Sven Engesser

Empfehlungen zur Datenauswertung und Ergebnisdarstellung

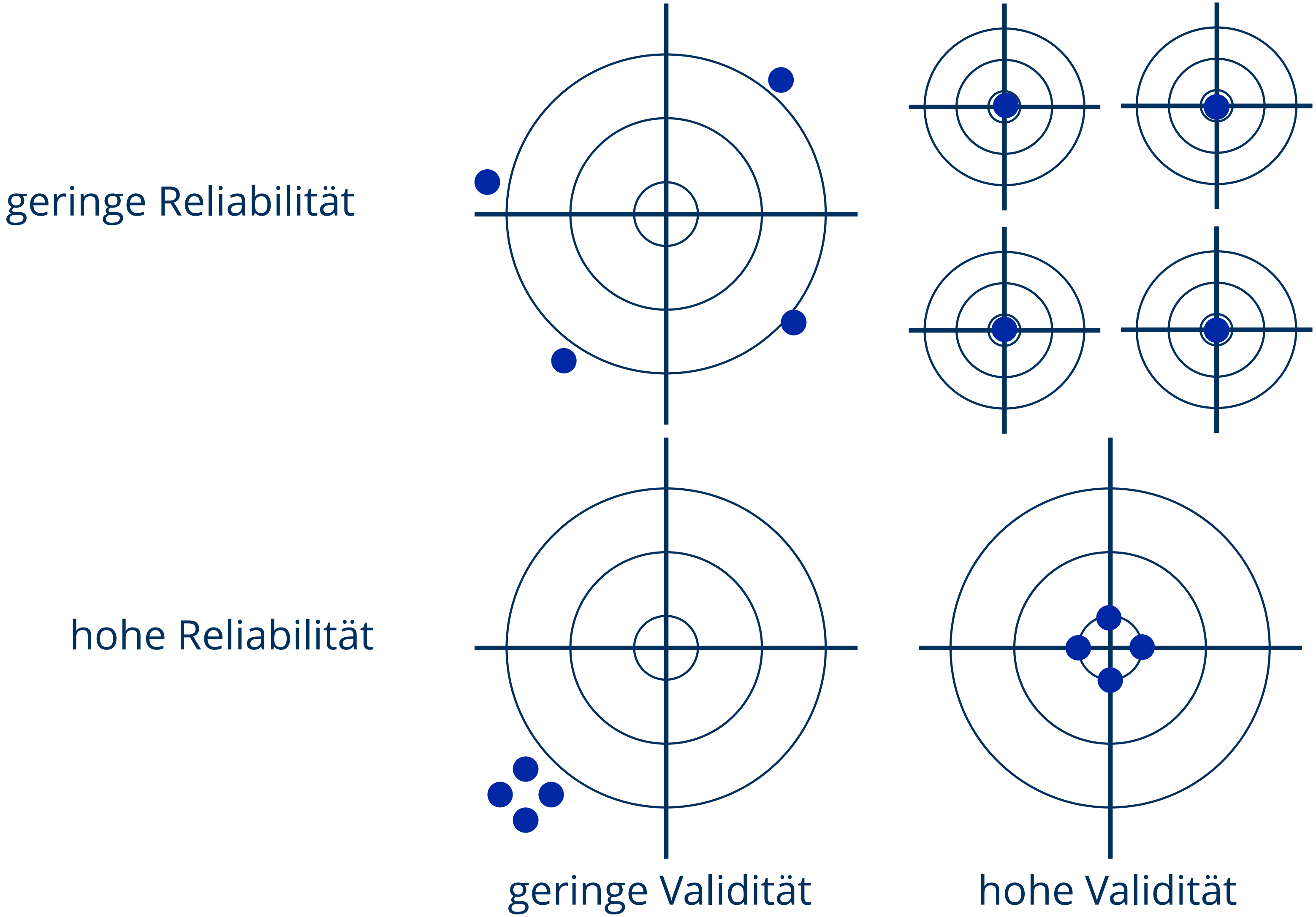
Gütekriterien der Messung

Validität und Reliabilität: Begriffsbestimmung

— **Validität** (Gültigkeit)
  – Def.: Das Instrument misst das, was es messen sollte.
  – Beispiele: Intelligenz, Pressefreiheit
  – Test: Vergleich der Instrumente

— **Reliabilität** (Zuverlässigkeit)
  – Def.: Das Instrument misst unter gleichen Bedingungen gleich.
  – Beispiele: Blutdruck, Blutalkoholkonzentration
  – Test: Vergleich der Messungen

Validität und Reliabilität: Abgrenzung



Dimensionen der Validität I

— **Inhaltsvalidität:**
Vollständigkeit der Messung

— **Kriteriumsvalidität:**
Empirische Zusammenhänge (zwischen Messwerten)

— **Konstruktvalidität:**
Theoretische Zusammenhänge (zwischen Konzepten)

(Schnell et al. 2008, S. 154; Rössler, 2010, S. 205)

Dimensionen der Validität II

- **Interne Validität**
  Rückführbarkeit auf Stimulus, Ausschluss von Störfaktoren

- **Externe Validität**
  Wirklichkeitstreue, Übertragbarkeit auf die Außenwelt

(Schnell et al. 2008, S. 219)

Datenaufbereitung

Datenkontrolle

- Häufigkeitsauszählungen
  - Werte außerhalb der Skalen („wild codes“)
  - Ausreißer (Problem bei vielen Verfahren)
  - Widersprüche zwischen Variablen
- Test auf Normalverteilung (Voraussetzung bei vielen Verfahren)
  - Histogramm
  - Kolmogorov-Smirnov-Test
- Festlegung fehlender Werte

Datenumwandlung

— Umcodieren (z. B. Invertierung, Kategorisierung)

— Berechnen (z. B. Indexbildung)

— Aggregieren (z. B. auf Beitragsebene)

— Mehrfachantworten

Reliabilitätstest

Reliabilitätstest: Richtlinien

— Inter-Coder-Reliabilität (mindestens zwei Codierer)

— Vor Beginn der Hauptcodierung

— Parallele Stichprobe

— Umfang von 30 – 50 Einheiten

— Randomisiert (oder zumindest sinnvoll quotiert)

— Einschluss des Zugriffskriteriums (Identifikation der Einheiten)

— Separate Darstellung für jede Variable

(Lombard et al., 2002; Neuendorf, 2002, S. 141-166; Rössler, 2010, S. 197-205)

Reliabilitätstest: Koeffizienten

- Prozentuale Übereinstimmung: PÜ = *Ü* / *n* > ,8 – ,9
- Cohens Kappa: κ = (Ü – *E*) / (1 – *E*) > ,6 – ,7
- Standardisierter Lotus: > ,6 – ,7

(Fretwurst, 2015; Lombard et al., 2002; Neuendorf, 2002, S. 141-166; Rössler, 2010, S. 204)

Inter-Coder-Reliabilität

Codierer	Beitrag	Variable 1	Variable 2	Variable 3
1	1	0	0	0
2	1	0	1	0
3	1	0	2	1
1	2	1	2	1
2	2	1	1	1
3	2	1	0	1
1	3	2	1	2
2	3	2	0	2
3	3	2	2	1

Inter-Coder-Reliabilität

Paar	**Variable 1**	**Variable 2**	**Variable 3**	***M***
1, 2	1,0	0	1,0	0,67
1, 3	1,0	0	0,33	0,44
2, 3	1,0	0	0,33	0,44
M	1,0	0	0,55	0,52

Indizes

Indizes: Spezifikation

— Bei **reflektiver** Spezifikation bestimmen die latenten Konstrukte die manifesten Variablen (z. B. Kundenzufriedenheit). Die Manifestationen sind korreliert und beliebig austauchbar.

— Bei **formativer** Spezifikation bestimmen die manifesten Variablen die latenten Konstrukte (z. B. sozio-ökonomischer Status). Die Manifestationen werden theoretisch festgelegt.

— Bei reflektiver Spezifikation sind **Eindimensionalität** (→ Faktoren-analyse) und **interne Konsistenz** (→ Cronbachs α) erforderlich.

— Bei formativer Spezifikation ist die **Validität** entscheidend.

(Eberl, 2004)

Indizes: Umwandlung

— Bei der **Minimum-Maximum-Normalisierung** wird die Skala (0 – 1) vereinheitlicht: $I_n = (I - I_{MIN}) / (I_{MAX} - I_{MIN})$

Bsp.: Wert 4 auf 5er-Skala: 0,75 = 4 - 1 / (5 - 1)

— Bei der **z-Standardisierung** werden Mittelwert (M = 0) und Standardabweichung (s = 1) vereinheitlicht: $I_z = (I - M_I) / s_I$

Bsp.: Wert 4 auf 5er-Skala (M = 3, s = 1): 1 = (4 - 3) / 1
(SPSS: Deskriptive Statistiken → Deskriptive Statistik)

Indizes: Berechnung

— **Summen-Indizes** sind sensibler, empfindlich gegenüber fehlenden Werten, kumulieren die Effekte und verändern die Skala (SPSS: Sum(Variable1, Variable 2, Variable 3))

— **Mittelwert-Indizes** sind robuster, weniger empfindlich gegenüber fehlenden Werten, reduzieren die Effekte und erhalten die Skala (SPSS: Mean(Variable1, Variable 2, Variable 3))

Univariate Analysen

Häufigkeiten: Richtlinien

— Bei **nominalen und ordinalen Variablen** werden **relative** Häufigkeiten (als Prozentsätze) und **absolute** Häufigkeiten (Fallzahlen) für die verschiedenen Ausprägungen angegeben.

— Metrische Variablen mit mehr als fünf Ausprägungen werden vor der Auszählung in der Regel angemessen **kategorisiert.**

— Bei Prozentsätzen werden **keine Dezimalstellen** angegeben.

— Im Idealfall wird bei relativen Häufigkeiten ein **Konfidenzintervall** (95 %) angegeben: (95 % KI [Untergrenze, Obergrenze])

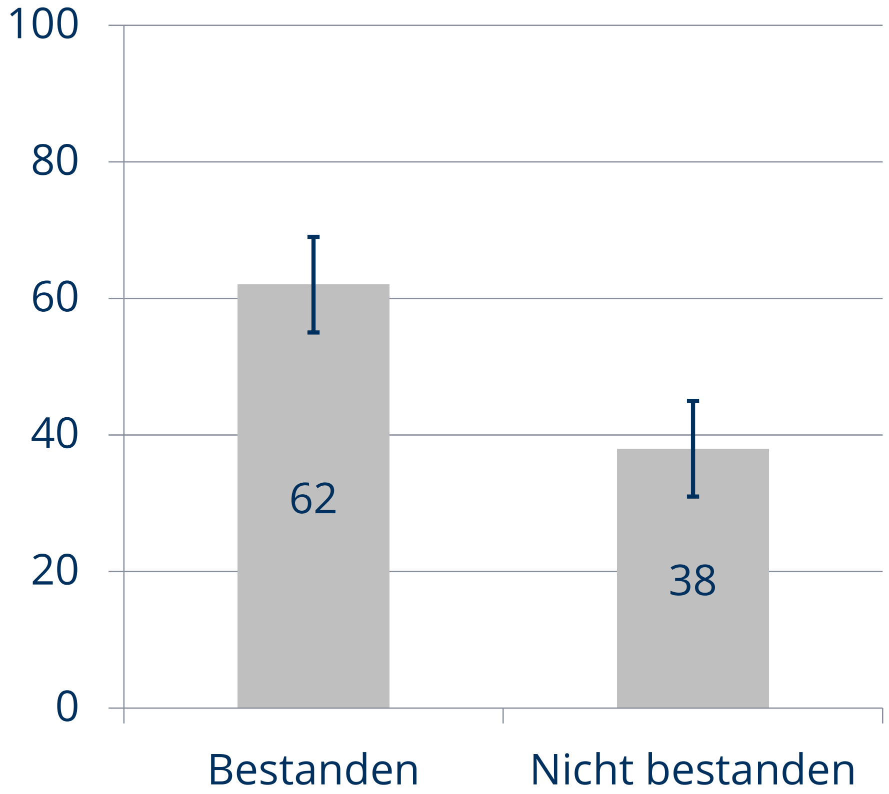
Häufigkeiten: Tabelle

Tabelle X: Ergebnis der Methoden-Klausur

Ergebnis	Prozent [95 % KI]	*n*
Bestanden	62 [55, 69]	115
Nicht bestanden	38 [31, 45]	70
Gesamt	100	185

Häufigkeiten: Säulendiagramm

Abbildung X: Ergebnis der Methoden-Klausur



Anmerkung: Prozentsätze, N = 185

Konfidenzintervalle (95 %) für Prozentsätze

— Formel:

$$P \pm 1{,}96 * \sqrt{\frac{P * (100 - P)}{N}}$$

— Beispiel:

$$62 \pm 1{,}96 * \sqrt{\frac{62 * 38}{185}} = 55,69$$

— Berechnung und Darstellung:

– Excel: Diagrammtools → Layout → Fehlerindikatoren

– SPSS: Balken → Optionen → Fehlerbalken

Mittelwert: Richtlinien

- Bei **(quasi-)metrischen Variablen** wird der **arithmetische Mittelwert (*M*)** mit **Standardabweichung (*s*)** und **Fallzahl (*n*)** angegeben.

- Bei Mittelwerten werden **zwei Dezimalstellen** angegeben.

- Im Idealfall wird bei Mittelwerten ein **Konfidenzintervall** (95 %) angegeben: (95 % KI [Untergrenze, Obergrenze])

- Bei nominalen Variablen wird der **Modalwert *Mo*** (häufigster Wert) und bei ordinalen Variablen der **Median *Mdn*** (Zentralwert) angeben.

Mittelwert: Tabelle

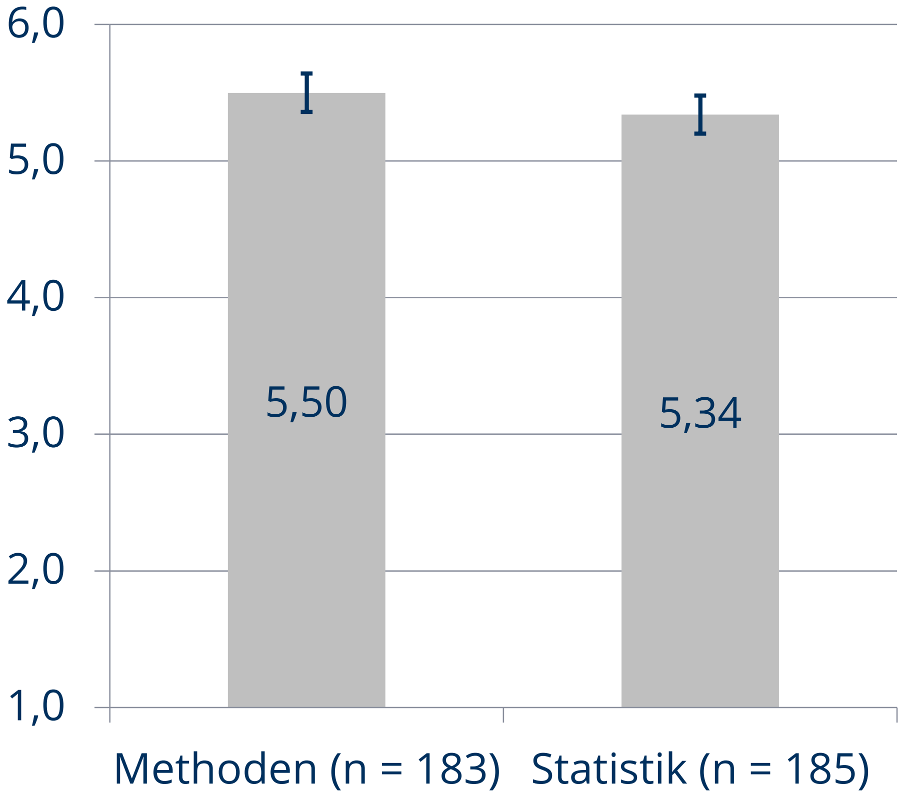
Tabelle X: Klausurnote

Klausur	*M* [95 % KI]	*s*	*N*
Methoden	5,50 [5,36, 5,64]	1,00	183
Statistik	5,34 [5,20, 5,48]	0,95	185

Anmerkung: Skala von 1 (= sehr schlecht) bis 6 (= sehr gut)

Mittelwert: Säulendiagramm

Abbildung X: Klausurnote



Anmerkung: Mittelwerte, Skala von 1 (= sehr schlecht) bis 6 (= sehr gut)

Konfidenzintervalle (95 %) für Mittelwerte

— Formel: $M \pm 1{,}96 * \left(\frac{s}{\sqrt{n}}\right)$

— Beispiel: $5{,}34 \ \pm 1{,}96 * \left(\frac{0{,}95}{\sqrt{185}}\right) = 5{,}20, 5{,}48$

— Berechnung und Darstellung:

- SPSS: Deskriptive Statistiken → Explorative Datenanalyse
- Excel – Berechnung: KONFIDENZ.NORM
- Excel – Darstellung: Diagrammtools → Layout → Fehlerindikatoren

Bivariate Analysen

Kreuztabelle: Richtlinien

— Bei **nicht-metrischen aV** werden relative und absolute Häufigkeiten in einer Kreuztabelle dargestellt.

— Die **Fallzahl** sollte in jeder Zelle ausreichend sein ($n \geq 5$)

— Zur Vereinfachung können aV und uV dichotomisiert werden.

— Die aV wird in den Zeilen und die **uV in den Spalten** dargestellt (SPSS: Prozentwerte → Spaltenweise)

— Optional werden die **Erwartungswerte (E)** angeben.
(SPSS: Häufigkeiten → Erwartet)

— Die Kennwerte des **χ2-Tests** werden wie folgt angegeben:
$\chi 2$ (df, N = 185) = 1,47, p = ,225

Kreuztabelle: Darstellung

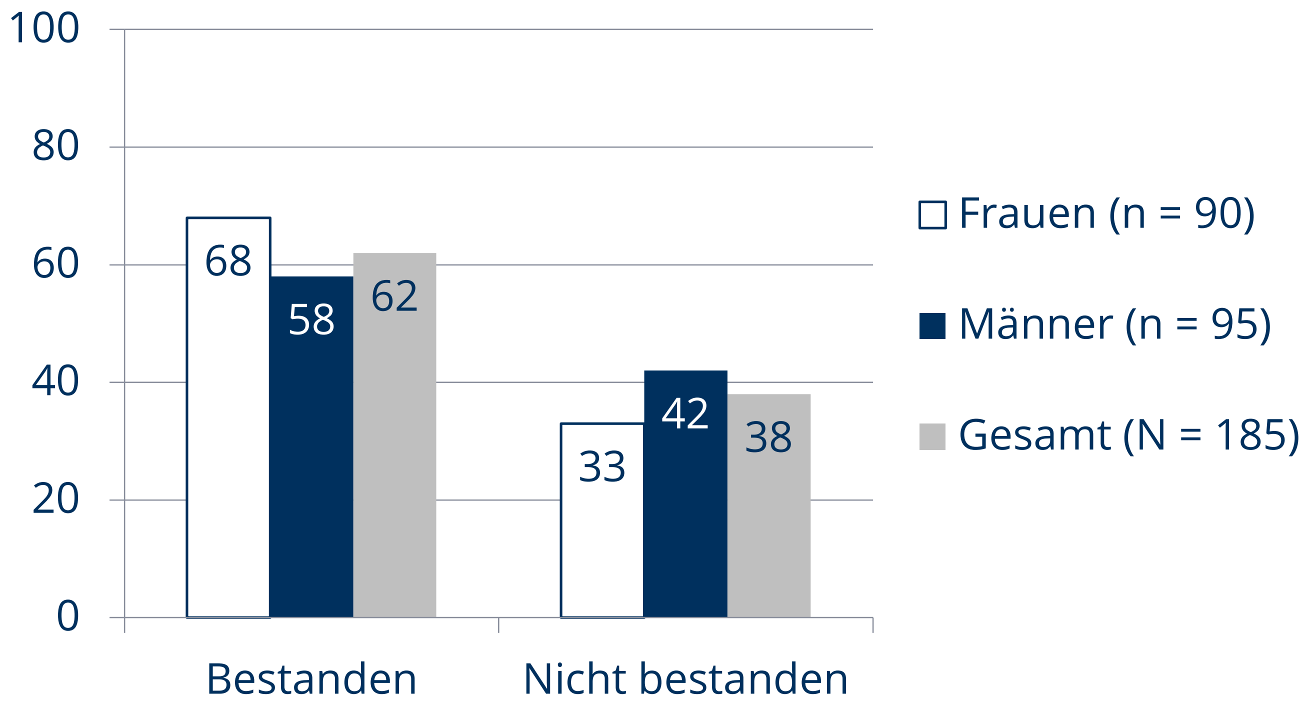
Tabelle X: Klausurergebnis nach Geschlecht

Klausur-ergebnis	Geschlecht				Gesamt	
	Weiblich		Männlich			
	Prozent	*n (E)*	Prozent	*n (E)*	Prozent	*n*
Bestanden	68	60 (56)	58	55 (59)	62	115
Nicht bestanden	33	30 (34)	42	40 (36)	38	70
Gesamt	100	90	100	95	100	185

Anmerkung: χ^2 (1, N = 185) = 1,47, p = ,225

Kreuztabelle: Säulendiagramm

Abbildung X: Klausurergebnis nach Geschlecht



Anmerkung: Prozentsätze, χ^2 (1, N = 185) = 1,47, p = ,225

Kreuztabelle: Erwartungswerte

Klausur-ergebnis	Geschlecht		
	Weiblich	Männlich	Gesamt
Bestanden	115 * 90 / 185 = 56	115 * 95 / 185 = 59	115
Nicht bestanden	70 * 90 / 185 = 34	70 * 95 / 185 = 36	70
Gesamt	90	95	185

Kreuztabelle: Freiheitsgrade

Formel: *df* = (Zeilenzahl - 1) * (Spaltenzahl - 1)

Freiheitsgrade (*df*)		χ^2-Wert	
1	3,84	6,64	10,83
2	5,99	9,21	13,82
3	7,82	11,34	16,27
4	9,49	13,28	18,47
…	…	…	…
Irrtumswahrscheinlichkeit (*p*)	,05	,01	,001

Kreuztabelle: Ausgabe in SPSS

Objektivität * Zeitungs-ID Kreuztabelle

			Zeitungs-ID		
			20min	TagesAnzeiger	Gesamt
Objektivität	Objektivität - Es werden nur Fakten präsentiert oder Fakten und Meinung sind klar getrennt	Anzahl	33	23	56
		Erwartete Anzahl	27.6	28.4	56.0
		% innerhalb von Zeitungs-ID	97.1%	65.7%	81.2%
	Keine Objektivität - Fakt und Meinung sind nicht auseinanderzuhalten	Anzahl	1	12	13
		Erwartete Anzahl	6.4	6.6	13.0
		% innerhalb von Zeitungs-ID	2.9%	34.3%	18.8%
Gesamt		Anzahl	34	35	69
		Erwartete Anzahl	34.0	35.0	69.0
		% innerhalb von Zeitungs-ID	100.0%	100.0%	100.0%

Kreuztabelle: χ^2-Test in SPSS

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	11.081[a]	1	.001	$N > 60$	
Kontinuitätskorrektur[b]	9.126	1	.003	$20 < N < 60$	
Likelihood-Quotient	12.752	1	.000		
Exakter Test nach Fisher				.001	$N < 20$ .001
Zusammenhang linear-mit-linear	10.921	1	.001		
Anzahl der gültigen Fälle	69				

a. 0 Zellen (0.0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 6.41.

b. Wird nur für eine 2x2-Tabelle berechnet

Mittelwertvergleich: Richtlinien

- Bei Unterschiedshypothesen, **(quasi-)metrischen aV und dichotomen uV** wird ein Mittelwertvergleich durchgeführt.
- Zu jedem Mittelwert werden **Standardabweichung** (*s*) und **Fallzahl** (*n*) angegeben.
- Zu jedem Mittelwertvergleich wird ein **t-Test** (für unabhängige Stichproben) durchgeführt (Levene-Test beachten!)
- Die Kennwerte des Mittelwertvergleichs werden angegeben: $t(df)$ = 2,14, p = ,033
- Tabellarisch wird die aV in Zeilen und die **uV in Spalten** dargestellt.

Mittelwertvergleich: Darstellung

Tabelle X: Klausurnote nach Geschlecht

	Geschlecht								
	Weiblich			Männlich			Gesamt		
	M	*s*	*n*	*M*	*s*	*n*	*M*	*s*	*N*
Klausurnote	5,50	0,90	90	5,20	1,00	95	5,34	0.95	185

Anmerkung: $t(183) = 2{,}14$, $p = {,}034$

Mittelwertvergleich: *t*-Wert

$$t = \frac{M_1 - M_2}{\sqrt{\left(\frac{n_1 + n_2}{n_1 * n_2}\right) * \left(\frac{(n_1 - 1) * s_1{}^2 + (n_2 - 1) * s_2{}^2}{n_1 + n_2 - 2}\right)}}$$

Mittelwertvergleich: Freiheitsgrade

Formel: $df = n_1 + n_2 - 2$

Freiheitsgrade (*df*)		*t*-Wert	
68	± 2,00	± 2,65	± 3,44
…	…	…	…
183	± 1,97	± 2,60	± 3,34
…	…	…	…
1000	± 1,96	± 2,58	± 3,10
Irrtumswahrscheinlichkeit (*p*)	,05	,01	,001

Mittelwertvergleich: Ausgabe in SPSS

Gruppenstatistiken

	Zeitungs-ID	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Verständlichkeit - Flesch-Index	20min	35	51.34	12.049	2.037
	TagesAnzeiger	35	46.23	10.494	1.774

Mittelwertvergleich: *t*-Test in SPSS

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-		
		F	Signifikanz	T	df	Sig. (2-seitig)
Verständlichkeit - Flesch-Index	Varianzen sind gleich	.850	.360	$p > ,05$	68	.063
	Varianzen sind nicht gleich			$p < ,05$	66.742	.063

Einfaktorielle Varianzanalyse: Richtlinien I

— Bei Unterschiedshypothesen, **(quasi-)metrischen aV und polytomen uV** wird eine einfaktorielle Varianzanalyse (ANOVA) durchgeführt.

— Zur Vermeidung einer **inflationären Irrtumswahrscheinlichkeit** (Bei drei Vergleichen steigt diese von 5 % auf 14 %.)

— Die Kennwerte des **F-Tests** werden wie folgt angegeben:
$F(df_{zwischen}, df_{innerhalb}) = 2{,}14$, $p = {,}033$

— Optional wird die **Effektstärke η²** angeben
(SPSS: ALM → Univariat → Optionen)

Einfaktorielle Varianzanalyse: Richtlinien II

— Es wird ein **Levene-Test** durchgeführt
(SPSS: Optionen → Test auf Homogenität der Varianzen)

— Bei Varianzgleichheit wird als Post-Hoc-Test **Gabriel**,
bei Varianzungleichheit **Games-Howell** durchgeführt (p = ,05).

— Signifikant unterschiedliche Mittelwerte (p < ,05) werden mit unterschiedlichen hochgestellten Buchstaben markiert.

— Tabellarisch wird die aV in Zeilen und die **uV in Spalten** dargestellt.

(Field, 2013)

Einfaktorielle Varianzanalyse: Darstellung

Tabelle X: Klausurnote nach Nationalität

	Nationalität											
	AT			DE			CH			Gesamt		
	M	*s*	*n*	*M*	*s*	*n*	*M*	*s*	*n*	*M*	*s*	*N*
Note	5,00[a]	0,90	90	4,90[a]	1,00	95	5,50[b]	1,00	100	5,14	0,95	285

Anmerkung: $F(2, 185) = 2{,}14$, $p = {,}033$; unterschiedliche Buchstaben markieren signifikant unterschiedliche Mittelwerte ($p < {,}05$, Gabriel)

Korrelation: Richtlinien

- Bei Zusammenhangshypothesen und **ordinalen Variablen** wird die Rangkorrelation (Kendalls τ) berechnet.
- Bei Zusammenhangshypothesen und **(quasi-)metrischen Variablen** wird die Produkt-Moment-Korrelation (Pearsons *r*) berechnet.
- Korrelationskoeffizienten erfordern zwei Dezimalstellen.
- Für jeden Korrelationskoeffizienten wird ein *t*-Test durchgeführt und angegeben: $t(df) = -4.31$, $p < ,001$
- Signifikante Korrelationskoeffizienten werden markiert:
  $^{+}p < ,1$, $^{*}p < ,05$, $^{**}p < ,01$, $^{***}p < ,001$

Korrelation: Darstellung

Tabelle X: Zusammenhänge zwischen Klausurnoten

	Methoden	Statistik	Einführung
Methoden		,60*	,30*
Statistik			,70*

Anmerkung: Pearsons *r; n* ≥ 90; markierte Werte sind signifikant (*p < ,05)

Korrelation: *t*-Wert

$$t = r * \sqrt{\frac{n-2}{1-r^2}}$$

Korrelation: Freiheitsgrade

Formel: $df = n - 2$

Freiheitsgrade (*df*)		*t*-Wert	
68	± 2,00	± 2,65	± 3,44
…	…	…	…
183	± 1,97	± 2,60	± 3,34
…	…	…	…
1000	± 1,96	± 2,58	± 3,10
Irrtumswahrscheinlichkeit (*p*)	,05	,01	,001

Korrelation: Ausgabe in SPSS

Korrelationen

		Einfluss	Betroffenheit	Aktualität
Einfluss	Korrelation nach Pearson	1	.351**	.139
	Signifikanz (2-seitig)		.003	.264
	N	70	70	66
Betroffenheit	Korrelation nach Pearson	.351**	1	-.040
	Signifikanz (2-seitig)	.003		.749
	N	70	70	66
Aktualität	Korrelation nach Pearson	.139	-.040	1
	Signifikanz (2-seitig)	.264	.749	
	N	66	66	66

**. Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Bivariate Auswertungsverfahren im Überblick

Hypothese	*aV*	*uV*	Auswertungsverfahren
Unterschied	Nicht-Metrisch	Dichotom	Kreuztabelle
	(Quasi-)Metrisch	Dichotom	Mittelwertvergleich
	(Quasi-)Metrisch	Polytom	Varianzanalyse
Zusammenhang	Ordinal		Rangkorrelation
	Metrisch		Produkt-Moment-Korrelation

(Diekmann, 2010, S. 703)

Multivariate Analysen

Lineare Regression: Voraussetzungen

— Bei linearen Zusammenhangshypothesen, **einer (quasi-) metrischen aV und mehreren (quasi-)metrischen uV** wird eine lineare Regressionsanalyse (OLS) durchgeführt.

— Zur **Vermeidung von Scheinkorrelationen.**

— Das Verhältnis zwischen Fällen (*N*) und Regressoren (*k*) sollte angemessen sein: $N / k > 30 – 50$

— Alle theoretisch und empirisch relevanten Regressoren müssen berücksichtigt werden.

— Regressoren dürfen nicht zu stark korrelieren (Toleranzen > ,1 – ,25) (sonst zuvor explorative Faktorenanalyse)

Lineare Regression: Angaben

— Für die Regression wird das **Bestimmtheitsmaß (R^2)** angegeben

— Für die Regression wird ein ***F*-Test** durchgeführt:
$F(k, N - k) = 42{,}64$, $p < {,}001$

— Für jeden Regressor werden **nicht standardisierte (*B*)** und **standardisierte (beta)** Koeffizienten angegeben: $\beta_k = (B_k * s_k) / s_{aV}$

— Für jeden Regressor wird ein ***t*-Test** durchgeführt: $df = n - k - 1$

— Im Idealfall wird für jeden Regressionskoeffizienten ein Konfidenz-intervall (95 %) angegeben: (95 % KI [Untergrenze, Obergrenze])

— Signifikante Regressionskoeffizienten werden markiert:
$^{+}p < {,}1$, $^{*}p < {,}05$, $^{**}p < {,}01$, $^{***}p < {,}001$

Lineare Regression: Tabelle

Tabelle X: Regressionsmodel zum Ergebnis der Methoden-Klausur

Regressor	*B* [95 % KI]	β [95 % KI]
Sitzungen	,30* [,25, ,35]	,21* [,11, ,32]
Vorbereitungszeit	,20* [,30, ,10]	,11* [,19, ,03]

Anmerkung: R^2 = ,12, F(2, 183) = 42,64, p < ,001; markierte Werte sind signifikant (*p < ,05)

Explorative Faktorenanalyse: Voraussetzungen

— Zur **Reduktion der Variablen** und **Aufdeckung von Dimensionen** wird eine explorative Faktorenanalyse (EFA) durchgeführt

— Das Verhältnis zwischen Fällen (N) und Items (k) sollte angemessen sein: N > 30 – 50, N / k > 3 – 5

— Items müssen korrelieren (KMO ≥ ,5 – ,6)

— Items sollten metrisch skaliert sein (sonst Korrespondenzanalyse)

— Items werden einer z-Standardisierung unterzogen

(MacCallum et al., 1999)

Explorative Faktorenanalyse: Extraktionsmethoden

— **Hauptkomponentenanalyse (PCA):** Faktoren als Sammelbegriff, Items können durch Faktoren vollständig erklärt werden ($h2$ = 1)

— **Hauptachsenanalyse (PFA):** Faktoren als Ursache, Items können durch Faktoren nicht vollständig erklärt werden

Explorative Faktorenanalyse: Extraktionskriterien

— **Kaiser-Kriterium:** Alle Faktoren mit Eigenwert $\lambda \geq 1$

— **Ellbogen-Kriterium:** Alle Faktoren links des Knicks auf einer Abbildung der abnehmenden Eigenwerte (Scree-Plot)

— **Varianz-Kriterium:** Alle Faktoren, die zusammen mindestens 50 – 90 Prozent der Gesamtvarianz erklären

Explorative Faktorenanalyse: Rotationsmethoden

— **Rechtwinklige Rotation (Varimax):** Faktoren nicht korreliert, zur Weiterverwendung in Regressions- und Clusteranalysen

— **Schiefwinklige Rotation (oblique):** Faktoren leicht korreliert, Option auf Faktorenanalyse zweiter Ordnung

Explorative Faktorenanalyse: Gütekriterien

— Erklärte Gesamtvarianz ≥ 50 %

— Mindestens 2 – 3 Items pro Faktor

— Keine gleichwertigen Mehrfachladungen (Differenz *a* > ,3)

— Faktorladung: *a* ≥ ,4 – ,5

— Kommunalität: *h2* ≥ ,6

— Cronbachs α ≥ ,6 – ,7

(MacCallum et al., 1999)

Explorative Faktorenanalyse: Angaben

- Für die Faktorenanalyse werden **KMO**, **Anzahl der Faktoren**, **aufgeklärte Gesamtvarianz** (in %) angegeben.
- Auch werden **Extraktionsmethode**, **Extraktionskriterium** (ggf. Scree-Plot) und **Rotationsmethode** angegeben und begründet.
- Für Faktoren werden **Eigenwert** (λ), **aufgeklärte Varianz** (in %) und **interne Konsistenz** (Cronbachs α) angegeben
- Für Items werden **Faktorladung** (*a*) und **Kommunalität** (*h2*) angegeben.

Explorative Faktorenanalyse: Tabelle

Tabelle X: Ergebnis der Faktorenanalyse

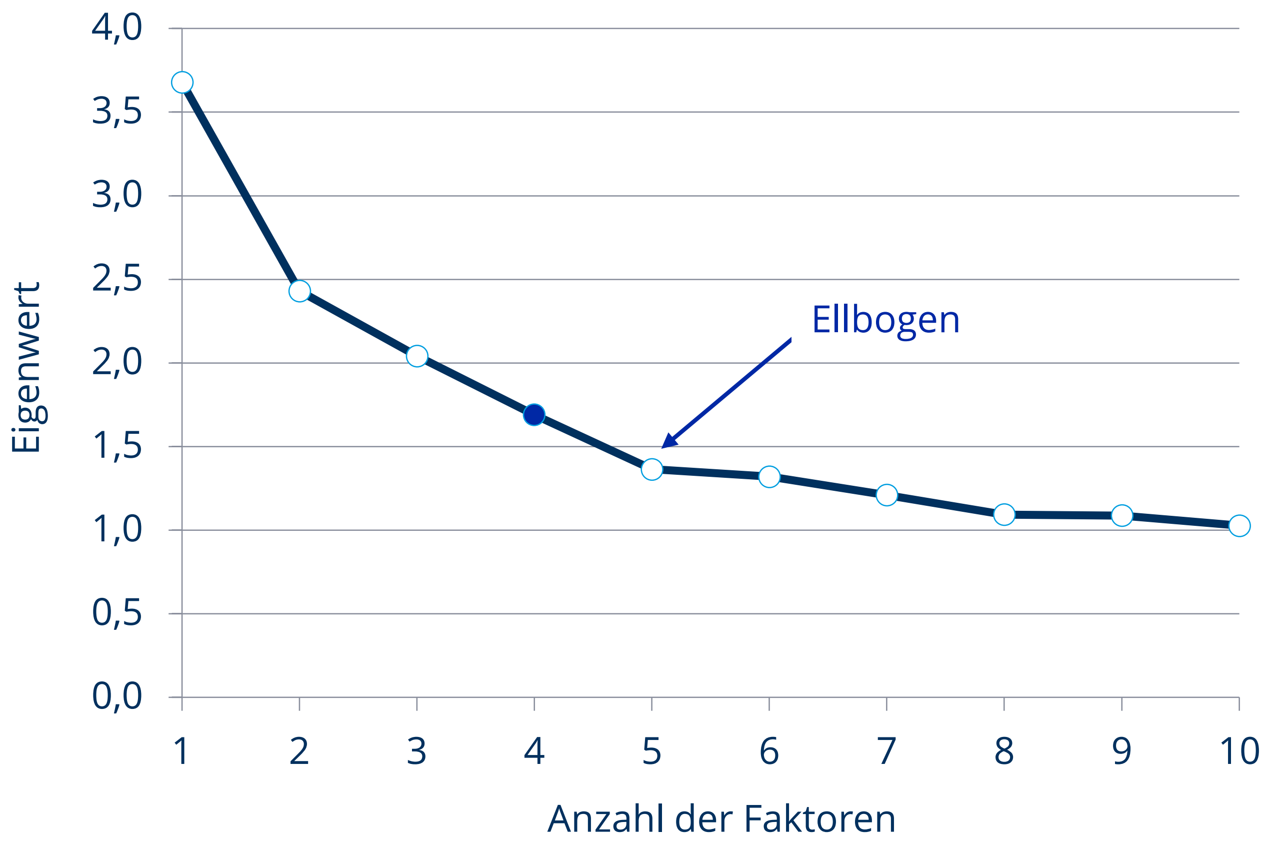
Items	Faktor 1	Faktor 2	Faktor 3	*h2*
Item 1	,83			,74
Item 2	,81			,71
Item 3	,74			,58
Item 4	,67			,67
Item 5	,67			,57
Item 6	,66			,71
Item 7		,85		,80
Item 8		,83		,74
Item 9		,74		,66
Item 10			,95	,92
Eigenwert	4,89	1,10	1,02	
Erklärte Varianz (%)	34	26	11	
Cronbachs α	,88	,80		
Erklärte Gesamtvarianz (%)		71		

grenzwertig niedrig

zu wenige Items

Explorative Faktorenanalyse: Scree-Plot

Abbildung X: Scree-Plot der Faktorenanalyse



Hierarchische Clusteranalyse: Voraussetzungen

— Zur **Klassifikation der Fällen** wird eine hierarchische Clusteranalyse durchgeführt

— Merkmale sollten **gleich skaliert** sein

— Merkmale sollten **nicht zu stark korrelieren** (sonst zuvor explorative Faktorenanalyse)

— Merkmale werden einer *z*-Standardisierung unterzogen

Hierarchische Clusteranalyse: Richtlinien I

— Als Algorithmus wird **Ward** ausgewählt.

— Als Maß wird die **quadrierte Euklidische Distanz** ausgewählt.

— Zur Bestimmung der Clusterzahl dient das **Ellbogen-Kriterium** und bei ausreichender Übersichtlichkeit das **Dendogramm**

— Optional wird der **Test von Mojena** durchgeführt.

(Milligan & Cooper, 1985; Mojena, 1977)

Hierarchische Clusteranalyse: Richtlinien II

- Als Validierung wird eine ***K*-Means-Clusteranalyse** (quasi-metrisches Niveau erforderlich!) durchgeführt und die auftretenden Unterschiede werden diskutiert

- Zur Benennung, Beschreibung und Interpretation der Cluster wird eine **einfaktorielle Varianzanalyse** der Clustervariablen mit der Clusterzugehörigkeit als Faktor durchgeführt.

- Zur Bestimmung der Cluster-Homogenität wird für jedes Merkmal ein ***F*-Wert** berechnet: $F = s_{cluster} / s_{gesamt}$ (möglichst $F < 1$)

(Milligan & Sokal, 1980)

Hierarchische Clusteranalyse: Tabelle

Tabelle X: Fusionsstufen der Clusteranalyse

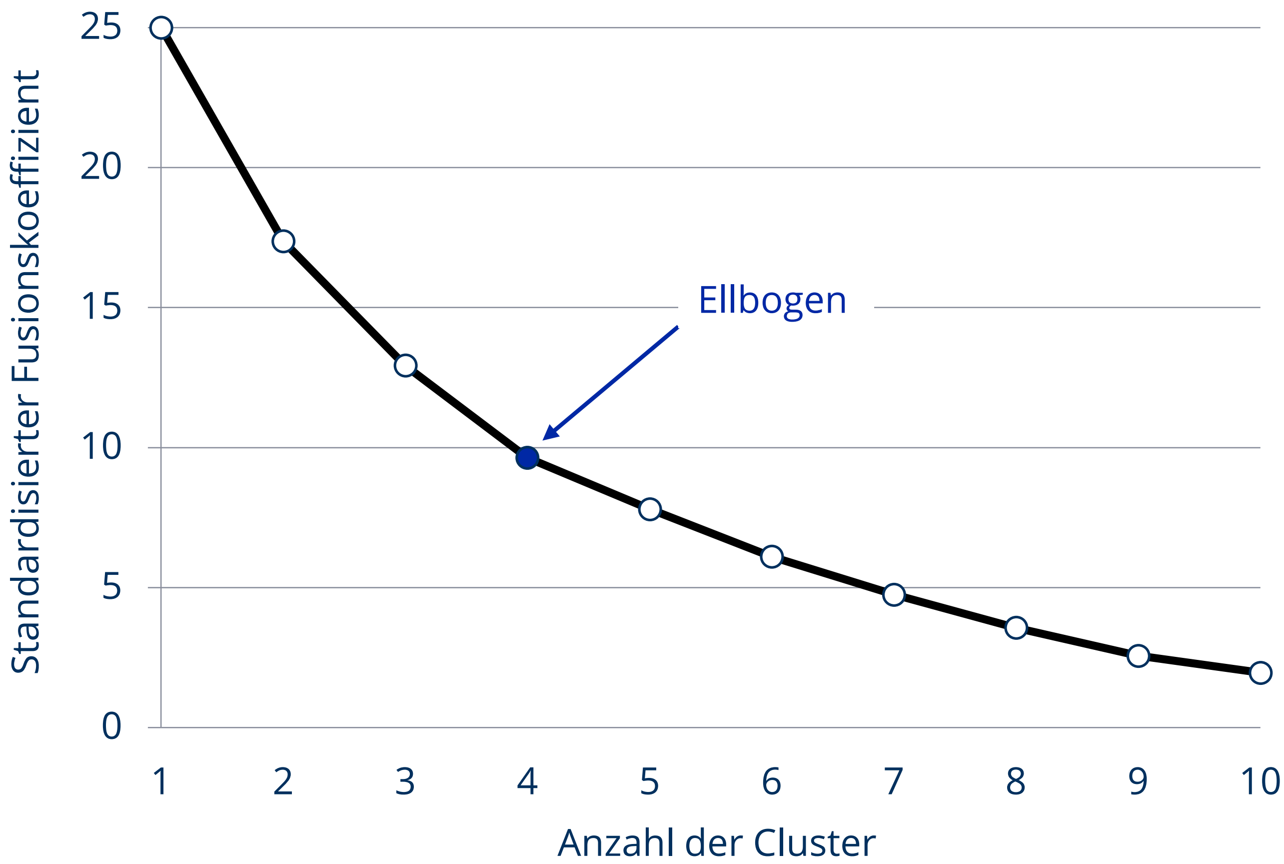
Anzahl Cluster	Fusionskoeffizient (*FK*)	*ΔFK*	FK_z(Mojena)
16	0,39	0,48	-0,82
15	0,87	0,72	-0,81
14	1,59	1,09	-0,78
13	2,68	1,33	-0,74
12	4,01	1,42	-0,69
11	5,43	2,14	-0,64
10	7,57	2,31	-0,56
9	9,88	3,83	-0,48
8	13,70	4,61	-0,34
7	18,31	4,91	-0,17
6	23,23	6,89	0,01
5	30,12	7,18	0,26
4	37,30	**12,23**	0,53
3	49,53	16,77	0,97
2	66,29	29,71	**1,59**
1	96,00		2,67

Zunahme

≥ Schwellenwert 1,25

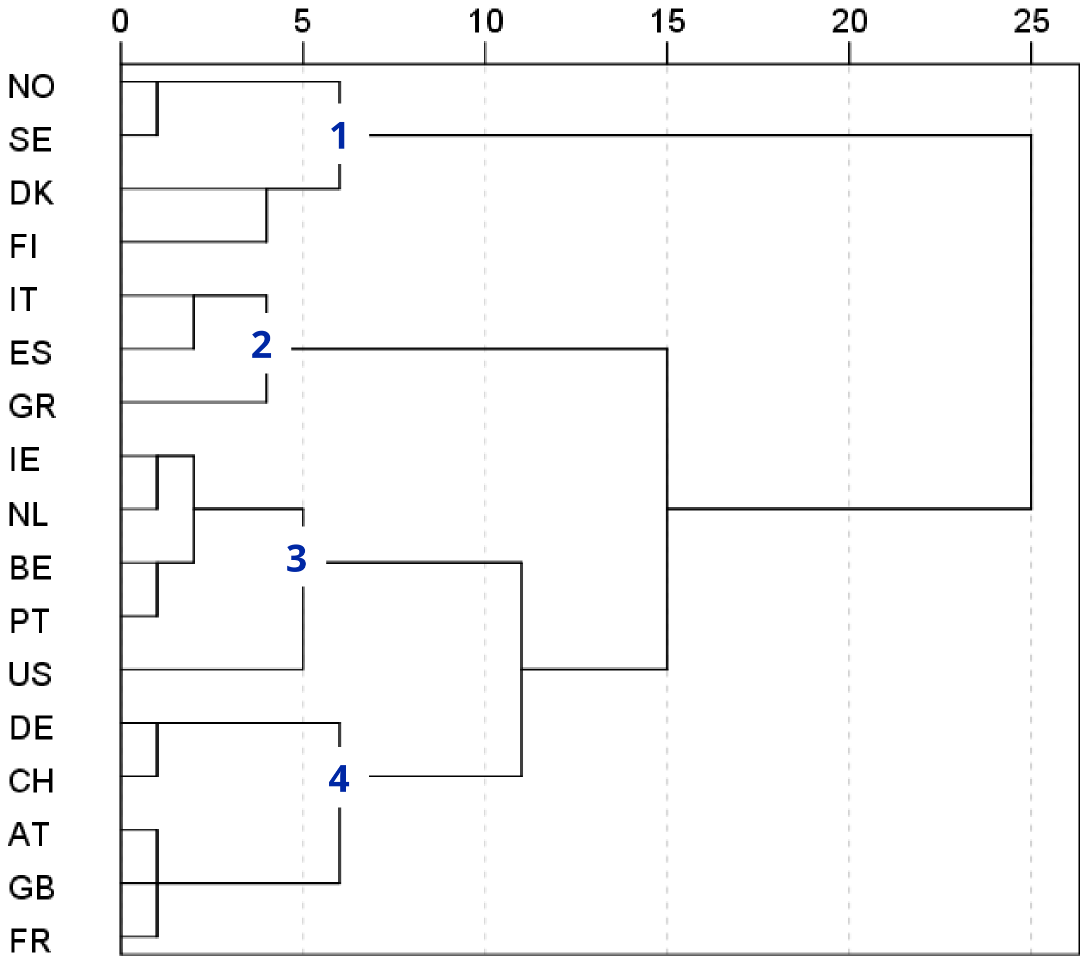
Hierarchische Clusteranalyse: Scree-Plot

Abbildung X: Scree-Plot der Clusteranalyse



Hierarchische Clusteranalyse: Dendogramm

Abbildung X: Dendogramm der Clusteranalyse



Hierarchische Clusteranalyse: Test von Mojena

— z-standardisierter Fusionskoeffizient: $FK_z = (FK - M_{FK}) / s_{FK}$

— Schwellenwert: $FK_z \geq 1{,}25 – 2{,}75$