

# Statistik und Datenanalyse: Aufbau

## 8. Sitzung – Dimensionsreduktion

Benjamin Fretwurst

▶ PDF-Version der Folien



# Inhalt

- 1 Messung und Analyse latenter Faktoren
  - 1.1 Multikollinearität und Dimensionsreduktion
  - 1.2 R-Angstskala in unserer Befragung
  - 1.3 Vorgehen der PCA und Faktorenanalyse
  - 1.4 Faktorrotation
- 2 Literaturbeispiel (Übung 3 – siehe LEF)
- Take Home – Ausblick – Vokabeln
- 3 Übung 3
  - 3.1 Laden Sie die Daten unserer Befragung
  - 3.2 Führen Sie eine Faktorenanalyse für unsere RAQ aus mit Hilfe des Paketes **psych**
- LEF 8
- Literatur

Orga



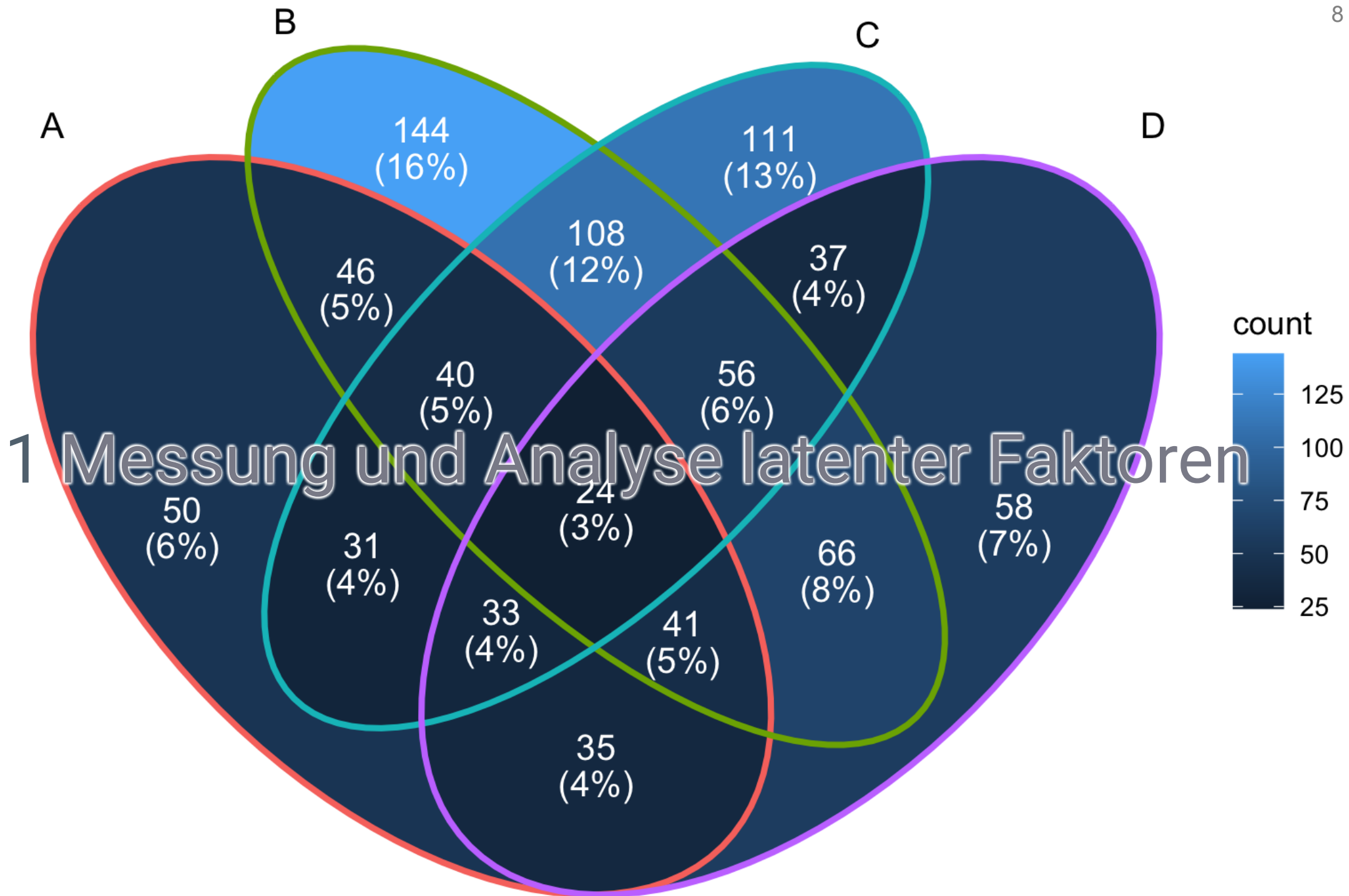
# R-Probleme und allgemeine Fragen

- Allgemeine Fragen bitte ins Forum ↩ entsprechende Teams-Anfragen oder Emails beantworte ich nicht mehr.
- NICHT BLOSS SCREENSHOTS bei R-Problemen!!!
- R-Übung 3 am Ende der Folien und auf unserer Seite unter 9 «Übung: Dimensionsreduktion».
- LEF 8: Wesentlich zum Text Song et al. (2004).

# Lernziele

## Faktorenanalyse

- Ziele von Faktorenanalysen
  - stark multikollineare UVs zusammenfassen
  - Indexbildung
  - latente Konstrukte identifizieren
  - latente Konstrukte prüfen
- Faktorenextraktion explorativ
- Indexbildung mit Faktorenanalyse
- konfirmatorische Faktorenanalyse
- Lesen von EFA und Regressionsdokumentationen



# Was geht? ... mit Faktorenanalysen!

- Mit Faktorenanalysen können latente Einflüsse explorativ gefunden werden.
- Die Messung latenter Konstrukte kann (konfirmatorisch) geprüft werden.
- Mit Faktorenanalysen können Indices gebaut werden.
- Wenn UVs in Regressionsmodellen hoch multikollinear sind, können sie zu unkorrelierten Faktoren zusammengefasst werden.

# Methoden – Einführung – Indices

## Bsp. Medienqualität

Home

Indizes

Dimension

Kombination

Skalen

TH



LEF 5

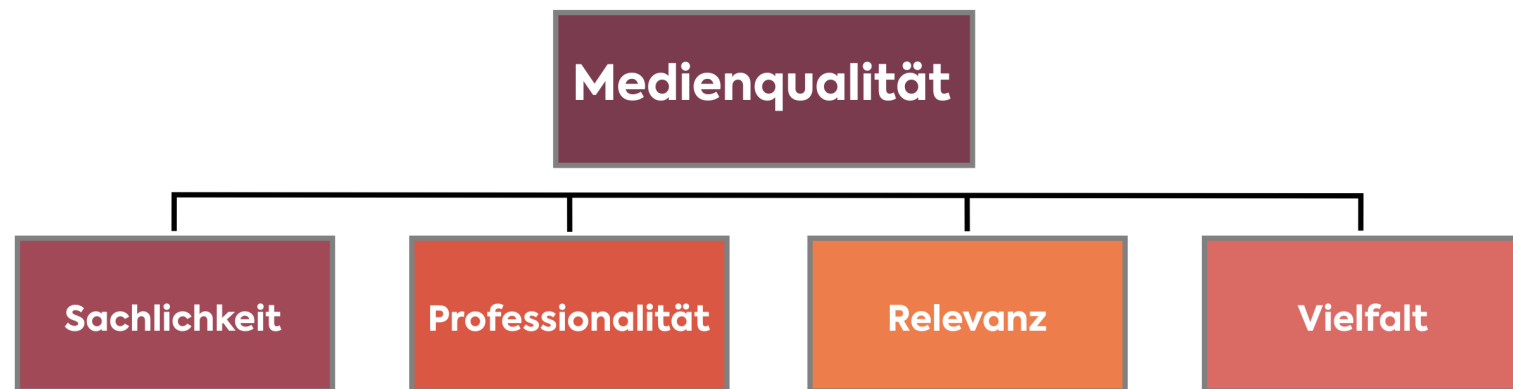
Vokabeln

Literatur

Notizen

Medienqualität als latente Variable

Medienqualität ist nicht direkt messbar. Sie kann aber anhand Indikatoren wie Sachlichkeit, Relevanz, Quellentransparenz usw. gemessen werden.



Quelle: Eigene Darstellung.



# Dimensionen



# 1.1 Multikollinearität und Dimensionsreduktion

## Multikollinearität

In einem Koordinationsystem macht jede Variable eine Dimension auf, also 3 Variablen, 3 Achsen → 3D.

Korrelieren die Variablen stark miteinander, liegen sie auf einer Geraden, die nur eine Dimension hat. Man kann die Variablen auf diese eine Dimension reduzieren. Dann müssen wir uns nur noch um eine Variable (Index) «kümmern», statt um drei! Wir versuchen also eine Dimensionsreduktion.

# 1.2 R-Angstskala in unserer Befragung

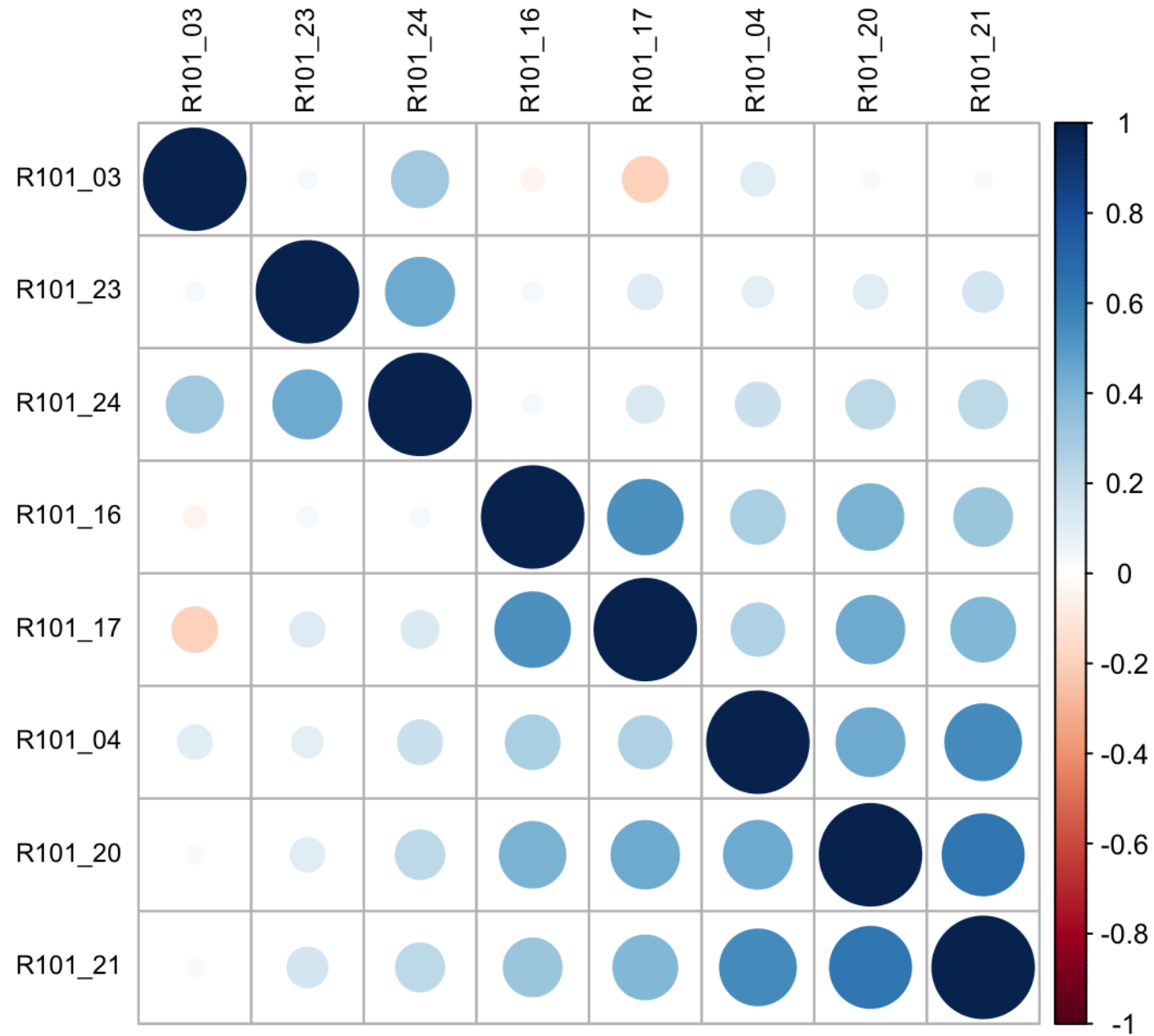
Wenn wir mehrere Variablen (zB einer Skala) zu einer reduzieren können, können wir damit Multikollinearitätsprobleme lösen können.

## R-Angstskala in DE-Adaption

**x**

R101_16	RAQ: Ich weine offen, wenn von zentraler Tendenz die Rede ist.
R101_17	RAQ: Ich falle in ein Koma, wenn ich eine Gleichung sehe.
R101_03	RAQ: Standardabweichungen begeistern mich.
R101_04	RAQ: Ich träume davon, dass Pearson mich mit Korrelationskoeffizienten angreift.
R101_20	RAQ: Ich kann nicht schlafen, weil ich an Signifikanzen denke.
R101_21	RAQ: Ich wache unter meiner Bettdecke auf und denke, dass ich unter einer Normalverteilung gefangen bin.
R101_23	RAQ: Wenn ich gut in Statistik bin, werden die Leute denken, ich sei ein Streber.
R101_24	RAQ: Ich mag Statistik, würde das aber nie vor meinen Freunden zugeben.

# Korrelationsmatrix



# Fragen, die die FA beantwortet

## **Wie viel geht bei der Dimensionsreduktion durch die Faktoren verloren, bzw. was wird abgebildet?**

Mit der Gesamtlösung kann man schauen, welchen Anteil der Varianz aller Faktoren durch die Faktorenlösung abgebildet wird.

## **Wie gut werden die Variablen durch die Faktoren abgebildet?**

Die Kommunalitäten und «Uniqueness» geben an, wie gut jede Variable durch die gebildeten Faktoren repräsentiert werden.

## **Was bedeuten die Faktoren?**

Faktorladungen geben die Korrelationen der Faktoren mit jeder Variable an. Also welche Faktoren, welche Variablen repräsentieren? Dadurch kann den Faktoren ein Sinn und ein Name gegeben werden.

# Faktoren in 3D

- Drei Variablen (3D)
- je mehr n, desto dunkler
- Zwei Faktoren
  - orthogonal
  - unkorreliert
- rotiert

WebGL is not supported by your browser - visit <https://get.webgl.org> for more info

# 1.3 Vorgehen der PCA und Faktorenanalyse

## 1. Prüfen, ob ein Set an Variablen für eine Faktorenanalyse geeignet ist

- Korrelationsanalyse
- KMO

## 2. Feststellen, wie viele latente Faktoren extrahiert werden sollten

- Scree-Plot
- Parallelanalyse

## 3. Interpretation der Faktoren

- Berechnung der Faktorladungen
- Identifikation der Faktoren (Interpretation)



# KMO

Das Kaiser-Meyer-Olkin-Kriterium prüft, ob ein Set von Variablen geeignet ist für eine Faktorenanalyse.

Kommt es nicht gut raus, kann man Variablen mit kleinen Werten ausschliessen.

► In R `psych::KMO()`

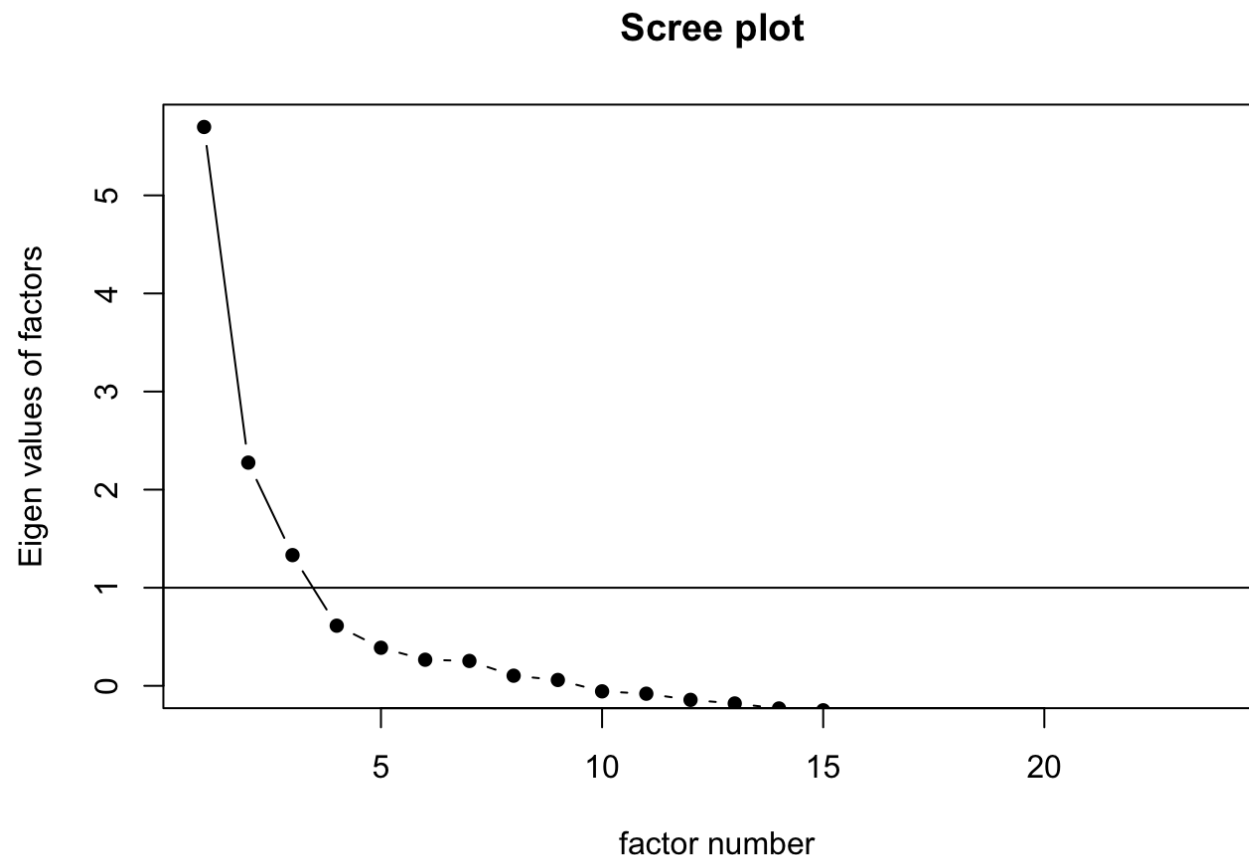
## Kaiser-Meyer-Olkin factor adequacy

- < .5 inakzeptabel
- .50-.59 miserabel
- .60-.69 mässig
- .70-.79 mittel ok
- .80-.89 gut
- .90-1.00 irre super

# Scree Plot der Eigenwerte (eigen values)

Die «factor number» über der 1-Linie («Eigen values» > 1) ist eine Empfehlung für die Anzahl an Faktoren, bei denen jeder Faktor mehr Varianz (Eigenwert > 1) auf sich vereint als die ursprünglichen Dimensionen.

► In R `psych::scree()`

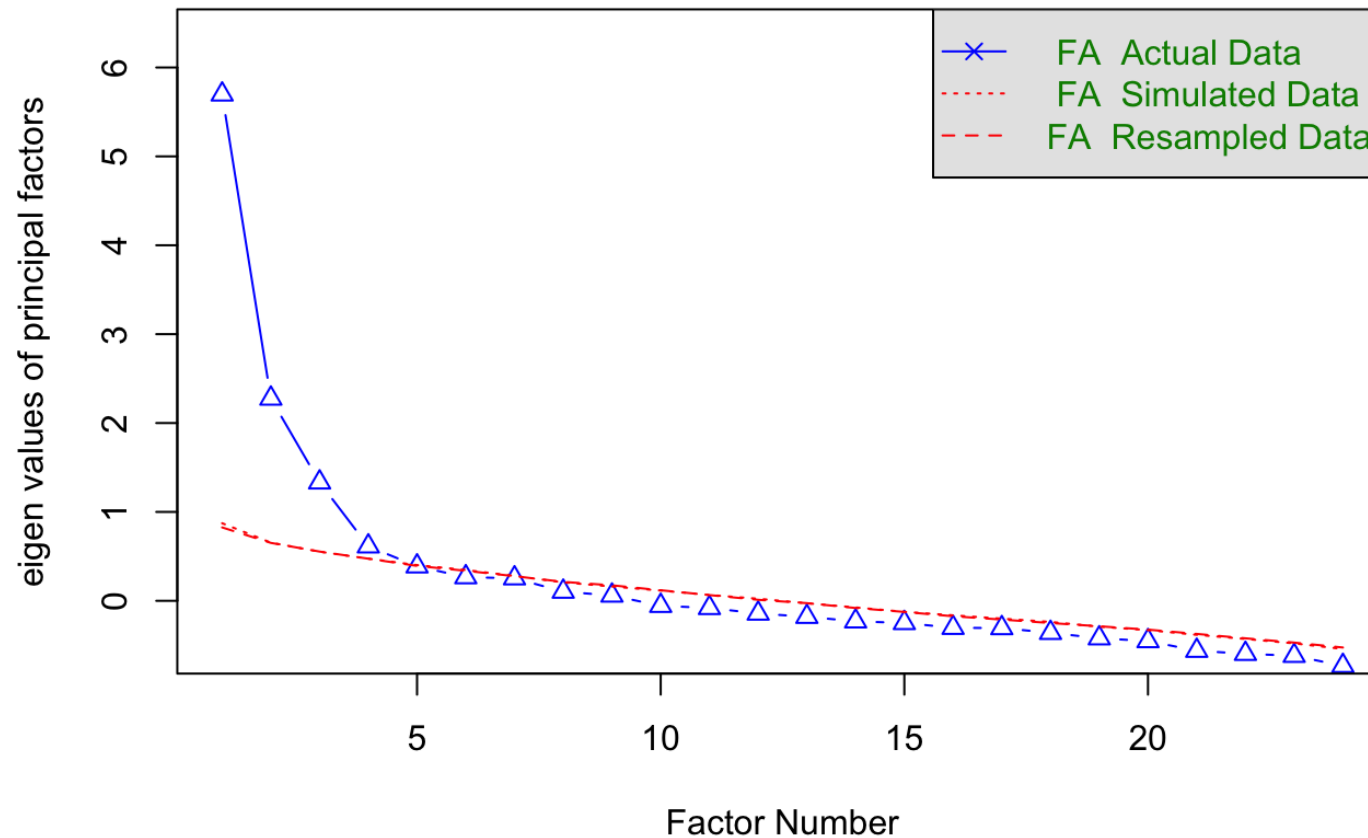


# Parallel Analysis als Alternative zum Scree Plot

Wie viele Faktoren soll es geben? Wie viele haben einen "Eigenvalue" > 1?

```
## Parallel analysis suggests that the number of factors = 4 and the number of components = NA
```

**Parallel Analysis Scree Plots**



# Skalenreliabilität der Faktoren (Cronbachs $\alpha$ )

## Skalenreliabilität

Skalenreliabilität ist die innere Konsistenz einer Skala. Der gängigste Koeffizient für die Skalenreliabilität ist Cronbachs  $\alpha$ . Er geht von  $-\infty$  bis 1, wobei 1 wäre perfekt.

$$\alpha_{st} = \frac{K \cdot \bar{r}}{1 + (K - 1) \cdot \bar{r}}$$

$\alpha_{st}$ : standardisiertes Cronbach  $\alpha$

K: Anzahl Elemente (Variablen) einer Skala

$\bar{r}$ : Durchschnittliche Korrelation





# 1.4 Faktorrotation

## Unrotiert

Beim Verfahren der Faktorenanalyse wird erst ein Faktor in die Variablen gelegt, der alle am besten erklärt. Dann kommt der zweite und optimiert den Rest der Varianz usw. Das ergibt ein Ungleichgewicht zwischen den Faktoren. Darum wird rotiert.

## Orthogonale und oblique Rotation

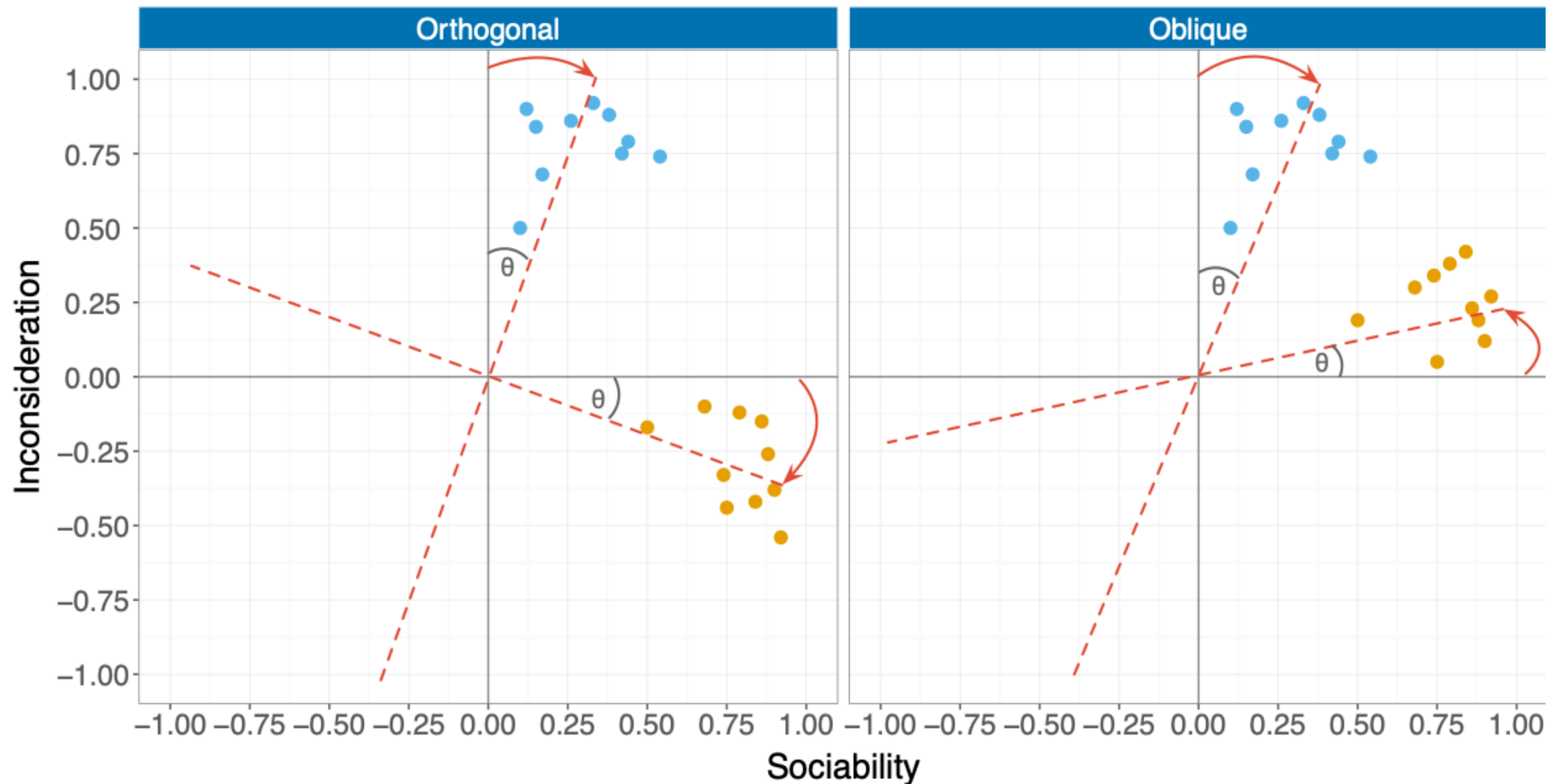
Faktoren werden rotiert, damit jeder optimal Varianz erklärt. Wenn orthogonal rotiert wird, sind die Faktoren 100% unkorreliert. Nach obliquen Rotation sind die Faktoren leicht korreliert, geben aber die Variablen besser wieder.

# Faktorrotation

Bei orthogonaler Rotation sind die Faktoren unkorreliert.

Orthogonal ist klarer interpretierbar ...

Bei der obliquen dürfen sie leicht korrelieren.  
... oblique gibt realistischere Ergebnisse.



# Faktorladungen

$$MR1 = b_1raq_6 + b_2raq_{18} + b_3raq_{13} + b_4raq_7 + b_5raq_{10} + b_6raq_{15} + \dots \quad (1)$$

$$MR2 = b_1raq_{09} + b_2raq_{23} + b_3raq_{19} + b_4raq_{22} + b_5raq_{02} \quad (2)$$

- Die Bs sind die Faktorladungen.
- Faktorladungen geben an, welches Gewicht (Bedeutung) die einzelnen Variablen für den jeweiligen Faktor haben.
- Jeder Faktor wird anhand der Variablen mit den höchsten Ladungen auf diesem Faktor interpretiert.



# Faktorladungen RAQ

# Variableneignung – Kommunalitäten & Uniqueness

## Kommunalitäten

Die Kommunalität einer Variable ist der Varianzanteil, den sie mit den extrahierten Faktoren teilt. Kommunalitäten unter .4 sind eher dürftig.

Uniqueness = 1 - Kommunalität

## Uniqueness

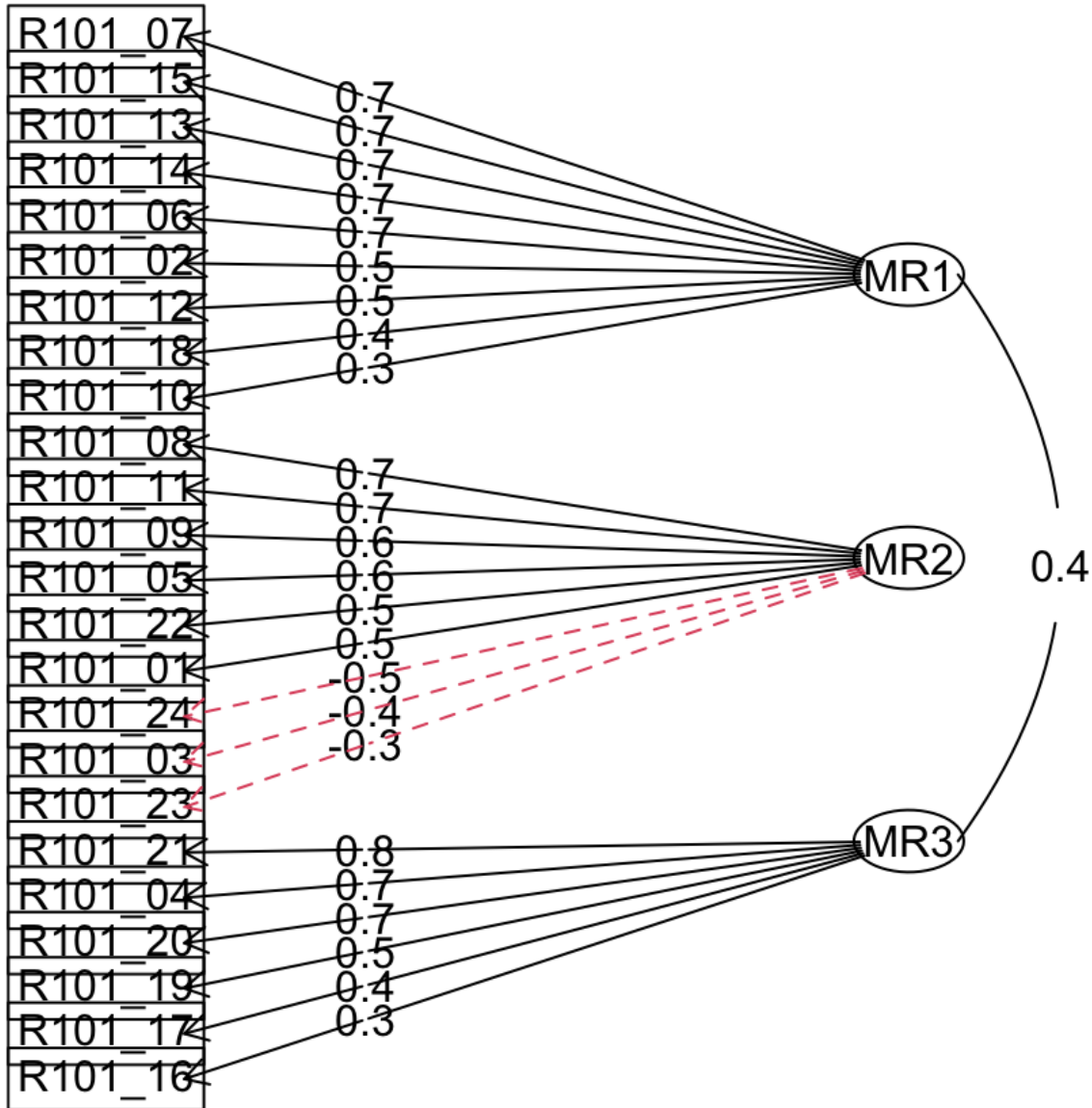
Die Uniqueness-Werte drücken aus, wie hoch der Varianzanteil ist, der **nicht** durch die Faktorenlösung erklärt werden konnte. Werte über .6 sind eher dürftig.

## Complexity

Die Komplexität je Variable gibt an, ob es Mehrfachladungen auf einer Variable gibt. Wenn sie 1 ist, dann ist das Ergebnis eindeutig, wenn sie nahe 2 ist, dann laden zwei Faktoren auf dieser Variable.

FA als SEM – is future music

# Factor Analysis





2 Literaturbeispiel (Übung 3 – siehe LEF)

# EFA Song et al. (2004)

## Internet Gratifications and Internet Addiction: On the Uses and Abuses of New Media

INDEOK SONG, B.A.,<sup>1</sup> ROBERT LAROSE, Ph.D.,<sup>1</sup> MATTHEW S. EASTIN, Ph.D.,<sup>2</sup> and  
CAROLYN A. LIN, Ph.D.<sup>3</sup>

### ABSTRACT

Internet addiction has been identified as a pathological behavior, but its symptoms may be found in normal populations, placing it within the scope of conventional theories of media attendance. The present study drew upon fresh conceptualizations of gratifications specific to the Internet to uncover seven gratification factors: Virtual Community, Information Seeking, Aesthetic Experience, Monetary Compensation, Diversion, Personal Status, and Relationship Maintenance. With no parallel in prior research, Virtual Community might be termed a “new” gratification. Virtual Community, Monetary Compensation, Diversion, and Personal Status gratifications accounted for 28% of the variance in Internet Addiction Tendency. The relationship between Internet addiction and gratifications was discussed in terms of the formation of media habits and the distinction between content and process gratifications.

A scenic landscape featuring rolling hills and a valley. In the foreground, a herd of black and white cows is grazing in a field of tall, dry grass. The middle ground shows a valley with green fields and scattered trees. The background consists of distant, hazy mountains under a cloudy sky. The text "Take Home – Ausblick – Vokabeln" is overlaid on the image.

# Take Home – Ausblick – Vokabeln

# Take Home

## Interaktionen

- Sie wissen, wofür man Faktorenanalysen alle einsetzen kann
- Sie können die Dokumentation einer Faktorenanalyse lesen und verstehen
- Sie können selbst eine Faktorenanalyse ausführen und interpretieren



# Ausblick

Wir gehen gemeinsam den Text von Song et al. (2004) durch und die R-Übung, die auf der nächsten Folie beschrieben ist.

# 3 Übung 3

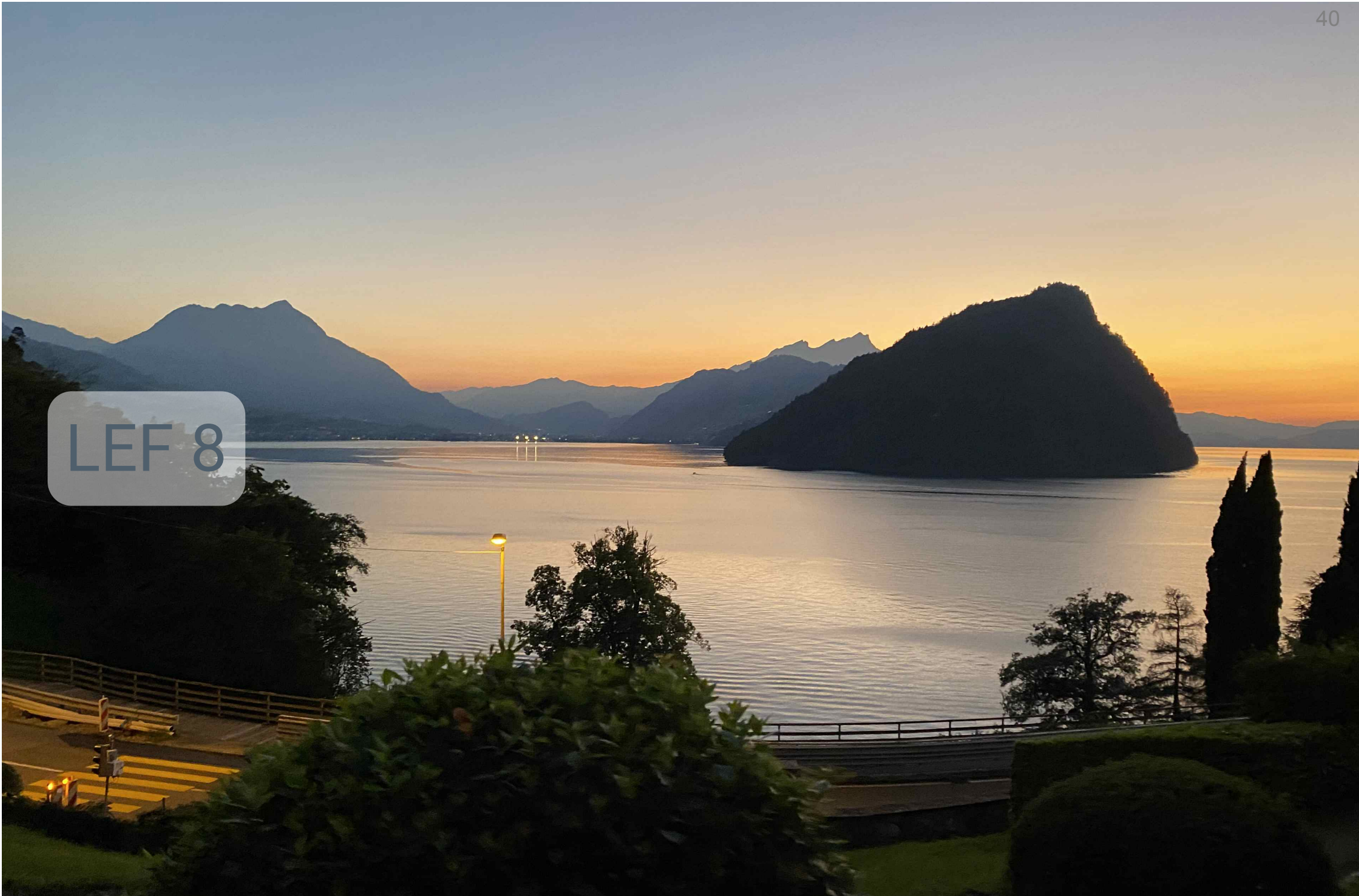
## 3.1 Laden Sie die Daten unserer Befragung

(Für den Fall, dass es Probleme gibt, laden Sie die Daten [hier](#) herunter. Wenn das auch zu Problemen führt, melden Sie sich im Forum.)

## 3.2 Führen Sie eine Faktorenanalyse für unsere RAQ aus mit Hilfe des Paketes **psych**

Ziehen Sie dazu die Hilfe dieser Website heran: <https://md.psych.bio.uni-goettingen.de/mv/unit/fa/efa.html>

LEF 8



# Essayfragen 8

E8.1 Bitte schauen Sie sich die folgenden drei (mistigen) Tabellen an, die aus fiktiven Publikationen stammen. Jedes mal werden die Koeffizienten als "beta" bezeichnet, aber es könnten nur die unstandardisierten Regressionskoeffizienten  $b$  sein oder die standardisierten Regressionskoeffizienten  $\text{std. } b$ . Was ist wohl was? a) Begründen Sie Ihre Entscheidung! b) Erklären Sie, warum es nicht  $\beta$ s sein können.

	A		
	beta	t	p
Var 1	0.02	1.4	0.168
Var 2	0.35	2.9	0.005
Var 3	0.43	3.1	0.003
Var 4	-0.21	-2.2	0.032
Var 5	0.03	1.8	0.078

	B		
	beta	t	p
pred 1	3.45	3.25	0.002
pred 2	1.23	2.34	0.023
pred 3	0.001	4.1	0.000
pred 4	2.5	-3.2	0.002
pred 5	0.3	1.8	0.078

	C		
	beta	t	p
UV 1	0.3	3.6	0.001
UV 2	0.5	1.28	0.206
UV 3	-0.2	-3.9	0.000
UV 4	-0.08	-4.2	0.000
UV 5	0.04	1.8	0.078

# Essayfragen 8.2-8.3

**E8.2** Schreiben Sie zu jedem der folgenden Schlagworte in ganzen Sätzen auf, zu welchen Zwecken Faktorenanalysen alles eingesetzt werden können: a) latente Variablen entdecken b) Multikollinearität c) Indices d) Messung latenter Konstrukte

**E8.3** a) Erklären Sie, was Faktorenadungen sind. b) Wie gehen Sie vor, wenn Sie Faktorladungen interpretieren wollen?

# Essayfragen 8.4-8.10

Lesen Sie Song et al. (2004) und beantworten Sie dann folgende Fragen und die anschliessenden MCs.

**E8.4** Was wird mit Cronbachs  $\alpha$  beschrieben?

**E8.5** Was für eine Faktorenanalyse haben Song et al durchgeführt?

**E8.6** Mit welcher Methoden wurden die Faktoren rotiert?

**E8.7** Wie viele initiale Faktoren hat die Faktorenanalyse herausgegeben?

**E8.8** a) Nach welchen Kriterien wurden die Faktoren ausgewählt? b) Wie bewerten Sie das Vorgehen?

**E8.9** a) Wie viele Items (Fragen) wurden mit der Faktorenanalyse ausgewertet? b) Auf wie viele Faktoren wurden die Items reduziert?

**E8.10** Wie viel Varianz erklären die Faktoren, die von Song et al ausgewählt wurden?



# MC-Fragen 8

# MC 8.1.

MC 8.1: Was sagen die Zahlen im Text von Song et al. (2004)?

richtig	falsch	Aussagen
<input type="radio"/>	<input type="radio"/>	Die Variablen erklären 67.4 Prozent der extrahierten Faktoren.
<input type="radio"/>	<input type="radio"/>	Die Faktoren wurden mit Hilfe der Varimax-Methode rotiert.
<input type="radio"/>	<input type="radio"/>	Alle Faktoren zeigen ein hohes Cronbachs $\alpha$ .
<input type="radio"/>	<input type="radio"/>	Der "Factor 5" erklärt etwa das Vierfache einer einfachen Variable.

Punkte: 0

## MC 8.2.

### MC 8.2: Was sagen die Zahlen im Text von Song et al. (2004)?

- | richtig               | falsch                | Aussagen  |
|-----------------------|-----------------------|---|
| <input type="radio"/> | <input type="radio"/> | Der "Factor 1" erklärt die meiste Varianz der ursprünglichen Variablen.   |
| <input type="radio"/> | <input type="radio"/> | Da auf "Factor 3" und "Factor 4" gleich viele Variablen laden, haben Sie dieselbe Varianzaufklärung.  |
| <input type="radio"/> | <input type="radio"/> | Wenn der Eigenwert eines Faktors für einen Faktor A grösser ist als bei Faktor B, muss bei A Cronbach $\alpha$ auch grösser sein als bei B. |
| <input type="radio"/> | <input type="radio"/> | In der Korrelationstabelle sind die Faktoren abgebildet, die teilweise stark korrelieren.   |

Punkte: 0

## MC 8.3.

### MC 8.3: Was sagen die Zahlen im Text von Song et al. (2004)?

- | richtig               | falsch                | Aussagen  |
|-----------------------|-----------------------|---|
| <input type="radio"/> | <input type="radio"/> | In Tabelle 3 sind die $\beta$ s der GG angegeben, die mit den b's in der Stichprobe geschätzt wurden. |
| <input type="radio"/> | <input type="radio"/> | Die standardisierten BETAS sind mit den Korrelationswerten der Tabelle 2 vergleichbar.                |
| <input type="radio"/> | <input type="radio"/> | Die Varianzaufklärung der Regression beträgt 54 Prozent.  |
| <input type="radio"/> | <input type="radio"/> | Der p-Wert von "Virutal community" ist genau 0.   |

**Punkte: 0**

## MC 8.4.

### MC 8.4: Was sagen die Zahlen im Text von Song et al. (2004)?

- | richtig               | falsch                | Aussagen   |
|-----------------------|-----------------------|--|
| <input type="radio"/> | <input type="radio"/> | Je stärker Studierende versuchen sich in eine «Virtuelle community» zu integrieren, desto eher werden sie internetsüchtig. |
| <input type="radio"/> | <input type="radio"/> | Multikollinearität in den UVs kann ausgeschlossen werden.  |
| <input type="radio"/> | <input type="radio"/> | Wer das Internet nutzt, um Informationen zu suchen, gerät eher nicht in eine Internetsucht.                                |
| <input type="radio"/> | <input type="radio"/> | Wer das Internet nutzt, um seinen «Personal status» zu verbessern, neigt stark zu Internetsucht.                           |

Punkte: 0

## MC 8.5.

### MC 8.5: Was sagen die Zahlen im Text von Song et al. (2004)?

- | richtig               | falsch                | Aussagen   |
|-----------------------|-----------------------|--|
| <input type="radio"/> | <input type="radio"/> | Die signifikante Korrelation zwischen «Personal status» und «Internetaddiction» ist eher durch Drittvariablen zustande gekommen. |
| <input type="radio"/> | <input type="radio"/> | Das Konfidenzintervall des BETA von «Monetary compensation» schliesst die 0 mit ein.   |
| <input type="radio"/> | <input type="radio"/> | Da «Personal status» einen p-Wert $> .05$ hat, ist $\beta$ in der GG 0 und daher deutlich kleiner als das in der Stichprobe.     |
| <input type="radio"/> | <input type="radio"/> | Da in der Tabelle standardisierte BETAs angezeigt werden, sind die t-Werte eher redundant.                                       |

Punkte: 0

Insgesamt (nur MCs) 0 von 10 Punkten, was 0% und etwa einer 1 entspricht.

# Vokabeln [↗](#)

 Search: 

Nr	Sitzung	Inhalt	Deutsch	Englisch	Erläuterung
<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>
87	7	EFA, CFA, Lesekompetenz	Cronbachs $\alpha$	Cronbachs $\alpha$	Cronbachs $\alpha$ ist für die innere Ko einer Skala. Grur die durchschnitt Korrelation aller Items. Cronbach bis max. 1 gehen währe dann perf
71	6	Faktorenanalysen	Dimensionsreduktion	dimensional reduction	Dimensionsredu bedeutet, dass c der Merkmalsdir (Variablen) redu.
72	6	Faktorenanalysen	Eigenwerte	eigenvalues	Varianz, die ein f sich vereint.

Nr	Sitzung	Inhalt	Deutsch	Englisch	Erläuterung
73	6	Faktorenanalysen	Faktorenanalyse	factor analysis	Verfahren zur Dimensionsreduktion. Mehrere Variablen werden in weniger Faktoren zusammengefasst.
76	6	Faktorenanalysen	Faktorladung	factor loadings	Die Korrelationen der Variablen mit den Faktoren.
77	6	Faktorenanalysen	Faktorlösung	Solution	Die Faktorlösung, bei der die Anzahl der Faktoren optimal ist für die Darstellung der Daten.
81	6	Faktorenanalysen	Faktorrotation	factor rotation	Die Faktorrotation, bei der die Faktoren so an die Variablen angepasst werden, so dass die Varianz besser erklärt wird.



Nr	Sitzung	Inhalt	Deutsch	Englisch	Erläuterung
78	6	Faktorenanalysen	Hauptkomponentenanalyse	Principle Component Analysis (PCA)	Die Hauptkomponenten fasst Variablen zusammen. Das latente Konzept interessiert dabei.
79	6	Faktorenanalysen	Kommunalitäten	communalities	Kommunalitäten Varianzanteil einer Variable, der durch Faktorenlösung ist. Unter .4 ist ein
84	6	Faktorenanalysen	Scree Plot	Scree Plot	Eine graphische Darstellung der Eigenwerte, die zeigt, welche Faktorenlösung
88	7	EFA, CFA, Lesekompetenz	Skalenreliabilität	scale reliability	Skalenreliabilität bezeichnet die interne Konsistenz einer Skala, also wie stark die Items eines Fragebogens

Nr	Sitzung	Inhalt	Deutsch	Englisch	Erläuterung
					zusammengehö Cronbachs $\alpha$ )
85	6	Faktorenanalysen	Strukturgleichungsmodell (SEM)	structural equation model (SEM)	Strukturgleichun sind Modelle über vielfältige Zusammenhäng mehrerer Variab Konfirmatorisch Faktorenanalyse der Regel teil von
86	6	Faktorenanalysen	Uniqueness	Uniqueness	Uniqueness ist 1 Kommunalität. V Variable viel Var die sie nicht mit Faktoren teilt, ist einzigartig.
74	6	Faktorenanalysen	explorative Faktorenanalyse	explorative factor analysis (EFA)	Es werden Faktore Fülle von Variablen (exploriert).

Nr	Sitzung	Inhalt	Deutsch	Englisch	Erläuterung
75	6	Faktorenanalysen	konfirmatorische Faktorenanalyse	confirmatory factor analysis (CFA)	Faktoren werden vorgegeben und ob die Daten sie (confirm).
80	6	Faktorenanalysen	latente Variablen	latent variable	Latente Variable solche, die nicht sondern von gemessenen Variablen gemessen werden
82	6	Faktorenanalysen	oblique Rotation	oblique rotation	Die Faktoren können korrelieren, stehen nicht ganz senkrecht zueinander.
83	6	Faktorenanalysen	orthogonale Rotation	orthogonal rotation	Die Faktoren stehen bei der Rotation senkrecht (orthogonal) zueinander. Die Korrelation zwischen den Faktoren ist

# Literatur

Song, Indeok, Robert Larose, Matthew S. Eastin, and Carolyn A. Lin. 2004. "Internet Gratifications and Internet Addiction: On the Uses and Abuses of New Media." *CyberPsychology & Behavior* 7 (4): 384–94. <https://doi.org/10.1089/cpb.2004.7.384>.