

X, Y Latente kategoriale Variablen

C, D, E, F Manifeste kategoriale Variablen

*** INPUT ***

res 2

lat 2

man 4

dim 2 2 2 2 2 2 2 2

lab R S X Y C D E F

sub CDEF CD EF -

mod X Y | X {XY}

C | X D | X

E | Y F | Y

R S | R

dat wr.dat

cas 336

*** LATENT CLASS OUTPUT ***

			Nein	X	1	X	1	X	2	X	2	Ja
			Nein	Y	1	Y	2	Y	1	Y	2	Ja
					0.4909		0.1491		0.0056		0.3544	
Ja	C	1			0.0851		0.0851		0.9563		0.9563	
Nein	C	2			0.9149		0.9149		0.0437		0.0437	
Ja	D	1			0.1988		0.1988		0.8669		0.8669	
Nein	D	2			0.8012		0.8012		0.1331		0.1331	
Ja	E	1			0.1520		0.8904		0.1520		0.8904	
Nein	E	2			0.8480		0.1096		0.8480		0.1096	
Ja	F	1			0.2924		0.9248		0.2924		0.9248	
Nein	F	2			0.7076		0.0752		0.7076		0.0752	

LCA Tabelle (Beispiel berechnet mit Latent Gold)

DFactor	Level	Size	DFactor1		DFactor2	
			Level1	Level2	Level1	Level2
Indicators			0,6532	0,3468	0,6165	0,3835
t1_money						
0			0,9135	0,0265	0,4397	0,8731
1			0,0865	0,9735	0,5603	0,1269
Mean			0,0865	0,9735	0,5603	0,1269
t1_time						
0			0,7934	0,1249	0,4009	0,8198
1			0,2066	0,8751	0,5991	0,1802
Mean			0,2066	0,8751	0,5991	0,1802
t2_money						
0			0,6598	0,1350	0,2813	0,7937
1			0,3402	0,8650	0,7187	0,2063
Mean			0,3402	0,8650	0,7187	0,2063
t2_time						
0			0,5290	0,1115	0,0584	0,9080
1			0,4710	0,8885	0,9416	0,0920
Mean			0,4710	0,8885	0,9416	0,0920

		Y		
		1	2	
X	1	0,4909	0,1491	0,64
	2	0,0056	0,3544	0,36
		$P(XY)$		$P(X)$

		Y		
		1	2	
X	1	0,767	0,233	1,0
	2	0,016	0,984	1,0
		$P(Y X)$		

$X = 1$ [2], wenn C, D = 2 [1]

$Y = 1$ [2], wenn E, F = 2 [1]

$$P(XY) = P(X) \cdot P(Y|X)$$

$$P(Y|X) = \frac{P(XY)}{P(X)}$$

***** STATISTICS *****

X-squared = 13.2637 (0.2093)

Degrees of freedom = 10

Number of parameters = 14 (+1)

*** PSEUDO R-SQUARED MEASURES ***

$$0,4570 = \frac{0,6931 - 0,3763}{0,6931}$$

* P(Y|X) *

	baseline	fitted	R-squared
entropy	0.6931	0.3763	0.4570

* P(C|X) *

	baseline	fitted	R-squared
entropy	0.6723	0.2508	0.6269

* P(D|X) *

	baseline	fitted	R-squared
entropy	0.6857	0.4604	0.3285

Entropie $H = -\sum_{i=1}^I \frac{n_i}{n} \ln \frac{n_i}{n}$

Aus dem pdf-File:
„Datenmatrix, Mittlere Werte
und Streuungsmaße“ (Statistik I)

Zahlen-
beispiele ..

	p_i	p_i	p_i	p_i
1	0,5	0,4	0,3	0,2
2	0,5	0,6	0,7	0,8
	0,69	0,67	0,61	0,50

.. unter Ver-
wendung des
natürlichen
Logarithmus \ln

z.B.: $-1 \times [(0,5 \times \log_e(0,5)) + (0,5 \times \log_e(0,5))] = 0,69$

z.B.: $-1 \times [(0,4 \times \log_e(0,4)) + (0,6 \times \log_e(0,6))] = 0,67$

	Gruppe 1	Gruppe 2	p_i
1	0,75	0,17	0,4
2	0,25	0,83	0,6
Entropie	0,56	0,46	0,67
Entropie (gewichtet)	$0,56 \cdot 0,40$ $= 0,22$	$0,46 \cdot 0,60$ $= 0,28$	
n	40	60	100

$$Pseudo - R^2 = \frac{S_y^2 - S_e^2}{S_y^2} = \frac{0,67 - (0,22 + 0,28)}{0,67} = \frac{0,17}{0,67} = \underline{\underline{0,25}}$$

*** INPUT ***

res 2

lat 2

man 4

dim 2 2 2 2 2 2 2 2

lab R S X Y C D E F

sub CDEF CD EF -

mod X Y | X {XY}

C | X D | X

E | Y F | Y

R S | R

dat wr.dat

cas 336

C	D	E	F	freq
1	1	1	1	49
1	1	1	2	5
1	1	2	1	7
1	1	2	2	2
1	2	1	1	9
1	2	1	2	1
1	2	2	1	3
1	2	2	2	4
2	1	1	1	7
2	1	1	2	1
2	1	2	1	10
2	1	2	2	8
2	2	1	1	22
2	2	1	2	12
2	2	2	1	18
2	2	2	2	45

CD

[R=1
S=2]

C	D	E	F	freq
1	1	?	?	13
1	2	?	?	5
2	1	?	?	6
2	2	?	?	19

EF

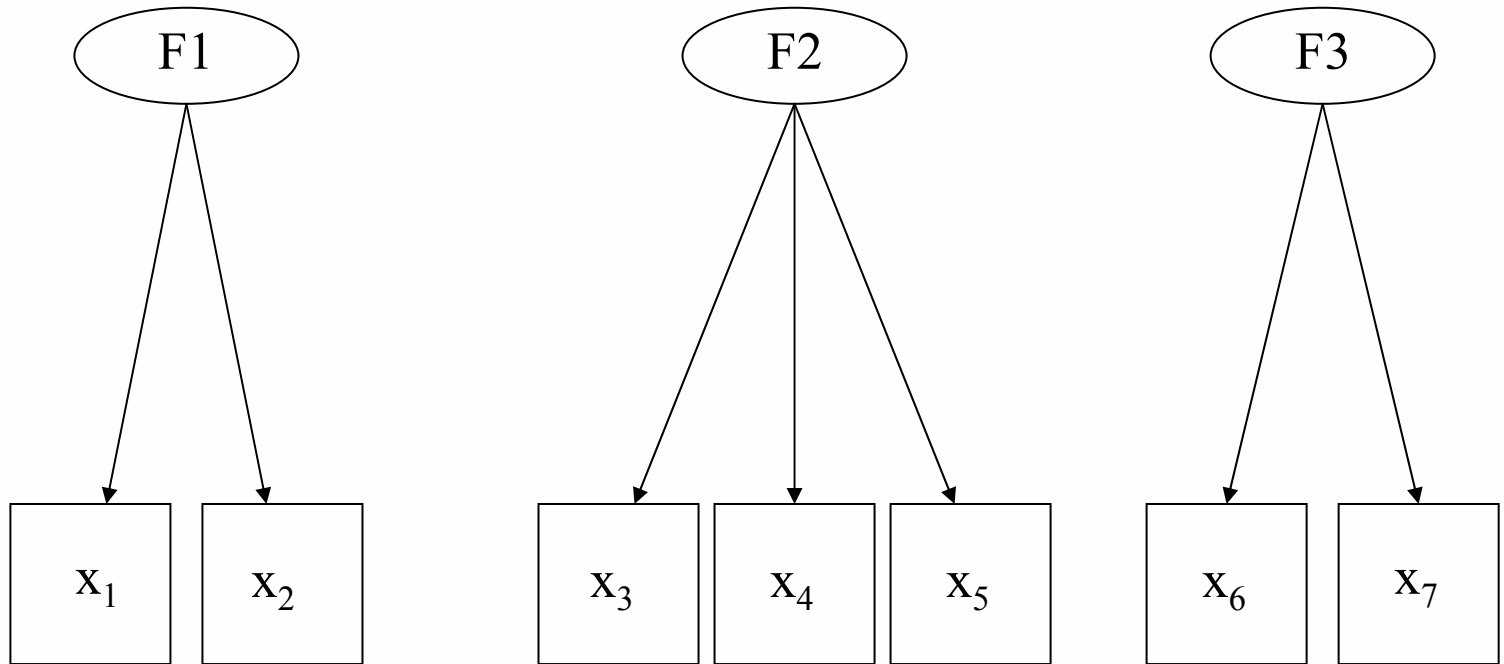
[R=2
S=1]

C	D	E	F	freq
?	?	1	1	8
?	?	1	2	0
?	?	2	1	0
?	?	2	2	7

-

[R=2
S=2]

C	D	E	F	freq
?	?	?	?	75



Grundgedanke der Faktorenanalyse

Schritte bei einer Faktorenanalyse

1. Berechnung der Korrelationsmatrix
2. Extraktion der Faktoren
3. Bestimmung der Kommunalitäten
4. Zahl der Faktoren
5. Faktorinterpretation
6. Bestimmung der Faktorwerte

Extraktion der Faktoren

Beschreibung von standardisierten Ausgangsvariablen z_j
als Funktion hypothetischer Faktoren

$$z_{kj} = a_{j1} \cdot p_{k1} + a_{j2} \cdot p_{k2} + \dots + a_{jQ} \cdot p_{kQ}$$

Variablen z_j $j=1,2,\dots,J$ Faktoren: $q= 1, 2, \dots, Q$ Fall k

a_{jq} **Faktorladung:** Gewicht (Einfluss) von Fak. q auf Var. j
[zur Schätzung der beobachteten Variablen als Funktion der hyp. Faktoren]

p_{kQ} **Faktorwert:** Ausprägung von Q bei Fall k
(Wert von k in Q)

Faktorwerte p und Faktorwertkoeffizienten w_{jk} [in der „Koeffizientenmatrix der Faktorwerte“]

Bestimmung von Faktorwerten p über Faktorwertkoeffizienten w :
[Schätzung der hypothetischen Faktoren q als Funktion der beobachteten Variablen j]

$$p_{kq} = w_{1q}z_{k1} + w_{2q}z_{k2} + \dots + w_{Jq}z_{kJ}$$

p_{kq}	Faktorwert von Person k in Faktor q ($q=1, \dots, Q$)
w_{jq}	standardisierte Regressionskoeffizienten (Faktorwertkoeffizient für Variable j und Faktor q)
z_{kj}	z-standardisierter Wert von Person k in Var. j ($j= 1, .., J$)

Beispiel: Wertrationalität

Kommunalitäten

	Anfänglich	Extraktion	
FR_49_1 taegl. Weg: fuer Umwelt hoehere Kosten	1,000 *	,525	$=0,725^2$
FR_49_2 taegl. Weg: umweltfreundliches Verkehrsmittel trotz hoeheren Zeitaufwand	1,000 *	,587	$=0,766^2$
FR_49_4 taegl. Weg: fuer Umwelt mehr Unbequemlichkeiten	1,000 *	,607	$=0,779^2$
FR_49_6 taegl. Weg: OEPNV nutzen, da Umweltschutz jeden angeht	1,000 *	,458	$=0,677^2$

Extraktionsmethode: Hauptkomponentenanalyse.

 $\Sigma: 4,0 \quad \Sigma: 2,177$ $2,177/4,0 = 0,544$ [54,4% erklärte Varianz]

* Varianz gleich Eins, da standardisierte Variablen

[Faktorladungen]

Komponentenmatrix^a

	Komponente
	1
FR_49_1 taegl. Weg: fuer Umwelt hoehere Kosten	,725
FR_49_2 taegl. Weg: umweltfreundliches Verkehrsmittel trotz hoeheren Zeitaufwand	,766
FR_49_4 taegl. Weg: fuer Umwelt mehr Unbequemlichkeiten	,779
FR_49_6 taegl. Weg: OEPNV nutzen, da Umweltschutz jeden angeht	,677

Extraktionsmethode: Hauptkomponentenanalyse.

a. 1 Komponenten extrahiert

 Ψ Ψ standard. Regressionskoeff. [bei uncorr. Faktoren wie einfache Korrelationen interpretierbar]

Erklärte Gesamtvarianz

Komponente	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2,179	54,463	54,463	2,179	54,463	54,463
2	,749	18,720	73,183			
3	,583	14,572	87,755			
4	,490	12,245	100,000			

Extraktionsmethode: Hauptkomponentenanalyse.

Koeffizientenmatrix der Komponentenwerte

	Komponente
	1
FR_49_1 taegl. Weg: fuer Umwelt hoehere Kosten	,333
FR_49_2 taegl. Weg: umweltfreundliches Verkehrsmittel trotz hoeheren Zeitaufwand	,352
FR_49_4 taegl. Weg: fuer Umwelt mehr Unbequemlichkeiten	,358
FR_49_6 taegl. Weg: OEPNV nutzen, da Umweltschutz jeden angeht	,311

Extraktionsmethode: Hauptkomponentenanalyse.
Rotationsmethode: Varimax mit Kaiser-Normalisierung.
Komponentenwerte.

Eigenwert eines Faktors
= dessen Varianzbeitrag in Hinblick auf die Varianz aller Variablen
(Summe der quadrierten Faktorladungen über betreffenden Faktor)

Hauptachsen-Faktorenanalyse

Kommunalitäten

	Anfänglich	Extraktion	
FR_49_1 taegl. Weg: fuer Umwelt hoehere Kosten	,275*	,367	[=0,606 ²]
FR_49_2 taegl. Weg: umweltfreundliches Verkehrsmittel trotz hoeheren Zeitaufwand	,297*	,447	[=0,669 ²]
FR_49_4 taegl. Weg: fuer Umwelt mehr Unbequemlichkeiten	,317*	,485	[=0,696 ²]
FR_49_6 taegl. Weg: OEPNV nutzen, da Umweltschutz jeden angeht	,222*	,287	[=0,536 ²]

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

 $\Sigma: 1,586$

[Faktorladungen]

Faktorenmatrix^a

	Faktor
	1
FR_49_1 taegl. Weg: fuer Umwelt hoehere Kosten	,606
FR_49_2 taegl. Weg: umweltfreundliches Verkehrsmittel trotz hoeheren Zeitaufwand	,669
FR_49_4 taegl. Weg: fuer Umwelt mehr Unbequemlichkeiten	,696
FR_49_6 taegl. Weg: OEPNV nutzen, da Umweltschutz jeden angeht	,536

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

a. 1 Faktoren extrahiert. Es werden 7 Iterationen benötigt.

R^2 bei Regression auf alle anderen Var's;
[Mit R^2 statt 1 in den Elementen der Hauptdiagonalen der zu analysierenden Korrelationsmatrix]

 $1,586/4 = 0,397$ [39,7% erkl. Varianz]

Erklärte Gesamtvarianz

aftern

Faktor	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2,179	54,463	54,463	1,586	39,660	39,660
2	,749	18,720	73,183			
3	,583	14,572	87,755			
4	,490	12,245	100,000			

Extraktionsmethode: Hauptachsen-Faktorenanalyse.

Koeffizientenmatrix der Faktorwerte

	Faktor
	1
FR_49_1 taegl. Weg: fuer Umwelt hoehere Kosten	,247
FR_49_2 taegl. Weg: umweltfreundliches Verkehrsmittel trotz hoeheren Zeitaufwand	,332
FR_49_4 taegl. Weg: fuer Umwelt mehr Unbequemlichkeiten	,368
FR_49_6 taegl. Weg: OEPNV nutzen, da Umweltschutz jeden angeht	,195

Extraktionsmethode: Hauptachsen-Faktorenanalyse.
 Rotationsmethode: Varimax mit Kaiser-Normalisierung.
 Methode für Faktorwerte: Regression.

Komponentenmatrix^a

	Komponente	
	1	2
FR_49_1 taegl. Weg: fuer Umwelt hoehere Kosten	,721	-,210
FR_49_2 taegl. Weg: umweltfreundliches Verkehrsmittel trotz hoeheren Zeitaufwand	,765	-5,252E-02
FR_49_4 taegl. Weg: fuer Umwelt mehr Unbequemlichkeiten	,778	2,581E-02
FR_49_6 taegl. Weg: OEPNV nutzen, da Umweltschutz jeden angeht	,687	,114
FR_49_5 taegl. Weg: Mensch handelt zu eigenem Vorteil	-3,044E-02	,825
FR_49_7 taegl. Weg: Bevoelkerung wenig umweltgerecht	,147	,807

Extraktionsmethode: Hauptkomponentenanalyse.

a. 2 Komponenten extrahiert

Rotierte Komponentenmatrix^a

	Komponente	
	1	2
FR_49_1 taegl. Weg: fuer Umwelt hoehere Kosten	,731	-,172
FR_49_2 taegl. Weg: umweltfreundliches Verkehrsmittel trotz hoeheren Zeitaufwand	,767	-1,171E-02
FR_49_4 taegl. Weg: fuer Umwelt mehr Unbequemlichkeiten	,776	6,722E-02
FR_49_6 taegl. Weg: OEPNV nutzen, da Umweltschutz jeden angeht	,680	,151
FR_49_5 taegl. Weg: Mensch handelt zu eigenem Vorteil	-7,434E-02	,822
FR_49_7 taegl. Weg: Bevoelkerung wenig umweltgerecht	,103	,814

Extraktionsmethode: Hauptkomponentenanalyse.

Rotationsmethode: Varimax mit Kaiser-Normalisierung.

a. Die Rotation ist in 3 Iterationen konvergiert.

Kommunalitäten

	Anfänglich	Extraktion
FR_49_1 taegl. Weg: fuer Umwelt hoehere Kosten	1,000	,564
FR_49_2 taegl. Weg: umweltfreundliches Verkehrsmittel trotz hoeheren Zeitaufwand	1,000	,588
FR_49_4 taegl. Weg: fuer Umwelt mehr Unbequemlichkeiten	1,000	,606
FR_49_6 taegl. Weg: OEPNV nutzen, da Umweltschutz jeden angeht	1,000	,485
FR_49_5 taegl. Weg: Mensch handelt zu eigenem Vorteil	1,000	,682
FR_49_7 taegl. Weg: Bevoelkerung wenig umweltgerecht	1,000	,673

Extraktionsmethode: Hauptkomponentenanalyse.

$$[=0,721^2 + (-0,210)^2]$$

(Quadrate der Faktorladungen addieren)

-
-
-
-
-

$$\Sigma: 6,0 \quad \Sigma: 3,598$$

$$3,598/6,0 = 0,5997 \text{ [59,97\% erklärte Varianz]}$$

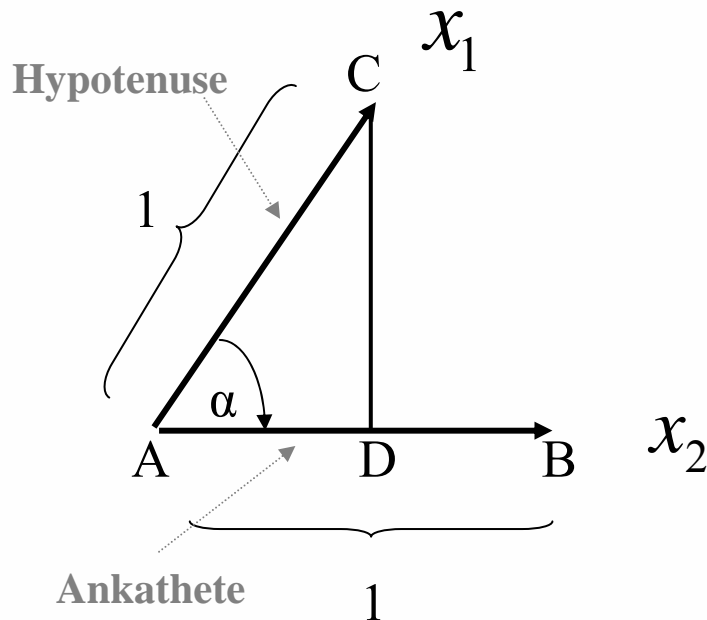
Erklärte Gesamtvarianz

Komponente	Anfängliche Eigenwerte			Summen von quadrierten Faktorladungen für Extraktion			Rotierte Summe der quadrierten Ladungen		
	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %	Gesamt	% der Varianz	Kumulierte %
1	2,205	36,743	36,743	2,205	36,743	36,743	2,202	36,705	36,705
2	1,394	23,226	59,969	1,394	23,226	59,969	1,396	23,264	59,969
3	,716	11,926	71,895						
4	,631	10,510	82,405						
5	,574	9,559	91,964						
6	,482	8,036	100,000						

Extraktionsmethode: Hauptkomponentenanalyse.

Eigenwert eines Faktors
 = dessen Varianzbeitrag in Hinblick
 auf die Varianz aller Variablen
 (Summe der quadrierten Faktorladungen
 über betreffenden Faktor)

Vektordarstellung einer Korrelation zwischen zwei Variablen



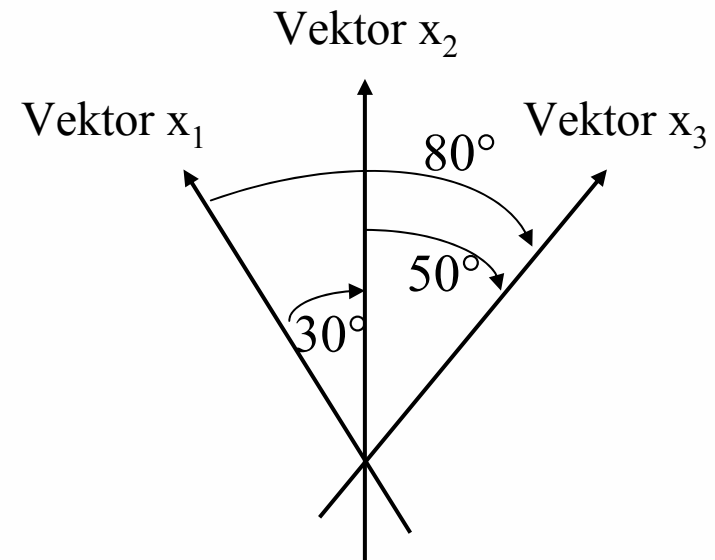
$$r_{(x_1x_2)} = \cos(\alpha) = \frac{AD}{AC}$$

Beispiel: Korrelationsmatrix

$$R = \begin{pmatrix} 1 & & \\ 0,8660 & 1 & \\ 0,1736 & 0,6428 & 1 \end{pmatrix}$$

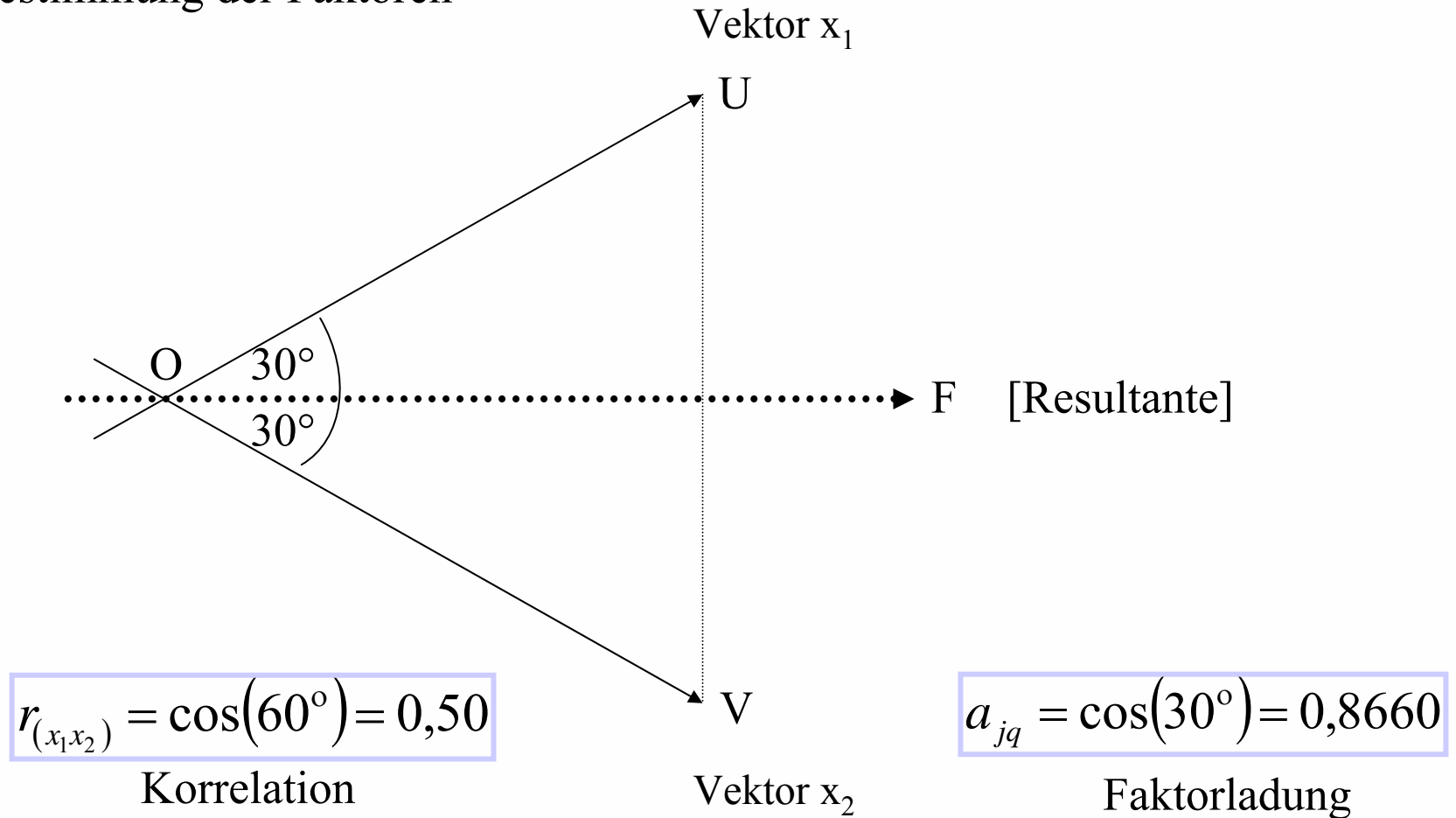
Dieselbe Korrelationsmatrix
mit Winkelausdrücken

$$R = \begin{pmatrix} 0^\circ & & \\ 30^\circ & 0^\circ & \\ 80^\circ & 50^\circ & 0^\circ \end{pmatrix}$$

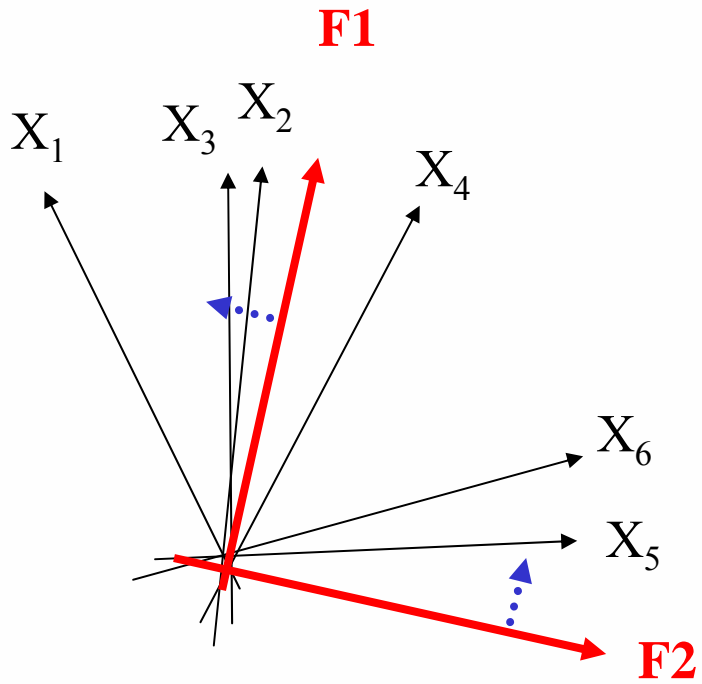


.. und in der Vektordarstellung

Bestimmung der Faktoren



Vor Rotation



Nach Rotation

