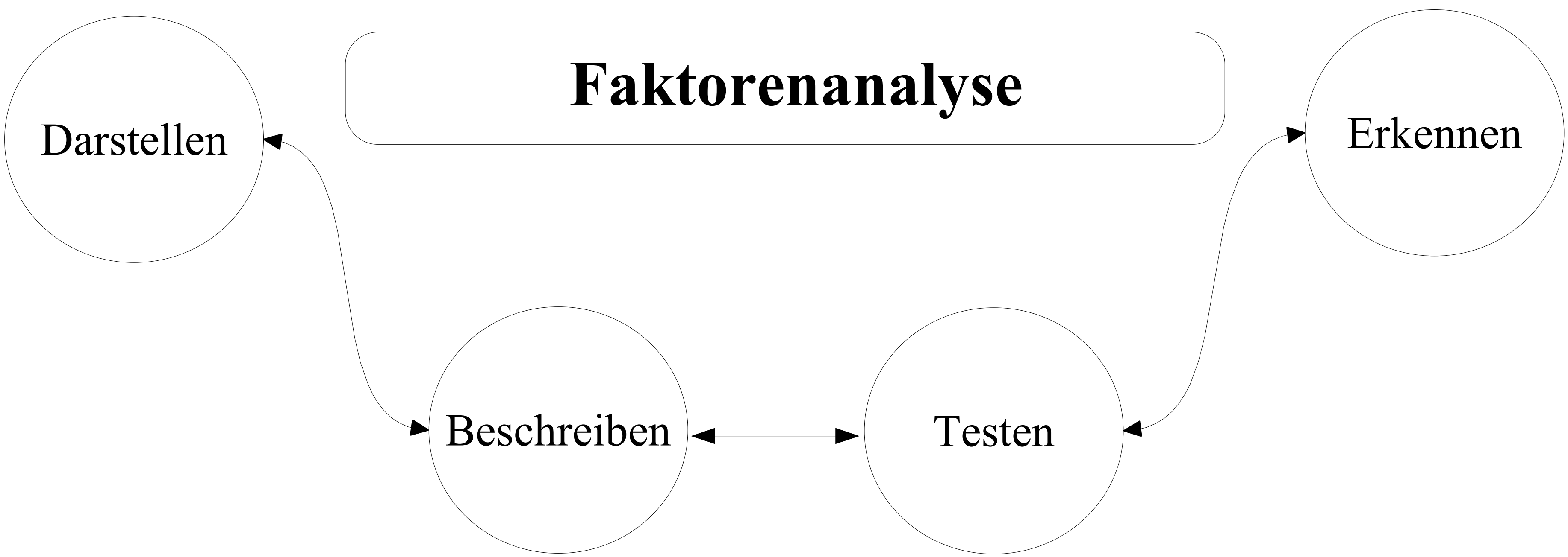
Vertiefungsrichtung Marktforschung

Sommersemester 2006

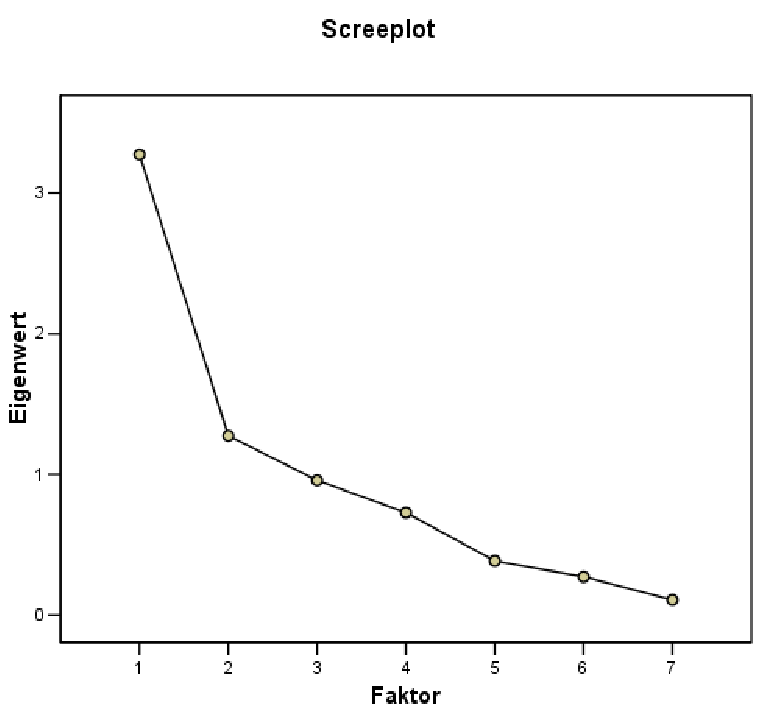
Dipl.-WiInf.(FH) Christian Reinboth

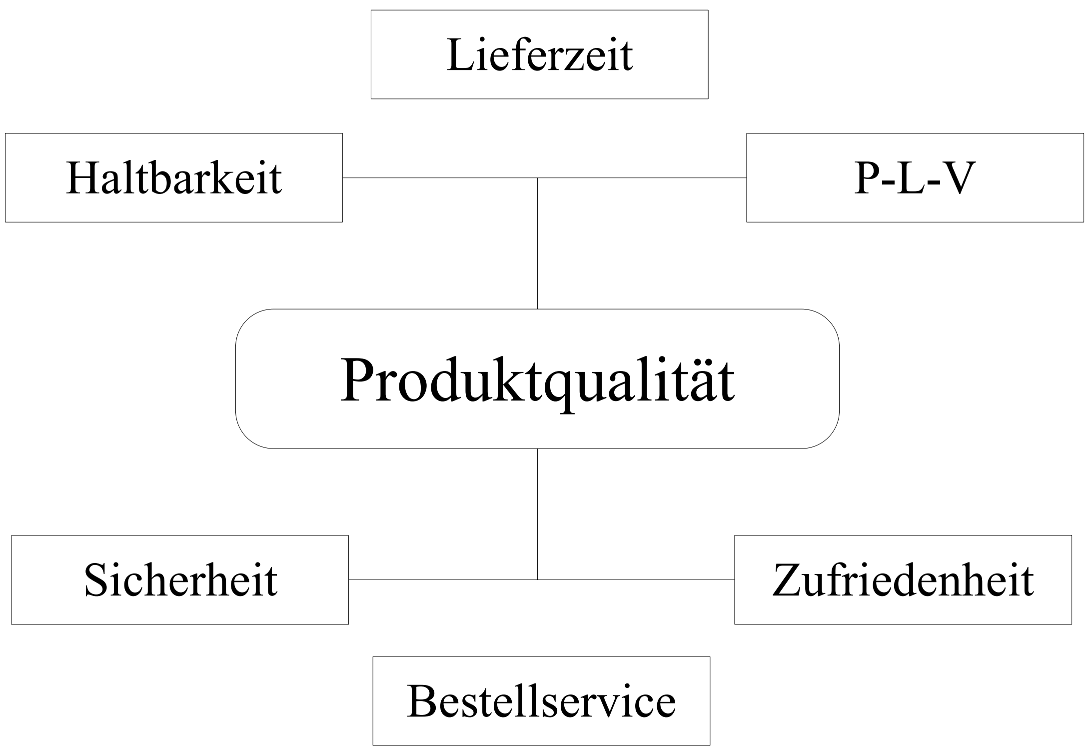
Faktorenanalyse



Faktorenanalyse

- In der Marktforschung hat man es häufig mit komplexen Begriffen und Sachverhalten zu tun
- Begriffe wie „Nutzen“ oder „Qualität“ lassen sich nicht durch eine einzige Variable ausdrücken
- Um beispielsweise die Qualität abzubilden, wird ein ganzes Bündel von Variablen benötigt:
  - Haltbarkeit, Preis-Leistungs-Verhältnis, Zuverlässigkeit, Zufriedenheit...
- Ziel: Reduktion von vielen Variablen auf komplexere Hintergrundvariablen
- Die Faktorenanalyse wird daher auch als dimensionsreduzierendes Verfahren bezeichnet





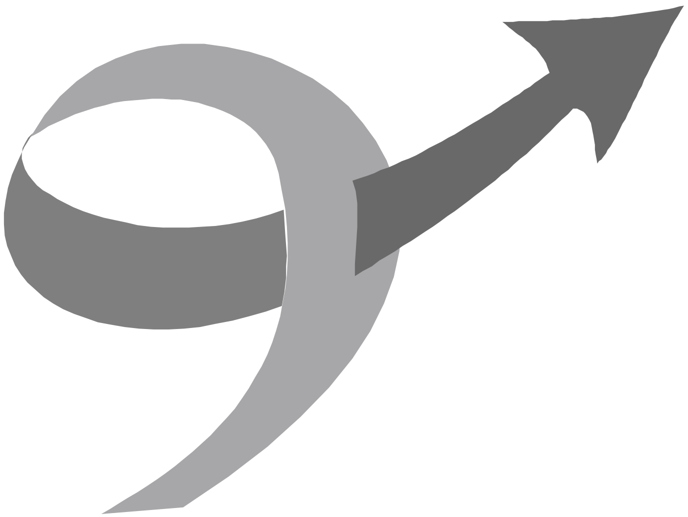
Inhalte: Faktorenanalyse

- Hintergründe der Faktoranalyse
- Was ist eine Faktorenanalyse?
- Anwendungsbeispiel: Frühgeburten
- Explorativ oder konfirmatorisch?
- Zielkonflikt der Faktorenanalyse
- Kausale Interpretation von Korrelationen
- Schritt 1: Variablenauswahl
  - Kriterien der Variablenauswahl
  - Eignung der Korrelationsmatrix
    - Prüfung auf Normalverteilung
    - Signifikanzniveau der Korrelationen
    - Struktur der Inversen
    - Bartlett-Test auf Spherizität
    - Anti-Image-Kovarianz-Matrix
    - Kaiser-Meyer-Oklin-Kriterium
- Schritt 2: Faktorextraktion
  - Das Fundamentaltheorem
  - Grafische Interpretation der Faktoren
  - Die Extraktionsproblematik
  - Bestimmung der Kommunalitäten
  - Die Hauptachsenanalyse
  - Die Hauptkomponentenanalyse
  - Bestimmung der Faktoranzahl
    - Kaiser-Kriterium
    - Screeplot / Scree-Test
- Schritt 3: Interpretation und Rotation
  - Interpretation der Faktorladungen
  - Einführung in die Rotation
  - Ablauf der Rotation
  - Orthogonale vs. Oblique Rotation
  - Rotationsmethoden
- Schritt 4: Bestimmung der Faktorwerte
  - Bestimmung der Faktorwerte
  - Interpretation der Faktorwerte
  - Grafische Darstellung der Faktorwerte
- Rechenschritte der Faktorenanalyse

Hintergründe der Faktorenanalyse

- Für viele Fragestellungen ist die **Untersuchung eines Wirkungszusammenhangs** zwischen **einer abhängigen und einer oder mehreren unabhängigen Variablen** wichtig

- Bei einer geringen Anzahl von unabhängigen Variablen lassen sich zu diesem Zweck problemlos **Korrelations- und Regressionsanalysen** durchführen:
  - Fragestellungen aus dem Bereich der Physik
  - Fragestellungen aus dem Bereich der Ingenieurswissenschaft



- Existieren dagegen zu viele Variablen, wird die Auswertung komplizierter, zudem ist in solchen Fällen häufig unklar, welche der vielen Variablen **unabhängig voneinander zum Erklärungsmodell beitrage**n:
  - Fragestellungen aus dem Bereich der Marktforschung
  - Fragestellungen aus dem Bereich der Sozialwissenschaften

- Um aus einer Masse von Variablen **voneinander unabhängige Beschreibungs- und Erklärungsfaktoren** zu extrahieren, kann das Verfahren der Faktorenanalyse eingesetzt werden

- Dies führt nicht nur zu einer **Vereinfachung bei der Auswertung** durch Reduktion der Variablen auf Hintergrundfaktoren, sondern erlaubt es den Durchführenden auch, zunächst **alle interessant erscheinenden Merkmale zu erheben** und im Anschluss **irrelevante Variablen wieder auszuschließen**

Was ist eine Faktorenanalyse?

- Die Faktorenanalyse gehört zu den **strukturen-entdeckenden Verfahren**
  - Ziel dieser Verfahren ist die **Aufdeckung von Zusammenhängen zwischen Variablen**
  - Auf eine **vorausgehende Aufteilung in abhängige und unabhängige Variablen** wird daher verzichtet

- Sie wird angewendet, wenn eine **Bündelung von Variablen** von Interesse ist
  - Dies kann der Fall sein, wenn eine große Vielzahl von Variablen zu einer Fragestellung erhoben wurde
  - Es stellt sich die Frage, ob die vielen Variablen auf **einige wenige zentrale Faktoren** reduziert werden können
  - Das Ziel ist also eine **Verdichtung von Informationen** bzw. die **Erkennung erklärungsrelevanter Variablen**

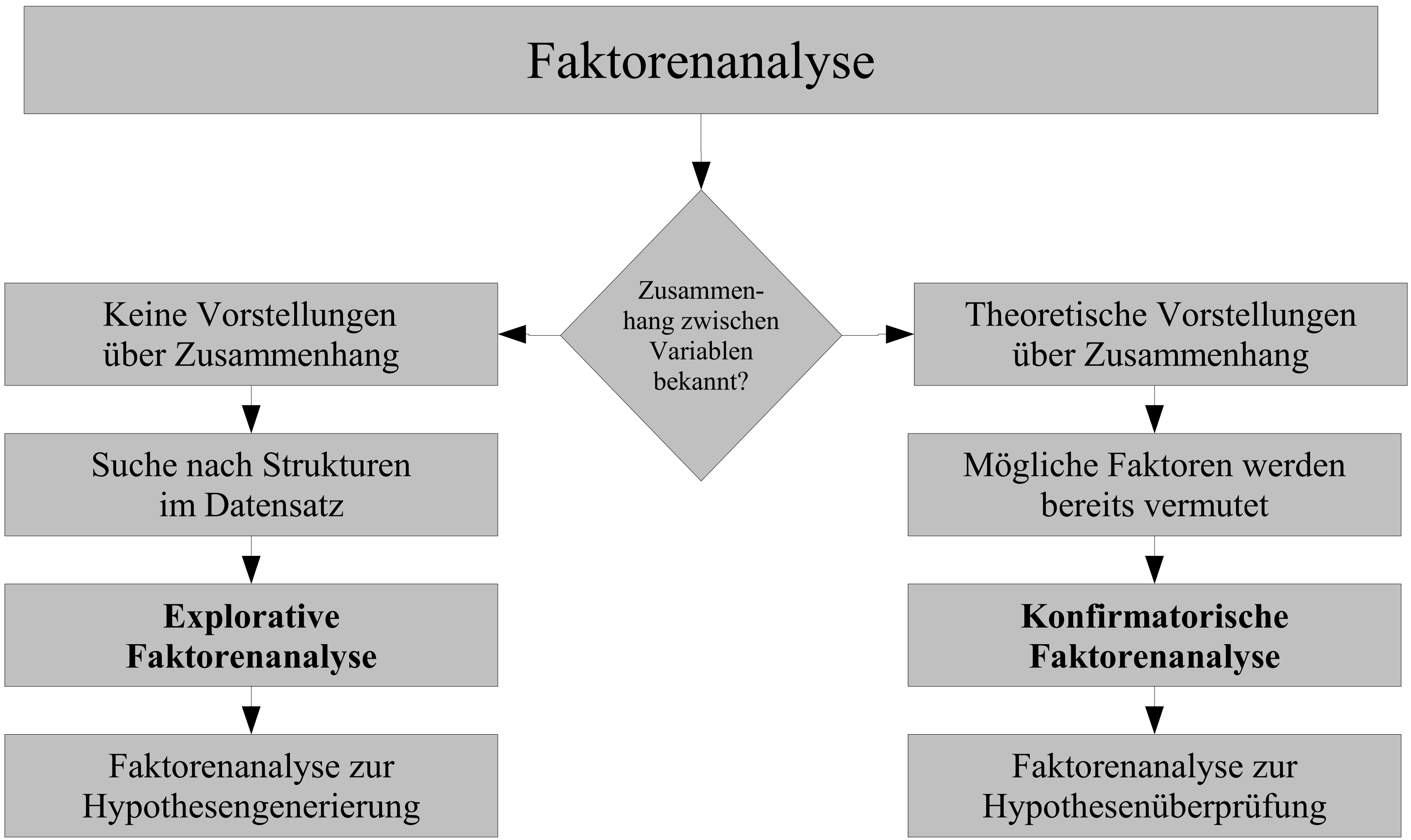
- Beispiel: Verdichtung einer Vielzahl von technischen Eigenschaften beim Auto auf wenige Dimensionen:
  - Technische Eigenschaften: maximale Drehzahl, PS, Höchstgeschwindigkeit, Maximalbeschleunigung...
  - Verdichtete Dimensionen: Leistung, Sicherheit, Größe...

- Wesentlicher Anwendungsbereich der Faktorenanalyse sind **Positionierungsanalysen** (faktorielle Positionierung)
  - **Subjektive Eigenschaftsbeurteilungen** werden auf **zugrundeliegende Beurteilungsdimensionen** verdichtet
  - Bei einer Verdichtung auf zwei oder drei Dimensionen ist eine **grafische Positionierungsdarstellung** möglich
  - Das Verfahren kann beispielsweise bei Qualitätsmarken, Unternehmen oder politischen Parteien angewandt werden

Anwendungsbeispiel: Frühgeburten

- Laufende Studie an der Hamburger Universitätsklinik Eppendorf (UKE)
- Frühgeborene aus den Jahren 1983 – 1986 werden regelmäßig untersucht
  - Als Frühgeburten gelten alle Säuglinge mit einem Gewicht von maximal 1.500g
  - Für diese Säuglinge wurden bei der Geburt verschiedene Daten erhoben
  - Seitdem werden sie regelmäßig psychologisch und neurologisch untersucht
- Warum ist diese Untersuchung wichtig?
  - Intensivmedizin hat dazu beigetragen, dass inzwischen 65% statt 15% aller Frühgeburten überleben
  - Folge davon ist vermutlich ein gestiegener Anteil neurologisch oder psychosozial gestörter Kinder
  - Ziel der Untersuchung ist es, diesen Zusammenhang zu bestätigen oder zu verwerfen
- Zur Untersuchung gehören auch verschiedene Intelligenztests
  - Im Rahmen dieser Tests fällt eine Vielzahl von Variablen an:
    - Wortergänzungsfähigkeit, Grammatikkenntnisse, Zahlenfolgegedächnis, Konzentrationsfähigkeit....
  - Diese Variablen lassen sich auf zwei wesentliche Faktoren reduzieren:
    - Allgemeine Intelligenz (AI) und sprachliche Intelligenz (SI)
  - Das Verfahren für diese Reduktion ist die Faktorenanalyse

Explorativ oder konfirmatorisch?



Zielkonflikt der Faktorenanalyse

Anzahl der zu extrahierenden Faktoren

Geringe Reduktion
Geringer Analysenutzen

Hohe Reduktion
Hoher Analysenutzen

Hohe Faktor-anzahl

Geringe Faktor-anzahl

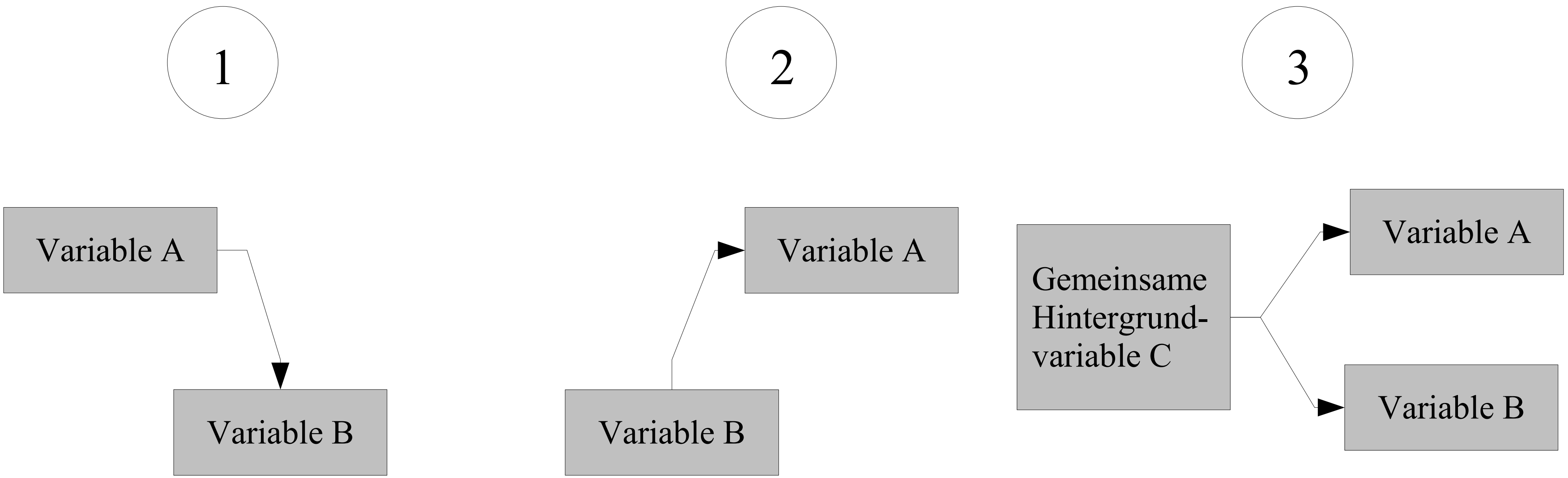
Geringer Informationsverlust

Hoher Informationsverlust

Die Entscheidung ist dem Anwender überlassen!

Kausale Interpretation von Korrelationen

Drei Interpretationsmöglichkeiten:



Variable A beeinflusst Variable B

Variable B beeinflusst Variable A

Beeinflussung beider Variablen durch C

In der Faktorenanalyse wird immer die dritte Möglichkeit unterstellt!

Nur wenn dies sachlogisch möglich erscheint, ist eine Faktorenanalyse zulässig!

Ablauf einer Faktorenanalyse

Schritt 1 Auswahl der Variablen und Erstellung der Korrelationsmatrix	Für alle in die Faktorenanalyse einbezogenen Variablen wird die Korrelationsmatrix erstellt. Aus dieser kann abgelesen werden, welche Variablen für die weitere Analyse unberücksichtigt bleiben sollten, da sie mit den übrigen Variablen nur minimal korrelieren.
Schritt 2 Faktorextraktion	Dieser Schritt wird auch als „Ziehen“ von Faktoren bezeichnet. Anhand verschiedener statistischer Kennwerte kann entschieden werden, ob das gefundene Faktorenmodell geeignet ist, die vorliegenden Variablen auf Hintergrundvariablen zurückzuführen.
Schritt 3 Interpretation der Faktoren und Faktorrotation	Die im zweiten Schritt gefunden Faktoren sind in der Regel nur sehr schwer oder gar nicht zu interpretieren. Um ihre Auswertung zu erleichtern, werden sie einer Transformation unterzogen, die als Rotation bezeichnet wird.
Schritt 4 Bestimmung der Faktorwerte	Im letzten Schritt wird bestimmt, welche Werte die untersuchten Objekte hinsichtlich der extrahierten Faktoren annehmen.

Variablenauswahl & Korrelationsmatrix

- Die Untersuchungsmerkmale sind **sorgfältig auszuwählen**
  - **Auslassung irrelevanter Merkmale**
  - **Zusammenfassung ähnlicher Merkmale**
- Die Stichprobe sollte zudem **möglichst homogen** sein
  - Homogenitätsgrad der Stichprobe beeinflusst Korrelation zwischen den Variablen

- Faktoren sind als **„hinter den Variablen" stehende Größen** aufzufassen
  - Sie repräsentieren den **Zusammenhang zwischen verschiedenen Ausgangsvariablen**
- Durch eine Korrelationsrechnung werden diese Zusammenhänge meßbar gemacht
  - Korrelationen zeigen Z**usammenhänge zwischen Variablenpaaren**
  - Variablen können als **voneinander abhängig** und damit „bündelungsfähig" erkannt werden

- Die Korrelation zwischen zwei Variablen wird mittels Bravais-Pearson berechnet:

$$r_{(x1,x2)} = \frac{\left(\sum_{k=1}^{K}\left((x_{k1}-\bar{x}_1)*(x_{k2}-\bar{x}_2)\right)\right)}{\left(\sqrt{\left(\sum_{k=1}^{K}(x_{k1}-\bar{x}_1)^2\right)*\left(\sum_{k=1}^{K}(x_{k2}-\bar{x}_2)^2\right)}\right)}$$

x_{k1} = Ausprägung der Variablen 1 bei Objekt k

$\bar{x}_1$ = Mittelwert der Ausprägungen von Variable 1 über alle Objekte k

x_{k2} = Ausprägung der Variablen 2 bei Objekt k

$\bar{x}_2$ = Mittelwert der Ausprägungen von Variable 1 über alle Objekte k

Variablenauswahl & Korrelationsmatrix

- Vor der Berechnung der Korrelationen ist ggf. die **Ausgangsdatenmatrix zu standardisieren**, da dadurch:
  - die Korrelationsrechnung und die Rechenschritte für die Faktorenanalyse erleichtert werden
  - die Interpretation der Ergebnisse der Faktorenanalyse vereinfacht wird
  - Variablen mit unterschiedlichen Maßeinheiten vergleichbar gemacht werden

- Das **Standardisierungsverfahren** ist bereits aus Statistik II bekannt: $Z_{kj}=\frac{(x_{kj}-\bar{x}_j)}{s_j}$
  - Bildung der **Differenz zwischen Mittelwert und Beobachtungswert einer Variablen**
  - Anschließende **Division durch die Standardabweichung**
- Dadurch ist sichergestellt, dass der neue **Mittelwert Null** und die neue **Standardabweichung Eins** ist

- Die Korrelationsmatrix der Ausgangsdaten ist identisch mit der Korrelationsmatrix der standardisierten Daten
- Im Falle der Korrelationsmatrix aus standardisierten Daten sind Varianz-Kovarianzmatrix und Korrelationsmatrix identisch

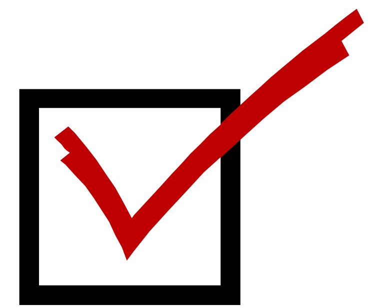
- Der Korrelationskoeffizient ist daher auch: $r_{(x_1,x_2)}=\frac{S_{(x_1,x_2)}}{(S_{(x_1)}S_{(x_2)})}\ mit: S_{(x_1,x_2)}=\frac{1}{(K-1)}\sum_k (x_{(k_1)}-\bar{x}_1)(x_{(k_2)}-\bar{x}_2)$
- Da wegen der Standardisierung beide Varianzen im Nenner Eins sind, sind Korrelationskoeffizient und Kovarianz gleich

- Die Korrelationsmatrix...
  - zeigt, welche Variablen der Ausgangsdaten offenbar mit welchen anderen Variablen **„irgendwie zusammenhängen“**
  - zeigt nicht, ob Variablen sich **gegenseitig bedingen** oder die Korrelation durch **Hintergrundfaktor(en)** zustandekommt

Eignung der Korrelationsmatrix

- Die **Eignung der Ausgangsdaten für die Faktorenanalyse** spiegelt sich in der Korrelationsmatrix wieder
- Eine **Überprüfung der Eignung anhand verschiedener Prüfkriterien** ist anzuraten

- Insgesamt stehen **sechs Prüfkriterien** zur Auswahl:
  - Prüfung der Variablen auf **Normalverteilung** (Explorative Datenanalyse)
  - Überprüfung des **Signifikanzniveaus** der Korrelationen
  - Analyse der **Struktur der Inversen** der Korrelationsmatrix
  - Durchführung eines **Bartlett-Tests auf Spherizität**
  - Analyse der **Anti-Image-Kovarianz-Matrix**
  - Überprüfung des **Kaiser-Meyer-Oklin-Kriteriums**



- **Nicht alle Kriterien** müssen vor Weiterführung der Analyse zwingend geprüft werden
- Anzuraten ist aber die Überprüfung anhand **mehr als nur eines Kriteriums**
- Insbesondere die **Signifikanzniveaus** der Korrelationen und das **Kaiser-Meyer-Olkin-Kriterium** sind zu beachten

Sind die Daten für die Faktorenanalyse geeignet?

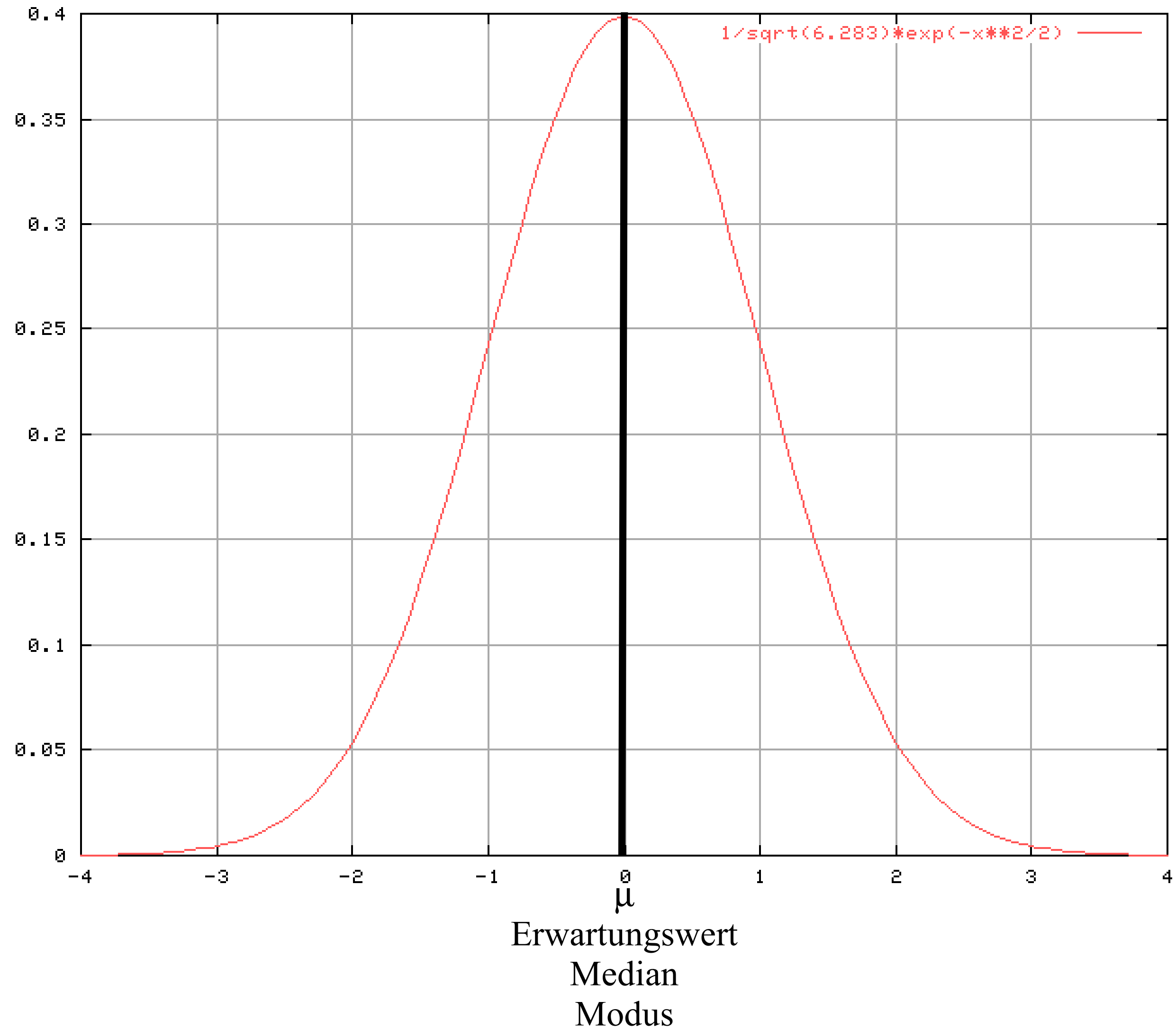
Normalverteilungsprüfung: Einführung

- Die Gauß- oder Normalverteilung ist die **wichtigste kontinuierliche Wahrscheinlichkeitsverteilung** $f(x)=\frac{1}{(\sigma\sqrt{(2\pi)})}e^{((\frac{-1}{2}(\frac{(x-\mu)}{\sigma})^2))}$

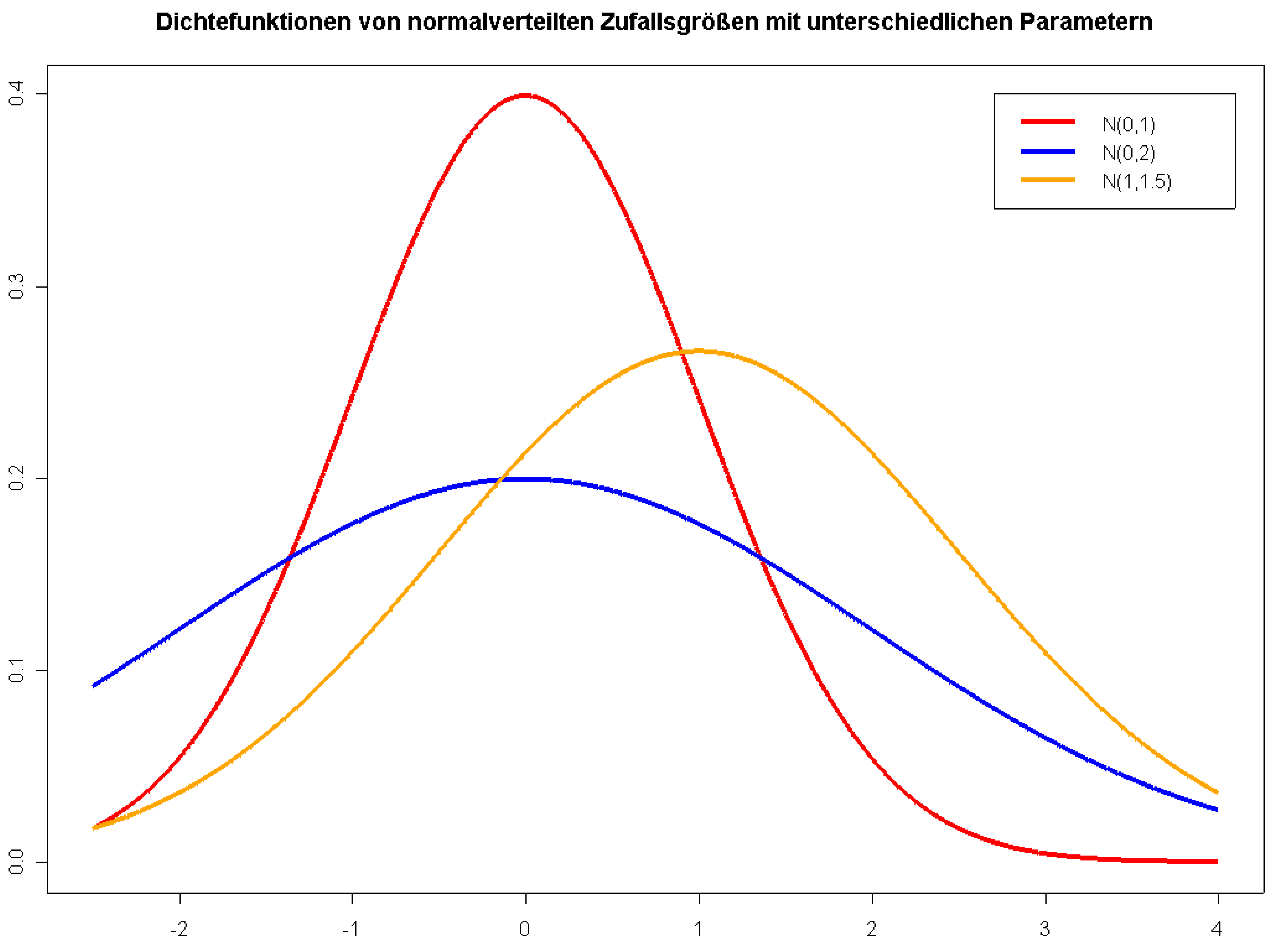
- Die zugehörige **Dichtefunktion** ist als **Gaußsche Glockenkurve** bekannt

- Eigenschaften**:**
  - Dichtefunktion ist **glockenförmig** und **symmetrisch**
  - **Erwartungswert, Median** und **Modus** sind **gleich**
  - **Zufallsvariable** hat eine **unendliche Spannweite**

- Viele statistische Verfahren **setzen die Normalverteilung** der Daten in der Grundgesamtheit **voraus**
- Es ist daher häufig zu **prüfen**, ob von einer solchen Verteilung ausgegangen werden kann (auch näherungsweise)



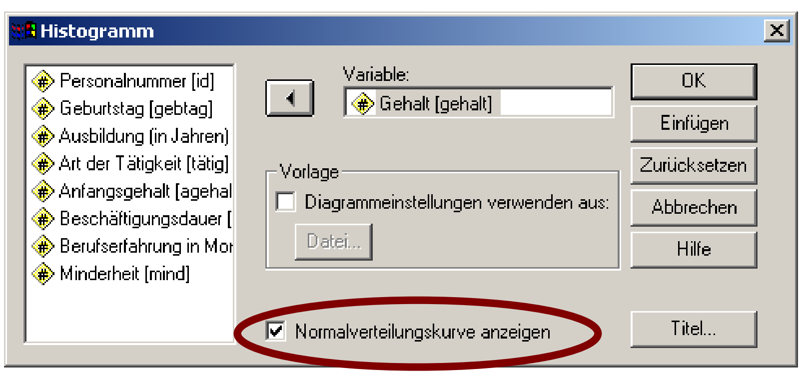
Normalverteilungsprüfung: Dichtefunktion

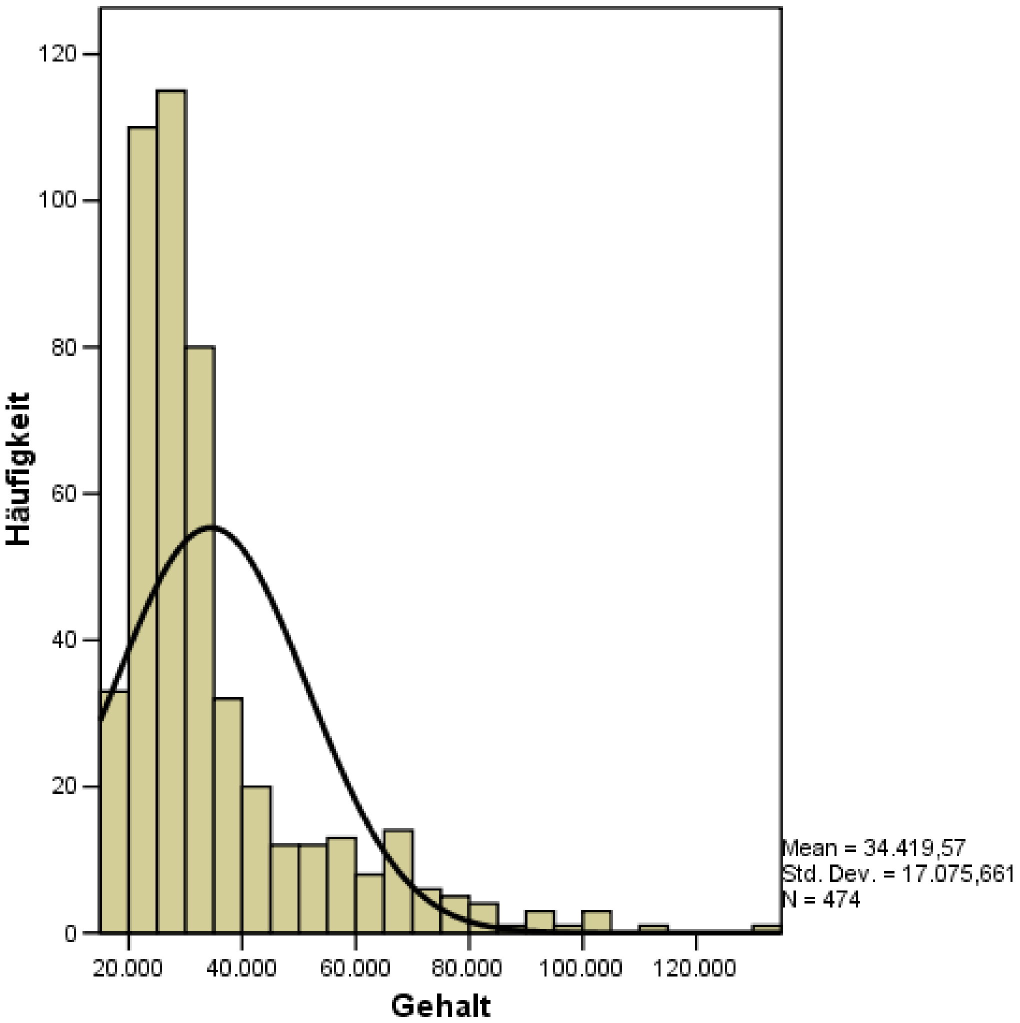


Normalverteilungsprüfung: Histogramm

- Grafische Analyse mit **Histogramm und überlagerter Normalverteilungskurve**

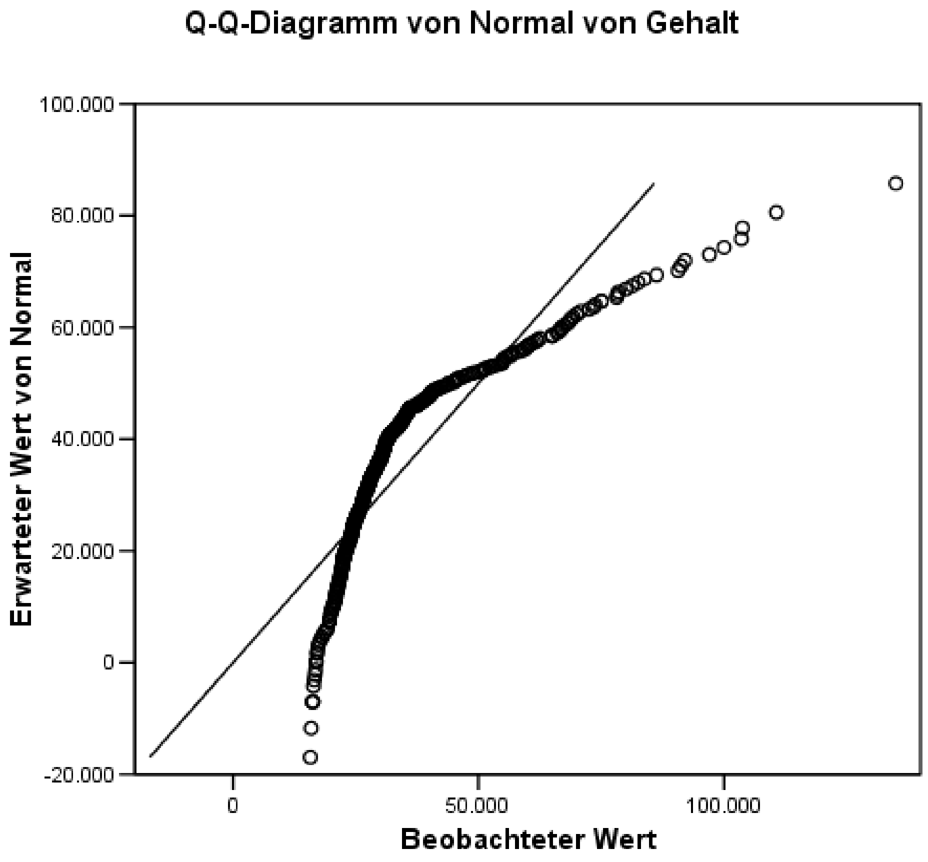
- Die Balken des Histogramms spiegeln die **Breite der Wertebereiche** wieder – da zudem für leere Wertebereiche ein Freiraum ausgegeben wird, kommt im Histogramm die **gesamte empirische Verteilung** der Variablen zum Ausdruck
- Dies ermöglicht den **direkten Vergleich** mit einer **eingezeichneten theoretischen Verteilung**, wie beispielsweise der Normalverteilung
- Der Grad der Abweichung einer Normalverteilung lässt sich auch anhand verschiedener Maßzahlen wie **Exzeß** (Kurtosis) und **Schiefe** bestimmen

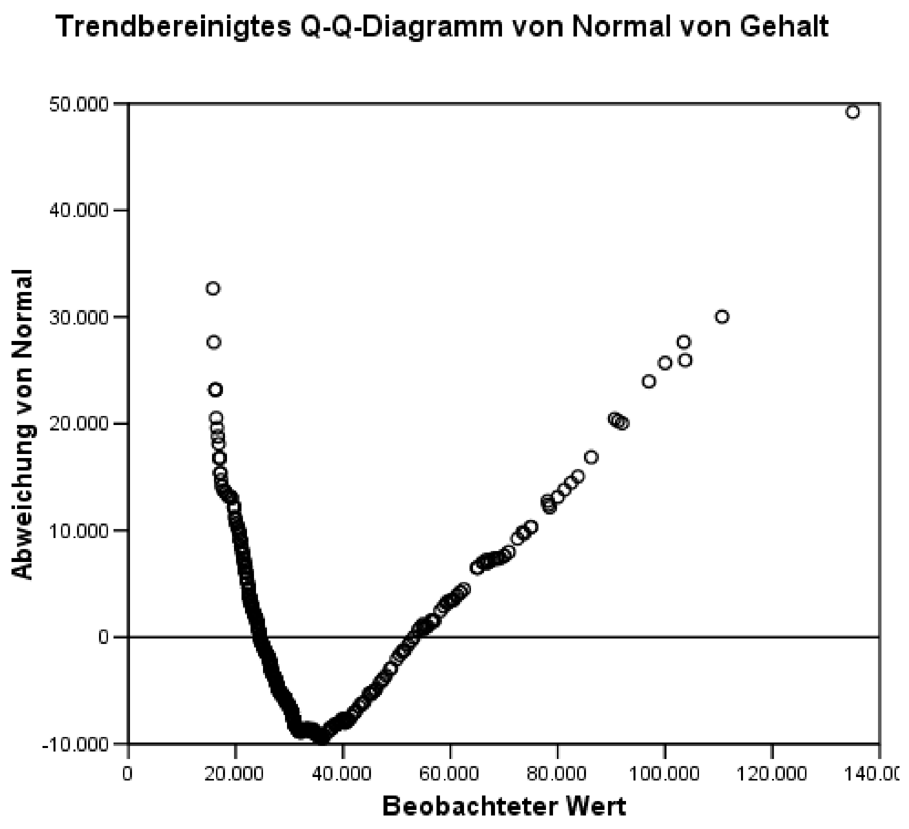




Normalverteilungsprüfung: Q-Q

- Grafische Analyse mit Q-Q-Diagramm und trendbereinigtem Q-Q-Diagramm





Normalverteilungsprüfung: K-S-A

- Die Prüfung auf Vorliegen einer Normalverteilung kann auch mit einem **Anpassungstests** erfolgen
- In SPSS lässt sich dazu beispielsweise der **Kolmogorov-Smirnov-Anpassungstest** nutzen
- Der Test arbeitet mit der **kumulierten empirischen** und der **kumulierten erwarteten Referenzverteilung**
- Die maximale Differenz zwischen beiden Verteilungen wird zur Berechnung der Prüfgröße Z nach Kolmogorov-Smirnov verwendet, mit der dann aus einer Tabelle der für einen Stichprobenumfang n kritische Wert für die maximale Differenz bei einem gegebenen Signifikanzniveau abgelesen werden kann

- **Nullhypothese H_0** des SPSS-Tests: **die Werte der untersuchten Variablen sind normalverteilt**
- Berechnet wird die Wahrscheinlichkeit, mit der das Zurückweisen dieser Hypothese falsch ist (**Signifikanzwert**)
- **Je größer diese Wahrscheinlichkeit ausfällt**, desto eher ist von einer **Normalverteilung der Werte auszugehen**

- Im nebenstehenden Beispiel eines Kolmogorov-Smirnov-Tests fällt der Signifikanzwert mit 0,00 so niedrig aus, dass die Annahme der Normalverteilung zurückzuweisen ist
- Bei der Interpretation ist zu beachten, dass es sich um einen **Test auf perfekte Normalverteilung** handelt
- Anzuraten ist daher die **Kombination** mit einem der **grafischen Prüfverfahren**

Kolmogorov-Smirnov-Anpassungstest

		Gehalt
N		474
Parameter der Normalverteilung[a,b]	Mittelwert	34.419,57
	Standardabweichung	17.075,661
Extremste Differenzen	Absolut	,208
	Positiv	,208
	Negativ	-,143
Kolmogorov-Smirnov-Z		4,525
Asymptotische Signifikanz (2-seitig)		,000

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Signifikanzniveau der Korrelationen

- Der Signifikanzwert gibt die **Wahrscheinlichkeit** an, **mit der eine zuvor formulierte Hypothese zutrifft**
- Für **alle Korrelationskoeffizienten** lassen sich die **Signifikanzniveaus** ausgeben
- Die **Nullhypothese H_0** besagt, dass in der Grundgesamtheit **kein Zusammenhang zwischen den Variablen** besteht (r=0)
- Das Signifikanzniveau des Korrelationskoeffizienten gibt die **Irrtumswahrscheinlichkeit bei Ablehnung von H_0** an
- Ein Signifikanzwert von 0,00 bedeutet, dass der Analytiker mit einer Wahrscheinlichkeit von beinahe(!) 0% einen Fehler macht, wenn er die Nullhypothese verwirft, und von einem signifikanten Zusammenhang zwischen den Variablen ausgeht

Korrelationsmatrix[a]

		Hubraum (cu. inches)	PS	Gewicht (lbs.)	Beschleunigung von 0 auf 100 km/h (sec.)	Anzahl der Zylinder
Korrelation	Hubraum (cu. inches)	1,000	,899	,934	-,562	,952
	PS	,899	1,000	,865	-,709	,844
	Gewicht (lbs.)	,934	,865	1,000	-,438	,896
	Beschleunigung von 0 auf 100 km/h (sec.)	-,562	-,709	-,438	1,000	-,528
	Anzahl der Zylinder	,952	,844	,896	-,528	1,000
Signifikanz (1-seitig)	Hubraum (cu. inches)		,000	,000	,000	,000
	PS	,000		,000	,000	,000
	Gewicht (lbs.)	,000	,000		,000	,000
	Beschleunigung von 0 auf 100 km/h (sec.)	,000	,000	,000		,000
	Anzahl der Zylinder	,000	,000	,000	,000	

a. Determinante = 7,980E-04

- Im vorliegenden Fall ist beispielsweise festzustellen, dass sich die Korrelationen zwischen den Variablen mit einer Wahrscheinlichkeit nahe 100% von Null unterscheiden (0% und 100% nie erreichbar > asymptotischer Kurvenverlauf)

Struktur der Inversen der Korrelationsmatix

- Die Eignung einer Korrelationsmatrix für die Faktorenanalyse lässt sich auch an der **Struktur der Inversen** erkennen
- Es wird davon ausgegangen, dass die Daten geeignet sind, **wenn die Inverse eine Diagonalmatrix darstellt**, d.h. die nicht-**diagonalen Elemente der inversen Korrelationsmatrix relativ nahe bei Null liegen**
- Es existiert **kein eindeutiges Kriterium** dafür, wie stark bzw. wie häufig die nicht-diagonalen Elemente von Null abweichen dürfen, ohne dass die Eignung der Daten in Frage gestellt werden muss

Inverse Korrelationsmatrix

	Hubraum (cu. inches)	PS	Gewicht (lbs.)	Beschleunigung von 0 auf 100 km/h (sec.)	Anzahl der Zylinder
Hubraum (cu. inches)	20,440	-3,978	-6,242	,549	-10,218
PS	-3,978	9,396	-4,174	3,306	1,347
Gewicht (lbs.)	-6,242	-4,174	10,403	-2,551	-1,208
Beschleunigung von 0 auf 100 km/h (sec.)	,549	3,306	-2,551	2,764	,434
Anzahl der Zylinder	-10,218	1,347	-1,208	,434	10,903

Bartlett-Test auf Spherizität

- Der Bartlett-Test (test of sphericity) überprüft die **Nullhypothese H_0**, dass **alle(!) Variablen der Grundgesamtheit**, aus der die untersuchte Stichprobe stammt, **untereinander unkorreliert sind**
- Dies würde bedeuten, dass sich die gefundenen Korrelationen auf **Zufallseffekte bei der Stichprobenziehung** zurückführen lassen, und in der Grundgesamtheit **keinerlei Zusammenhang zwischen den Variablen** besteht
- Trifft dies zu, ist der **Datensatz für eine Faktorenanalyse ungeeignet**, da er keine Variablen enthält, die sich auf gemeinsame Hintergrundvariablen zurückführen ließen

H_0: Die Variablen sind in der Grundgesamtheit unkorreliert
H_1: Die Variablen sind in der Grundgesamtheit korreliert

- Der Bartlett-Test setzt voraus, dass...
  - die **Variablen in der Grundgesamtheit normalverteilt** sind
    - Diese Voraussetzung kann grafisch oder rechnerisch überprüft werden (Histogramm, Q-Q-Diagramm, Kolmogoroff-Smirnov)
  - die **Prüfgröße näherungsweise einer Chi²-Verteilung folgt**
    - Dies bedeutet, dass der Wert der Prüfgröße auch von der Größe der Stichprobe abhängig ist (!)

KMO- und Bartlett-Test

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,793
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	2821,259
	df	10
	Signifikanz nach Bartlett	,000

- Im Beispiel kommt der sehr hohe Chi²-Wert von 2821 mit einer Wahrscheinlichkeit von 1 – 0 = 100% nur zustande, wenn mindestens zwei (nicht alle!) Variablen in der Grundgesamtheit korreliert sind

- Wichtig: Der Bartlett-Test erlaubt **keine Rückschlüsse auf die Signifikanz der einzelnen Korrelationen(!)**
- Ein hoher Chi²-Wert bedeutet daher nicht, dass **alle Koeffizienten in der Korrelationsmatrix** qualitativ zutreffend sind
- Um dies zu überprüfen, muss für **jeden einzelnen Koeffizienten** ein Signifikanztest durchgeführt werden (wie gesehen)

Anti-Image-Kovarianz-Matrix

- Dem Anti-Image liegt folgende Idee (nach Guttmann) zugrunde:
  - Wenn zwei Variablen miteinander korrelieren, lässt sich die **Varianz jeder der beiden Variablen** wenigstens **teilweise durch die andere Variable erklären**
  - Je stärker die Korrelation ist, desto größer ist der **Anteil an Varianz**, der durch die Korrelation erklärt werden kann
  - Solange die Korrelation nicht perfekt ist, wird es aber auch immer noch einen **unerklärbaren Varianzanteil** geben
  - Die Gesamtvarianz einer Variablen lässt sich also aufteilen in:
    - **Image** (durch die korrelierende Variable **erklärbarer Teil**)
    - **Anti-Image** (durch die korrelierende Variable **nicht erklärbarer Teil**)
- Ein Variablenpaar mit **niedrigem Anti-Image-Wert** korreliert daher stark miteinander

- Bei der Faktorenanalyse ist zu beachten, dass **mehr als zwei Variablen** betrachtet werden
  - Alle Variablen im Datensatz können theoretisch **wechselseitig miteinander korrelieren** (Korrelationsmatrix)
  - Daher muss nicht die einfache Korrelation sondern die **partielle Korrelation** betrachtet werden
  - Die partielle Korrelation ist die **Korrelation zwischen zwei Variablen bei Ausschaltung aller anderen Variablen**

- Das Anti-Image eines Variablenpaars bei der Faktorenanalyse ist daher
    - der Teil der Varianz einer Variablen
    - der sich nicht durch die korrelierende Variable erklären lässt
    - wenn der Einfluss aller übrigen Variablen ausgeschaltet wurde

Anti-Image-Kovarianz-Matrix

- Variablen sind nur dann für eine Faktorenanalyse geeignet, wenn die **Anti-Image-Werte möglichst gering ausfallen**

- Idealerweise ergibt sich für die Anti-Image-Kovarianz-Matrix eine **Diagonalmatrix**
  - In der Realität ist kaum mit dem Zustandekommen einer **perfekten Diagonalmatrix** zu rechnen
  - Es stellt sich daher die Frage, wenn das **Kriterium der Diagonalmatrix** näherungsweise erfüllt ist
  - Dziuban & Shirkey: Anteil an nicht-diagonalen Elementen ungleich Null sollte **unter 25%** liegen
  - Ungleich Null wird dabei als (> 0,09) definiert

Anti-Image-Matrizen

		Hubraum (cu. inches)	PS	Gewicht (lbs.)	Beschleunigung von 0 auf 100 km/h (sec.)	Anzahl der Zylinder
Anti-Image-Kovarianz	Hubraum (cu. inches)	4,892E-02	-2,07E-02	-2,935E-02	9,714E-03	-4,585E-02
	PS	-2,071E-02	,106	-4,270E-02	,127	1,315E-02
	Gewicht (lbs.)	-2,935E-02	-4,27E-02	9,613E-02	-8,870E-02	-1,065E-02
	Beschleunigung von 0 auf 100 km/h (sec.)	9,714E-03	,127	-8,870E-02	,362	1,441E-02
	Anzahl der Zylinder	-4,585E-02	1,315E-02	-1,065E-02	1,441E-02	9,172E-02
Anti-Image-Korrelation	Hubraum (cu. inches)	,797[a]	-,287	-,428	7,301E-02	-,684
	PS	-,287	,799[a]	-,422	,649	,133
	Gewicht (lbs.)	-,428	-,422	,813[a]	-,476	-,113
	Beschleunigung von 0 auf 100 km/h (sec.)	7,301E-02	,649	-,476	,662[a]	7,908E-02
	Anzahl der Zylinder	-,684	,133	-,113	7,908E-02	,842[a]

a. Maß der Stichprobeneignung

- Wichtig: In der Anti-Image-Kovarianz-Matrix werden **nicht die partiellen Korrelationskoeffizienten**, sondern deren **invertierte negative Werte** ausgewiesen (!)

Kaiser-Meyer-Olkin-Kriterium

HOCHSCHULE harz

- Kaiser, Meyer & Olkin entwickelten auf der **Basis der Anti-Image-Kovarianz-Matrix** eine Prüfgröße
- Diese Prüfgröße wird als **KMO-Kriterium** oder MSA (**measure of sampling adequacy**) bezeichnet

- Das KMO-Kriterium gibt an, in welchem Umfang die **Variablen in der Grundgesamtheit korrelieren**
- Es ist somit ein geeigneter Indikator dafür, ob eine Faktorenanalyse durchgeführt werden sollte oder nicht

- Der Wertebereich des KMO-Kriteriums liegt **stets zwischen 0 und 1**
- Es kann für **einzelne Variablenpaare** oder die **gesamte Korrelationsmatrix** berechnet werden
- Kaiser & Rice schlagen für die Interpretation der Werte die folgende Skala vor:

KMO >= 0,9	marvelous	(„erstaunlich“)
KMO >= 0,8	meritorious	(„verdienstvoll“)
KMO >= 0,7	middling	(„ziemlich gut“)
KMO >= 0,6	mediocre	(„mittelmäßig“)
KMO >= 0,5	miserable	(„kläglich“)
KMO < 0,5	unacceptable	(„untragbar“)

- Eine Korrelationsmatrix ist also dann für eine Faktorenanalyse ungeeignet, wenn (KMO < 0,5)
- Wünschenswert ist dagegen ein Wert von (KMO >= 0,8)

Kaiser-Meyer-Olkin-Kriterium

- Das KMO-Kriterium (für die gesmate Korrelationsmatrix) berechnet sich: $KMO = \frac{(\sum\sum r_{ij}^2)}{(\sum\sum r_{ij}^2 + \sum\sum a_{ij}^2)}$

- Dabei steht r_{ij} für den Korrelationskoeffizienten der Variablen i und j und a_{ij} für die entsprechende partielle Korrelation
- Da Korrelationen einer Variablen mit sich selbst (stets Eins) unberücksichtigt bleiben müssen, wird i = j ausgeschlossen

- Ein **KMO-Wert nahe Eins** wird erreicht, wenn die **partiellen Korrelationskoeffizienten klein** ausfallen
- Bei **großen partiellen Korrelationskoeffizienten** fällt der **KMO-Wert entsprechend geringer** aus

KMO- und Bartlett-Test

Maß der Stichprobeneignung nach Kaiser-Meyer-Olkin.		,793
Bartlett-Test auf Sphärizität	Ungefähres Chi-Quadrat	2821,259
	df	10
	Signifikanz nach Bartlett	,000

↔

KMO >= 0,8 meritorious
KMO >= 0,7 middling

- Im Beispiel liegt der KMO-Wert nahe den gewünschten (>= 0,8) und spricht somit nicht gegen eine Faktorenanalyse

- Zu empfehlen ist auch die **Berechnung der KMO-Werte für einzelne Variablen**: $KMO_{sv} = \frac{(\sum r_{ij}^2)}{(\sum r_{ij}^2 + \sum a_{ij}^2)}$
- Die Werte sind in der **Diagonalen der Anti-Image-Kovarianz-Matrix** abzulesen
- Ein mittelmäßiger KMO-Wert für die gesamte Korrelationsmatrix kann unter Umständen gesteigert werden, indem **Variablen mit niedrigem Einzel-KMO** von der **weiteren Analyse ausgeschlossen** werden

Ablauf einer Faktorenanalyse

Schritt 1 Auswahl der Variablen und Erstellung der Korrelationsmatrix	Für alle in die Faktorenanalyse einbezogenen Variablen wird die Korrelationsmatrix erstellt. Aus dieser kann abgelesen werden, welche Variablen für die weitere Analyse unberücksichtigt bleiben sollten, da sie mit den übrigen Variablen nur minimal korrelieren.
Schritt 2 Faktorextraktion	Dieser Schritt wird auch als „Ziehen“ von Faktoren bezeichnet. Anhand verschiedener statistischer Kennwerte kann entschieden werden, ob das gefundene Faktorenmodell geeignet ist, die vorliegenden Variablen auf Hintergrundvariablen zurückzuführen.
Schritt 3 Interpretation der Faktoren und Faktorrotation	Die im zweiten Schritt gefunden Faktoren sind in der Regel nur sehr schwer oder gar nicht zu interpretieren. Um ihre Auswertung zu erleichtern, werden sie einer Transformation unterzogen, die als Rotation bezeichnet wird.
Schritt 4 Bestimmung der Faktorwerte	Im letzten Schritt wird bestimmt, welche Werte die untersuchten Objekte hinsichtlich der extrahierten Faktoren annehmen.

Faktorextraktion: Fundamentaltheorem

- Nachdem die **Eignung der Variablen** geprüft wurde, werden im nächsten Schritt die **Faktoren extrahiert**

- Die **Grundannahme der Faktorenanalyse** ist dabei:
  - Jeder Wert einer Ausgangsvariablen lässt sich als **Linearkombination hypothetischer Faktoren** beschreiben
  - Dieser Zusammenhang lässt sich mathematisch formulieren: $X_{kj} = a_{j1} * p_{k1} + a_{j2} * p_{k2} + \ldots + a_{jQ} * p_{kQ}$
  - Standardisiert man die Werte lautet der mathematische Ausdruck: $Z_{kj} = a_{j1} * p_{k1} + a_{j2} * p_{k2} + \ldots + a_{jQ} * p_{kQ} = \sum_{q=1}^{Q} a_{jq} * p_{kq}$
  - Die **Faktorladung** zeigt dabei die **Stärke des Zusammenhangs** zwischen Faktor und Variablen
  - In Matrixschreibweise lässt sich der standardisierte Ausdruck darstellen als: **Z = P * A'**
    - Die standardisierte Datenmatrix Z wird hier als Linearkombination von Faktoren dargestellt

- Ausgehend von dieser Grundannahme lässt sich das **Fundamentaltheorem von Thurstone** herleiten: **R = A * C * A'**
- Für unabhängige Faktoren lässt sich vereinfachen: **R = A * A'**

- **Aussage des Fundamentaltheorems**:
  - **Korrelationsmatrix R** lässt sich durch **Faktorladungen** A und **Korrelation zwischen den Faktoren** C reproduzieren
  - Für **unabhängige (unkorrelierte) Faktoren** entspricht C einer **Einheitsmatrix**
  - Da die **Multiplikation mit einer Einheitsmatrix** zur **Ausgangsmatrix** führt, kann gekürzt werden
  - Wichtig: Gekürztes Fundamentaltheorem setzt **Linearverknüpfung und Unabhängigkeit der Faktoren** voraus (!)

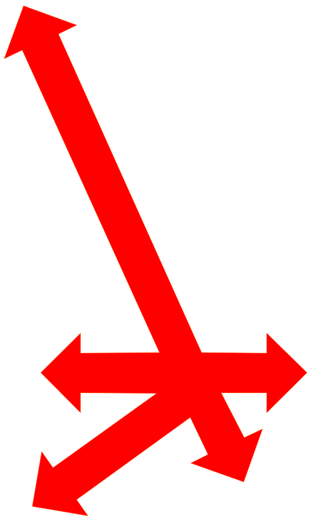
Faktorextraktion: Fundamentaltheorem

- Das (vereinfachte) Fundamentaltheorem lässt sich wie folgt herleiten:
  - Grundannahme der Faktorenanalyse: $X_{kj} = a_{j1} * p_{k1} + a_{j2} * p_{k2} + \ldots + a_{jQ} * p_{kQ}$
  - Grundannahme bei standardisierten x-Werten: $Z_{kj} = a_{j1} * p_{k1} + a_{j2} * p_{k2} + \ldots + a_{jQ} * p_{kQ} = \sum_{q=1}^{Q} a_{jq} * p_{kq}$
  - Auf Matrix-Schreibweise reduzierte Grundannahme: $Z = P * A'$
  - Die Korrelationsmatrix R lässt sich bei standardisierten Daten aus der Datenmatrix Z ermitteln: $R = \frac{1}{(K-1)} * Z' * Z$
  - Da Z im Rahmen der Faktorenanalyse durch P*A' beschrieben wird, lässt sich einsetzen: $R = \frac{1}{(K-1)} * (P * A')' * (P * A')$
  - Nach Auflösung der Klammern ergibt sich nach den Multiplikationsregeln für Matrizen: $R = A * (\frac{1}{(K-1)}) * P' * P * A'$
  - Da die Daten standardisiert sind lässt sich $\frac{1}{(K-1)} * P' * P$ auch als Korrelationsmatrix der Faktoren C bezeichnen
  - Daraus ergibt sich das Fundamentaltheorem nach Thurstone: $R = A * C * A'$
  - Werden die Faktoren als unkorreliert angenommen, so entspricht C einer Einheitsmatrix
  - Da die Multiplikation einer Matrix mit einer Einheitsmatrix die Ausgangsmatrix ergibt, kann gekürzt werden: $R = A * A'$
  - Das gekürzte Fundamentaltheorem gilt nur bei Linearverknüpfung und Unabhängigkeit der Faktoren (!)

$$R = A * A'$$

Faktorextraktion: Grafische Interpretation

- Der **Informationsgehalt einer Korrelationsmatrix** lässt sich **grafisch im Vektor-Diagramm** darstellen
- Zwei Vektoren sind **linear unabhängig**, wenn sie **senkrecht (orthogonal)** zueinander stehen
- Sind die Vektoren (= Variablen) korreliert, wird dies grafisch durch einen **Winkel** dargestellt

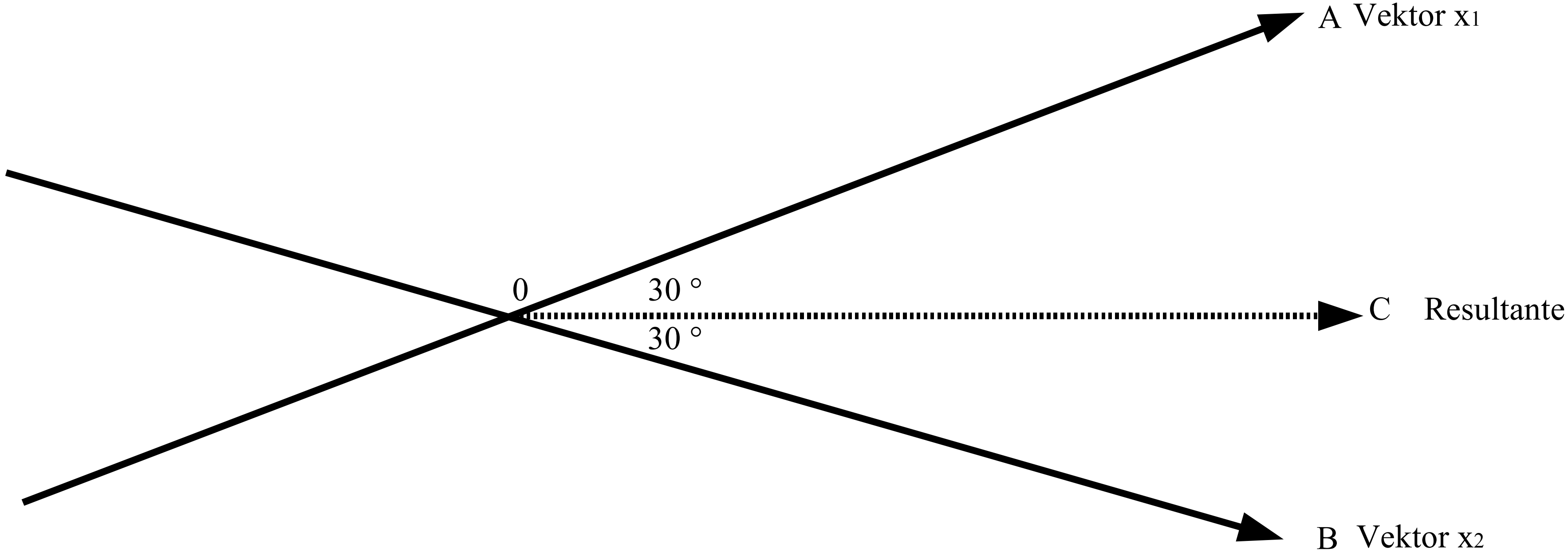


- Beispiel: Eine Korrelation von r = 0,5 würde sich in einem Winkel von 60° ausdrücken
- Wie berechnet sich der Winkel?
  - Der Korrelationskoeffizient wird durch den Cosinus ausgedrückt
  - Der Cosinus eines 60°-Winkels beträgt genau 0,5
  - Der Cosinus eines 90°-Winkels beträgt genau 0,0
    - Aus diesem Grund stehen die Vektoren für unabhängige Variablen senkrecht zueinander
  - Der Cosinus eines 0°-Winkels beträgt genau 1,0
    - Aus diesem Grund liegen die Vektoren für vollkommen abhängige Variablen übereinander

Grad	**cos**	**Grad**	**cos**
0	1,00	10	0,98
20	0,94	30	0,87
40	0,77	50	0,64
60	0,50	70	0,34
80	0,17	90	0,00

Faktorextraktion: Grafische Interpretation

- Je **mehr Variablen im Datensatz** sind, desto **mehr Dimensionen** werden für die grafische Darstellung benötigt
- Das **Ziel der Faktorenanalyse** ist es, das durch die Korrelationskoeffizienten gemessene Verhältnis der Variablen zueinander in einem **Raum mit möglichst wenig Dimensionen** darzustellen
- Die **Zahl der benötigten Achsen** dieser Darstellung entspricht der **Zahl der gefundenen Faktoren**

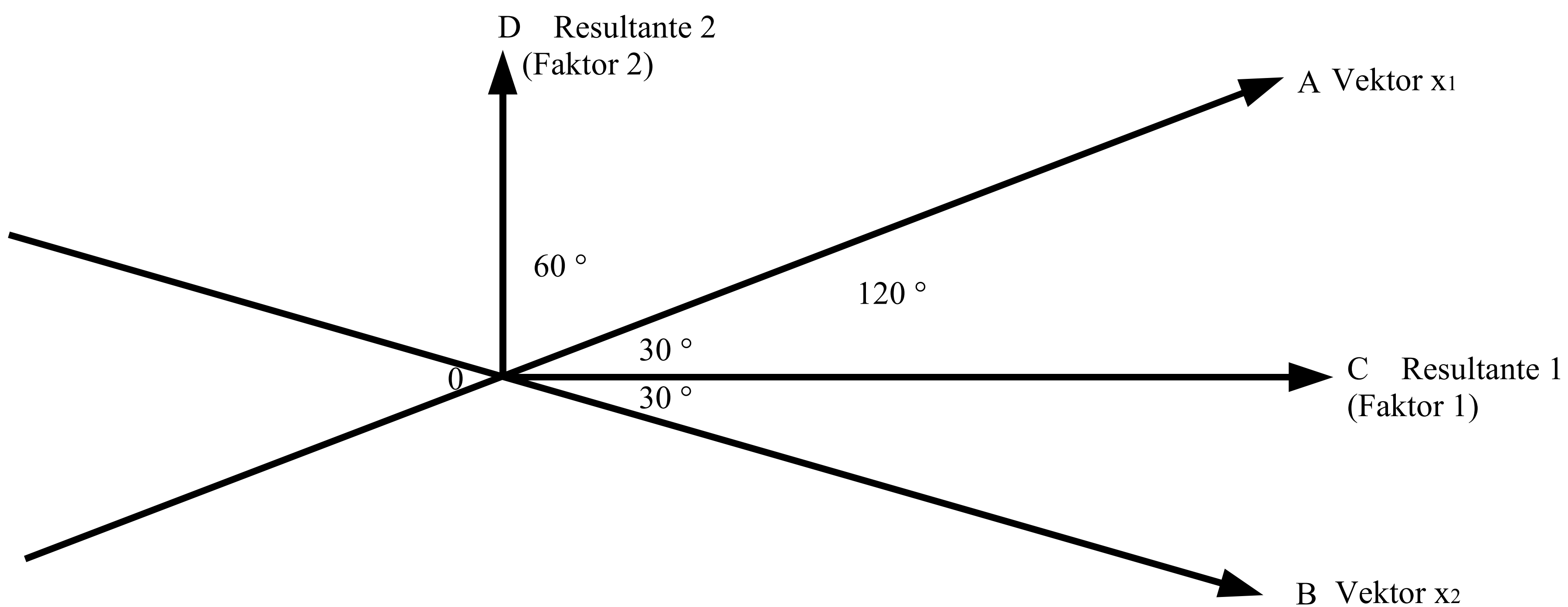


- Beispiel: Zwei Variablen mit einer Korrelation von $r = 0{,}5$ stehen als Vektoren (0A & 0B) im 60°-Winkel zueinander
- Verktor 0C (Resultante) ist eine zusammenfassende (faktorielle) Beschreibung der beiden anderen Vektoren
- Die beiden neu entstehenden 30°-Winkel zeigen den Zusammenhang zwischen dem Faktor und den beiden Variablen
- Sie repräsentieren ebenfalls Korrelationskoeffizienten – die zwischen den jeweiligen Variablen und dem Faktor
- Diese Korrelationskoeffizienten werden als Faktorladungen bezeichnet ($\cos 30° = 0{,}87$)

Faktorextraktion: Extraktionsproblem

- Wie lassen sich Vektoren (Faktoren) finden, die zusammenfassend für die Vektoren (Variablen) stehen?
  - **Erster Faktor** ist der **Schwerpunkt** aller aus den Variablen gebildeten Vektoren
  - Der **zweite Faktor** muss **rechtwinklig zum ersten Faktor** stehen (**Unkorreliertheit der Faktoren**) etc.
  - Anzahl der **benötigten Dimensionen** = Anzahl der **benötigten Achsen** = Anzahl der **extrahierten Faktoren**

- Erklären die extrahierten Faktoren die Variablen restlos, ist die **Summe der Ladungsquadrate** jeder Variablen gleich Eins



Bestimmung der Kommunalitäten

- Ziel der Faktorenanalyse: **Weniger Faktoren extrahieren als Variablen vorhanden sind**
  - Beispiel: 8 Variablen laden auf 3 Faktoren hoch
  - Problem: In der Praxis wird meist **nicht die gesamte Varianz durch die Faktoren erklärbar** sein
  - Es verbleibt eine **Restvarianz**, die durch **andere Faktoren oder Meßfehler** bedingt wird
  - Je mehr Faktoren extrahiert werden, desto mehr Varianz wird insgesamt durch die Faktoren erklärt
  - Der Teil der Gesamtvarianz, der durch die extrahierten Faktoren erklärt wird, wird als **Kommunalität** bezeichnet

- Das (gekürzte) Fundamentaltheorem ist daher um eine Unbekannten-Komponente zu erweitern: $\mathbf{R = A * A' + U}$
- In den Term U fließen die **spezifische Varianz** sowie die **potentiellen Meßfehler** ein (**Einzelrestfaktoren**)

- Problem: Die Kommunalität muss durch den Anwender geschätzt werden > Entscheidung des Anwenders
- Kommunalität = 0,7 > man vermutet, dass 70% der Ausgangsvarianz durch gemeinsame Faktoren erklärt werden können

- Zusammenhang zwischen **Variablenzahl** und korrekter **Einschätzung der Kommunalität**:
  - Je größer die Zahl der Variablen ist, desto unwichtiger ist die **exakte Schätzung der Kommunalitäten**
  - Bei **steigender Anzahl der Variablen** nimmt der **prozentuale Anteil der diagonalen Matrixelemente** ab
  - In einer 2x2-Matrix bilden die diagonalen Elemente 50% der Matrix, bei einer 100x100-Matrix noch 1%
  - Eine fehlerhafte Einschätzung im Fall der 100 Elemente hat daher **erheblich geringere Auswirkungen**

Bestimmung der Kommunalitäten

- In der Praxis haben vor allem **zwei Verfahren der Kommunalitätenschätzung** eine Bedeutung

- Möglichkeit 1: Man geht davon aus, dass die **gesamte Varianz der Ausgangsvariablen** durch die Faktoren erklärt wird
  - In diesem Fall ist die **Summe der Kommunalitäten gleich Eins**, da **keine Einzelrestfaktoren** auftreten
  - Eine **explizite Kommunalitätenschätzung** durch die Faktorenanalyse findet in diesem Fall nicht statt

- Möglichkeit 2: Aufgrund inhaltlicher Überlegungen wird ein **Schätzwert für die Kommunalitäten** vorgegeben
  - **Vorgabewert** ist häufig der **höchste quadrierte Korrelationskoeffizient** aus der Korrelationsmatrix
  - Die Faktoren zusammen liefern **mindestens den gleichen Erklärungsbeitrag** wie die höchste Korrelation
  - Wird dieser Vorgabewert zur Schätzung verwendet, fällt diese daher **in der Regel zu niedrig** aus

- Die Kommunalitätenschätzung wirkt sich unmittelbar auf die **Wahl des Faktorextraktionsverfahrens** aus
- Auch hier haben vor allem **zwei iterative Verfahren der Faktorextraktion** in der Praxis eine Bedeutung:
  - **Die Hauptachsenanalyse** > die Varianz spaltet sich immer in **Kommunalitäten und Einzelrestvarianz** auf
    - Der **Startwert der Kommunalitäten** in der Hauptachsenanalyse ist ein **Schätzwert von unter Eins**
  - **Die Hauptkomponentenanalyse** > die Varianz kann hier **vollständig durch die Faktoren erklärt** werden
    - Der **Startwert der Kommunalitäten in der Hauptkomponentenanalyse ist Eins** (bei Faktoren = Variablen)
- **Die Wahl des Faktorextraktionsverfahrens ist wichtig, da die Ergebnisse unterschiedlich interpretiert werden!**

Die Hauptachsenanalyse

- Annahme: Die Varianz einer Ausgangsvariablen **spaltet sich immer in Kommunalitäten und Einzelrestvarianz auf**

- Der Anwender hat während des laufenden Verfahrens **Eingriffsmöglichkeiten**:
  - Entweder er kommt **aufgrund inhaltlicher Überlegungen** zu einer **Einschätzung der wahren Kommunalität**
  - Oder er führt die **Schätzung der Kommunalität** im **Iterationsprozeß der Hauptachsenanalyse** durch

- Ziel: **Erklärung der Varianzen der Variablen in Höhe der Kommunalitäten durch die Faktoren**
- Aus diesem Grund ist die Hauptachsenanalyse das richtige **Verfahren zur inhaltlichen Interpretation**

- Die **Interpretation der Faktoren** aus der Hauptachsenanalyse erfolgt **kausal!**

„Wie lässt sich die <u>Ursache</u> bezeichnen, die für hohe Ladungen der Variablen auf diesen Faktor verantwortlich ist?“

Die Hauptkomponentenanalyse

- Annahme: Die Varianz einer Ausgangsvariablen **kann vollständig durch die Extraktion von Faktoren erklärt werden**

- Werden genauso **viele Faktoren extrahiert wie es Variablen gibt**, ist die **Kommunalität bei Eins**
  - Werden dagegen **weniger Faktoren extrahiert**, **sinkt auch die Kommunalität**
  - Weniger Faktoren bedeutet daher einen **bewußten Informationsverzicht** für ein **verwendbareres Modell**

- Ziel: **möglichst umfassende Reproduktion der Datenstruktur mit möglichst wenig Faktoren**
- Aus diesem Grund gibt es **keine Unterscheidung zwischen Kommunalitäten und Einzelrestvarianz**

- Die **Interpretation der Faktoren** aus der Hauptkomponentenanalyse erfolgt **nicht kausal!**

„Wie lassen sich die auf einen Faktor hochladenden Variablen durch einen <u>Sammelbegriff</u> (Komponente) zusammenfassen?“

Die Hauptkomponentenanalyse

- **Vorgehensweise** bei der Hauptkomponentenanalyse:
  - Bestimmung des ersten Faktors (**erste Hauptkomponente**) so, dass ein **möglichst großer Teil der Varianz** erklärt wird
  - Ermittlung des zweiten Faktors (**zweite Hauptkomponente**) so, dass er **orthogonal zum ersten Faktor** steht (also unkorreliert ist) und einen möglichst großen Teil der Restvarianz erklärt
  - Auf diese Weise lassen sich so lange weitere Faktoren ermitteln, wie beobachtete Variablen im Modell sind

- Dabei ist folgender Zusammenhang zu beachten:
  - Werden **n Variablen durch n Faktoren erklärt**, wird die **Varianz komplett aufgeklärt** werden
  - Ein solches Modell liefe jedoch dem **Ziel der Dimensionsreduktion** zuwider
  - Werden bei n Variablen **weniger als n Faktoren** gebildet, wird ein **Teil der Varianz unaufgeklärt** bleiben
  - Es gibt daher einen **Tradeoff zwischen dem Grad der Dimensionsreduktion und der Genauigekeit des Modells**

- Der Anwender muss daher entscheiden, **welche Faktoren aus dem Modell ausgeschlossen** werden sollen
- Dabei ist es logisch, Faktoren mit geringem Erklärungsgehalt eher auszuschließen als Faktoren mit hohem

Welche Faktoren sollten ins Modell aufgenommen werden?
Wie viele Faktoren sollte das finale Modell enthalten?

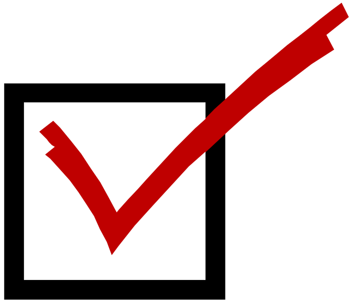
>>> Problem der Bestimmung der Faktoranzahl

Bestimmung der Faktoranzahl

- Zur Bestimmung der Faktoranzahl existieren **keine allgemeinverbindlichen Vorschriften**
- Gefragt ist die **subjektive Entscheidung des Anwenders**

- Die **maximal mögliche Anzahl der Faktoren** entspricht der **Anzahl der Variablen**
- Das Ziel ist es stets, eine **geringere Anzahl an Faktoren** im Modell zu belassen
- Die Entscheidung über die Anzahl der extrahierten Faktoren **liegt beim Anwender**

- Es bieten sich **sechs verschiedene Entscheidungskriterien** an:
  - **Extraktion bis x% der Varianz erklärt sind** (vorher festzulegen!)
  - **Kaiser-Kriterium** (Standardkriterium in SPSS)
  - **Screeplot / Scree-Test**
  - **Extraktion von n Faktoren** (Anzahl zuvor inhaltlich festgelegt)
  - **Zahl der Faktoren < Hälfte der Variablen**
  - **Alle interpretierbaren Faktoren extrahieren** (inhaltliche Entscheidung)



Welche Faktoren sollten ins Modell aufgenommen werden?
Wie viele Faktoren sollte das finale Modell enthalten?

>>> Problem der Bestimmung der Faktoranzahl

Faktoranzahl: Kaiser-Kriterium

- Kaiser-Kriterium: **Alle Faktoren mit Eigenwerten oberhalb von Eins werden extrahiert**

- Eigenwerte sind die **Summe aller quadrierten Faktorladungen eines Faktors über alle Variablen**
- Sie können als Richtmaß für die durch den jeweiligen Faktor erklärte Varianz aller Beobachtungswerte betrachtet werden

- Um die Einschätzung der Eigenwerte zu erleichtern, werden **alle Variablen einer Z-Transformation unterzogen**
  - Jede Variable weisst hinterher einen **Mittelwert von Null** und eine **Standardabweichung von Eins** auf
  - Die Gesamtstreuung der n Variablen beträgt daher ebenfalls n (Interpretation des Eigenwertes!)

- Jeder Faktor dessen Varianzerklärungsanteil oberhalb von Eins liegt, wird nach dem Kaiser-Kriterium extrahiert
  - Ein Faktor mit einem **Eigenwert von mehr als Eins erklärt mehr als eine einzelne Variable**
  - Ein Faktor mit einem **Eigenwert von weniger als Eins erklärt weniger als eine einzelne Variable**

Erklärte Gesamtvarianz

Komponente	Anfängliche Eigenwerte			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	1,741	43,534	43,534	1,642	41,057	41,057
2	1,131	28,283	71,818	1,230	30,760	71,818
3	,580	14,489	86,306			
4	,548	13,694	100,000			

Extraktionsmethode: Hauptkomponentenanalyse.

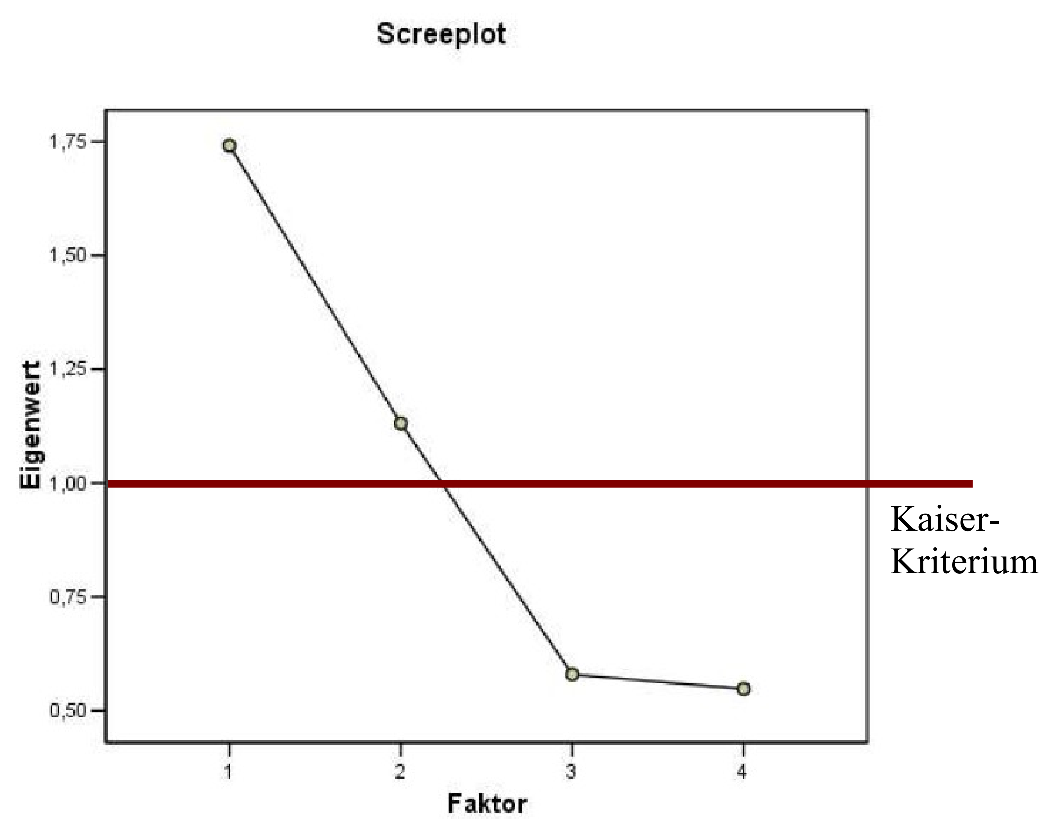
Faktoranzahl: Screeplot / Scree-Test

- Beim Scree-Test werden die **Eigenwerte der Faktoren im Screeplot abnehmend angeordnet**
- Bei **verbundenen Punkten** ergibt sich eine sich **asymptotisch der Abszisse annähernde Punktlinie**

- An der Stelle mit der **größten Differenz zwischen zwei Eigenwerten** ergibt sich ein **Knick** (elbow)
- Der **letzte Punkt links dieses Knicks** bestimmt die **Anzahl der zu extrahierenden Faktoren**
- Grund dafür ist, dass Faktoren mit kleinen Eigenwerten **wenig zur Erklärung beitragen**
- Sie werden daher als „**Geröll**“ (scree) betrachtet

- Das Verfahren liefert **nicht immer eindeutige Lösungen**, z.B. Bei ähnlichen Differenzen
- Eine **subjektive Entscheidung des Anwenders** ist beim Scree-Test nicht zu umgehen

- Der Screeplot kann auch zur **visuellen Unterstützung** einer **Entscheidung nach dem Kaiser-Kriterium** verwendet werden



Ablauf einer Faktorenanalyse

HOCHSCHULE harz

Schritt 1
Auswahl der Variablen und Erstellung der Korrelationsmatrix

Für alle in die Faktorenanalyse einbezogenen Variablen wird die Korrelationsmatrix erstellt. Aus dieser kann abgelesen werden, welche Variablen für die weitere Analyse unberücksichtigt bleiben sollten, da sie mit den übrigen Variablen nur minimal korrelieren.

Schritt 2
Faktorextraktion

Dieser Schritt wird auch als „Ziehen“ von Faktoren bezeichnet. Anhand verschiedener statistischer Kennwerte kann entschieden werden, ob das gefundene Faktorenmodell geeignet ist, die vorliegenden Variablen auf Hintergrundvariablen zurückzuführen.

Schritt 3
Interpretation der Faktoren und Faktorrotation

Die im zweiten Schritt gefunden Faktoren sind in der Regel nur sehr schwer oder gar nicht zu interpretieren. Um ihre Auswertung zu erleichtern, werden sie einer Transformation unterzogen, die als Rotation bezeichnet wird.

Schritt 4
Bestimmung der Faktorwerte

Im letzten Schritt wird bestimmt, welche Werte die untersuchten Objekte hinsichtlich der extrahierten Faktoren annehmen.

Interpretation der Faktorladungen

- Im Anschluss an die Extraktion der Faktoren sind diese noch **entsprechend zu interpretieren**
- Erster Schritt der Faktorinterpretation ist die **Analyse der (unrotierten) Faktorladungen**

- Bei der **Analyse der Faktorladungen** ist vor allem zu beachten
  - Die **inhaltliche Interpretation** von Faktoren erfordert **ausgeprägte Sachkenntnis** bezüglich des Untersuchungsfeldes
    - **Gute Methodenkenntnisse allein befähigen nicht** im ausreichenden Maße zur Sachinterpretation
  - Die **Unterschiede zwischen Hauptachsen-und Hauptkomponentenanalyse** sind zu bedenken:
    - Die **Interpretation der Faktoren** aus der **Hauptachsenanalyse** erfolgt **kausal!**
      - Welche Ursache ist für hohe Faktorladungen einer Variablen verantwortlich?
    - **Die Interpretation der Faktoren** aus der **Hauptkomponentenanalyse** erfolgt **nicht kausal!**
      - Welche Oberbegriffe lassen sich für die Faktoren ausmachen?

- Beispiel: IQ und Englisch werden hauptsächlich durch den ersten Faktor erklärt, Mathematik dagegen durch den zweiten
- Die Variable Sprache dagegen lädt etwa gleich gut auf die beiden extahierten Faktoren
- Sie wäre daher beiden Faktoren zuzuordnen, was wiederum zu Problemen bei der Interpretation der Faktoren führt

Komponentenmatrix[a]

	Komponente	
	1	2
iq	,793	-,104
sprache	,584	-,650
mathe	,396	,817
englisch	,783	,177

Extraktionsmethode: Hauptkomponentenanalyse.
a. 2 Komponenten extrahiert

Rotation: Einführung

- Der Anwender hat bei der Interpretation der Faktorladungen teils **erheblichen subjektiven Spielraum**
  - Entscheidung, **wie stark eine Variable auf einen Faktor laden muss**, um diesem **zugeordnet zu werden**
  - Grundsätzliche Regel: Bei **Ladungshöhen ab 0,5** findet eine Zuordnung statt
  - Aber: Bei **Ladungen einer Variablen von über 0,5 auf mehrere Faktoren** gehört sie zu **jedem der Faktoren**
  - In solchen Fällen ist **keine sinnvolle Interpretation der Faktoren mehr möglich**

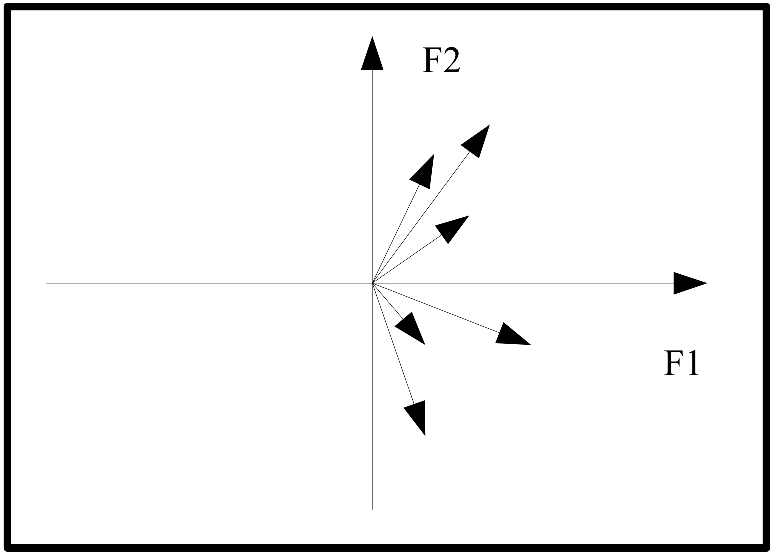
- Theoretischer Ansatz hinter den **Rotationsverfahren**:
  - Vorüberlegung der Faktorenanalyse: Beobachtete **Variablen sind Ausdruck komplexer Hintergrundvariablen**
  - Die **Beziehung der Variablen zu diesen Hintergrundvariablen** (Faktoren) **zeigt sich an den Faktorladungen**
  - Große Faktorladungen zeigen große, kleine Faktorladungen geringe Bedeutung des Faktors für die Variable
  - Ein Faktor ist dann **leicht zu interpretieren**, wenn die auf ihn ladenden Variablen **untereinander homogen sind**
  - Die **Interpretation ist dagegen schwer**, wenn ein Faktor mit **sehr vielen oder allen Variablen stark korreliert**
  - Nach der Faktorextraktion ist eine solche Situation jedoch **in der Praxis nicht unwahrscheinlich**
  - Aus diesem Grund werden die Faktoren einer als **Rotation bezeichneten Transformation** unterworfen

- Die **Signifikanz der Ladungsgrößen** ist auch **vom Stichprobenumfang abhängig**

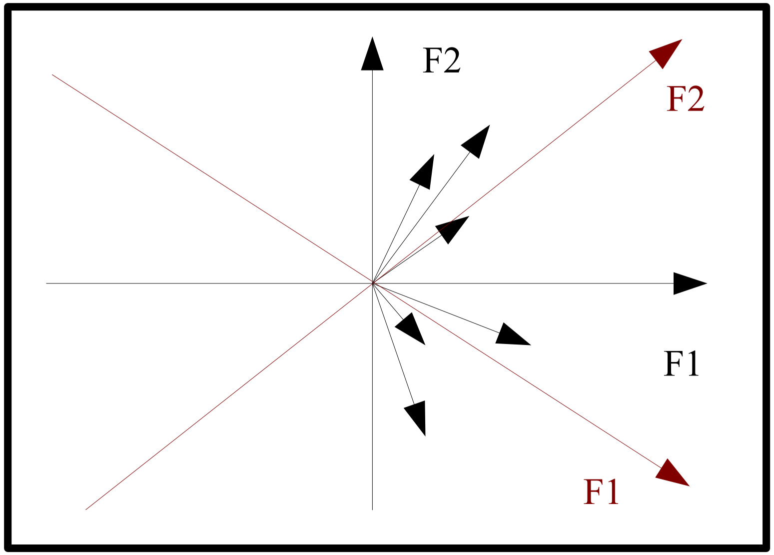
Rotation: Ablauf

- Grundgedanke der Rotation: **Die Faktorladungen lassen sich in einem Koordinatensystem darstellen**
- **Rotiert man die Achsen** dieses Koordinatensystems, lassen sich die **Faktorladungen besser auf die Faktoren verteilen**

- Es ist in **zwei Methoden der Rotation** zu unterscheiden:
  - **Orthogonale (rechtwinklige) Rotation**: Varimax, Quartimax und Equimax
  - **Oblique (schiefwinklige) Rotation**: Direktes Oblimin und Promax



Unrotierte Faktorladungen



Rotierte Faktorladungen

Rotation: Drehung des Koordinatenkreuzes in seinem Ursprung zur Erleichterun der Interpretation

Rotation: Orthogonal vs. oblique

- **Orthogonale (rechtwinklige) Rotation:**
  - Annahme: **Faktoren korrelieren nicht untereinander** (sind voneinander unabhängig)
  - Voneinander **unabhängige Faktoren stehen im rechten Winkel** (orthogonal) zueinander
  - Die Faktorachsen bleiben daher während der Rotation ebenfalls im rechten Winkel zueinander

- **Oblique (schiefwinklige) Rotation:**
  - Annahme: **Faktoren korrelieren untereinander** (sind nicht voneinander unabhängig)
  - Voneinander **abhängige Faktoren stehen in beliebigen Winkeln** (nicht-orthogonal) zueinander
  - Die Faktorachsen bleiben daher während der Rotation nicht im rechten Winkel zueinander
  - Vorteil: **Wesentlich bessere Aufteilung der Faktorladungen** auf die Faktoren möglich
  - Nachteil: **Faktoren nicht mehr unabhängig voneinander** > Ziel der Faktoranalyse verfehlt

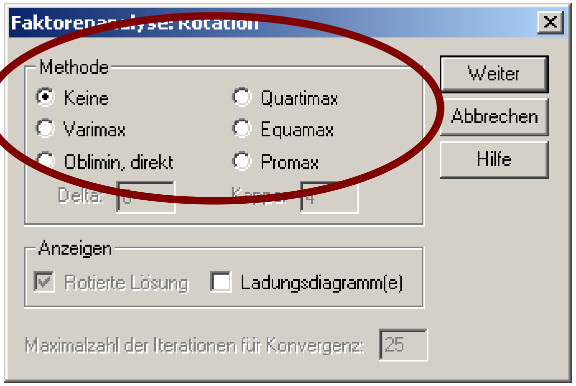
- Ergebnis der Rotation: **Verbesserte Zuordnung der einzelnen Variablen zu den Faktoren**
- Rotation **verändert die Faktorladungen und Eigenwerte**, **nicht aber die Kommunalitäten** des Modells
- Die **Aussagekraft einer Hauptachsenanalyse** wird durch die Rotation des Koordinatenkreuzes **nicht verändert!**
- Alle Formen der Rotation sind daher **nicht als Änderungen** sondern als **Nachoptimierungen** zu verstehen

- In der Praxis wird zumeist **nur noch die rotierte Faktorladungsmatrix** inhaltlich interpretiert

Rotation: Methoden

- **Orthogonale (rechtwinklige) Rotation:**
  - Varimax-Methode: **Einfache Interpretation der Faktoren**
    - Rotation der Achsen zur Reduktion Variablen mit hoher Faktorladung reduziert wird
  - Quartimax-Methode: **Einfache Interpretation der Variablen**
    - Rotation der Achsen zur Erklärung einer Variablen mit möglichst wenig Faktoren
  - Equimax-Methode: **Mischform aus Varimax und Quartimax**
  - Die **gebräuchlichste orthogonale Rotations-Methode ist Varimax**

- **Oblique (schiefwinklige) Rotation:**
  - Direktes Oblimin: **Grad der Schiefwinkligkeit kann vorgegeben werden** (inhaltliche Überlegungen)
  - Promax: **Grad der Schiefwinkligkeit wird durch das Iterationsverfahren bestimmt**
  - Die **gebräuchlichste oblique Rotations-Methode ist das direkte Oblimin**
  - Promax findet in der Regel **nur bei sehr umfangreichen Stichproben** Anwendung



Ablauf einer Faktorenanalyse

Schritt 1
Auswahl der Variablen und Erstellung der Korrelationsmatrix

Für alle in die Faktorenanalyse einbezogenen Variablen wird die Korrelationsmatrix erstellt. Aus dieser kann abgelesen werden, welche Variablen für die weitere Analyse unberücksichtigt bleiben sollten, da sie mit den übrigen Variablen nur minimal korrelieren.

Schritt 2
Faktorextraktion

Dieser Schritt wird auch als „Ziehen“ von Faktoren bezeichnet. Anhand verschiedener statistischer Kennwerte kann entschieden werden, ob das gefundene Faktorenmodell geeignet ist, die vorliegenden Variablen auf Hintergrundvariablen zurückzuführen.

Schritt 3
Interpretation der Faktoren und Faktorrotation

Die im zweiten Schritt gefunden Faktoren sind in der Regel nur sehr schwer oder gar nicht zu interpretieren. Um ihre Auswertung zu erleichtern, werden sie einer Transformation unterzogen, die als Rotation bezeichnet wird.

Schritt 4
Bestimmung der Faktorwerte

Im letzten Schritt wird bestimmt, welche Werte die untersuchten Objekte hinsichtlich der extrahierten Faktoren annehmen.

Bestimmung der Faktorwerte

- Nach der Extraktion stellt sich die Frage, **welche Werte die untersuchten Objekte bezüglich der Faktoren annehmen**
- Die Analyse dieser Fragestellung wird auch als **Problem der Bestimmung der Faktorwerte bezeichnet**

- Typisches Fallbeispiel: Befragung von Personen zu verschiedenen Eigenschaften von Automobilen
  - Die Faktorenanalyse zeigt, dass sich Eigenschaften durch gemeinsame Faktoren erklären lassen
  - So gehören PS-Zahl, Drehmoment und Höchstgeschwindigkeit zum Faktor „Technik“ etc.
  - Interessant ist nun die Frage, wie die verschiedenen Marken bezüglich der Faktoren beurteilt wurden

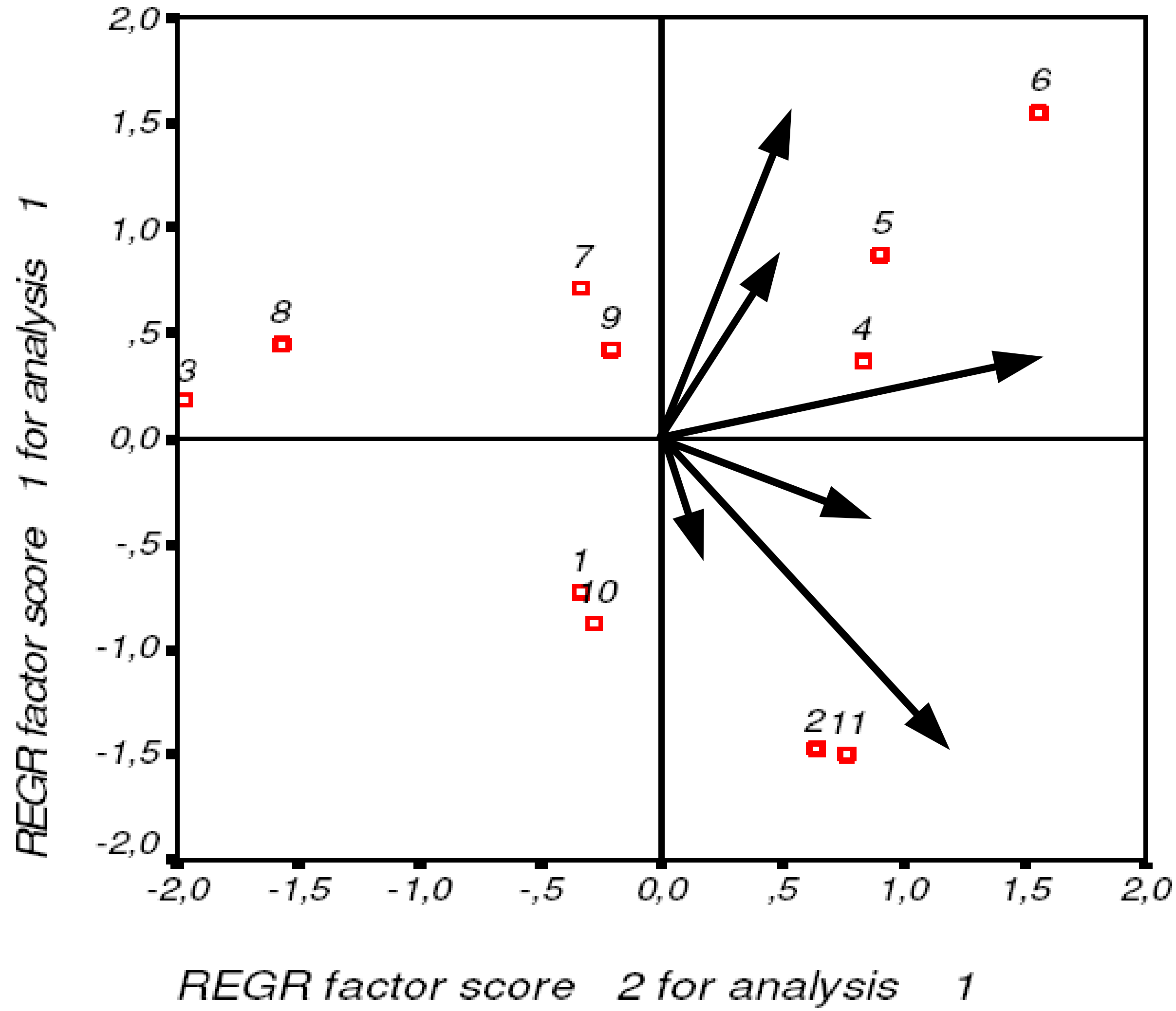
- **Zielfunktion der Faktorenanalyse**: $Z = P * A`$
  - Zur Bestimmung der Faktorwerte ist diese Gleichung nach P aufzulösen
  - Multiplikation von rechts mit der inversen Matrix: $Z * (A')^{-1} = P * A' * (A')^{-1}$
  - Da $A' * (A')^{-1}$ definitionsgemäß die Einheitsmatrix E ergibt, folgt: $Z * (A')^{-1} = P * E$
  - Da $P * E = P$ ist, ergibt sich weiterhin $P = Z * (A')^{-1}$
- Für das meist **nicht-quadratische Faktormuster A** ist die einfache Inversion nicht möglich
  - In diesem Fall wird von rechts mit A multipliziert: $Z * A = P * A' * A$
  - Die Matrix $(A' * A)$ ist quadratisch und daher invertierbar: $Z * A * (A' * A)^{-1} = P * (A' * A) * (A' * A)^{-1}$
  - Da $(A' * A) * (A' * A)^{-1}$ per Definition der Einheitsmatrix entspricht, ergibt sich: $P = Z * A * (A' * A)^{-1}$
- Zur **Lösung dieser Gleichung** sind ggf. **Schätzverfahren** anzuwenden (Regression, Bartlett, Anderson-Ruth...)

Interpretation der Faktorwerte

- Faktorwerte können negativ oder positiv ausfallen bzw. (näherungsweise) bei Null liegen
- Sie werden unter Verwendung aller Faktorladungen aus der rotierten Faktorladungsmatrix berechnet
- Auch kleine Faktorladungen haben daher Einfluss auf die Größe der Faktorwerte

- **Negativer Faktorwert** = Objekt ist bezüglich des betrachteten Faktors und im Vergleich mit den anderen betrachteten Objekten **unterdurchschnittlich ausgeprägt**
- **Positiver Faktorwert** = Objekt ist bezüglich des betrachteten Faktors und im Vergleich mit den anderen betrachteten Objekten **überdurchschnittlich ausgeprägt**
- **Faktorwert nahe Null** = Objekt hat bezüglich des betrachteten Faktors und im Vergleich mit den anderen betrachteten Objekten eine **durchschnittliche Ausprägung**

Grafische Darstellung der Faktorwerte



Rechenschritte der Faktorenanalyse

HOCHSCHULE harz

X enthält die Ausprägungen der Personen / Objekte i.b.a. die abgefragten Variablen

In den Spalten stehen die Merkmale, in den Zeilen die Objekte

Ausgangs-daten-matrix X

Z enthält die standardisier-ten Ausprä-gungen der Personen / Objekte i.b.a. die abgefragten Variablen

In den Spalten stehen die Merkmale, in den Zeilen die Objekte

Standar-disierte Daten-matrix Z

R beschreibt die statistischen Zusammenhänge zwischen den Variablen

Die Matrix ist quadratisch, die Zahl der Zeilen und Spalten wird durch die Zahl der Merkmale in Z bestimmt

Korre-lations-matrix R

Kommunalitäten-problem

R enthält nun in der Hauptdiagonalen die geschätzten Kommunalitäten

Die Matrix ist quadratisch, die Zahl der Zeilen und Spalten wird durch die Zahl der Merkmale in Z bestimmt

Reduzierte Korrelations-matrix X

Extraktions-problem

A enthält die Korrelationen zwischen Variablen und Faktoren

Die Matrix ist in der Regel nicht quadratisch, da die Zahl der Faktoren (Spalten) kleiner sein soll als die der Merkmale (Zeilen)

Faktor-ladungs-matrix = Faktoren-struktur A

Rotations-problem

A* enthält die Korrelationen zwischen Variablen und Faktoren nach Drehung des Ko-ordinatenkreuzes

Die Matrix ist in der Regel nicht quadratisch, da die Zahl der Faktoren (Spalten) kleiner sein soll als die der Merkmale (Zeilen)

Rotierte Faktoren-struktur A*

Schätzung der Faktorenwerte

P enthält nicht mehr die Aus-prägungen der einzelnen Personen / Objekte i.b.a. die Ausgangsvariablen, sondern i.b.a. die ermittelten Faktoren

Die Matrix ist in der Regel nicht quadratisch. Sie enthält in den Zeilen die Objekte und in den Spalten die Faktoren

Faktor-werte-matrix P

Aus: Backhaus, Erichson, Plinke & Weiber

Gibt es noch Fragen?