interpretation der Koeffizienten.....	21
3.1.2	Beurteilung der Güte des Gesamtmodells .....	22
3.1.3	Ausreißerdiagnostik.....	24
3.1.4	Beurteilung einzelner Koeffizienten.....	25
3.2	Multinomiale logistische Regression.....	25
<b>4</b>	<b>Faktorenanalyse .....</b>	<b>27</b>
4.1	Variablenauswahl.....	29
4.2	Extraktion der Faktoren .....	30
4.2.1	Faktorladungen .....	30
4.2.2	Erklärungsanteil der Variablen / Kommunalitäten.....	32
4.2.3	Zahl der Faktoren / Eigenwerte.....	32
4.3	Faktorinterpretation .....	33
4.4	Bestimmung der Faktorwerte .....	34
<b>5</b>	<b>Clusteranalyse .....</b>	<b>34</b>
5.1	Proximitätsmaße .....	35
5.1.1	Proximitätsmaße für metrische Skalen.....	35
5.1.2	Proximitätsmaße für nominale Skalen .....	36
5.2	Fusionierung.....	38
5.2.1	Single Linkage.....	39
5.2.2	Complete-Linkage.....	39
5.2.3	Ward-Verfahren .....	40

5.2.4	Zahl der Cluster.....	41
5.3	Beurteilung der Ergebnisse.....	42
<b>6</b>	<b>Neuronale Netze.....</b>	<b>43</b>
6.1	Grundlagen neuronaler Netze.....	44
6.1.1	Aufbau neuronaler Netze .....	44
6.1.2	Netzwerktopologie .....	45
6.2	Funktionsweise eines einfachen KNN am XOR-Beispiel.....	46
6.3	Trainieren des KNN.....	48
	<b>Literaturverzeichnis.....</b>	<b>51</b>
	<b>Tabellen.....</b>	<b>52</b>

# 1 EINFÜHRUNG

## 1.1 Einordnung des Themas in den Ablauf einer Umfrage

Bei der Vorlesung „multivariate Verfahren für die Umfrageanalyse“ richtet sich das Hauptaugenmerk auf die Analyse von Daten, die in einer Umfrage gewonnen wurde. Wie diese Daten erhoben werden, ist nicht der Schwerpunkt dieser Veranstaltung. Dennoch soll das Thema in den Ablauf einer kompletten Umfrageanalyse, sozusagen als Exkurs zu Beginn, eingeordnet werden, um ein gewisses Grundverständnis für den gesamten Prozess einer Umfrage zu erwerben.

### Planung der Umfrage

- Ziele  
Müssen vor der Erstellung des Fragebogens festgelegt werden. Nur zur Klärung des Forschungsziels notwendige Fragen werden auf dem Fragebogen berücksichtigt. Hohe Anzahl der Fragen reduziert Rücklauf.
- Form der Umfrage  
z.B. Interview: Sehr genau, allerdings kostspielig. Direkt oder Telefonisch.  
Fragebogen: Verschicken, austeilern, auslegen?  
Online: Terminal, Internet, Mail? Kosten nicht unterschätzen. Repräsentativität je nach Zielgruppe fraglich.
- Ausführbarkeit  
Können Zielpersonen erreicht werden? Gibt es besondere Vorschriften?
- Zeitplan

### Entwurf und Gestaltung des Fragebogens

- Entwicklung der Fragen
- Gestaltung des Fragebogens (siehe Exkurs Kapitel 1.2)
- Evtl. Überprüfung des Fragebogens an Testgruppe
- Werden die Fragen richtig verstanden? Wie viel Zeit wird benötigt?  
Gibt es Fragen, die alle Personen/Befragten gleich beantworten?
- Begleitschreiben / Instruktionen

## **Bestimmung der Stichprobe und Durchführung**

- Vorbereitungen für Versand / Interviewer etc.
- Bestimmung der Stichprobe (Größe, Personenkreis etc.)
- Zielgruppe im Auge behalten. Hinreichende Größe wählen (Zentraler Grenzwertsatz beachten), hierbei auch Rücklaufquote berücksichtigen. Große Stichproben sind kein Garant für Repräsentativität.
- Erreichbarkeit: Zielgruppe vs. Bequemlichkeit
- Überlegungen zur Quotierung / Hochrechnung

## **Auswertung**

- Datenerfassung
- Häufigkeitsauszählungen
- Einfache deskriptive Statistik
- Dependenz / Interdependenzanalysen
- Hypothesentests

## **1.2 Übersicht über das Thema**

Die theoretischen Grundlagen der multivariaten Analyseverfahren wurden bereits in den sechziger Jahren gelegt, jedoch fehlten zunächst die praktischen Anwendungsmöglichkeiten aufgrund der fehlenden oder zu kostspieligen Computerrechenleistung. Mit dem Einzug der PCs sind die multivariaten Verfahren jedoch fester Bestandteil der Praxis vieler Fachgebiete wie beispielsweise Wirtschaft (Marktforschung, Wirtschaftsforschung etc.), Soziologie aber auch Medizin und Biologie.

### **1.2.1 Unterscheidungsmerkmale**

Der Titel der Veranstaltung „multivariate Analysemethoden“ weist zunächst auf Verfahren hin, denen mehr als zwei Variablen zugrunde liegen. Univariate Analysemethoden, wie zum Beispiel die einfache Häufigkeitsauszählung, betrachten lediglich eine Variable, bivariate Methoden schon zwei, was die Untersuchung von einfachen Zusammenhängen ermöglicht (Korrelation und einfache Regression). Multivariate Verfahren können noch mehr Variablen untersuchen. Somit stellen die bivariaten Verfahren Spezialfälle der multivariaten Verfahren dar, und werden in dieser Vorlesung zum Teil als einfache Beispiele für die vorgestellten Verfahren herangezogen.

Neben der Anzahl der untersuchten Variablen können auch die Methoden an sich in unterschiedliche Kategorien eingeteilt werden. Zum einen kann man zwischen Struktur-prüfenden und Struktur-entdeckenden Verfahren unterscheiden:

Bei den Struktur-prüfenden Verfahren handelt es sich um Methoden, die eine bestimmte Vermutung über die Art des (Kausal-)Zusammenhangs bestätigen und quan-

tifizieren sollen. Beispielsweise in welcher Weise der Absatz eines Produktes von dem Verkaufspreis, dem Werbeeinsatz und der Platzierung in den Supermärkten abhängt. Den Struktur-entdeckenden Verfahren liegt eine solche Vermutung nicht zugrunde. Hier sollen die Zusammenhänge zwischen den Variablen aufgedeckt werden, beispielsweise bei der Frage, wie verschiedene Produkte im subjektiven Wahrnehmungsraum der Konsumenten zueinander stehen.

Ein weiteres Unterscheidungskriterium ist die Einteilung in Dependenz- und Interdependenzanalysen. Bei den Dependenzanalysen gibt es stets mindestens eine abhängige Variable, deren Wert von anderen, unabhängigen Variablen bestimmt wird. Als Beispiel kann hier die Regression dienen, wenn beispielsweise der Absatz eines Produktes durch die Höhe des Verkaufspreises und den Werbeeinsatz erklärt wird. Bei Interdependenzanalysen werden die Wechselwirkungen zwischen den Variablen untersucht. Eine eindeutige Richtung des Zusammenhangs gibt es hier nicht. Bei der Berechnung eines Korrelationskoeffizienten zwischen der PS-Zahl eines Autos und der präferierten Durchschnittsgeschwindigkeit kann man schnell einen Zusammenhang feststellen. Es ist jedoch nicht ohne weiteres festzustellen, ob jemand schnell fährt, weil er oder sie ein genügend stark motorisiertes Auto besitzt, oder ob die Leistungsstärke der Autos aufgrund der bevorzugten Geschwindigkeit ausgewählt wird.

### 1.2.2 Skalenniveaus

Die richtige Wahl der unterschiedlichen Methoden richtet sich neben dem Ziel der Untersuchung auch nach dem Skalenniveau der zur Verfügung stehenden Daten. Das Skalenniveau beschreibt wie viele Informationen in den Ausprägungen der Variablen stecken. Zur kurzen Wiederholung hier eine kurze Übersicht:

Skalentyp		Merkmale	mögliche rechnerische Handhabung
nicht-metrisch	Nominalskala	Unterscheidung möglich	Häufigkeiten, Kontingenzkoeffizienten
	Ordinalskala	Rangfolge	Median, Rangkorrelationskoeffizienten
metrisch	Intervallskala	Konstanter Wertabstand ohne natürlichem Nullpunkt	Addition, Subtraktion
	Verhältnisskala	Konstanter Wertabstand mit natürlichem Nullpunkt	Addition, Subtraktion, Division, Multiplikation
	Absolutskala	Konstanter Wertabstand, natürlicher Nullpunkt und natürliche Einheit	Addition, Subtraktion, Division, Multiplikation

Allerdings können viele nominal- oder ordinalskalierte Merkmale mit Hilfe von Dummies (1/0-Variablen) in annähernd metrische Hilfsvariablen überführt werden,

in dem jede Ausprägung für sich in eine Dummyvariable überführt wird. So kann z.B. die Variable ‚Schultyp‘ (Hauptschule, Realschule, Gymnasium) in die Dummy-Variablen Hauptschule (1/0 – Ja/Nein) und Realschule (1/0 – Ja/Nein) transformiert werden. Die Möglichkeit des Gymnasiums ist bereits aus den anderen beiden Schulform-Dummys abzulesen (lineare Abhängigkeit: wenn beide den Wert Null haben, muss es sich um ein Gymnasium handeln), so dass das Gymnasium als Referenzkategorie dient und nicht eine weitere Dummyvariable bekommt.

### 1.2.3 Regressionsanalyse

Bei der Regressionsanalyse wird eine abhängige Variable durch einen funktionalen Zusammenhang aus einer Reihe unabhängiger Variablen geschätzt. Ziel ist es, die genaue Form des Zusammenhangs zu bestimmen, damit Prognosen angestellt werden können und die Bedeutung der einzelnen unabhängigen Variablen ermittelt werden kann.

Abhängige Variable  $\leftarrow$  unabhängige Variablen  
 $Y \leftarrow f(x_1, x_2, \dots, x_n)$   
(metrisch)  $\leftarrow$  (metrisch)

Bsp.: Absatz  $\leftarrow$  Werbung,

jedoch problematisch: Bekanntheitsgrad eines Produktes  $\leftarrow ? \rightarrow$  Absatz

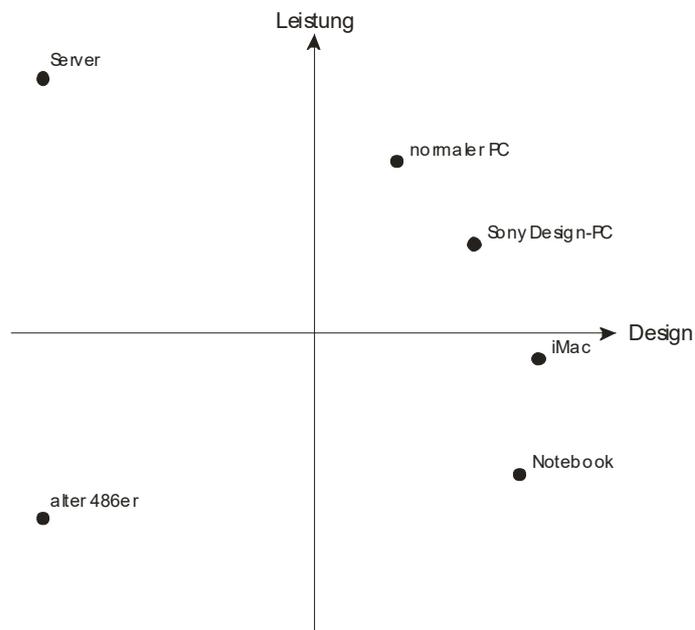
### 1.2.4 Faktorenanalyse

Die Faktorenanalyse versucht die Anzahl der vorhandenen Variablen (Merkmale) zu verringern, indem ähnliche Variablen zu Faktoren zusammengefasst werden. Auf diese Weise wird die Übersichtlichkeit des Datenmaterials gesteigert und einfache grafische Darstellungsformen ermöglicht. Zum anderen werden redundante Informationen beseitigt, so dass Faktoren in einer Regression unabhängige Variablen ersetzen können, die untereinander stark anhängig sind.

Bsp.:

<i>Variablen</i>	<i>Faktoren</i>
Prozessorleistung	} ‚Leistung‘
Festplattengröße	
Arbeitsspeicher	
Gehäusegröße	} ‚Design‘
Lautstärke	
Gewicht	

Nachdem die ursprünglichen sechs Variablen zu zwei Faktoren zusammengefasst wurden, kann nun ein einfaches Diagramm mit verschiedenen Produkten erstellt werden:



### 1.2.5 Clusteranalyse

Bei der Clusteranalyse handelt es sich ebenfalls um ein Verfahren, welches Kategorien aufdecken soll. Allerdings zielt es nicht auf die Merkmale wie die Faktorenanalyse, sondern auf die Merkmalsträger. Es sollen also Objekte zusammengefasst werden, so dass Gruppen entstehen, deren Mitglieder möglichst homogen sein sollen, die Gruppen untereinander jedoch möglichst unterschiedlich. Auf diese Weise können z.B. bestimmte Kundentypen in einem Unternehmen identifiziert werden, für die beispielsweise spezielle Werbung geschaltet werden soll.

### 1.2.6 Neuronale Netze

Neuronale Netze spielen bei den multivariaten Analyseverfahren eine besondere Rolle. Zum einen sind sie sehr neu und erst seit kurzem für den Anwender verfügbar. Zum anderen versuchen sie nicht, einen Zusammenhang zu erklären oder zu beschreiben, sie dienen lediglich dazu, bestimmte Ergebnisse vorherzusagen, ohne den Zusammenhang genau formuliert zu haben. So werden neuronale Netze mit einer bekannten Stichprobe trainiert, so dass später bei eingegebenen ‚unabhängigen‘ Variablen die Ergebnisse zuverlässig prognostiziert werden können.

## 1.3 Exkurs: Gestaltung von Fragebögen

### Glaubwürdige Kommunikation

- Einleitung im Anschreiben
- Vertrauen gewinnen (z.B. Hinweis auf Verwendungszweck und Datenschutz)
- Bedankung für die Teilnahme

### Neutrales aber ansprechendes Design

- Professionalität zeigen
- Aufwand für den Fragebogensteller sollte deutlich werden („Der Fragebogen ist für den Steller wichtig“)
- Unübersichtliches Design provoziert Fehler

### Gestaltung des Fragebogens: Auf Länge und Strukturierung achten

- Die erste Frage soll möglichst Interesse wecken
- Einfache Fragen zu Beginn, „heikle“ Fragen erst später
- Nach Themengebieten gruppieren
- Personenbezogene (demografische) Fragen besser zum Schluss
- Evtl. Konsistenzprüfende Fragen einbauen
- Evtl. Filterfragen um andere Fragen überspringen zu können (in Maßen einsetzen)

### Gestaltung der Fragen: Kurz, klar, einfach

- Kurze, präzise Fragen
- Einfache Sprache
- (keine doppelten Negationen, keine schweren Fremdwörter, Mehrdimensionalität vermeiden, keine komplizierten Berechnungen etc.)
- Der Befragte muss die Antwort kennen (können)
- Sensibilität bei „heiklen“ Fragen (z.B. Kriminalität, Sexualität, Einkommen ...)
- Möglichst exakte Informationen abfragen. Aber: auf Anonymität achten

### Gestaltung der Antwortmöglichkeiten

- Bei Vorgaben: Vollständige Antwortkategorien
- „keine Ahnung“ / „Non Attitude“-Feld = wertvolle Information; es soll keine Antwortmöglichkeit aufgezwungen werden
- Positive und negative (ausgewogene) Antwortoption
- Bei Rating: Überlegung gerade / ungerade Anzahl von Punkten

- Einheiten angeben
- Unterschiedliche Polung kann Aufmerksamkeit erhöhen
- Bei der Konzeption schon die Auswertung im Auge behalten

### Grundsätzlich: Nicht manipulieren

- Suggestive Fragen vermeiden
- Keine durch Vorurteile belasteten Begriffe verwenden (z.B. „Boss“, „Bürokrat“, ...)

## 2 REGRESSIONSANALYSE

Der Regressionsanalyse liegt die Vermutung zugrunde, dass eine abhängige Variable mit Hilfe einer mathematischen Formel durch eine beliebige Anzahl unabhängiger Variablen erklärt werden kann. Hierbei ist es wichtig fachlich fundierte Überlegungen zugrunde zu legen („erst denken, dann rechnen“).

Y	$x_1, x_2, \dots, x_j$
Abhängige Variable	unabhängige Variablen
Regressand	Regressoren
Erklärte Variable	erklärende Variablen
Endogene Variable	Exogene Variablen

Es wird oft unterstellt, dass es sich um einen linearen Zusammenhang handelt:

$$\hat{Y} = f(x_1, x_2, \dots, x_N)$$

$$\hat{Y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_j x_j$$

Dies ist zum einen der Einfachheit des Verfahrens geschuldet, zum anderen lassen sich viele Zusammenhänge, besonders in einem bestimmten Gültigkeitsraum tatsächlich, durch einen annähernd linearen Zusammenhang beschreiben. Zudem kann der allgemeine lineare Ansatz leicht in andere Formen, etwa einer quadratischen, oder polynomen Form umgewandelt werden (mehr dazu später).

Nachdem das Modell nun formuliert und die allgemeine Form gefunden ist, stellen sich zwei Probleme:

- Aus den Daten der Stichprobe muss eine Funktion gefunden werden, d.h. die  $b$ 's müssen bestimmt werden (Regressionsfunktion schätzen).
- Danach muss geprüft werden, ob der Zusammenhang, der in der Stichprobe gefunden wurde, auch für die Grundgesamtheit gilt (Prüfung des Regressionsmodells).

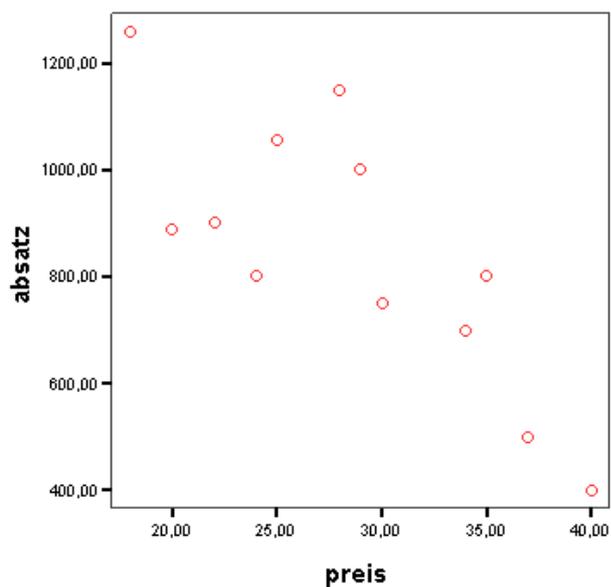
## 2.1 Schätzung der Regressionsfunktion

Als Beispiel für die Beobachtungen nehmen wir 15 verschiedene Kioske, die ein Produkt zu verschiedenen Preisen anbieten. Aus dem Preis resultiert die abgesetzte

Menge:  $\hat{y} = b_0 + b_1 x$

Nummer	Absatz	Preis
1	950,00	20,00
2	800,00	24,00
3	1150,00	28,00
4	1100,00	25,00
5	1200,00	18,00
6	500,00	37,00
7	700,00	34,00
8	400,00	40,00
9	800,00	35,00
10	1000,00	29,00
11	1500,00	18,00
12	800,00	20,00
13	900,00	22,00
14	750,00	30,00
15	1000,00	25,00

Da es nur zwei verschiedene Variablen gibt, lässt sich das Datenmaterial in einem Diagramm gut darstellen:



Ziel ist es nun, durch diese Punktwolke möglichst gut eine Gerade zu legen. Die Bedeutung des Begriffs ‚gut‘ muss hierbei noch geklärt werden. Die Lage einer Geraden wird über den Y-Achsenabschnitt ( $b_0$ ) und die Steigung ( $b_1$ ) bestimmt, es gilt also diese beiden Parameter zu schätzen.

Schon auf den ersten Blick auf die Punktwolke wird klar, dass es keine Gerade geben wird, auf der alle Beobachtungspunkte liegen werden. Dennoch soll die Gerade das Datenmaterial möglichst treffend beschreiben. Eine Möglichkeit die Güte der Geraden zu bestimmen liegt in den Abweichungen zwischen den beobachteten Werten und dem Punkt auf der Geraden:

$$\text{Residuen } e_k = y_k - \hat{y}_k$$

$e_k$  ist die nicht erklärte Abweichung für eine Beobachtung  $k$ ,  $y$  steht für die empirischen Beobachtungswerte,  $\hat{y}$  für die Schätzwerte der Regressionsfunktion.

Kommen wir auf den Begriff ‚gut‘ zurück. Die Gerade liegt dann gut in der Punktwolke, wenn alle Abweichungen zwischen der Geraden und den Beobachtungswerten möglichst klein sind. Ziel der Regression ist es also, die Abweichungen  $e$  zu minimieren:

$$\sum_{k=1}^K e_k^2 \rightarrow \min!$$

$$\sum_{k=1}^K (y_k - \hat{y}_k)^2 \rightarrow \min!$$

Die Formel für  $\hat{y}$  ist bekannt, da es sich hierbei um die Regressionsgleichung handelt, die Ausgangspunkt für diese Überlegungen war:

$$\sum_{k=1}^K (y_k - [b_0 + b_1 x_k])^2 \rightarrow \min!$$

Diese Funktion kann nun nach  $b_0$  und  $b_1$  gelöst werden. Für den Beispielfall erhalten wir:

$$\hat{y} = 1705 - 30 \cdot x_1$$

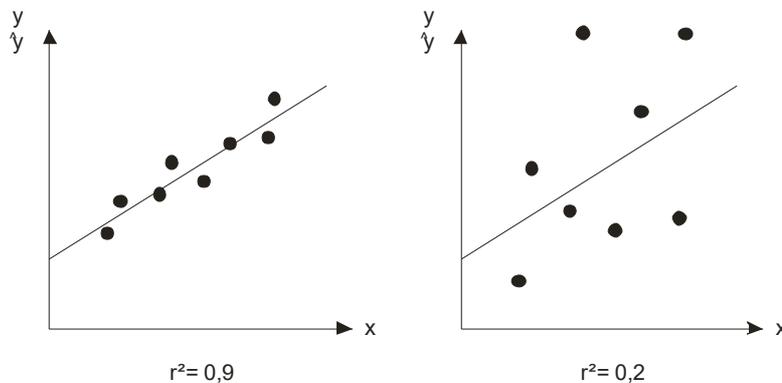
Schätzung der Regressionsfunktion heißt also: eine Funktion finden, die die Lage der Datenpunkte der Stichprobe möglichst gut beschreibt. Damit ist innerhalb der Stichprobe der Einfluss der unabhängigen Variablen (hier: Preis) auf die abhängige Variable (hier: Absatz) geklärt und darüber hinaus können Prognosen für die abhängige Variable angestellt werden, indem man für die unabhängigen Variablen Werte einsetzt. Bei einem Preis von 20 Euro würde ein Kiosk in diesem Beispiel also  $1705 - 30 \cdot 20 = 1105$  Stück absetzen.

## 2.2 Prüfung des Regressionsmodells

Durch die Schätzung der Regressionsfunktion haben wir die Parameter erhalten, die die lineare Funktion bestimmen, welche das Datenmaterial am besten beschreibt. Jedoch gibt es keinen Anhaltspunkt für die Güte der Funktion, weder in Bezug auf die Stichprobe, noch auf die Übertragbarkeit der Ergebnisse auf die Grundgesamtheit.

### 2.2.1 Prüfung der Regressionsfunktion

Im 2-Variablen-Fall ist die Gerade zwar optimal in die Punktwolke eingebettet (es gibt keine bessere, bei der die Residuen kleiner wären), jedoch ist es z.B. noch nicht möglich zu beurteilen, ob die tatsächlichen Beobachtungswerte gut durch die Gerade beschrieben werden können:



Daher ist es nötig die gefundenen Ergebnisse auf ihre Güte zu überprüfen. Ein Anhaltspunkt hierfür ist das Verhältnis von erklärter Streuung zur Gesamtstreuung.

$$\frac{\text{erklärte Streuung}}{\text{Gesamtstreuung}} = \frac{\sum_k (\bar{y} - \hat{y}_k)^2}{\sum_k (\bar{y} - y_k)^2} = r^2 = \text{Bestimmtheitsmaß}$$

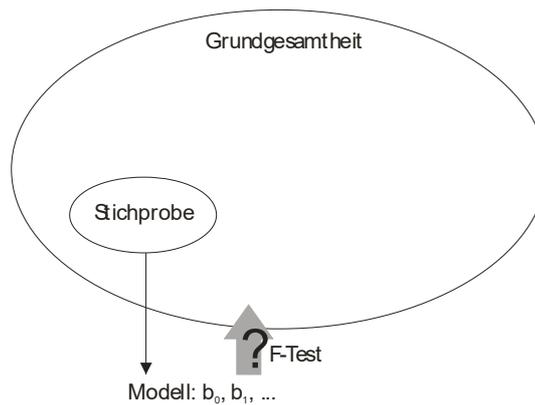
Das Bestimmtheitsmaß liegt zwischen 0 und 1 und gibt den relativen Anteil der erklärten Streuung an der Gesamtstreuung an. Ein hoher Wert bescheinigt der Regressionsfunktion also eine gute Beschreibung des Datenmaterials.

Da mit zusätzlichen unabhängigen Variablen das Bestimmtheitsmaß automatisch steigt, ist eine Korrektur erforderlich, die die Anzahl der unabhängigen Variablen berücksichtigt:

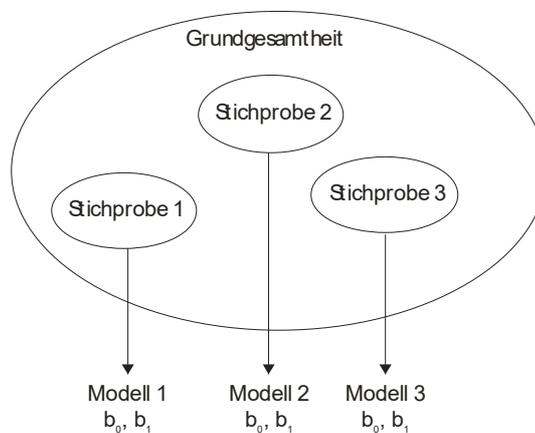
$$r_{\text{kor.}}^2 = r^2 - \frac{J \cdot (1 - r^2)}{K - J - 1}$$

J ist die Anzahl der Regressoren, K die der beobachteten Werte

Allerdings ist dies erst ein rein deskriptives Mittel um die Güte der Schätzfunktion innerhalb der Stichprobe zu beschreiben. Das  $r^2$  sagt nichts über die Grundgesamtheit aus. Da der Schluss von einer Stichprobe auf die Grundgesamtheit jedoch zumeist das gewünschte Ziel ist, benutzt man die F-Statistik, um diesen Zusammenhang zu beurteilen.



Die Auswahl der Daten aus der Grundgesamtheit, die die Stichprobe bilden erfolgt zufällig. Da sich die Parameter der Regressionsgleichung aus der zufällig gezogenen Stichprobe ergeben, sind auch die Parameter  $b_0$  und  $b_1$  (und aller weiteren  $b_i$ ) Zufallszahlen.



Nehmen wir an, die unten stehende Tabelle sei die (eigentlich unbekannte) Grundgesamtheit, aus der eine Stichprobe von 3 Fällen gezogen werden soll. Mehrere Ausgänge für die geschätzte Regressionsfunktion sind dann denkbar.

Beobachtung i	Alter (x)	Einkommen (y)
1	21	800
2	25	1200
3	24	1100
4	22	1000
5	26	1500

Stichprobe 1 (Beobachtung 1, 2 u. 3):  $b_0 = -1300$   $b_1 = 100$   
 Stichprobe 2 (Beobachtung 2, 3 u. 4):  $b_0 = -421$   $b_1 = 293$   
 Stichprobe 3 (Beobachtung 3, 4 u. 5):  $b_0 = -1800$   $b_1 = 125$

Die aus den Daten der Stichprobe berechneten Parameter der Regressionsfunktion sind also Zufallsrealisationen und schwanken um einen unbekannt Parameter der Grundgesamtheit  $\beta_0$  und  $\beta_1$  (im Beispiel ist die Grundgesamtheit ausnahmsweise bekannt und die wahren Parameter lauten  $\beta_0 = -1679$  und  $\beta_1 = 118$ ).

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j + \varepsilon \quad \varepsilon \text{ ist in diesem Falle eine zufällige Störgröße}$$

Nachdem der Einfluss des Zufalls auf die Regressionsfunktion geklärt ist, können wir uns wieder der ursprünglichen Frage bezüglich der Güte der Regressionsfunktion zuwenden: Können die Y-Werte durch die Regressionsfunktion mit den x-Werten erklärt werden oder nicht? Angenommen, die Parameter  $\beta_j$  seien alle Null. Die abhängige Variable Y ergäbe sich dann rein zufällig und frei von systematischen Einflüssen der unabhängigen Variablen – das Modell wäre dann nicht geeignet, die abhängige Variable zu erklären:

$$Y = \varepsilon$$

Diese Annahme kann mittels eines F-Tests überprüft werden:

$$H_0 : \beta_j = 0 \quad \text{mit } j=1\dots J$$

Die Vermutung, dass die unabhängigen Variablen keinen Einfluss haben, steht in der Nullhypothese. Kann diese durch den nun durchzuführenden Test verworfen werden, kann man mit einer festzulegenden Sicherheitswahrscheinlichkeit behaupten, die unabhängigen Variablen trügen signifikant zur Erklärung des Modells bei.

Das Sicherheitsniveau soll beispielsweise 95% betragen. Die Gegenwahrscheinlichkeit hiervon ist die Irrtumswahrscheinlichkeit (Signifikanzniveau), die besagt, dass die Nullhypothese in 5% der Fälle zu Unrecht verworfen wird. In 5 von 100 Fällen kann es also sein, dass der Test fälschlicherweise zu dem Schluss gelangt, ein Erklärungszusammenhang sei gegeben.

Um den Test durchzuführen wird eine empirische Prüfgröße ( $F_{emp}$ ) auf der Grundlage des Datenmaterials der Stichprobe berechnet und dann mit einem theoretischen F-Wert ( $F_{kritisch}$ ) verglichen:

$$F_{emp} = \frac{r^2 / J}{(1 - r^2) / (K - J - 1)}$$

J: Anzahl Regressoren  
 K: Anzahl Beobachtungen  
 J= Freiheitsgrade des Zählers  
 K-J-1= Freiheitsgrade des Nenners

Für das eingangs erwähnte Preis-Absatz-Beispiel ergibt sich ein empirischer F-Wert von 17,4. Der kritische F-Wert kann passend zum gewählten Signifikanzniveau direkt aus der dazugehörigen Tabelle abgelesen werden. Hierzu werden zusätzlich nur die Freiheitsgrade des Zählers und des Nenners benötigt. Im Beispiel beträgt der kritische F-Wert 4,75. Zur Entscheidung, ob die Nullhypothese verworfen werden kann, kommt man durch den Vergleich der beiden Werte. Ist der empirische Wert größer als der kritische, so wird  $H_0$  verworfen und man gelangt zu dem Schluss, wie auch in diesem Beispiel, dass die unabhängigen Variablen zur Erklärung des Modells signifikant beitragen.

Gängige Softwareprogramme können zur Klärung der Modellgüte auch den umgekehrten Weg gehen. Sie berechnen zunächst den empirischen F-Wert. Nun wird angenommen, dass der kritische F-Wert minimal kleiner sein soll, als der empirische Wert, so dass die Nullhypothese verworfen wird und das Modell als signifikant eingestuft wird. Aus dieser Wahl ergibt sich ein ‚empirisches Signifikanzniveau‘, welches auf den Ergebnisausdrucken der Statistikprogramme aufgeführt wird. Man kann nun direkt ablesen, mit welcher Wahrscheinlichkeit man sich irrt, wenn man die Nullhypothese verwirft und den Variablen einen signifikanten Erklärungsbeitrag unterstellt.

## 2.2.2 Prüfung der Koeffizienten

Nachdem geprüft wurde, ob die unabhängigen Variablen insgesamt einen Einfluss auf die abhängige Variable haben, und das Modell als Ganzes zur Erklärung des jeweiligen Sachverhalts herangezogen werden kann, ist es auch möglich einzelne Koeffizienten der Regressionsgleichung zu überprüfen. Auf diese Weise können einzelne Variablen ausfindig gemacht werden, die keinen signifikanten Erklärungsbeitrag leisten und evtl. aus dem Modell herausgenommen werden müssen.

Für diese Überprüfung verwendet man einen t-Test, der im Wesentlichen analog zu dem F-Test des vorangegangenen Abschnitts abläuft:

$H_0 : \beta_j = 0$	einzelner Koeffizient j wird getestet,
$\alpha = 0,10$	Signifikanzniveau: 10%
$t_{emp} = \frac{b_j}{s_{b_j}} = \frac{-29,7}{7,1} = -4,17$	Regressionskoeffizient / Standardfehler des Regressionskoeffizienten
$t_{kritisch} = 2,160$	Aus Tabelle mit $\alpha/2=0,05$ und $(K-J-1)= 13$ Freiheitsgraden

Wenn der empirische t-Wert betragsmäßig größer ist als der kritische, kann die Nullhypothese verworfen werden und ein Einfluss der einzelnen Variable unter dem festgelegten Signifikanzniveau unterstellt werden.

## 2.2.3 Prüfung der Grundannahmen für das Modell

Dem Schluss von der Stichprobe auf die Grundgesamtheit liegt ein stochastisches Modell zugrunde. Dieses Modell setzt einige Prämissen voraus, die für die Durchführung der Tests gegeben sein müssen. Nur so wird der BLUE (Best Linear Unbiased Estimator) geliefert. Allerdings muss auch gesagt werden, dass das lineare Regressionsmodell recht robust gegenüber kleineren Verletzungen der folgenden Prämissen ist und auch dann noch zuverlässige Ergebnisse liefert:

$$(a) \quad y_k = \beta_0 + \sum_{j=1}^J \beta_j x_{jk} + \varepsilon_k \quad \text{mit } k=1,2,\dots,K; J < K$$

Lineares Modell

Zahl der erklärenden Variablen  $J$  ist kleiner als Zahl der Beobachtungen  $K$

$$(b) \quad E(\varepsilon_k) = 0$$

Vollständigkeit des Modells: Störgrößen haben den Erwartungswert Null.

Falls nicht, sind Verzerrungen systematisch, d.h. wichtige Einflussgröße bleiben unberücksichtigt

$$(c) \quad \text{Var}(\varepsilon_k) = \sigma^2$$

Homoskedastizität: Konstante Varianz der Störgrößen

$$(d) \quad \text{Cov}(\varepsilon_k, \varepsilon_{k+r}) = 0$$

Keine Autokorrelation der Störgröße

(e) Zwischen den erklärenden Variablen  $x_j$  besteht keine lineare Abhängigkeit untereinander (keine [exakte] Multikollinearität)

(f)  $\varepsilon_k$  ist normalverteilt.

Um verlässliche Aussagen über die Ergebnisse der Regressionsanalyse machen zu können, müssen diese Prämissen nun auf mögliche Verletzungen hin überprüft werden.

### Zu (a): Nichlinearität

Die hier verwendete Methode der kleinsten Quadrate schätzt nur lineare Modelle. Linearität ist also die Hauptvoraussetzung für alle weiteren Schritte. In den sehr vielen praktischen Fällen kann zumindest für einen Teilbereich der Daten Linearität wenigstens annäherungsweise angenommen werden, jedoch ist auch eine Transformation der unabhängigen Variablen möglich, so dass andere Funktionsverläufe in eine lineare Form konvertiert werden können:

$$Y = \beta_0 + \beta_1 x' + \varepsilon \quad \text{mit } x' = f(x)$$

Die Transformationsfunktionen für  $x$  können vielseitig sein. Zum Beispiel:

$$x' = \ln(x)$$

$$x' = e^x$$

$$x' = x^2$$

Unabhängige Variablen werden also für die OLS-Methode umgewandelt, so dass wieder Linearität vorliegt. Die berechneten Koeffizienten beziehen sich dann jedoch auch auf die  $x'$ -Werte und müssen für eine Interpretation rücktransformiert werden.

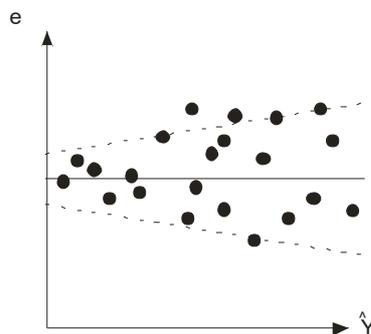
### Zu (b): Unvollständiges Modell

Sind nicht alle Einflussgrößen im Modell enthalten, so kommt es zu systematischen Verzerrungen. Dies ist auf ein unvollständiges theoretisches Wissen zurückzuführen, da schon bei der Modellformulierung wichtige Variablen nicht berücksichtigt wurden. Es liegt ein so genanntes ‚underfittig‘ vor. Falls aus unangebrachter Vorsicht zu viele Variablen Einzug in das Modell gefunden haben, führt dies zwar nicht zu Verzerrungen, jedoch erhält man ineffiziente Schätzer. Zufällige rein statistische Korrelationen können Signifikanzen verdecken.

### Zu (c): Heteroskedastizität

Die Varianz der Störgröße muss konstant sein. Andernfalls würde sie von den unabhängigen Variablen oder der Reihenfolge der Beobachtungen abhängen. Bei langwierigen Beobachtungen kann sinkende Konzentration z.B. zur Zunahme von Messfehlern führen. Dies führt zu ineffizienten Schätzern, zu falschen Standardfehlern der Koeffizienten und somit auch zu falschen Aussagen in den t-Tests.

Heteroskedastizitäten können zum einen durch eine einfache visuelle Inspektion erkannt werden, zum anderen kann der Goldfeld/Quandt-Test Aufschluss bieten:



Der Goldfeld/Quandt-Test nimmt zwei zufällige Unterstichproben und überprüft, ob deren Varianzen gleich groß sind. Das Verhältnis der beiden Varianzen ist F-verteilt und kann somit durch einen Test überprüft werden:

$$H_0 : \sigma_1 = \sigma_2$$

$$F_{emp} = \frac{s_1^2}{s_2^2} \quad \text{mit } s_i^2 = \frac{\sum_{k=1}^{K_i} e_k^2}{K_i - J - 1}; s_1^2 > s_2^2$$

J: Anzahl Regressoren  
K: Anzahl beobachtungen

$$F_{kritisch} = F(K_1 - J - 1; K_2 - J - 1)$$

Wenn der empirische Wert größer als der kritische Wert ist, so wird die Nullhypothese verworfen und Heteroskedastizität liegt vor.

### Zu (d): Autokorrelation

Sind die Störgrößen nicht unabhängig voneinander, spricht man von Autokorrelation. Die Störgröße einer Beobachtung hängt also beispielsweise von der Störgröße der vorangegangenen Beobachtung ab. Dies kann dazu führen, dass die Standardfehler der einzelnen Koeffizienten unterschätzt werden, und der t-Test eine größere Testgenauigkeit vortäuscht.

Zur Aufdeckung von Autokorrelationen kann entweder eine einfache visuelle Inspektion herangezogen werden, oder ein Durbin-Watson-Test durchgeführt werden.

$$d_{emp} = \frac{\sum_{k=2}^K (e_k - e_{k-1})^2}{\sum_{k=1}^K e_k^2}$$

Der Durbin-Watson-Test prüft, ob eine Autokorrelation erster Ordnung vorliegt, also ob ‚benachbarte‘ Beobachtungen voneinander abhängig sind. In der Nullhypothese wird unterstellt, dass keine Autokorrelation vorliegt.

Die empirische Prüfgröße erfasst die Differenzen zwischen zwei aufeinander folgenden Störgrößen. Sind die Differenzen sehr klein, so liegt eine positive Autokorrelation nahe, sind die Differenzen auf der anderen Seite sehr groß, so kann man eine negative Korrelation vermuten. Die Prüfgröße kann Durbin-Watson-verteilte Werte zwischen 0 und 4 annehmen. Je nach dem, in welchen Abschnitt dieses möglichen Wertebereichs die Prüfgröße fällt, kann eine Entscheidung über die Nullhypothese getroffen werden:



Die kritischen Werte  $d_u$  und  $d_o$  sind Durbin-Watson-verteilt mit  $d(K; J; \alpha)$ , wobei K die Fallzahl und J die Zahl der unabhängigen Variablen ist.

### **Zu (e): Multikolarität**

Liegt ein starker linearer Zusammenhang zwischen zwei unabhängigen Variablen vor, so werden die Schätzer zunehmend ineffizient. Es kommt zu ungenauen Standardfehlern und somit zu einem falschen Ergebnis bei den t-Tests. Liegt sogar eine exakte lineare Abhängigkeit vor, so kann das Modell rechnerisch nicht gelöst werden. Zur Überprüfung eignet sich eine Korrelationsmatrix mit Korrelationskoeffizienten nach Bravais-Pearson. Falls auf diesem Wege Korrelationen zum Vorschein kommen, so muss jeweils eine der betroffenen Variablen mit starker Korrelation aus dem Modell entfernt werden. In einigen Fällen kann der Einsatz von Faktoren (siehe Kapitel 4) die betroffenen Variablen zusammenfassen und im Modell ersetzen. In einigen Fällen verringert sich die Korrelationen mit steigendem Stichprobenumfang.

### **Zu (f): fehlende Normalverteilung**

Die Annahme über die Normalverteilung der Störgröße ist eine der wichtigsten im linearen Regressionsmodell. Ohne diese Annahme wären sämtliche t- und F-Statistiken nicht anwendbar. In vielen Fällen kann man jedoch aus sachlogischen Überlegungen von einer Normalverteilung ausgehen. In den übrigen Fällen hilft der zentrale Grenzwertsatz, der besagt, dass sich die Summe zufallsverteilter Variablen mit steigender Fallzahl an die Normalverteilung annähert. In der Praxis kann man eine ausreichende Annäherung ab  $n=30$  unterstellen.

## **3 LOGISTISCHE REGRESSION**

Das Ziel der logistischen Regressionsanalyse ist es die Regression auch für nominalskalierte abhängige Variable zu ermöglichen und somit z.B. eine Gruppenzugehörigkeit auf der Grundlage einer Reihe von unabhängigen Variablen (metrisch skaliert) zu schätzen. Beispielsweise könnte es von Interesse sein, eine Kaufentscheidung (Ja/Nein) zu modellieren oder das präferierte Verkehrsmittel (Bahn/Bus/Flugzeug) zu schätzen.

### **3.1 Binomiale Logistische Regression**

Grundsätzlich wird auch bei der logistischen Regression an der Regressionsgleichung  $b_0 + b_1x_1 + b_2x_2 + \dots$  festgehalten. Allerdings liefert diese Funktion keine nominalskalier-

te Variable als Ergebnis, so dass nicht die Merkmalsausprägung ( $y$ ) direkt geschätzt werden soll, sondern stattdessen die Wahrscheinlichkeit dieser bestimmten Merkmalsausprägung ( $y^* = P(y=1)$ ). Ist diese Wahrscheinlichkeit genügend hoch, so kann die Zugehörigkeit zu einer bestimmten Ausprägung angenommen werden. Da diese Wahrscheinlichkeit nicht beobachtet werden kann (sondern nur die konkrete Ausprägung von  $y$ ), spricht man von einer latenten Variable. Wahrscheinlichkeiten können nur Werte zwischen 0 und 1 annehmen, daher bedarf es einiger Umformungen um sie für eine Regressionsfunktion nutzbar zu machen. Man setzt als abhängige Variable nicht das  $y^*$  ein, sondern nimmt zunächst das Chancenverhältnis („Odd“), welches sich aus dem Quotienten von Wahrscheinlichkeit zu Gegenwahrscheinlichkeit ergibt. Dies ist nur eine andere Schreibweise für eine Wahrscheinlichkeit, wie man sie eventuell aus dem Wettgeschäft kennt („Die Chancen stehen 2:4 für Silberblitz“). Allerdings erweitert sich der Wertebereich durch diese Umformung schon einmal auf  $[0; +\infty]$ . Nun kann man dieses Chancenverhältnis logarithmieren und erhält das so genannte ‚Logit‘, welches einen Wertebereich von  $-\infty$  bis  $+\infty$  hat und somit von der Regressionsgleichung geschätzt werden kann:

$$\ln\left(\frac{P(y_i = 1)}{1 - P(y_i = 1)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$$

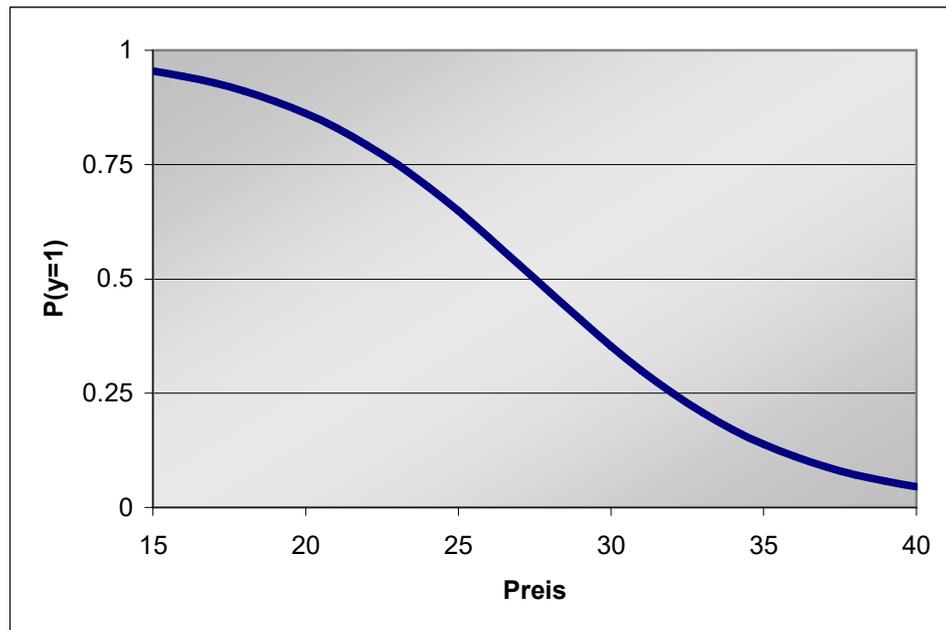
Dieser Term kann nun nach dem interessierenden  $y^*$  – der Wahrscheinlichkeit das  $y=1$  ist – aufgelöst werden:

$$y^* = P(y_i = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots)}}$$

Neben dieser mathematisch-praktischen Sichtweise können aber auch inhaltliche Gründe für die Wahl des logarithmischen Ansatzes angeführt werden. Bei der linearen Regression galt die Annahme, dass die Störgröße normalverteilt ist. Da sich die Verteilungsfunktionen der Normalverteilung und der logarithmischen Verteilung stark ähneln, ist die Annahme erlaubt, dass die Störgrößen annähernd logarithmisch verteilt sind. Somit darf insgesamt eine logarithmische Verteilung angenommen werden.

### 3.1.1 Interpretation der Koeffizienten

Über einen Maximum-Likelihood-Ansatz können nun die Koeffizienten auf der Grundlage der Beobachtungen geschätzt werden (dazu im nächsten Abschnitt mehr). Anders als bei der linearen Regressionsanalyse können die gefundenen Koeffizienten jedoch nicht so anschaulich interpretiert werden, da der Funktionsverlauf nicht mehr linear ist, und eine Änderung der unabhängigen Variable um eine Einheit an jeder Stelle der Funktion andere Änderungen der abhängigen Variable verursacht.



Lediglich das Vorzeichen kann als Richtung des Zusammenhangs interpretiert werden. Beispielsweise kann in der obigen Abbildung nur gesagt werden, dass mit höherem Preis die Wahrscheinlichkeit eines Kaufes fällt. Eine allgemeingültige Aussage über die Größe der Veränderung ist jedoch nicht möglich. Steigt der Preis um einen Euro von 15 auf 16 Euro, so sinkt die Kaufbereitschaft um 1,1 Prozentpunkte. Steigt der Preis von einem anderen Preisniveau, beispielsweise von 28 ebenfalls um einen Euro auf 29, so fällt die Wahrscheinlichkeit des Kaufes um 6,6 Prozentpunkte.

Für das Chancenverhältnis lässt sich aus den Koeffizienten zwar eine konstante Änderungsrate herleiten ( $e^b$ ), jedoch ist eine Interpretation dieser Änderungsrate schwierig.

### 3.1.2 Beurteilung der Güte des Gesamtmodells

Um die Güte des Gesamtmodells beurteilen zu können, müssen wir noch einmal einen Blick auf die Entstehung der Koeffizienten werfen. Wie bereits erläutert sind die Koeffizienten das Ergebnis einer Optimierung einer Maximum-Likelihood-Funktion. Hierbei sollen diejenigen  $b$  gefunden werden, die eine größtmögliche Trennung zwischen den beiden Gruppen verursachen:

$$L = \prod_{\text{für } y_i=1} P(y_i=1) \cdot \prod_{\text{für } y_i=0} (1-P(y_i=1))$$

$$= \prod_{k=1}^K \left( \frac{1}{1+e^{-(\beta_0+\beta_1x_1+\beta_2x_2+\dots)}} \right)^{y_k} \left( 1 - \frac{1}{1+e^{-(\beta_0+\beta_1x_1+\beta_2x_2+\dots)}} \right)^{1-y_k}$$

Es werden alle geschätzten Wahrscheinlichkeiten für Gruppe 1 miteinander multipliziert bei denen die Gruppe 1 auch tatsächlich beobachtet wurde und anschließend mit dem Produkt aller Wahrscheinlichkeiten für Gruppe 0 bei denen die Gruppe 0 beobachtet wurde. Auf diese Weise geht L gegen 1, wenn die geschätzten Wahrscheinlichkeiten möglichst nahe an der tatsächlich beobachteten Ausprägung der

abhängigen Variable liegen. Durch die Maximierung von L mit dem Maximum-Likelihood-Ansatz erhält man diejenigen Koeffizienten, die die Gruppen maximal trennen.

Man weiß also, dass es keine anderen Koeffizienten gibt, die für das Datenmaterial der Stichprobe eine bessere Trennung produzieren würden, jedoch muss trotzdem die Güte des Modells untersucht werden um festzustellen, ob das Modell tatsächlich geeignet ist. Hierzu verwendet man Gütemaße, die häufig auf dem logarithmierten Likelihood-Wert aufbauen – dem Log-Likelihood-Wert.

**(a) Devianz**

Die Devianz errechnet sich aus dem mit -2 multiplizierten Log-Likelihood-Wert. Sie ist umso kleiner, je besser das Modell ist. Da sie annähernd  $\chi^2$ -verteilt ist, kann leicht ein Test durchgeführt werden. Jedoch wird die Größe des LL-Wertes nicht nur von der reinen Trennkraft des Modells beeinflusst, sondern auch von der Verteilung der Beobachtungen auf die Gruppen. Wenn in einem extremen Beispiel nur ein Fall in Gruppe 0, jedoch 99 Fälle in Gruppe 1 vorliegen, so kann man allein durch das  $b_0$  eine hohe Trennqualität erreichen. Die Devianz berücksichtigt dieses Phänomen nicht, und wird deshalb nur selten verwendet, da im Grunde genommen beide Gruppen gleich stark besetzt sein müssten, um den Test durchführen zu können.

**(b) Likelihood-Ratio-Test**

Der LR-Test vergleicht den Likelihood-Wert nicht mit einem absoluten Wert, sondern mit dem L-Wert der sich aus einem Modell ohne Koeffizienten (Null-Modell, restringiertes Modell) ergeben hätte. Liegen die beiden zu vergleichenden Likelihood-Werte nah beieinander, so tragen die Koeffizienten, die nur in dem einen Modell zu finden sind, nicht bedeutsam zur Erklärung der latenten Variablen bei, und der Einfluss der unabhängigen Variablen muss in Frage gestellt werden:

$$H_0: L_u = L_r$$

(der Likelihood-Wert des unrestringierten Modells ist gleich dem des restringierten Modells und somit sind  $\beta_j = 0$  mit  $j=1,2,\dots,J$ )

Der mit J Freiheitsgraden annähernd  $\chi^2$ -verteilte Prüfwert des LR-Testes berechnet sich wie folgt:

$$LR = -2 \cdot \ln \frac{L_r}{L_u} = -2 \cdot (LL_r - LL_u)$$

Die meisten Statistikprogramme geben auch bei diesem Test ein empirisches Signifikanzniveau aus. Der Likelihood-Ratio-Test führt also eine Überprüfung der Signifikanz des Modells durch, deren Ergebnisse auch auf die Grundgesamtheit übertragen werden können.

### (c) McFadden's $R^2$

Ähnlich wie der LR-Test benutzt auch McFadden's Pseudo  $R^2$  die LL-Werte des unrestringierten und des restringierten Modells. Jedoch wird lediglich ein Maß für die Trennkraft des Modells berechnet. Ein Test, der eine Aussage über die Grundgesamtheit zulässt, wird nicht durchgeführt. Somit ist diese Größe eher zum Vergleich zweier Modelle in Bezug auf Ihre Trennkraft geeignet als eine allgemeingültige Aussage in Bezug auf die Grundgesamtheit zu treffen.

$$R_{McFadden}^2 = 1 - \frac{LL_u}{LL_r}$$

### (d) Count $R^2$

Dieser Parameter bildet eine Quote der mit dem Modell korrekt vorgenommenen Gruppenzuordnungen für die Stichprobe und ist somit ein sehr einfaches, leicht nachzuvollziehendes Maß für die Nutzbarkeit des Modells. Bei der Beurteilung des  $R^2$  muss ein Vergleich mit einer rein zufälligen Gruppenzuordnung angestellt werden – das Modell soll schließlich besser sein als der bloße Zufall. Im binären Fall beispielsweise käme eine rein zufällige Gruppenzuordnung zu einer Trefferquote von 50%.

### 3.1.3 Ausreißerdiagnostik

Sollte sich herausstellen, dass die Gesamtgüte des Modells nicht ausreichend ist, so kann man zu dem Schluss gelangen, dass das gewählte Modell die Wirklichkeit nicht genau nachbilden kann und somit unpassend ist. In diesem Falle müsste man seine theoretischen Vorüberlegungen überdenken und das Modell verbessern bzw. neu aufbauen. Allerdings ist es auch möglich, dass einige wenige Ausreißer für die schlechten Ergebnisse der Überprüfung der Modellgüte verantwortlich sind. Dies kann zum einen ein Hinweis darauf sein, dass das Modell nicht genug spezifiziert ist und eventuell Einflussfaktoren fehlen, die für auffällige Ausprägungen der Ausreißer verantwortlich sind. Allerdings kann es auch möglich sein, dass ein atypisches Antwortverhalten vorliegt, wenn z.B. die Fragen des Fragebogens nicht richtig verstanden wurden oder manche Fragen nicht ernsthaft ausgefüllt wurden. In diesem Falle kann man in Betracht ziehen, diese offensichtlich falschen Beobachtungen aus der Untersuchung auszuschließen. Hierzu müssen die Ausreißer identifiziert werden. Dazu eignen sich am besten die ‚Residuen‘ – in diesem Falle Abweichungen zwischen der latenten Variablen und den beobachteten Werten der abhängigen Variablen. Um eine klarere Identifikation zu ermöglichen werden die Residuen zu den Pearson-Residuen normiert:

$$r_k = \frac{y_k - y_k^*}{\sqrt{y_k^* \cdot (1 - y_k^*)}}$$

Liegt ein Pearson-Residuum betragsmäßig über eins, so handelt es sich um einen Ausreißer, der eventuell entfernt werden kann.

### 3.1.4 Beurteilung einzelner Koeffizienten

Wie bei der linearen Regression auch, können die einzelnen Koeffizienten hinsichtlich Ihres Erklärungsbeitrags untersucht werden.

Zum einen kann man einen Likelihood-Ratio-Test für einzelnen Koeffizienten durchführen. In diesem Falle wird das unrestringierte Modell jedoch nicht mit einem komplett restringierten Modell verglichen, sondern mit einem Modell, welches um den zu testenden Koeffizienten gekürzt wurde (Reduced Coefficient). Der empirische LR-Wert ist wiederum  $\chi^2$ -verteilt und berechnet sich analog zu dem LR-Test für das Gesamtmodell aus der Formel:

$$LR = -2 \cdot (LL_{rc} - LL_u)$$

Gängiger ist es jedoch einen Wald-Test durchzuführen, der stark an den t-Test des linearen Modells angelehnt ist. Mit Hilfe der Nullhypothese  $H_0: \beta_j=0$  können einzelne Koeffizienten auf Ihre Signifikanz überprüft werden. Die Testgröße ist  $\chi^2$ -verteilt mit  $J$  Freiheitsgraden und berechnet sich aus

$$W = \left( \frac{\beta_j}{s_{\beta_j}} \right)^2$$

Die Koeffizienten und der Standardfehler können oft aus den Ergebnissen der Statistikprogramme entnommen werden. Üblich ist jedoch auch, dass der Wald-Wert und ein empirisches Signifikanzniveau hierfür direkt mit ausgegeben wird.

## 3.2 Multinomiale logistische Regression

Oft liegen bei nominal skalierten Variablen nicht nur zwei, sondern eine Vielzahl von Ausprägungen vor. In diesem Falle reicht das in Kapitel 3.1 vorgestellte binomiale Logit Modell nicht aus, sondern muss auf den Mehr-Gruppen-Fall erweitert werden.

Bislang wurde durch die Regression das Logit geschätzt zur Gruppe 1 und nicht zur Gruppe 0 zu gehören:

$$\ln \left( \frac{P(y_i = 1)}{1 - P(y_i = 1)} \right) = \ln \left( \frac{P(\text{Gruppe 1})}{P(\text{Gruppe 0})} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$$



Welches die Gruppe 1 und welches die Gruppe 0 ist, ist letztendlich eine Definitionsfrage, die von dem Forscher festgelegt wird. Es ist also genauso denkbar, Koeffizien-

ten für den umgekehrten Fall zu schätzen: für die Wahrscheinlichkeit zur Gruppe 0 zu gehören und dementsprechend nicht zur Gruppe 1.

**Gruppe 0** **Gruppe 1**



$$\ln\left(\frac{1-P(y_i=1)}{P(y_i=1)}\right) = \ln\left(\frac{P(\text{Gruppe 0})}{P(\text{Gruppe 1})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$$

Bei einer solchen Richtungsänderung der Schätzung vertauschen sich lediglich die Vorzeichen der Koeffizienten, die Beträge bleiben natürlich gleich. Es macht also für die praktischen Ergebnisse keinen Unterschied, ob die eine oder die andere Richtung gewählt wird.

Kommen nun weitere Gruppen hinzu, so müssen Logits für die einzelnen ‚Paarverbindungen‘ geschätzt werden. Auf diese Weise kommt man zu latenten Variablen, die jeweils angeben, wie Wahrscheinlich es ist in die eine und nicht in die andere Gruppe des Paares zu gehören. Dieser Wahrscheinlichkeitswert gilt nur in der bilateralen Beziehung der beiden Gruppen und sagt nichts über Zugehörigkeit zu anderen Gruppen aus. Man erhält also mehrere Sätze von Koeffizienten, die jeweils für eine Paarbeziehung gelten.

**Gruppe 1** **Gruppe 2**



**Gruppe 3**

$$\ln\left(\frac{P(\text{Gruppe 1})}{P(\text{Gruppe 3})}\right) = \beta_{01} + \beta_{11} x_1 + \beta_{21} x_2 + \dots$$

$$\ln\left(\frac{P(\text{Gruppe 2})}{P(\text{Gruppe 3})}\right) = \beta_{02} + \beta_{12} x_1 + \beta_{22} x_2 + \dots$$

Der Übergang von der ersten zur zweiten Gruppe braucht nicht selbst geschätzt werden, da diese Wahrscheinlichkeit über die beiden anderen Beziehungen konstruierbar ist:

$$\ln\left(\frac{P(\text{Gruppe 1})}{P(\text{Gruppe 2})}\right) = \ln\left(\frac{P(\text{Gruppe 1})}{P(\text{Gruppe 3})}\right) - \ln\left(\frac{P(\text{Gruppe 2})}{P(\text{Gruppe 3})}\right)$$

Es ist also nur nötig die Koeffizienten der Logits einer Referenzgruppe zu schätzen (in diesem Falle Gruppe 3). Nach wie vor lassen sich die Koeffizienten nur hinsichtlich ihres Vorzeichens interpretieren, und zwar in der Weise, dass mit steigendem  $x$  die Wahrscheinlichkeit zur geschätzten Gruppe zu gehören (und nicht zur Referenzgruppe) steigt (bei positivem Koeffizienten) bzw. sinkt (bei negativem Koeffizienten).

Durch das Maximum-Likelihood-Verfahren erhält man alle Koeffizienten um die Wahrscheinlichkeit für eine bestimmte Gruppenzugehörigkeit zu schätzen:

$$P(y = i) = \frac{e^{x_k \beta_i}}{1 + \sum_{q=1}^{m-1} e^{x_k \beta_q}} \quad \begin{array}{l} i = \text{Gruppe} \\ m = \text{Anzahl Gruppen} \end{array}$$

Im Drei-Gruppen-Fall sieht dies beispielsweise so aus:

$$P(y = 1) = \frac{e^{\beta_{01} + \beta_{11}x_1 + \beta_{21}x_2}}{1 + e^{\beta_{01} + \beta_{11}x_1 + \beta_{21}x_2} + e^{\beta_{02} + \beta_{12}x_1 + \beta_{22}x_2}}$$

Auf diese Weise kann die Wahrscheinlichkeit der Zugehörigkeit zu jeder Gruppe bestimmt werden. Diejenige Gruppe, für die  $y_i^*$  am größten ist, wird für  $\hat{y}$  gewählt.

Zur Beurteilung der Gesamtmodellgüte und der einzelnen Koeffizienten stehen dem multinomialen Modell die gleichen Parameter zur Verfügung wie im binomialen Fall. Allerdings ist bei der Beurteilung der einzelnen Koeffizienten auffällig, dass nun für jede unabhängige Variable mehrere Koeffizienten (nämlich für jedes Logit eine Koeffizient) geschätzt werden, und daher für jede unabhängige Variable auch mehrere Wald-Tests durchgeführt werden, die jeweils die Trennung zwischen der jeweiligen Gruppe und der Referenzgruppe beurteilen.

## 4 FAKTORENANALYSE

Anders als die bisher kennen gelernten Verfahren der Regressionsanalyse, in denen festgelegte Strukturen überprüft und quantifiziert wurden, handelt es sich bei der Faktorenanalyse um ein Struktur- entdeckendes Verfahren, bei dem die vorhandenen Variablen zu Faktoren verdichtet werden sollen. Auf diese Weise wird das Datenmaterial zur Interpretation übersichtlicher, Redundanzen durch Multikollinearitäten werden beseitigt, so dass Faktoren anstelle der eigentlichen Variablen als unabhängige Größen in eine Regressionsanalyse einfließen können.

Als durchgehendes Beispiel sollen für dieses Kapitel die Bewertungen (Punkte von 1 bis 10) von 500 Kunden bezüglich ihrer Computer dienen. Es wurden 6 Merkmale erhoben ( $x_1$  bis  $x_6$ ), die die Geräusentwicklung, die Gehäusegröße und das Gewicht des Gehäuses, sowie die Prozessorleistung, die Festplattenkapazität und die Größe des Arbeitsspeichers erfassen. Insgesamt waren unter den Beobachtungen 10 Computermodelle, die in den Auswertungen verglichen werden sollen. Daher wurden für die 10 Modelle jeweils die durchschnittlichen Bewertungen berechnet, so dass nun eine 10x6 Matrix für die Daten vorliegt:

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
Apple	10	8	10	6	4	3
Sony	10	9	7	8	7	9
Aldi	3	2	1	7	8	9
...						

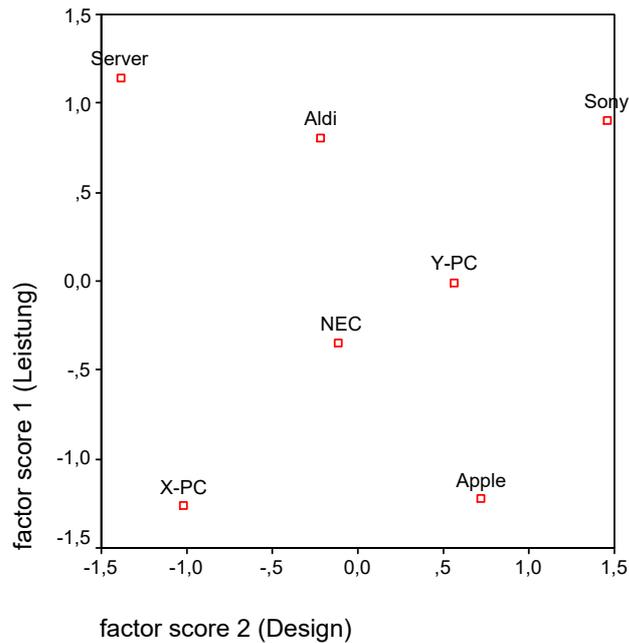
Wie eben bereits geschildert, ist es das Ziel der Faktorenanalyse die Variablen zusammenzufassen, die miteinander zusammenhängen. Spricht man im statistischen Sinne von Zusammenhängen, so sind dies stets Korrelationen. Korrelationen können die Stärke und die Laufrichtung eines Zusammenhangs angeben, jedoch keine kausale Richtung, so dass verschiedene Interpretationen möglich sind. Beispielhaft soll hier der Zusammenhang zwischen der Geräusentwicklung und der Gehäusegröße betrachtet werden:

- (1) Gehäuse  $\rightarrow$  Geräusch      Durch ein kleines Gehäuse muss der PC stärker gekühlt werden, wodurch mehr Lüftergeräusche entstehen.
- (2) Gehäuse  $\leftarrow$  Geräusch      Da man einen möglichst leisen PC bauen möchte, wurde ein großes Gehäuse gewählt, das gut schallgedämmt ist.

Zudem gibt es jedoch auch eine dritte Möglichkeit, deren Annahme die Grundlage der Faktoranalyse ist:

- (3) Gehäuse  $\leftarrow$  dahinter stehende Größe  $\rightarrow$  Geräusch  
 Eine dritte, hinter den beiden beobachteten Variablen stehende Größe, beeinflusst diese. Eine gleichzeitige Änderung der Größe und Lautstärke hat liegt in der Veränderung dieser dritten Größe begründet.

Diese ‚dahinter stehende Größe‘ sind die Faktoren, die gefunden werden sollen, um die Variablen, die durch diese Größe beeinflusst werden, zu beschreiben. In dem genannten Beispiel könnte diese Größe ‚Design‘ heißen. Ein Computerkäufer hat eine bestimmte Vorstellung über das Design des Computers, zum Beispiel abhängig davon, ob er diesen in einem Büro oder im Wohnzimmer einsetzen möchte. Diese Vorstellung lässt ihn dementsprechend eine passende Gehäusegröße und Lautstärke wählen. Des Weiteren könnte der Käufer den Prozessor, den Arbeitsspeicher und die Festplatte nach Leistungskriterien auswählen. Eine Faktoranalyse käme dann zum Beispiel zu folgendem Bild:



Selbstverständlich bedeutet eine Informationsverdichtung immer auch ein gewisses Maß an Informationsverlust. In manchen Situationen ist es jedoch sinnvoll, diesen kalkulierten Informationsverlust, den man bei der Faktoranalyse auch konkret beziffern kann, hinzunehmen.

Für die folgenden Rechenoperationen empfiehlt es sich, die ursprünglichen (X-) Variablen in ein standardisiertes Format zu bringen. Eine solche Normierung zwischen 0 und 1 vereinfacht zum einen die Rechenschritte, erlaubt eine einfachere Interpretation und ermöglicht den Vergleich zwischen den Varianzen der verschiedenen Variablen. Daher wird im Folgenden anstelle der ursprünglichen Variablen die standardisierte Variable z verwendet:

$$z_{kj} = \frac{x_{kj} - \bar{x}_j}{s_j}$$

k=Beobachtung k  
j=Merkmal j

## 4.1 Variablenauswahl

Bevor mit der Faktoranalyse begonnen werden kann ist es notwendig diejenigen Variablen zu identifizieren, die für eine Verdichtung in Frage kommen. Eine Korrelationsmatrix zeigt an, welche Variablen mit anderen Variablen genügend stark korrelieren. Variablen, die keine Verbindung zu anderen Variablen haben, können aus der Faktoranalyse ausgeschlossen werden. Die gefundenen Korrelationskoeffizienten lassen sich gegen die Nullhypothese  $r=0$  (kein Zusammenhang) testen, und lassen somit auch eine Beurteilung hinsichtlich der Grundgesamtheit zu.

Neben dem sehr detaillierten Mittel der Korrelationsmatrix gibt es auch weitere Maße, die bei der Variablenauswahl hilfreich sein können. So z.B. der Barlett-Test, der

die Korrelationsmatrix gegen die Einheitsmatrix testet. Kann diese Hypothese abgelehnt werden, so kann man von vorhandenen Korrelationen unter den Variablen ausgehen.

Der Anti-Image-Test berechnet die Anteile der Varianz, die unabhängig von einem systematischen Einfluss ist. Sind die Anti-Image-Werte klein, so kann von Korrelationen ausgegangen werden.

Eine Maßzahl, welches das Datenmaterial als Ganzes beurteilt, ist das Kaiser-Meyer-Olkin Kriterium (auch MSA – Measure of Sampling Adequacy – genannt). Auf einen Blick lässt sich die Eignung der Variablen grob anhand folgender Skala einschätzen:

KMO-Wert	Beurteilung
<0,5	untragbar
≥0,5	kläglich
≥0,6	mittelmäßig
≥0,7	ziemlich gut
≥0,8	verdienstvoll
≥0,9	erstaunlich

## 4.2 Extraktion der Faktoren

Der Grundsatz der Faktorenanalyse legt fest, dass jeder Beobachtungswert als Linearkombination mehrerer (zunächst hypothetischer) Faktoren beschrieben werden kann. Aus den Faktoren lassen sich also die Merkmalsausprägungen der ursprünglichen Variablen wiederherstellen.

$$z_{kj} = \sum_{i=1}^I a_{ji} p_{ki}$$

k=Beobachtung k

j=Merkmal j

i=Faktor i

a=Faktorladung

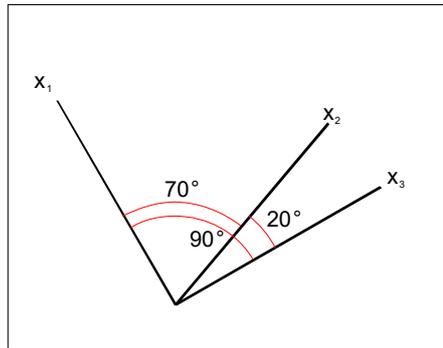
p=Faktorwert

Ein ursprünglicher Wert lässt sich also aus dem Produkt der Faktorladungen mit den jeweiligen Faktorwerten aller Faktoren darstellen. Die Faktorladungen geben das Gewicht, die Bedeutung des Faktors für die ursprüngliche Variable an. Es geht also um einen Zusammenhang zwischen Faktor und ursprünglicher Variable. Zusammenhänge lassen sich, wie bereits bei den grundlegenden Überlegungen zu diesem Kapitel erwähnt, statistisch gesehen als Korrelationen ausdrücken. Es handelt sich bei der Faktorladung also um den Korrelationskoeffizienten zwischen Variable und Faktor.

### 4.2.1 Faktorladungen

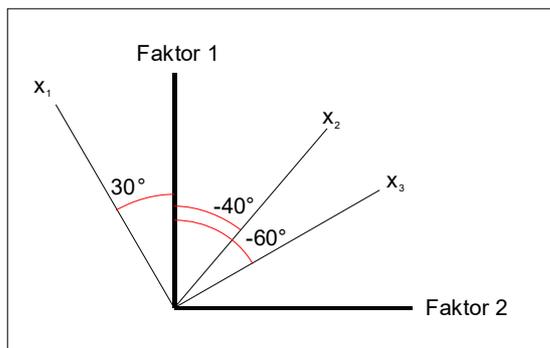
Diesen Grundgedanken kann man graphisch veranschaulichen, wenn man die Korrelationen in einem Vektordiagramm als Winkel darstellt. Über die Arcus Kosinus-Funktion ( $\cos^{-1}$ ) lassen sich Korrelationskoeffizienten in Winkel umrechnen. Ein

Winkel von  $0^\circ$  entspricht also einer vollen Korrelation ( $r=1$ ), während ein Winkel von  $90^\circ$  einer völligen Unabhängigkeit entspricht ( $r=0$ ).



Ziel der Faktorenanalyse ist es, das Verhältnis der Variablen zueinander in einem möglichst gering dimensionierten Raum darzustellen. Es können also nun nach und nach Faktoren erzeugt werden, bis die Variablen mit diesen hinreichend gut dargestellt werden können. Der erste Faktor entspricht der ‚Richtung‘, die allen vorhandenen Variablen am meisten gerecht wird. Würde man Seile an einen Gegenstand binden, und an diesen gleichmäßig in die Richtungen der Variablen ziehen, so würde sich der Gegenstand in Richtung des Faktors bewegen.

Der zweite Faktor wird nun auf der Basis der Überlegung eingeführt, dass die Faktoren unter sich keine Zusammenhänge haben sollen um Redundanzen zu vermeiden. Völlige Unabhängigkeit entspricht einem Winkel von 90 Grad:



Auf diese Weise kann nun eine Tabelle mit den Faktorladungen (jeweils als Winkel und Koeffizient) erstellt werden:

	<b>Faktor 1</b>	<b>Faktor 2</b>
$x_1$	$30^\circ$ 0,87	$-120^\circ$ -0,50
$x_2$	$-40^\circ$ 0,77	$-50^\circ$ 0,64
$x_3$	$-60^\circ$ 0,50	$-30^\circ$ 0,87

In diesem Beispiel können alle Variablen durch nur zwei Faktoren komplett dargestellt werden, da sich die Variablen alle auf einer zweidimensionalen Ebene befinden. Würde ein Vektor der Variablen etwas aus dem Papier ‚herausragen‘, also zu einem gewissen Anteil in eine dritte Dimension zeigen, so wäre ein dritter Faktor notwendig, um die Variablen wieder 100-prozentig beschreiben zu können.

#### 4.2.2 Erklärungsanteil der Variablen / Kommunalitäten

Wie viel von den jeweiligen Variablen durch die Faktoren erklärt werden kann, wird durch die Kommunalitäten gemessen. Wird eine Variable vollständig erklärt, so gilt:

$$\text{Kommunalität}_j = \sum_{i=1}^I a_{ij}^2 = 1 \quad \begin{array}{l} j=\text{Merkmal } j \\ i=\text{Faktor } i \end{array}$$

Da jedoch möglichst wenige Faktoren gebildet werden sollen, können die Variablen in aller Regel nicht vollständig erklärt werden. Dieser Umstand kann auf zwei Weisen interpretiert werden: Die Hauptkomponentenanalyse geht davon aus, dass die Variablen durch eine entsprechende Anzahl von Komponenten (im Fall der Hauptkomponentenanalyse spricht man nicht von Faktoren – rechnerisch gibt es jedoch keinen Unterschied zu Faktoren) vollständig erklärt werden können. Werden weniger Komponenten extrahiert, so wird der daraus entstehende Informationsverlust bewusst zugunsten der Informationsverdichtung hingenommen. Die Komponenten können dann als „Sammelbegriff“ für diejenigen Variablen interpretiert werden, die hoch auf den jeweiligen Faktor laden.

Die Hauptachsenanalyse unterstellt den Variablen ein gewisses Maß an Restvarianz. Man geht also davon aus, dass die Variablen zwar durch die Faktoren bestimmt werden, jedoch nicht vollständig von diesen abhängen. Es bleibt neben der Kommunalität eine Restvarianz, die nicht durch die Faktoren erklärt werden kann. Im Falle der Hauptachsenanalyse wird daher ein Ausgangsvarianzniveau von vornherein festgelegt; der Forscher gibt also aufgrund inhaltlicher Überlegungen die Erklärungskraft der Faktoren für die Variablen vor. Die gefundenen Faktoren werden kausal interpretiert, d.h. als Ursache für die hohen Ausprägungen der ursprünglichen Merkmale (z.B. der Faktor „Choleriker sein“ ist die *Ursache* für „hohen Blutdruck“ und „rote Gesichtsfarbe“ – allerdings wird z.B. der Blutdruck nicht allein und vollständig durch die Neigung zu Wutausbrüchen erklärt. Die Kommunalität wird also unter 1 liegen).

#### 4.2.3 Zahl der Faktoren / Eigenwerte

Wie oben bereits beschrieben, können mit jedem zusätzlichem Faktor die ursprünglichen Merkmale besser erklärt werden. Spätestens wenn ebenso viele Faktoren wie ursprüngliche Merkmale extrahiert wurden, können die Merkmale zu 100 Prozent durch die Faktoren ohne Informationsverluste ersetzt werden. Daher stellt sich die Frage, wie viele Faktoren sinnvoll sind.

Eine Möglichkeit ist, die Zahl der Faktoren von vornherein aus inhaltlichen oder pragmatischen Gründen (etwa der Wunsch 2-dimensionale Diagramme erstellen zu können) festzulegen.

Sinnvoller ist es jedoch das Kaiser-Kriterium anzuwenden, welches über die Höhe der Eigenwerte entscheidet wie viele Faktoren zu extrahieren sind. Als Eigenwert

bezeichnet man die Summe der quadrierten Faktorladungen eines Faktors über alle ursprünglichen Variablen:

$$EV_i = \sum_{j=1}^J a_{ij}^2$$

j=Merkmal j  
i=Faktor i

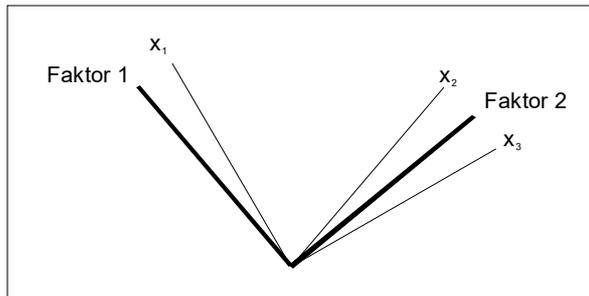
Die Eigenwerte geben das Verhältnis der Erklärungskraft eines einzelnen Faktors im Verhältnis zu einer ursprünglichen Variablen (theoretischer EV=1) an. Die Summe der Eigenwerte aller möglichen Faktoren (wenn ebenso viele Faktoren extrahiert werden, wie ursprüngliche Merkmale vorhanden sind) entspricht der Summe der Eigenwerte der ursprünglichen Merkmale. Ist der Eigenwert eines Faktors also größer als eins, so vermag er mehr Varianz zu erklären als eines der ursprünglichen Merkmale. In diesem Falle wird der Faktor nach dem Kaiser-Kriterium extrahiert.

Auf den Eigenwerten baut auch der Scree-Plot auf, der die Zahl der Faktoren gegen die Eigenwerte in einem Diagramm abträgt. In diesem Diagramm ist ein Knick zu suchen, der den Berg vom Geröll (Scree) trennt. Für die Faktoranalyse sollen nach diesem Kriterium nur Faktoren aus der „Bergregion“ extrahiert werden.

### 4.3 Faktorinterpretation

Nachdem die Anzahl der Faktoren festgelegt und die Faktoren extrahiert wurden, erfolgt nun die Interpretation der Faktoren je nach dem Verfahren der Hauptkomponenten- oder Hauptachsenanalyse aus der Sachlogik heraus. Dabei werden diejenigen Variablen einem Faktor zugeordnet, die hoch auf einem Faktor laden. In der Regel geht man von einer Faktorladung von mindestens 0,5 aus. Diese Merkmale bestimmen dann die Interpretation des Faktors.

Allerdings ist die Interpretation ohne weitere Transformationen schwierig, da die Extraktion der Faktoren lediglich auf rein mathematische Weise erfolgte ohne die Sachlogik zu berücksichtigen, und somit nicht für die Interpretation optimiert wurde. Es ist jedoch möglich, die alle Faktoren gemeinsam zu rotieren, so dass die Faktoren möglichst gut mit Gruppen von Variablen übereinstimmen. Durch diese Rotation werden zwar die Faktorladungen verändert, jedoch bleiben die Kommunalitäten und Verhältnisse der Faktoren untereinander konstant. Man kann diesen Vorgang mit der Drehung eines Achsenkreuzes vergleichen: werden X- und Y-Achse in einem Diagramm leicht gedreht, so kann man die in dem Diagramm liegenden Punkte trotzdem genauso gut wie zuvor beschreiben. Die Eigenwerte werden durch die Rotation ebenfalls beeinflusst, jedoch bleibt die Summe der erklärten Varianz bei der festgestellten Faktorzahl ebenfalls unverändert. Die Rotation ist also ein reines Hilfsmittel zur Interpretationserleichterung.



Die tatsächliche Rotation kann in den Ausgaben des Statistikprogramms SPSS an der Hauptdiagonale der Rotationsmatrix abgelesen werden. Dort findet man den Kosinuswert des Rotationswinkels.

#### 4.4 Bestimmung der Faktorwerte

Die Faktorwerte können als Ausprägungen der Faktoren bei bestimmten Merkmals-trägern gesehen werden. Wie viel „Leistung“ steckt beispielsweise in einem Apple-Computer, wie viel „Design“ bietet ein Sony-Notebook?

In den meisten Fällen können die Faktorwerte aus den gefundenen Faktoren und den ursprünglichen Merkmalsausprägungen von den Statistikprogrammen intern durch eine einfache Regressionsschätzung (beispielsweise nach der Methode der kleinsten Quadrate) berechnet und zu den ursprünglichen Daten hinzugefügt werden.

Mit den Regressionskoeffizienten lassen sich die Faktorwerte einfach berechnen:

$$p_{ki} = \sum_{j=1}^J z_{kj} b_{kji}$$

- p=Faktorwert
- b=Regressionsfaktor
- k=Merkmalsträger
- i=Faktor i
- j=Merkmal j

Mit Hilfe dieser Werte lassen sich dann z.B. übersichtliche Grafiken (Faktorwert-plots) erstellen, die die Daten mit wenigen Dimensionen abbilden, da j Merkmale nun auf i Faktoren reduziert wurden.

## 5 CLUSTERANALYSE

Die Clusteranalyse ist ein Verfahren um in einer heterogenen Grundgesamtheit ho-mogene Gruppen zu entdecken. Dabei sollen die Gruppen so zusammengesetzt werden, dass innerhalb der Gruppen möglichst viel Ähnlichkeit zwischen den Grup-penelementen besteht und sich die Gruppen gegeneinander möglichst gut abgrenzen.

Auf diese Weise ist es zum Beispiel möglich eventuell vorhandene verschiedene Persönlichkeitstypen in einer Zielgruppe zu identifizieren oder Testmärkte in Gruppen aufzuteilen.

## 5.1 Proximitätsmaße

Das wesentliche Element der Clusteranalyse ist die Beurteilung von Ähnlichkeiten zwischen Objekten (Merkmalsträgern). Diese Ähnlichkeiten zwischen allen Merkmalsträgern lassen sich in einer Distanz- oder Ähnlichkeitsmatrix abtragen, die die Grundlage der Gruppierung bildet. Bei der Wahl dieses Proximitätsmaßes ist zunächst das Skalenniveau der zu Grunde liegenden Merkmale ausschlaggebend.

### 5.1.1 Proximitätsmaße für metrische Skalen

Für metrische Skalen kommen im Wesentlichen zwei Proximitätsmaße in Frage. Zum einen kann die Distanz zwischen zwei Merkmalsträgern über die Minkowski-Metrik bestimmt werden. Die Distanz ist immer dann von Interesse, wenn die absolute Entfernung zwischen zwei Objekten ausschlaggebend für die Beurteilung der Ähnlichkeit ist. Ganz anders arbeiten Ähnlichkeitsmaße wie der Q-Korrelationskoeffizient. Hier werden große Ähnlichkeiten festgestellt, wenn Gemeinsamkeiten im Profil (unabhängig von der Distanz) vorhanden sind. Möchte man Gruppen von Unternehmen bilden, die in den letzten Jahren die gleiche Umsatzentwicklung erfahren haben, so wählt man den Q-Korrelationskoeffizienten. Ist lediglich die Höhe des Umsatzes von Interesse, so wählt man die Minkowski-Metrik, die kleine und große Unternehmen (unabhängig von der Umsatzentwicklung) in verschiedene Gruppen einteilt.

#### Minkowski-Metrik

$$d_{k,l} = \left( \sum_{j=1}^J |x_{kj} - x_{lj}|^r \right)^{\frac{1}{r}}$$

d=Distanz

k=Merkmalsträger k

l=Merkmalsträger l

j=Merkmal j

r=Minkowski-Konstante ( $r \geq 1$ )

Die Minkowski-Konstante r bestimmt die genaue Form der Distanzberechnung. Bei der Clusterung von Standorten wird häufig  $r=1$  verwendet, was zur so genannten City-Block-Metrik führt. In diesem Falle werden die Distanzen der einzelnen Merkmale einfach addiert. Das am häufigsten verwendete Distanzmaß ist die euklidische Distanz (teilweise auch quadrierte euklidische Distanz), die man erhält, wenn  $r=2$  gewählt wurde. Dies ist die direkte Verbindung zwischen zwei Datenpunkten.

Wichtig ist in jedem Falle die Vergleichbarkeit der Maßeinheiten. Falls diese nicht übereinstimmen können die Variablen standardisiert werden.

## Q-Korrelationskoeffizient

$$r_{k,l} = \frac{\sum_{j=1}^J (x_{kj} - \bar{x}_k)(x_{lj} - \bar{x}_l)}{\sqrt{\sum_{j=1}^J (x_{kj} - \bar{x}_k)^2 \cdot \sum_{j=1}^J (x_{lj} - \bar{x}_l)^2}}$$

$r$ =Ähnlichkeit  
 $k$ =Merkmalsträger  $k$   
 $l$ =Merkmalsträger  $l$   
 $j$ =Merkmal  $j$   
 $\bar{x}_k$ =Durchschnittswert aller Eigenschaften des Merkmalsträgers  $k$

Der Korrelationskoeffizient bestimmt die Ähnlichkeit zwischen zwei Merkmalsträgern unter Berücksichtigung des Profils aller Merkmale und kann Werte zwischen -1 (vollkommen gegensätzlich) bis +1 (vollkommen ähnlich) annehmen. Ergibt sich für  $r$  ein Wert von 0, so gibt es keine Ähnlichkeit.

### 5.1.2 Proximitätsmaße für nominale Skalen

Für Merkmale auf nominalem Skalenniveau muss in jedem Falle eine Zerlegung in Dummy-Variablen erfolgen. Die hier vorgestellten Proximitätsmaße bestimmen die Ähnlichkeit von zwei Merkmalsträgern unter Berücksichtigung aller Variablen. Bei jedem Paarvergleich eines Merkmals gibt es nun drei verschiedene Möglichkeiten:

- Beide Merkmalsträger besitzen die Eigenschaft (Fall A)
- Nur einer der beiden Merkmalsträger besitzt die Eigenschaft (Fall B oder C)
- Keiner der beiden merkmalsträger besitzt die Eigenschaft (Fall D)

		Merkmalsträger 1	
		Eigenschaft vorhanden (1)	Eigenschaft nicht vorh. (0)
Merkmals-träger k	Eigenschaft vorhanden (1)	A	B
	Eigenschaft nicht vorh. (0)	C	D

Die Ähnlichkeiten zwischen zwei Merkmalsträgern können nun durch verschiedene Maßzahlen ermittelt werden, die sich hauptsächlich darin unterscheiden, welches Gewicht den verschiedenen Übereinstimmungsmöglichkeiten zugemessen wird. Alle Maßzahlen sind zwischen 0 (unähnlich) und 1 (ähnlich) normiert:

Maßzahl	Formel	Bemerkungen
Tanimoto	$S_{k,l} = \frac{A}{A+B+C}$	Die Ähnlichkeit nach Tanimoto bewertet Merkmalsträger dann als ähnlich, wenn bei beiden die Eigenschaft vorhanden ist. Ist bei beiden die Eigenschaft nicht vorhanden, so ist dies ein Fall, der überhaupt nicht berücksichtigt wird, also weder als Ähnlichkeit noch als Unähnlichkeit behandelt wird.
Simple Matching	$S_{k,l} = \frac{A+D}{m}$ mit $m=A+B+C+D$	Bei dem Simple Matching-Verfahren gelten Objekte dann als ähnlich, wenn sie entweder die Eigenschaft gemeinsam besitzen, oder nicht besitzen. Dies wird ins Verhältnis zur Zahl aller Eigenschaften gesetzt.
Russel & Rao	$S_{k,l} = \frac{A}{m}$ mit $m=A+B+C+D$	Bei Russel and Rao werden nur diejenigen Merkmalsträger als Ähnlich eingestuft, die möglichst viele Eigenschaften gemeinsam besitzen. Das gemeinsame Fehlen einer Eigenschaft gilt nicht als Ähnlichkeit.
Dice	$S_{k,l} = \frac{2A}{2A+B+C}$	Die Dice-Maßzahl basiert auf der gleichen Idee wie der Tanimoto-Parameter, jedoch wird dem gemeinsamen Vorhandensein ein doppelt so hohes Gewicht beigemessen. Durch den leicht gekrümmten Kurvenverlauf dieser Maßzahl steigt die Ähnlichkeit bei wenigen Übereinstimmungen schnell an, nimmt bei vielen Übereinstimmungen aber nur langsam zu.

Die wichtigste Unterscheidung, die bei der Auswahl der Maßzahl zu beachten ist, richtet sich danach, ob das Vorhandensein einer Eigenschaft den gleichen Informationsgehalt besitzt wie das Nichtvorhandensein. Ist das gemeinsame Fehlen einer Eigenschaft nicht unbedingt eine Ähnlichkeit (nicht deutsch zu sein, kann bedeuten, dass der eine Merkmalsträger Engländer, der andere Franzose ist – also nicht wirklich eine Gemeinsamkeit), so ist eine Maßzahl zu wählen, die Fall D nicht im Zähler berücksichtigt. Wenn das gemeinsame Fehlen einer Eigenschaft (z.B. nicht weiblich, sprich männlich zu sein) den gleichen Ähnlichkeitsgrad beschreibt wie das Vorhandensein, wählt man beispielsweise das Simple Matching Verfahren. Die Sachlogik bestimmt die Wahl der Maßzahl.

Grundsätzlich ist darauf zu achten, dass eine Überführung einer nominalskalierten Variable mit vielen Ausprägungen eine hohe Anzahl von gemeinsam nicht vorhandenen Eigenschaften generiert, was schnell zu starken Verzerrung führen kann. Dies gilt besonders dann, wenn mehrere Variablen mit verschieden vielen Ausprägungen in Dummies zerlegt werden.

## 5.2 Fusionierung

Die Ähnlichkeit zwischen den Objekten ist die Grundlage für die Entscheidung, wie die Gruppen gebildet werden sollen. Aus der Informatik sind zwei große Verfahren bekannt.

Die partitionierenden Verfahren beginnen mit einer Startverteilung und versuchen durch Austauschprozesse eine Verbesserung der Zielfunktion (hier: Maximierung der Ähnlichkeit) zu erreichen. Dieses einleuchtende Verfahren bringt allerdings das Problem eines sehr großen Rechenaufwandes mit sich. Die Anzahl der zu probierenden Kombinationsmöglichkeiten ergibt sich aus Anzahl der Gruppen hoch Anzahl der Merkmalsträger. Möchte man also 500 Kunden in 3 Gruppen aufteilen, so müssten  $3,636 \cdot 10^{238}$  Möglichkeiten (mit jeweils einer kompletten Ähnlichkeitsmatrix) ausprobiert werden<sup>1</sup>.

Praktikabel sind hingegen die hierarchischen Verfahren, die zunächst davon ausgehen, dass jeder Merkmalsträger eine Gruppe für sich bildet und daraufhin Schritt für Schritt diejenigen Gruppen miteinander verschmolzen werden, die die größte Ähnlichkeit zueinander aufweisen. Der Ablauf sieht also folgendermaßen aus:

- 1.) Jeder Merkmalsträger ist zunächst eine Gruppe (Cluster).
- 2.) Berechnung der Ausgangsdistanz- bzw. -ähnlichkeitsmatrix.
- 3.) Fusionierung der beiden Cluster, die die größte Ähnlichkeit besitzen.
- 4.) Berechnung einer neuen (reduzierten) Distanz- bzw. Ähnlichkeitsmatrix. (solange mit Schritt 3 fortfahren, wie die Anzahl der Gruppen weiter reduziert werden kann/soll).

Die grundsätzliche Vorgehensweise ist also recht einfach. Allerdings muss entschieden werden, wie mit den Clustern verfahren werden soll, die aus mehreren Objekten bestehen – wie werden in diesem Fall die Distanzen (bzw. Ähnlichkeiten) zu anderen Clustern bestimmt?

An folgendem Beispiel sollen die verschiedenen Möglichkeiten hierfür veranschaulicht werden. Es ist eine Distanzmatrix von 5 Personen gegeben:

	Karl-Heinz	Edwin	Bigi	Herta	Uschi
Karl-Heinz	0	4	6	12	26
Edwin		0	20	13	14
Bigi			0	15	17
Herta				0	11
Uschi					0

---

<sup>1</sup> Ein moderner Pentium 4 Prozessor schafft etwa 7500 MFLOPS (also 7.500 Millionen Fließkomma-Rechenoperationen pro Sekunde). Selbst wenn die Berechnung der Ähnlichkeitsmatrix nur eine Rechenoperation benötigen würde, dauerte die gesamte Berechnung  $1,5 \cdot 10^{221}$  Jahre.

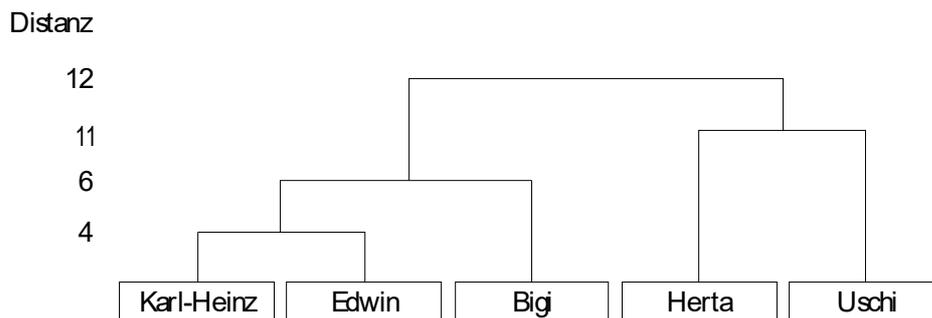
### 5.2.1 Single Linkage

Das einfachste Verfahren ist das Single-Linkage-Verfahren, welches die Distanz zwischen einem Cluster und einem anderen Objekt als die niedrigste Distanz zwischen den Clustermitgliedern und dem Vergleichsobjekt bestimmt:

$$D_{p+q,r} = \min \{ D_{p,r}; D_{q,r} \}$$

p+q=Cluster mit Objekten p und q  
r=Vergleichsobjekt

Aus dem oben genannten Beispiel ergibt sich folgende Fusionierungsreihenfolge, die zusammen mit den Distanzen in dem Dendogramm<sup>2</sup> deutlich wird:



Zunächst werden die beiden Objekte mit der geringsten Distanz zu einem neuen Cluster zusammengefasst (Karl-Heinz und Edwin mit einem Distanzwert von 4). Danach stößt Bigi zu diesem Cluster hinzu. Sie hat zwar zu Edwin einen Distanzwert von 20, jedoch zählt hier nur der kleinere Distanzwert zu Karl-Heinz von 6. Alle anderen potenziellen Zusammenschlüsse haben größere Distanzwerte.

Das Single Linkage Verfahren (auch „nearest neighbour“-Verfahren genannt) neigt leider dazu viele kleine, aber nur wenige große Gruppen zu bilden. Durch Verkettungseffekte kann es vorkommen, dass die Distanzen innerhalb eines Clusters größer werden als die Distanzen zwischen zwei Clustern (vgl. Beispiel letzter Fusionierungsschritt. Die Distanz zwischen den Gruppen beträgt 12, obwohl zwischen Edwin und Bigi 20 Distanzpunkte liegen). Schlecht getrennte Gruppen werden mit diesem verfahren also nicht erkannt, weswegen es kaum dazu zu gebrauchen ist, sinnvolle Gruppen zu bilden. Allerdings können durch dieses Verfahren Ausreißer sehr gut identifiziert werden, die andere Verfahren stören könnten.

### 5.2.2 Complete-Linkage

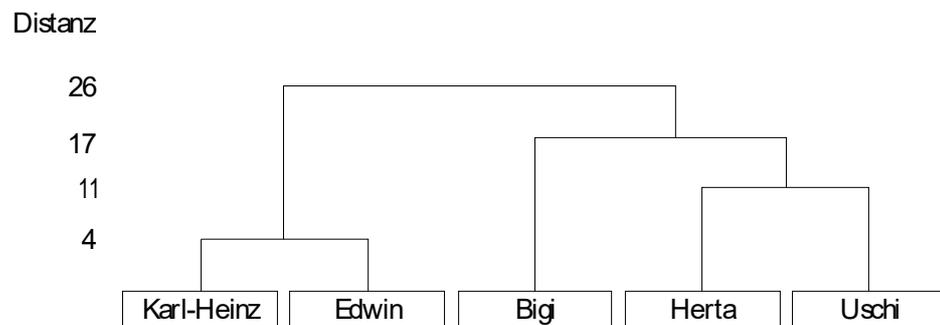
Bei dem Complete-Linkage-Verfahren handelt es sich um eine Modifikation des Single-Linkage-Verfahrens. Um den Effekten der Verkettung entgegenzutreten, wurde die Distanzfunktion so verändert, dass nun der jeweils größte Einzelabstand ausschlaggebend ist:

<sup>2</sup> Das Statistikpaket SPSS plottet das Dendrogramm auf eine auf 25 normierte Achse. Auf diese Weise können je nach gewähltem Verfahren die relativen Distanzen (bzw. Fehlerquadratsummen) abgelesen werden. Je größer der Höhenunterschied der Verbindungslinien zweier Gruppen bei einem Fusionierungsschritt ist, desto ungleicher sind die Gruppen.

$$D_{p+q,r} = \max \{ D_{p,r}; D_{q,r} \}$$

p+q=Cluster mit Objekten p und q  
r=Vergleichsobjekt

Auf diese Weise werden eher viele kleine Gruppen erzeugt. Jedoch verursachen Ausreißer bei dem Complete-Linkage-Verfahren zum Teil starke Verzerrungen, da einzelne Objekte mit hohen Distanzwerten die anderen Gruppenmitglieder stark dominieren können. Das Dendrogramm zeigt den unterschiedlichen Fusionsverlauf:



### 5.2.3 Ward-Verfahren

Das Ward-Verfahren ist dasjenige Verfahren, welches sich in der Praxis durchgesetzt hat. Es führt zu verlässlichen Ergebnissen mit etwa gleich großen Gruppen, fordert allerdings ein Distanzmaß (also metrisch skalierte Variablen) als Basis.

Die Distanz zu einem neuen Cluster wird mit folgender Formel bestimmt:

$$D_{p+q,r} = \frac{1}{n_p + n_q + n_r} \left( (n_p + n_r) \cdot D_{p,r} + (n_q + n_r) \cdot D_{q,r} - n_r \cdot D_{p,q} \right)$$

$n_p, n_q, n_r$ =Anzahl der Objekte in Cluster p,q oder r

$D_{p,r}$ = Distanz zwischen dem Einzelobjekt q und Objekt r etc.

Allerdings ist das Distanzmaß selbst nun nicht mehr der bestimmende Faktor bei der Frage der Fusionierung. Hierfür wird vielmehr die Varianz zu Grunde gelegt. Um genau zu sein werden stets diejenigen Objekte fusioniert, die die Fehlerquadratsumme ( $n \cdot$ Varianz) am wenigsten erhöhen. Auf diese Weise werden möglichst homogene Cluster mit möglichst geringer Varianz gebildet.

V=Fehlerquadratsumme

i=Cluster i

j=Merkmal j

k=Merkmalsträger k

$x_{kji}$  =Ausprägung des Merkmals j des Merkmalsträgers k in Cluster i

$\bar{x}_{ji}$  =Mittelwert des Merkmals j in Cluster i

$$V_i = \sum_{k=1}^{K_i} \sum_{j=1}^J (x_{kji} - \bar{x}_{ji})^2$$

Wählt man die quadrierte euklidische Distanz als Proximitätsmaß, so ist die Zunahme der Fehlerquadratsumme stets genau halb so hoch wie die Distanz, wie in folgendem Beispiel deutlich wird:

In der Ausgangsdistanzmatrix (vgl. Tabelle XX) besteht jedes Cluster aus genau einem Merkmalsträger. Die Fehlerquadratsumme beträgt also Null, da es noch keine Varianzen geben kann. Nun werden diejenigen Objekte miteinander vereinigt, die die Fehlerquadratsumme am wenigsten erhöhen – da die Distanzen genau doppelt so hoch sind wie die Fehlerquadratsumme, werden Karl-Heinz und Edwin mit einer Distanz von 4 (Fehlerquadratsumme  $V=2$ ) zu einem neuen Cluster vereinigt. Daraufhin muss eine neue Distanzmatrix berechnet werden. Die neue Distanz zwischen dem Cluster „Karl-Heinz & Edwin“ und Bigi berechnet sich beispielsweise nach oben genannter Formel aus:

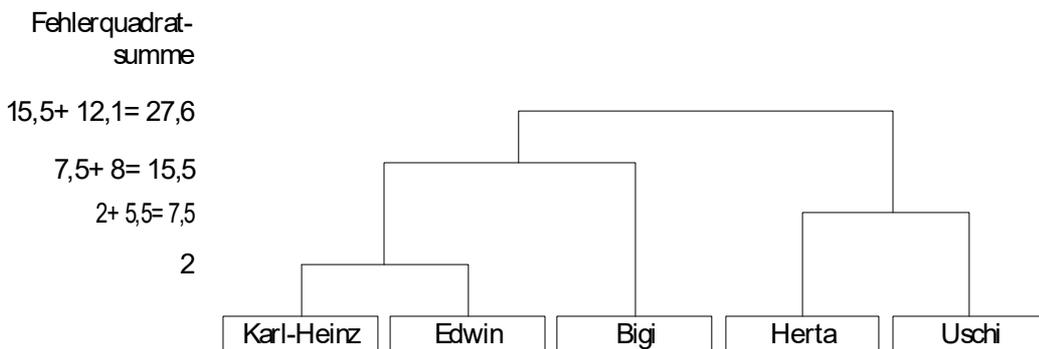
$$D_{K-H\&E,B} = \frac{1}{n_{K-H} + n_E + n_B} \left( (n_{K-H} + n_B) D_{p,r} + (n_E + n_B) D_{E,B} - n_B D_{K-H,E} \right)$$

$$= \frac{1}{1+1+1} (2 \cdot 6 + 2 \cdot 20 - 1 \cdot 4) = 16$$

Es ergibt sich folgende Distanzmatrix:

	K-H & E	Bigi	Herta	Uschi
K-H & E	0	16	15,33	25,33
Bigi		0	15	17
Herta			0	11
Uschi				0

Bei einer weiteren Fusionierung würde die Fehlerquadratsumme am wenigsten durch die Zusammenbringung von Uschi und Herta ( $D=11$ ,  $\Delta V=5,5$ ) steigen. Den gesamten Verlauf des Algorithmus zeigt das Dendrogramm:



#### 5.2.4 Zahl der Cluster

Nachdem der Fusionierungsalgorithmus durchlaufen ist muss geklärt werden, an welcher Stelle auf weitere Fusionierungen verzichtet werden soll, d.h. wie viele Cluster generiert werden. Leider gibt es keine eindeutigen Kriterien, wie etwa das Kaiser-Kriterium der Faktorenanalyse, jedoch ist es möglich anhand der Zunahme des Hete-

rogenitätsmaßes eine Entscheidung zu treffen. Ist hier ein Sprung (beim Ward-Verfahren beispielsweise in der Fehlerquadratsumme) festzustellen, so bedeutet dies, dass an dieser Stelle zwei Cluster „zwangsverbunden“ wurden, obwohl diese an sich sehr unterschiedlich waren. Diese statistischen Hinweise müssen mit den praktischen Anforderungen an die Gruppenzahl in Einklang gebracht werden. Eventuell ist es sinnvoll eine Zahl von 12 Clustern der Handhabbarkeit zuliebe weiter zu reduzieren, auch wenn die Gruppen dadurch heterogener werden. Diese Entscheidung liegt stets in der Abwägung des Forschers.

### 5.3 Beurteilung der Ergebnisse

Nachdem die Cluster gefunden wurden, müssen die Ergebnisse zum einen auf ihre Güte hin überprüft, zum anderen die gefundenen Cluster charakterisiert werden.

Um zu überprüfen, ob die gefundenen Gruppen homogen sind, und somit die Gruppierung sinnvoll ist, kann der F-Wert herangezogen werden. Dies ist lediglich ein deskriptives Maß, welches trotz seines Namens keinen Test über die F-Verteilung zulässt – dies ist auch nicht nötig, da im Falle der Clusteranalyse nicht von der Stichprobe auf eine Grundgesamtheit geschlossen wird.

$$F_{ji} = \frac{s_{ji}^2}{s_j^2} \quad \begin{array}{l} i=\text{Cluster } i \\ j=\text{Merkmal } j \end{array}$$

Der F-Wert prüft, ob die Varianz eines Merkmals j innerhalb eines Clusters i größer ( $F > 1$ ) oder kleiner ( $F < 1$ ) ist als die Varianz des gleichen Merkmals in der gesamten Erhebung. Auf diese Weise kann für jedes Merkmal in jedem Cluster überprüft werden, ob die Einteilung in Gruppen zu einer Steigerung der Homogenität geführt hat. Ein Cluster wird dann als vollkommen homogen bezeichnet, wenn sich für alle Merkmale dieses Clusters eine kleinere Varianz innerhalb des Clusters ergibt als in der gesamten Erhebung.

Eine Hilfe zur Interpretation der Cluster liefert hingegen der t-Wert. Auch dieser hat einen trügerischen Namen, denn es liegt keine überprüfbare Student-t-Verteilung vor. Vielmehr ist der t-Wert in diesem Falle lediglich ein deskriptiver Vergleich normierter Mittelwerte:

$$t_{ji} = \frac{\bar{x}_{ji} - \bar{x}_j}{s_j} \quad \begin{array}{l} i=\text{Cluster } i \\ j=\text{Merkmal } j \end{array}$$

Der Mittelwert eines Merkmals j in einem Cluster i wird mit dem Mittelwert dieses Merkmals ohne die Clustereinteilung verglichen und über die Standardabweichung des Merkmals normiert. Ergibt sich ein t-Wert von größer Null, ist das jeweilige Merkmal überdurchschnittlich stark in dem entsprechenden Cluster ausgeprägt.

# 6 NEURONALE NETZE

Die Rolle der neuronalen Netze nimmt unter den bisher vorgestellten Analyseverfahren eine Sonderrolle ein. Es ist ein Verfahren, welches ursprünglich als Modell entwickelt wurde, um die Vorgänge im menschlichen Gehirn verstehen und nachbilden zu können. Es handelt sich, abstrakt formuliert, um informationsverarbeitende Systeme, die aus einer großen Zahl einfacher Einheiten (Neuronen) bestehen, die sich Informationen in Form der Aktivierung der Zellen über gerichtete Verbindungen zusenden. (A. Zell, 1994, S.23) Warum neuronale Netze auch für analytische Zwecke derart interessant sind, macht ein Vergleich zwischen dem menschlichen Gehirn und einem modernen Computer deutlich.

	<b>Gehirn</b>	<b>Rechner</b>
Anzahl Schaltelemente	$10^{11}$ Neuronen	$10^{10}$ Transistoren
Art	Massiv parallel	i.d.R. seriell
Schaltgeschwindigkeit	1 ms ( $10^{-3}$ Sekunden)	1 ns ( $10^{-9}$ Sekunden)
Theoretisch mögliche Schaltvorgänge pro Sekunde	$10^{14}$	$10^{19}$

Der erste Blick lässt beide Kontrahenten relativ gleich gut aussehen – von der reinen Rechenleistung ist der Computer dem menschlichen Gehirn sogar weit überlegen. Welchen Unterschied die massive parallele (ein Neuron ist mit bis zu 10.000 anderen Neuronen vernetzt) Arbeitsweise des Gehirns jedoch ausmacht, wird mit folgendem Beispiel deutlich. Ein Mensch erkennt ein bekanntes Objekt oder eine Person nach etwa 0,1 Sekunden. Ausgehend von der Schaltgeschwindigkeit, sind in dieser Zeitspanne bis zu 100 Schaltschritte im menschlichen Gehirn möglich. (Über die tatsächliche Zahl der involvierten Zellen sagt jedoch dies nichts aus). Mit 100 Schaltschritten kann das Gehirn also ein komplexes Objekt oder gar eine Person erkennen – ein herkömmlicher (von Neumann-)Rechner schafft mit 100 Assemblerbefehlen (einfachste Operationsanweisungen für den Prozessor) dagegen nichts Beeindruckendes: - z.B. das Addieren aller Zahlen von 1 bis 100.

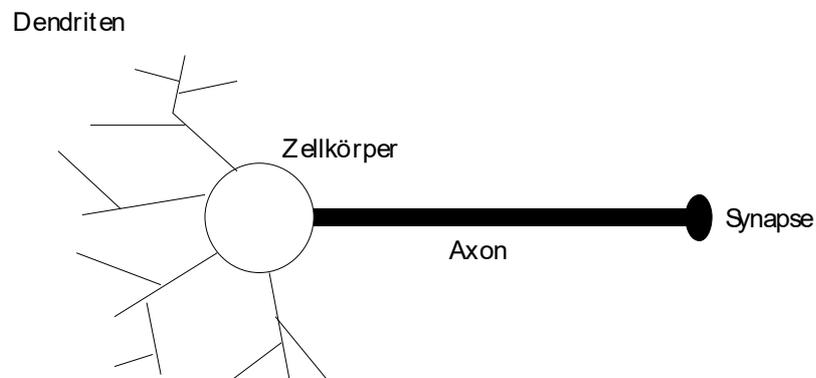
Künstliche Neuronale Netzwerke (KNN) lassen sich daher gut für Analysezwecke einsetzen, etwa im Bereich der Klassifikation (vgl. Clusteranalyse) oder der Prognosen (vgl. Regression). Die größten Vorteile liegen in der Lernfähigkeit (es bedarf also keiner genau für ein bestimmtes Problem definierten Algorithmen), der hohen Fehlertoleranz (Robustheit gegen Rauschen), des Fehlens der Linearitätsbeschränkung sowie des Arbeitens ohne Hypothesen. Allerdings sind mit diesem Ansatz auch einige Nachteile verbunden. Der größte Unterschied zu den bisherigen Verfahren liegt darin, dass keine Introspektion möglich ist. Das neuronale Netz ist eine „Black Box“ – das in Ihr gespeicherte Wissen lässt sich nicht durch einfache Regeln beschreiben. Des Weiteren ist ein Lernprozess erforderlich, der auch auf schnellen modernen

Rechnern sehr langwierig sein kann und je nach Komplexität des Problems große Lerndatenmengen erfordert.

## 6.1 Grundlagen neuronaler Netze

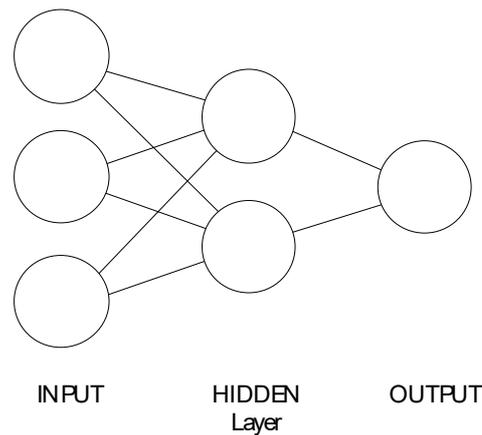
### 6.1.1 Aufbau neuronaler Netze

Das biologische Vorbild der Neuronen, gliedern sich in drei wesentliche Teile: Über zahlreiche Dendriten werden Signale, die von anderen Nervenzellen ausgesandt werden aufgenommen und diese Information an den Zellkörper weitergeleitet. Dort werden die Informationen verarbeitet und um dann bei einem gewissen Schwellenwert ein Ausgabesignal auszulösen, welches über das Axon, an dessen Ende die Synapse sitzt, an andere Zellen weitergegeben wird.



Die künstlichen Neuronen in neuronalen Netzen funktionieren ähnlich. Natürlich sind es in diesem Falle nur Konstrukte von Software, die die Funktionsweise der biologischen Vorlagen kopieren sollen. Der Aufbau dieser Softwareelemente ist jedoch genauso: Eingabesignale von verschiedenen anderen Zellen werden empfangen und entsprechend der Gewichte der Eingangsleitungen zu einem einheitlichen Eingangssignal verdichtet. Dies wird von der Propagierungsfunktion übernommen. Der Wert der Propagierungsfunktion wird wiederum an die Aktivierungsfunktion weitergegeben, die darüber entscheidet, ob (bzw. wie stark) das Neuron ‚feuern‘ soll.

Es gibt verschiedene Typologien von künstlichen neuronalen Netzwerken. Hier sollen lediglich Feed-Forward Preceptrons vorgestellt werden, die über einen überwachten Lernprozess trainiert werden sollen. Die Feedforward-Eigenschaft besagt, dass lediglich Verknüpfungen zwischen den Neuronen in Flussrichtung bestehen. Es findet keine Rückkopplung zwischen Neuronen verschiedener Schichten statt. Als Preceptron bezeichnet man einen Netzwerktyp, der sich durch eine Eingabeschicht, (mehrere) versteckte Schichten und eine Ausgabeschicht auszeichnet.

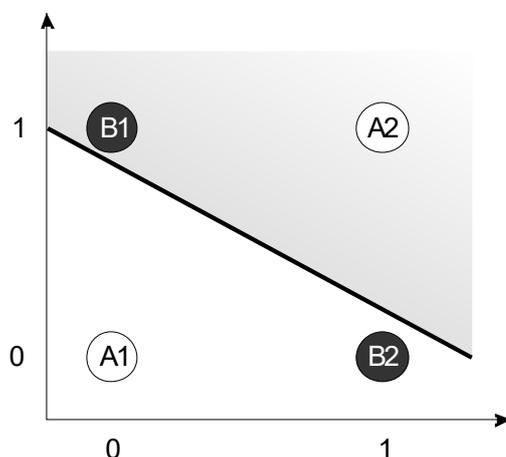


In die drei Eingabeneuronen könnten zum Beispiel Informationen über einen Kreditnehmer einfließen, etwa die Anzahl der Kinder, das Einkommen, und ob ein negativer Eintrag in einer Schutzgemeinschaft (z.B. Schufa) vorliegt. Das neuronale Netz verarbeitet diese Informationen und aufgrund des vorangegangenen Lernprozesses kann das Ausgabeneuron signalisieren, ob ein Kredit gewährt werden kann, oder dies zu risikoreich ist.

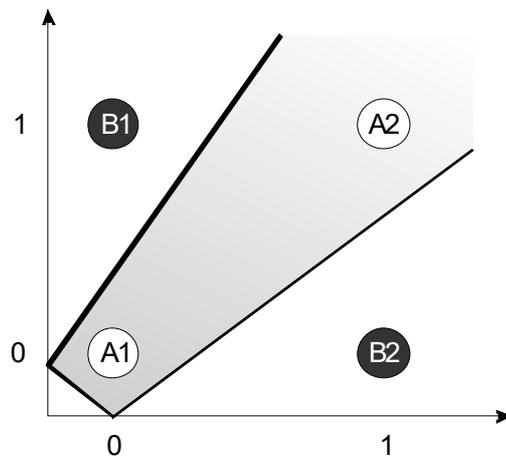
### 6.1.2 Netzwerktopologie

Die Zahl der versteckten Ebenen bestimmt dabei in herausragender Weise die Leistungsfähigkeit der KNN. Ein Perzeptron kann in einer endlichen Zeit alles lernen, was es aufgrund seiner Topologie repräsentieren kann. Die Repräsentierbarkeit bezeichnet die Fähigkeit des Netzes eine gegebene Funktion realisieren zu können. An dieser Stelle liegt das Problem. Einfache Perzeptrons können nur linear separierbare Mengen repräsentieren. Es gibt also viele Funktionen, die sich nicht durch ein einfaches Perzeptron lösen lassen können. Ein Beispiel hierfür ist die Exklusiv-Oder-Funktion (XOR), die als Ergebnis dann „1“ liefert, wenn zwei binäre Variablen unterschiedlich sind, und „0“ wenn diese gleich sind.

Zur Lösung dieses Problems müssen die Punkte A1 und A2 in einer Menge, die Punkte B1 und B2 in einer zweiten Menge zugeordnet werden. Dies ist mit einer einzigen Geraden nicht möglich (Problem der linearen Separierbarkeit).



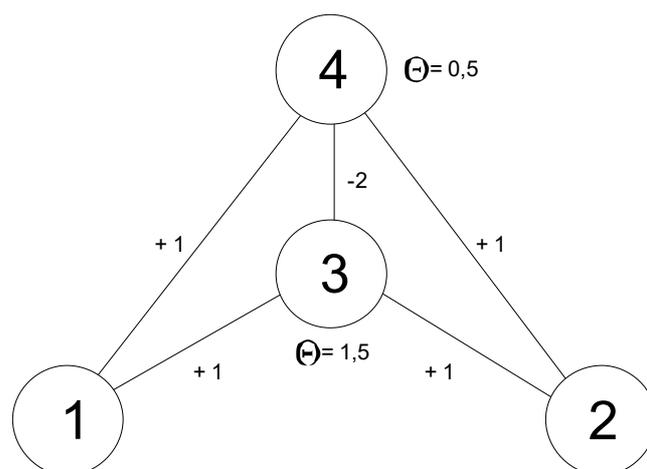
Um komplexere Probleme zu lösen müssen weitere Stufen in das Netz eingeführt werden. Zweitstufige Perzeptrons können konvexe Polygone klassifizieren, dreistufige Perzeptrons beliebige Formen.



## 6.2 Funktionsweise eines einfachen KNN am XOR-Beispiel

Anhand eines einfachen Beispiels sollen in diesem Abschnitt die wichtigsten Bestandteile eines KNN sowie die Funktionsweise eines bereits trainierten Netzes erläutert werden. Der Trainingsprozess wird separat im nächsten Abschnitt beleuchtet.

Die Aufgabe dieses Netzes soll die Ausführung einer logischen Exklusiv-Oder-Verknüpfung (XOR) sein. Es gibt zwei binäre Eingabewerte und einen binären Ausgabewert, der genau dann ‚wahr‘ (also 1) anzeigen soll, wenn sich beide Eingabewerte voneinander unterscheiden. Das Netz soll ‚falsch‘ (also 0) zurückgeben, wenn die Eingabewerte gleich sind.



Das oben gezeigte Netz setzt sich aus folgenden Bestandteilen zusammen:

**Gewichte zwischen den Neuronen  $w_{ij}$ :** Die Gewichte zwischen einem Neuron  $i$  und einem anderen Neuron  $j$  geben an, wie stark das Ausgangssignal des Neurons  $i$  an das zweite Neuron  $j$  weitergegeben wird. Aus einer technischen Sichtweise könnte man sie als die Leitfähigkeit des Signals verstehen. Zu Beginn eines Lernprozesses

sind diese Gewichte rein zufällig gesetzt und werden dann im Laufe des Trainings modifiziert, so dass das Netz seine Aufgabe lösen kann. Aus den einzelnen Gewichten kann eine Gewichtsmatrix  $w$  erstellt werden:

$$w_{ij} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Da es sich in diesem Beispiel um ein Feedforward-Netz handelt ist nur das obere Dreieck der Matrix gefüllt. Rückwärtsgerichtete Verbindungen gibt es in diesem Falle nicht, so dass diese Gewichte auf Null gesetzt werden.

**Aktivierungszustand  $a_j(t)$ :** Grad der Aktivierung eines Neurons zum Zeitpunkt  $t$ . In diesem Beispiel sind nur binäre Aktivierungszustände (1 und 0) möglich. Jedoch sind im Allgemeinen alle reellen Zahlen als Aktivierungszustand gültig. In diesem Falle wird auf die Fuzzy-Logik zurückgegriffen

**Ausgabewert und Ausgabefunktion  $o_j = f_{out}(a_j)$ :** Die Ausgabefunktion bestimmt anhand des Aktivierungszustandes, welcher Output  $o$  die Zelle verlässt. In diesem Beispiel wird einfach der Aktivierungszustand als Ausgabewert ohne Berechnungen verwendet:  $o_j(t) = a_j(t)$

**Propagierungsfunktion  $net_j(t)$ :** Die Propagierungsfunktion fasst die eingehenden Signale anderer Neuronen zu einem einheitlichen Eingabewert zusammen. In der Regel wird eine gewichtete Summe der Ausgabewerte anderer Neuronen gebildet. Es sind jedoch auch andere Formen, etwa das gewichtete Produkt, denkbar.

$$net_j(t) = \sum_i o_i(t) w_{ij}$$

**Aktivierungsfunktion  $f_{act}$ :** Die Aktivierungsfunktion legt anhand des derzeitigen Aktivierungszustandes und der Signaleingänge aus der Propagierungsfunktion fest, welchen Aktivierungszustand das Neuron annimmt. Bei binären Neuronen, wird der Aktivierungszustand dann auf 1 gesetzt, wenn die Aktivierungsfunktion einen gewissen Schwellenwert  $\Theta$  überschreitet.

$$a_j(t+1) = f_{act}(a_j(t), net_j(t), \Theta_j)$$

in diesem Beispiel:

$$a_j(t+1) = \begin{cases} 1 & \text{wenn } net_j(t) \geq \Theta_j \\ 0 & \text{sonst} \end{cases}$$

Mit diesen Informationen kann das Neuronale Netz schrittweise für alle erdenklichen Eingabewerte an  $o_1$  und  $o_2$  durchlaufen werden:

$\mathbf{o}_1$	$\mathbf{o}_2$	$\mathbf{net}_3$	$\Theta_3$	$\mathbf{o}_3$	$\mathbf{net}_4$	$\Theta_4$	$\mathbf{o}_4$
0	0	$0 \cdot 1 + 0 \cdot 1 = 0$	1,5	0	$0 \cdot 1 + 0 \cdot 1 + 0 \cdot -2 = 0$	0,5	0
0	1	$0 \cdot 1 + 1 \cdot 1 = 1$	1,5	0	$0 \cdot 1 + 1 \cdot 1 + 0 \cdot -2 = 1$	0,5	1
1	0	$1 \cdot 1 + 0 \cdot 1 = 1$	1,5	0	$1 \cdot 1 + 0 \cdot 1 + 0 \cdot -2 = 1$	0,5	1
1	1	$1 \cdot 1 + 1 \cdot 1 = 2$	1,5	1	$1 \cdot 1 + 1 \cdot 1 + 1 \cdot -2 = 0$	0,5	0

### 6.3 Trainieren des KNN

Ohne Lernregeln ist es einem neuronalen Netz nicht möglich, gegebene Aufgaben anhand von Beispielen zu erlernen. Wäre das KNN aus dem vorangegangenen Abschnitt nicht trainiert worden, so hätte es die XOR-Aufgabe nicht lösen können. Daher kommt dem Training neuronaler Netze eine besondere Bedeutung zu. Grundsätzlich gibt es viele Möglichkeiten, auf das Verhalten von neuronalen Netzen Einfluss zu nehmen, und es somit etwas zu lehren:

- Entwicklung neuer Verbindungen
- Löschen existierender Verbindungen
- Verändern der Gewichte zwischen Neuronen
- Verändern der Schwellenwerte
- Verändern der Funktionen  $f_{\text{out}}$ ,  $f_{\text{act}}$ ,  $\text{net}_j$
- Hinzufügen von neuen Neuronen
- Löschen von Neuronen

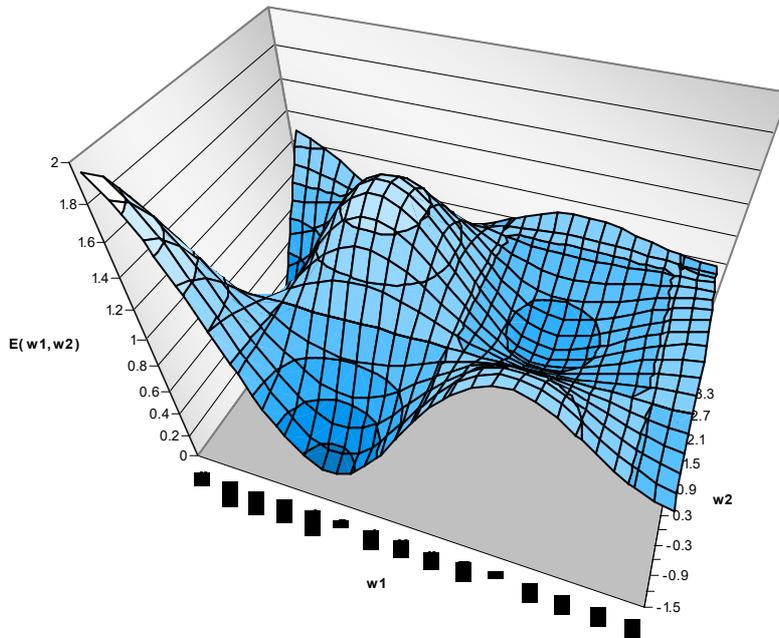
In der Praxis konzentriert sich die Steuerung des Lernens auf die Veränderung der Gewichte zwischen den Neuronen. Auf diese Weise können auch Verbindungen praktisch ‚gelöscht‘ (Verbindungsgewicht=0) oder wieder ins Leben gerufen werden (Verbindungsgewicht von Null verschieden setzen). Auch die Modifikation des Schwellenwertes kann über den ‚Umweg‘ einer Veränderung der Gewichte realisiert werden. Das Abändern der Aktivierungs-, Propagierungs- und Ausgabefunktion ist gängigen Modellen nur selten anzutreffen und hat sich als wenig Erfolgreich herausgestellt. Allerdings wird die Veränderung der Neuronenzahl zur Laufzeit als ein Gebiet mit hohem Optimierungspotenzial angesehen.

An dieser Stelle soll lediglich ein einzelnes Lernverfahren beispielhaft vorgestellt werden. Es handelt sich hierbei um das so genannte Backpropagation-Verfahren.

Um effektiv lernen zu können, muss eine Möglichkeit geschaffen werden, das Ergebnis des neuronalen Netzes mit dem tatsächlichen Wert der Trainingsdaten zu vergleichen. Die Differenz zwischen Schätzung des KNN und tatsächlichen Wert kann man als Fehler bezeichnen, der sich aus den verschiedenen Verbindungsgewichten ergibt:

$$E(W) = E(w_1, w_2, \dots, w_n)$$

Oft wird für die Fehlerfunktion die quadrierte Differenz zwischen erwarteter und tatsächlicher Ausgabe verwendet. In einem Beispiel mit nur zwei Verbindungen kann man den Fehler grafisch darstellen, indem man die möglichen Ausprägungen der Verbindungsgewichte auf der X- und Y-Achse, den sich hieraus ergebenden Fehler auf der Z-Achse darstellt:



Der Fehler hängt also von den Verbindungsgewichten ab und lässt sich daher durch eine Veränderung dieser beeinflussen. Ziel muss es also sein, für die Fehlersumme den globalen Minimalpunkt zu finden.

Das Backpropagation-Verfahren beruht dabei auf der Idee der Gradientenverfahren. Ein Gradient ist ein Vektor, der an einem beliebigen Punkt in die Richtung des steilsten Anstieges zeigt und dessen Länge ein Maß für die Steilheit ist. Ein Minimum soll hier dadurch gefunden werden, dass alle Gewichte um einen Bruchteil des negativen Gradienten verändert werden. Auf dieser Grundlage aufbauend, wurde die Backpropagation-Regel dahingehend erweitert, dass Netze mit mehreren Ebenen und semilinearen Aktivierungsfunktionen (monoton und differenzierbar, im Allgemeinen jedoch nichtlinear) trainiert werden können. Auf eine Herleitung dieser Regel soll an dieser Stelle jedoch verzichtet werden. Es ergibt sich für die Änderung der Gewichte für ein Lernmuster p:

$$\Delta_p w_{ij} = \eta o_{pi} \delta_{pj}$$

mit:

$$\delta_{pj} = \begin{cases} f'_{act}(net_{pj})(t_{pj} - o_{pj}) & \text{falls } j \text{ Ausgabeszelle ist} \\ f'_{act}(net_{pj}) \sum_k \delta_{pk} w_{jk} & \text{falls } j \text{ verdeckte Zelle ist} \end{cases}$$

t = Eingabe  
(teaching input)  
bei einem Muster p.

bzw. falls eine logistische Aktivierungsfunktion der Form

$$f(x) = \frac{1}{1 + e^{-x}}$$

verwendet wird, ergibt sich:

$$\delta_{pj} = \begin{cases} o_{pj}(1-o_{pj})(t_{pj}-o_{pj}) & \text{falls } j \text{ Ausgabeszelle ist} \\ o_{pj}(1-o_{pj}) \sum_k \delta_{pk} w_{jk} & \text{falls } j \text{ verdeckte Zelle ist} \end{cases}$$

$\eta$  ist dabei der Lernfaktor; gewissermaßen die Schrittweite, mit der auf das Minimum zugesteuert werden soll.

Auch wenn die Backpropagation-Regel meist zu guten Ergebnissen führt, gibt es dennoch einige Hindernisse, die bei der Anwendung dieser Regel im Auge behalten werden müssen. Zum einen kann es vorkommen, dass aufgrund einer ungünstigen Schrittweite globale Minima nicht gefunden werden können. Dies kann passieren, wenn sich der Algorithmus in einem lokalen Minimum verfängt (a), flache Plateaus im Fehlergebirge die Annäherung an ein Minimum praktisch zum Erliegen bringen (b), Schluchten mit dem globalen Minimum übersprungen werden (c), oder der Algorithmus an den Wänden einer Schlucht oszilliert (d). Für alle diese Fehler spielt die Wahl des Lernfaktors eine entscheidende Rolle. Wird  $\eta$  zu groß gewählt, so ist die Gefahr der oben beschriebenen Fehler größer. Je kleiner der Wert von  $\eta$  jedoch ist, desto größer ist der Rechenaufwand beim trainieren des Netzes. An dieser Stelle ist etwas Fingerspitzengefühl des Forschers gefragt.

Des Weiteren ist das Backpropagation-Verfahren nicht in der Lage Symmetrien der Gewichte in der der Ausgabeschicht vorgelagerten Schicht zu brechen. Dieses Problem kann allerdings leicht beseitigt werden, indem die Ausgangsgewichte zufällig gewählt werden und alle voneinander verschieden sind.

Nachdem der Trainingsprozess mit den Musterdaten abgeschlossen wurde, kann das neuronale Netz damit beginnen aufgrund der ‚Erfahrungen‘ des Trainings die gewünschten Aufgaben selbständig mit neuen, unbekanntem Daten lösen.

# LITERATURVERZEICHNIS

- Bacher, J.: Clusteranalyse: Anwendungsorientierte Einführung, 2. Auflage, Oldenbourg Verlag, 2002
- Backhaus, K; Erichson, B.; Plinke, W.; Weiber, R.: Multivariate Analysemethoden, 9. Auflage, Springer Verlag, Berlin 2000
- Bühl, A.; Zöfel, P.: SPSS Version 10 – Einführung in die moderne Datenanalyse unter Windows, 7. Auflage, Addison Wesley Verlag, München 2000
- Greene, W.H.: Econometric Analysis, 5. Auflage, Prentice Hall, Upper Saddle River, NJ, USA 2003
- Hartung, J.; Elpelt, B.: Multivariate Statistik, Oldenbourg-Verlag, München 1999
- Zell, A.; Simulation neuronaler Netze, Addison-Wesley (Deutschland), Bonn 1994

# TABELLEN

## $\chi^2$ -Tabelle

Vertrauenswahrscheinlichkeit	0.5	0.25	0.20	0.1	0.05	0.025	0.01	0.005	0.001
Freiheitsgrade									
1	0.455	1.323	1.642	2.706	3.841	5.024	6.635	7.879	10.827
2	1.386	2.773	3.219	4.605	5.991	7.378	9.210	10.597	13.815
3	2.366	4.108	4.642	6.251	7.815	9.348	11.345	12.838	16.266
4	3.357	5.385	5.989	7.779	9.488	11.143	13.277	14.860	18.466
5	4.351	6.626	7.289	9.236	11.070	12.832	15.086	16.750	20.515
6	5.348	7.841	8.558	10.645	12.592	14.449	16.812	18.548	22.457
7	6.346	9.037	9.803	12.017	14.067	16.013	18.475	20.278	24.321
8	7.344	10.219	11.030	13.362	15.507	17.535	20.090	21.955	26.124
9	8.343	11.389	12.242	14.684	16.919	19.023	21.666	23.589	27.877
10	9.342	12.549	13.442	15.987	18.307	20.483	23.209	25.188	29.588
11	10.341	13.701	14.631	17.275	19.675	21.920	24.725	26.757	31.264
12	11.340	14.845	15.812	18.549	21.026	23.337	26.217	28.300	32.909
13	12.340	15.984	16.985	19.812	22.362	24.736	27.688	29.819	34.527
14	13.339	17.117	18.151	21.064	23.685	26.119	29.141	31.319	36.124
15	14.339	18.245	19.311	22.307	24.996	27.488	30.578	32.801	37.698
16	15.338	19.369	20.465	23.542	26.296	28.845	32.000	34.267	39.252
18	17.338	21.605	22.760	25.989	28.869	31.526	34.805	37.156	42.312
20	19.337	23.828	25.038	28.412	31.410	34.170	37.566	39.997	45.314
22	21.337	26.039	27.301	30.813	33.924	36.781	40.289	42.796	48.268
24	23.337	28.241	29.553	33.196	36.415	39.364	42.980	45.558	51.179
26	25.336	30.435	31.795	35.563	38.885	41.923	45.642	48.290	54.051
28	27.336	32.620	34.027	37.916	41.337	44.461	48.278	50.994	56.892
30	29.336	34.800	36.250	40.256	43.773	46.979	50.892	53.672	59.702
32	31.336	36.973	38.466	42.585	46.194	49.480	53.486	56.328	62.487
34	33.336	39.141	40.676	44.903	48.602	51.966	56.061	58.964	65.247
36	35.336	41.304	42.879	47.212	50.998	54.437	58.619	61.581	67.985
38	37.335	43.462	45.076	49.513	53.384	56.895	61.162	64.181	70.704
40	39.335	45.616	47.269	51.805	55.758	59.342	63.691	66.766	73.403
42	41.335	47.766	49.456	54.090	58.124	61.777	66.206	69.336	76.084
44	43.335	49.913	51.639	56.369	60.481	64.201	68.710	71.892	78.749
46	45.335	52.056	53.818	58.641	62.830	66.616	71.201	74.437	81.400
48	47.335	54.196	55.993	60.907	65.171	69.023	73.683	76.969	84.037
50	49.335	56.334	58.164	63.167	67.505	71.420	76.154	79.490	86.660
55	54.335	61.665	63.577	68.796	73.311	77.380	82.292	85.749	93.167
60	59.335	66.981	68.972	74.397	79.082	83.298	88.379	91.952	99.608
70	69.334	77.577	79.715	85.527	90.531	95.023	100.425	104.215	112.317
80	79.334	88.130	90.405	96.578	101.879	106.629	112.329	116.321	124.839
90	89.334	98.650	101.054	107.565	113.145	118.136	124.116	128.299	137.208
100	99.334	109.141	111.667	118.498	124.342	129.561	135.807	140.170	149.449
120	119.334	130.055	132.806	140.233	146.567	152.211	158.950	163.648	173.618
200	199.334	213.102	216.609	226.021	233.994	241.058	249.445	255.264	267.539

# Durbin-Watson-Tabelle

Aus: Griffith, Hill, Judge: Learning and Practicing Econometrics, 1993

**Table 5** Critical Values for the Durbin-Watson Test: 5% Significance Level<sup>a</sup>

T	K = 2		K = 3		K = 4		K = 5		K = 6		K = 7		K = 8		K = 9		K = 10		K = 11	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$														
6	0.610	1.400																		
7	0.700	1.356	0.467	1.896																
8	0.763	1.332	0.559	1.777	0.368	2.287														
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588												
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822										
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.316	2.645	0.203	3.005								
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.379	2.506	0.268	2.832	0.171	3.149						
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.445	2.390	0.328	2.692	0.230	2.985	0.147	3.266				
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360		
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.472	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257	0.502	2.461	0.407	2.667	0.321	2.873	0.244	3.073
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.692	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.732	2.124	0.637	2.290	0.547	2.460	0.461	2.633	0.380	2.806
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.734
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.751	2.174	0.666	2.318	0.584	2.464	0.506	2.613
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.012	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.958	0.874	2.071	0.798	2.188	0.723	2.309	0.650	2.431
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278	0.682	2.396
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251	0.712	2.363
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226	0.741	2.333

<sup>a</sup>K refers to the number of columns in X, including the constant term.

**Table 5** (Continued)

T	K = 2		K = 3		K = 4		K = 5		K = 6		K = 7		K = 8		K = 9		K = 10		K = 11	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$														
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.102	0.836	2.203	0.769	2.306
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181	0.795	2.281
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.080	1.891	1.015	1.979	0.950	2.069	0.885	2.162	0.821	2.257
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144	0.845	2.236
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.877	1.053	1.957	0.991	2.041	0.930	2.127	0.868	2.216
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112	0.891	2.198
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.970	2.098	0.912	2.180
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085	0.932	2.164
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072	0.945	2.149
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022	1.038	2.088
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.156	1.986	1.110	2.044
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959	1.170	2.010
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939	1.222	1.984
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923	1.266	1.964
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.837	1.369	1.873	1.337	1.910	1.305	1.948
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901	1.339	1.935
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893	1.369	1.925
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886	1.396	1.916
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881	1.420	1.909
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877	1.442	1.903
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874	1.462	1.898
150	1.720	1.746	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.847	1.608	1.862	1.594	1.877
200	1.758	1.778	1.748	1.789	1.738	1.799	1.728	1.810	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863	1.665	1.874

## Student-t-Tabelle

Alpha	0.5	0.25	0.20	0.1	0.05	0.025	0.01	0.005	0.001
Freiheitsgrade (K-J-1)									
1	1.000	2.414	3.078	6.314	12.706	25.452	63.656	127.321	636.578
2	0.816	1.604	1.886	2.920	4.303	6.205	9.925	14.089	31.600
3	0.765	1.423	1.638	2.353	3.182	4.177	5.841	7.453	12.924
4	0.741	1.344	1.533	2.132	2.776	3.495	4.604	5.598	8.610
5	0.727	1.301	1.476	2.015	2.571	3.163	4.032	4.773	6.869
6	0.718	1.273	1.440	1.943	2.447	2.969	3.707	4.317	5.959
7	0.711	1.254	1.415	1.895	2.365	2.841	3.499	4.029	5.408
8	0.706	1.240	1.397	1.860	2.306	2.752	3.355	3.833	5.041
9	0.703	1.230	1.383	1.833	2.262	2.685	3.250	3.690	4.781
10	0.700	1.221	1.372	1.812	2.228	2.634	3.169	3.581	4.587
11	0.697	1.214	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	0.695	1.209	1.356	1.782	2.179	2.560	3.055	3.428	4.318
13	0.694	1.204	1.350	1.771	2.160	2.533	3.012	3.372	4.221
14	0.692	1.200	1.345	1.761	2.145	2.510	2.977	3.326	4.140
15	0.691	1.197	1.341	1.753	2.131	2.490	2.947	3.286	4.073
16	0.690	1.194	1.337	1.746	2.120	2.473	2.921	3.252	4.015
18	0.688	1.189	1.330	1.734	2.101	2.445	2.878	3.197	3.922
20	0.687	1.185	1.325	1.725	2.086	2.423	2.845	3.153	3.850
22	0.686	1.182	1.321	1.717	2.074	2.405	2.819	3.119	3.792
24	0.685	1.179	1.318	1.711	2.064	2.391	2.797	3.091	3.745
26	0.684	1.177	1.315	1.706	2.056	2.379	2.779	3.067	3.707
28	0.683	1.175	1.313	1.701	2.048	2.368	2.763	3.047	3.674
30	0.683	1.173	1.310	1.697	2.042	2.360	2.750	3.030	3.646
32	0.682	1.172	1.309	1.694	2.037	2.352	2.738	3.015	3.622
34	0.682	1.170	1.307	1.691	2.032	2.345	2.728	3.002	3.601
36	0.681	1.169	1.306	1.688	2.028	2.339	2.719	2.990	3.582
38	0.681	1.168	1.304	1.686	2.024	2.334	2.712	2.980	3.566
40	0.681	1.167	1.303	1.684	2.021	2.329	2.704	2.971	3.551
42	0.680	1.166	1.302	1.682	2.018	2.325	2.698	2.963	3.538
44	0.680	1.166	1.301	1.680	2.015	2.321	2.692	2.956	3.526
46	0.680	1.165	1.300	1.679	2.013	2.317	2.687	2.949	3.515
48	0.680	1.164	1.299	1.677	2.011	2.314	2.682	2.943	3.505
50	0.679	1.164	1.299	1.676	2.009	2.311	2.678	2.937	3.496
55	0.679	1.163	1.297	1.673	2.004	2.304	2.668	2.925	3.476
60	0.679	1.162	1.296	1.671	2.000	2.299	2.660	2.915	3.460
70	0.678	1.160	1.294	1.667	1.994	2.291	2.648	2.899	3.435
80	0.678	1.159	1.292	1.664	1.990	2.284	2.639	2.887	3.416
90	0.677	1.158	1.291	1.662	1.987	2.280	2.632	2.878	3.402
100	0.677	1.157	1.290	1.660	1.984	2.276	2.626	2.871	3.390
120	0.677	1.156	1.289	1.658	1.980	2.270	2.617	2.860	3.373
200	0.676	1.154	1.286	1.653	1.972	2.258	2.601	2.838	3.340

## F-Verteilungstabelle

Irrtumswkt. $\alpha =$	Freiheitsgrade des Zählers J (Zahl der erklärenden Variablen)													
	1	2	3	4	5	6	7	8	9	10	12	14	16	
Freiheitsgrade des Nenners K-J-1 (K=Zahl der Beobachtungen)	1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.07	61.35
	2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.43
	3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.20
	4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.88	3.86
	5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.25	3.23
	6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.88	2.86
	7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.64	2.62
	8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.48	2.45
	9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.35	2.33
	10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.26	2.23
	12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.12	2.09
	14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.02	2.00
	16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.95	1.93
	18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.90	1.87
	20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.86	1.83
	25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.79	1.76
	30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.74	1.71
	35	2.85	2.46	2.25	2.11	2.02	1.95	1.90	1.85	1.82	1.79	1.74	1.70	1.67
	40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.68	1.65
	45	2.82	2.42	2.21	2.07	1.98	1.91	1.85	1.81	1.77	1.74	1.70	1.66	1.63
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73	1.68	1.64	1.61	
75	2.77	2.37	2.16	2.02	1.93	1.85	1.80	1.75	1.72	1.69	1.63	1.60	1.57	
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66	1.61	1.57	1.54	
150	2.74	2.34	2.12	1.98	1.89	1.81	1.76	1.71	1.67	1.64	1.59	1.55	1.52	
200	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.58	1.54	1.51	
300	2.72	2.32	2.10	1.96	1.87	1.79	1.74	1.69	1.65	1.62	1.57	1.53	1.49	
400	2.72	2.32	2.10	1.96	1.86	1.79	1.73	1.69	1.65	1.61	1.56	1.52	1.49	
500	2.72	2.31	2.09	1.96	1.86	1.79	1.73	1.68	1.64	1.61	1.56	1.52	1.49	

Irrtumswkt.		Freiheitsgrade des Zählers J (Zahl der erklärenden Variablen)													
$\alpha = 0.05$		1	2	3	4	5	6	7	8	9	10	12	14	16	
Freiheitsgrade des Nenners K-J-1 (K=Zahl der Beobachtungen)	1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.90	245.36	246.47	
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.33	
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.25	
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.18	
	25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.07	
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	1.99	
	35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.04	1.99	1.94	
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.90	
	45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	1.97	1.92	1.87	
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.89	1.85		
75	3.97	3.12	2.73	2.49	2.34	2.22	2.13	2.06	2.01	1.96	1.88	1.83	1.78		
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.79	1.75		
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.82	1.76	1.71		
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.74	1.69		
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.78	1.72	1.68		
400	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.78	1.72	1.67		
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.77	1.71	1.66		

Irrtumswkt. $\alpha = 0.025$		Freiheitsgrade des Zählers J (Zahl der erklärenden Variablen)													
		1	2	3	4	5	6	7	8	9	10	12	14	16	
Freiheitsgrade des Nenners K-J-1 (K=Zahl der Beobachtungen)	1	647.79	799.48	864.15	899.60	921.83	937.11	948.20	956.64	963.28	968.63	976.72	982.55	986.91	
	2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.44	
	3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.28	14.23	
	4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.68	8.63	
	5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.46	6.40	
	6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.30	5.24	
	7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.60	4.54	
	8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.13	4.08	
	9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.80	3.74	
	10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.55	3.50	
	12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.21	3.15	
	14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.98	2.92	
	16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.82	2.76	
	18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.70	2.64	
	20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.60	2.55	
	25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.44	2.38	
	30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.34	2.28	
	35	5.48	4.11	3.52	3.18	2.96	2.80	2.68	2.58	2.50	2.44	2.34	2.27	2.21	
	40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.21	2.15	
	45	5.38	4.01	3.42	3.09	2.86	2.70	2.58	2.49	2.41	2.35	2.25	2.17	2.11	
50	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32	2.22	2.14	2.08		
75	5.23	3.88	3.30	2.96	2.74	2.58	2.46	2.37	2.29	2.22	2.12	2.05	1.99		
100	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.24	2.18	2.08	2.00	1.94		
150	5.13	3.78	3.20	2.87	2.65	2.49	2.37	2.28	2.20	2.13	2.03	1.95	1.89		
200	5.10	3.76	3.18	2.85	2.63	2.47	2.35	2.26	2.18	2.11	2.01	1.93	1.87		
300	5.07	3.73	3.16	2.83	2.61	2.45	2.33	2.23	2.16	2.09	1.99	1.91	1.85		
400	5.06	3.72	3.15	2.82	2.60	2.44	2.32	2.22	2.15	2.08	1.98	1.90	1.84		
500	5.05	3.72	3.14	2.81	2.59	2.43	2.31	2.22	2.14	2.07	1.97	1.89	1.83		

Irrtumswkt.		Freiheitsgrade des Zählers J (Zahl der erklärenden Variablen)												
$\alpha = 0.01$		1	2	3	4	5	6	7	8	9	10	12	14	16
Freiheitsgrade des Nenners K-J-1 (K=Zahl der Beobachtungen)	1	4052.18	4999.34	5403.53	5624.26	5763.96	5858.95	5928.33	5980.95	6022.40	6055.93	6106.68	6143.00	6170.01
	2	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.42	99.43	99.44
	3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.05	26.92	26.83
	4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.25	14.15
	5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.77	9.68
	6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.60	7.52
	7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.36	6.28
	8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.56	5.48
	9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	5.01	4.92
	10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.60	4.52
	12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.05	3.97
	14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.70	3.62
	16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.45	3.37
	18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.27	3.19
	20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.13	3.05
	25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.89	2.81
	30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.74	2.66
	35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.74	2.64	2.56
	40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.56	2.48
	45	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74	2.61	2.51	2.43
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.56	2.46	2.38	
75	6.99	4.90	4.05	3.58	3.27	3.05	2.89	2.76	2.65	2.57	2.43	2.33	2.25	
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.37	2.27	2.19	
150	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.31	2.20	2.12	
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.27	2.17	2.09	
300	6.72	4.68	3.85	3.38	3.08	2.86	2.70	2.57	2.47	2.38	2.24	2.14	2.06	
400	6.70	4.66	3.83	3.37	3.06	2.85	2.68	2.56	2.45	2.37	2.23	2.13	2.05	
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.22	2.12	2.04	

Irrtumswkt. $\alpha = 0.005$		Freiheitsgrade des Zählers J (Zahl der erklärenden Variablen)													
		1	2	3	4	5	6	7	8	9	10	12	14	16	
Freiheitsgrade des Nenners K-J-1 (K=Zahl der Beobachtungen)	1	16212.46	19997.36	21614.13	22500.75	23055.82	23439.53	23715.20	23923.81	24091.45	24221.84	24426.73	24572.01	24683.77	
	2	198.50	199.01	199.16	199.24	199.30	199.33	199.36	199.38	199.39	199.39	199.42	199.42	199.45	
	3	55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.88	43.68	43.39	43.17	43.01	
	4	31.33	26.28	24.26	23.15	22.46	21.98	21.62	21.35	21.14	20.97	20.70	20.51	20.37	
	5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.21	13.09	
	6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.88	9.76	
	7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	8.03	7.91	
	8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.87	6.76	
	9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.09	5.98	
	10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.53	5.42	
	12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.77	4.67	
	14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.43	4.30	4.20	
	16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.10	3.97	3.87	
	18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.86	3.73	3.64	
	20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.55	3.46	
	25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.37	3.25	3.15	
	30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.06	2.96	
	35	8.98	6.19	5.09	4.48	4.09	3.81	3.61	3.45	3.32	3.21	3.05	2.93	2.83	
	40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.95	2.83	2.74	
	45	8.71	5.97	4.89	4.29	3.91	3.64	3.43	3.28	3.15	3.04	2.88	2.76	2.66	
50	8.63	5.90	4.83	4.23	3.85	3.58	3.38	3.22	3.09	2.99	2.82	2.70	2.61		
75	8.37	5.69	4.63	4.05	3.67	3.41	3.21	3.05	2.93	2.82	2.66	2.54	2.45		
100	8.24	5.59	4.54	3.96	3.59	3.33	3.13	2.97	2.85	2.74	2.58	2.46	2.37		
150	8.12	5.49	4.45	3.88	3.51	3.25	3.05	2.89	2.77	2.67	2.51	2.38	2.29		
200	8.06	5.44	4.41	3.84	3.47	3.21	3.01	2.86	2.73	2.63	2.47	2.35	2.25		
300	8.00	5.39	4.36	3.80	3.43	3.17	2.97	2.82	2.69	2.59	2.43	2.31	2.21		
400	7.97	5.37	4.34	3.78	3.41	3.15	2.95	2.80	2.68	2.57	2.41	2.29	2.20		
500	7.95	5.35	4.33	3.76	3.40	3.14	2.94	2.79	2.66	2.56	2.40	2.28	2.19		