Varimax Rotation & Kaiser Criterion

A step-by-step visual companion to Lesson 3

Prof. Dr. Jörg Osterrieder

School of Engineering and Computer Sciences

Part 1

Rotating factors to read them

A visual walk-through in 14 steps

The problem — messy loadings

Raw factor loadings often split across both factors — hard to interpret.

Item	Factor 1	Factor 2
Q1	+0.668	+0.529
Q2	+0.598	+0.541
Q3	+0.614	+0.430
Q4	-0.418	+0.684
Q5	-0.348	+0.672
Q6	-0.488	+0.696

- Q1–Q6 each load on BOTH F1 and F2.
- Some items load negatively on F1 — this is normal in real data.
- Which factor do they belong to?

We need a clearer picture — that's what rotation gives us.

After Varimax — the same 6 items

Varimax rotates the axes to push each item toward ONE factor.

Before

Item	Factor 1	Factor 2
Q1	+0.668	+0.529
Q2	+0.598	+0.541
Q3	+0.614	+0.430
Q4	-0.418	+0.684
Q5	-0.348	+0.672
Q6	-0.488	+0.696

After Varimax

Item	Factor 1	Factor 2
Q1	+0.850	+0.045
Q2	+0.800	+0.096
Q3	+0.750	-0.004
Q4	+0.054	+0.800
Q5	+0.104	+0.749
Q6	+0.005	+0.850

- Each item now loads strongly on ONE factor, near zero on the other (real Varimax output, angle $= 35.32^\circ$).
- Row sums of squares (h^2) didn't change — only the split.

Same items, same communalities — but a much cleaner picture.

What stayed, what changed

Stayed constant: communalities h_i^2

Item	h^2 before	h^2 after	match?
Q1	0.725	0.725	✓
Q2	0.650	0.650	✓
Q3	0.563	0.563	✓
Q4	0.643	0.643	✓
Q5	0.573	0.573	✓
Q6	0.723	0.723	✓

Changed: per-factor SS and V criterion

Quantity	Before	After	Shift
SS(F1)	1.714	1.940	↑
SS(F2)	2.161	1.935	↓
Total SS	3.875	3.875	=
V	0.026	0.207	⇑ **7.9**×

- Communalities h_i^2 are **rotation-invariant**.
- Per-factor SS redistributes; total SS unchanged; **V increases**.

Rotation moves loadings between factors — but cannot create or destroy variance.

The rotation matrix — what Varimax actually computed

A 35.32° rotation: every item's coordinates are multiplied by R.

Rotation matrix R

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

$$\theta = +35.32^\circ$$

$$\cos\theta = 0.8159, \quad \sin\theta = 0.5781$$

$$R = \begin{pmatrix} +0.8159 & -0.5781 \\ +0.5781 & +0.8159 \end{pmatrix}$$

Worked example for Q1

$$\text{Q1 raw} = (+0.6676,\ +0.5285)$$

$$\text{F1}_{\text{new}} = 0.6676 \times 0.8159 + 0.5285 \times 0.5781 = 0.5447 + 0.3055 = \mathbf{0.8503}$$

$$\text{F2}_{\text{new}} = 0.6676 \times (-0.5781) + 0.5285 \times 0.8159 = -0.3860 + 0.4312 = \mathbf{0.0452}$$

$$\text{Q1 rotated} = (+0.8503,\ +0.0452)$$

- R is orthogonal: $R^\top R = I$ (preserves h^2).
- Repeat for Q2–Q6: same matrix, six row-vector multiplications.

The rotation matrix IS what Varimax found — it's a single 35.32° turn applied to all items.

A different rotation — not Varimax

A -10° rotation also keeps h^2 fixed — but doesn't simplify.

-10° rotation (NOT Varimax)

Item	Factor 1	Factor 2
Q1	+0.566	+0.636
Q2	+0.495	+0.636
Q3	+0.530	+0.530
Q4	-0.530	+0.601
Q5	-0.460	+0.601
Q6	-0.601	+0.601

V before $= 0.026$

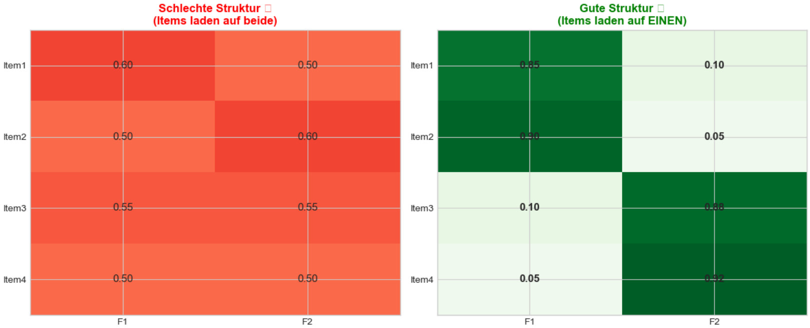
V at $-10^\circ = 0.004$

V Varimax $= 0.207$

- Loadings still split across both factors — worse than the start, not better.
- -10° is one arbitrary choice — V *drops* $6.5\times$ from 0.026 to 0.004 (wrong direction).

Any orthogonal rotation preserves h^2. Only Varimax maximizes simple structure.

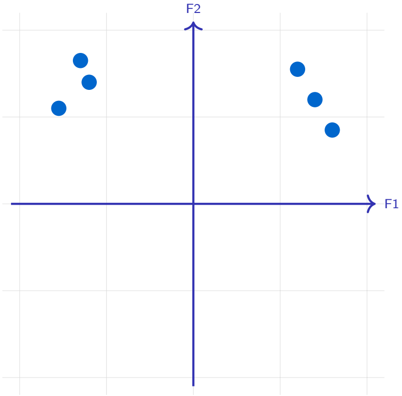
Goal: each item loads strongly on ONE factor, near zero on the other.



- Q1–Q3 on F1.
- Q4–Q6 on F2.

This “simple structure” is what we aim for.

The data stays still. We turn the axes.



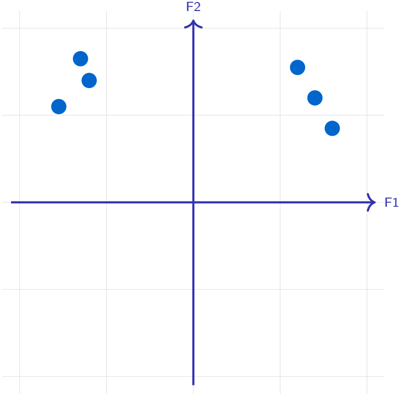
angle = 0°

- Dots (loadings) don't move.
- Axes F1, F2 look for a better orientation.

Rotation = changing our viewpoint, not the data.

Rotation step 1 — 0 degrees

Starting position. 0 degrees.



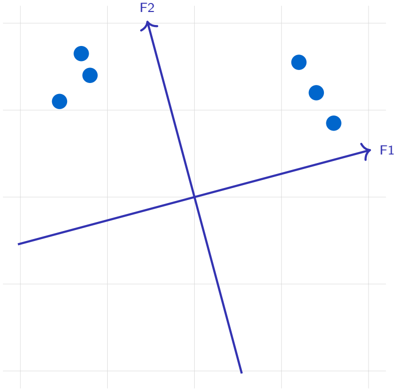
angle = 0°

- Dots spread across the plane.
- No axis captures them well.

Start here. No rotation yet.

Rotation step 2 — 15 degrees

Turn 15 degrees. . .



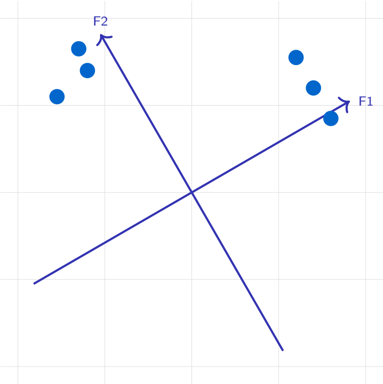
angle = 15°

- Dots unchanged.
- Getting warmer?

Still searching.

Rotation step 3 — 30 degrees

Turn 30 degrees. . .



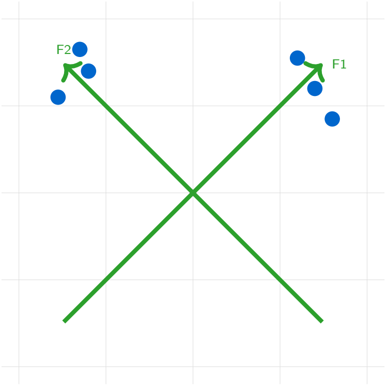
angle = 30°

- Dots on the plane are fixed.
- F1 is nearly aligned with a dot cluster.

Closer — notice dots lining up with F1.

Rotation step 4 — 45 degrees (the sweet spot)

At 45 degrees each dot sits (nearly) on ONE axis.



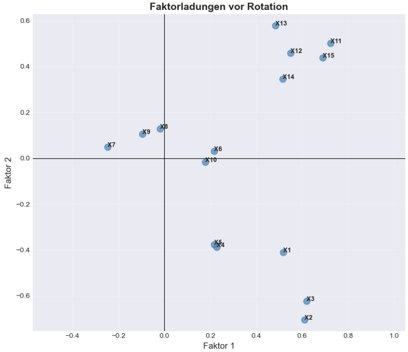
angle = 45° (sweet spot)

- Three dots hug F1.
- Three dots hug F2.

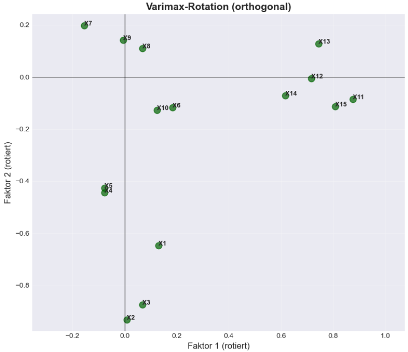
This is what simple structure looks like.

Before vs after

Same loadings, two viewpoints.



Before rotation



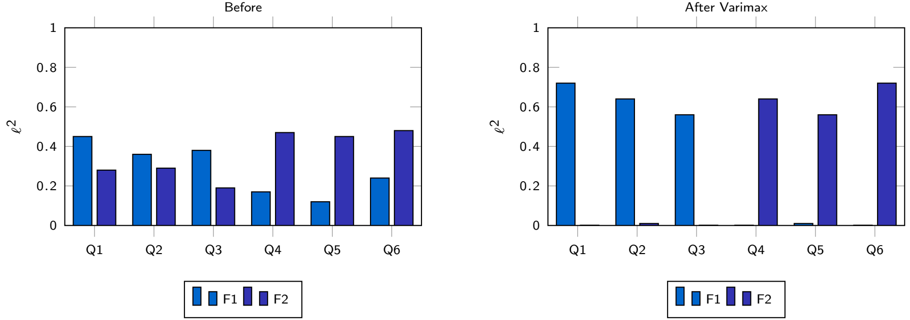
After Varimax

- Left: messy before rotation.
- Right: clean after Varimax.

Rotation doesn't change variance explained — it changes how we read it.

Varimax picks the angle — the idea in one picture

Varimax finds the angle that maximises the variance of squared loadings in each column.



- Before: spread out → low variance.
- After: peaks + zeros → high variance.

"Varimax" = Variance Maximisation.

Maximise this number, one factor-column at a time.

$$V(\Lambda) = \sum_{j=1}^{k} \mathrm{Var}\left(\ell_{ij}^{2}\right)$$

formula → tall F1 bars + near-zero F2 bars

- ℓ_{ij} = loading of item i on factor j. Squaring removes sign.
- Our 6-item example: $V_{\text{before}} = 0.026$, $V_{\text{after}} = 0.207$ (next slide shows the arithmetic).

One number, one goal: make each column polarised.

Worked example — computing V by hand

Plug the 6 squared loadings into $\mathrm{Var}(\ell^2)$ for each factor, then sum.

Factor 1 column

$$\ell^2_{i,F1} = (0.72,\ 0.64,\ 0.56,\ 0.00,\ 0.01,\ 0.00)$$

$$\text{mean} = 0.32$$

$$\mathrm{Var}(\ell^2_{F1}) = \tfrac{1}{6}\Big[(0.72-0.32)^2 + (0.64-0.32)^2 + (0.56-0.32)^2 + (0.00-0.32)^2 + (0.01-0.32)^2 + (0.00-0.32)^2 \Big]$$

$$= \tfrac{1}{6}[0.16 + 0.10 + 0.06 + 0.10 + 0.10 + 0.10]$$

$$\approx \mathbf{0.104}$$

Factor 2 column

$$\ell^2_{i,F2} = (0.00,\ 0.01,\ 0.00,\ 0.64,\ 0.56,\ 0.72)$$

$$\text{mean} = 0.32$$

$$\mathrm{Var}(\ell^2_{F2}) \approx \mathbf{0.104}$$

Total V

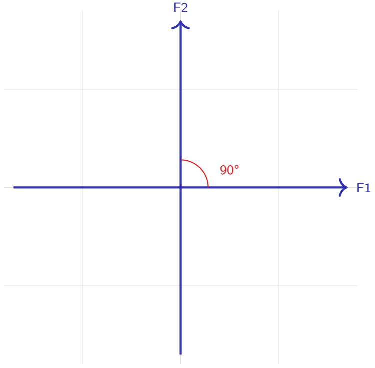
$$V = 0.104 + 0.104 = 0.207$$

- Each column has 3 tall + 3 near-zero $\rightarrow$ high variance.
- Before: each column $\approx$ uniform ($V = 0.026$). After: polarised ($V = 0.207$). $7.9\times$ jump.

That variance jump is what Varimax is hunting — and you can check it with a calculator.

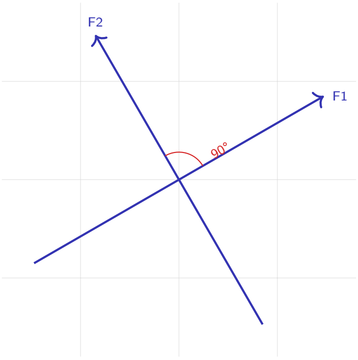
Varimax keeps the axes perpendicular

Varimax rotates, but keeps F1 $\perp$ F2.



Before: 0°

orthogonal

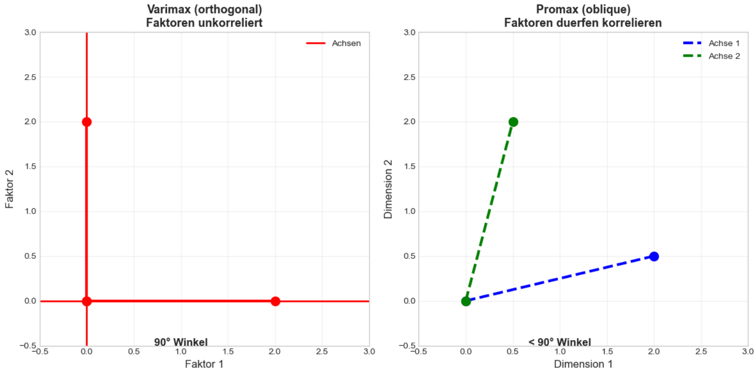


After: 30°

- Factors stay uncorrelated.
- Each factor is its own dimension.

Orthogonal rotation = factors remain independent.

Promax relaxes the 90° rule — for when real factors correlate.



- Varimax: orthogonal.
- Promax: oblique.

Use Promax only if you expect factors to be correlated.

Three takeaways

- Rotate the axes, not the data.
- Maximise column variance of squared loadings.
- Keep the axes perpendicular (orthogonal).

Simple structure = readable factors.

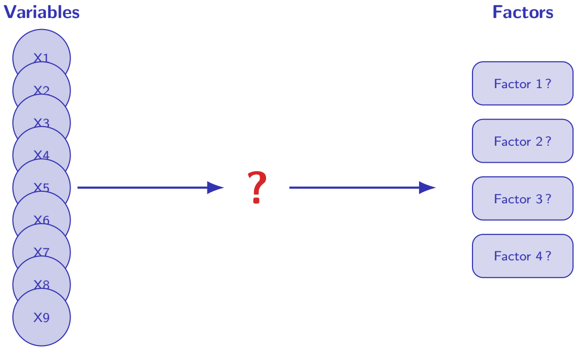
Part 2

How many factors should we keep?

Step-by-step with the Kaiser rule ($\lambda > 1$)

The question

We have 9 variables. How many factors do we really need?



- Too few → lose information.
- Too many → overfit noise.

We need a rule to decide.

Recap — what is an eigenvalue?

An eigenvalue = how much variance one direction captures.

Kovarianzmatrix Σ:

$$\begin{bmatrix} 2.0 & 0.8 \\ 0.8 & 1.0 \end{bmatrix}$$

Schritt 1: Eigenwertgleichung aufstellen

$$\det(\Sigma - \lambda I) = 0$$

- Big $\lambda \rightarrow$ important direction.
- Small $\lambda \rightarrow$ noise direction.

Eigenvalue = variance along a direction.

Compute step 1 — standardize variables

First, rescale every variable to mean 0 and variance 1.

Before

Variable	Mean	Var
Height (cm)	170	25
Weight (g)	72000	1600
Score (0–10)	7.2	0.04

standardize →

After

Variable	Mean	Var
Height	0	1
Weight	0	1
Score	0	1

- Different units would distort eigenvalues.
- After standardization: each variable carries exactly 1 unit of variance.

Why variance = 1 matters — watch the next slides.

Compute step 2 — correlation matrix

Now build the correlation matrix R.

	V1	V2	V3	V4
V1	1.00	0.82	0.18	0.12
V2	0.82	1.00	0.22	0.15
V3	0.18	0.22	1.00	0.78
V4	0.12	0.15	0.78	1.00

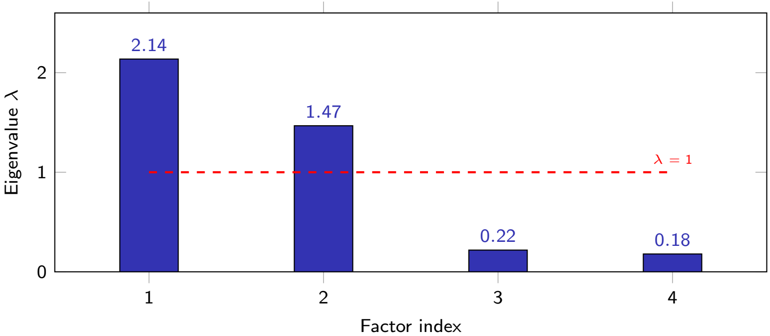
Correlation matrix R (4×4) — C.5 will compute its 4 eigenvalues

- Diagonal = 1 (variable with itself).
- Off-diagonal = correlations between variables.

Eigenvalues will be computed on R.

Compute step 3 — extract eigenvalues

Solve $\det(R - \lambda I) = 0 \Rightarrow 4$ eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_4$.



- $\lambda_1 = 2.137 > 1$ ✓ $\lambda_2 = 1.467 > 1$ ✓
- $\lambda_3 = 0.218 < 1$, $\lambda_4 = 0.179 < 1$ × (drop) — Kaiser keeps 2 factors.

The bars tell us how important each factor is.

The eigenvectors — where the factors point

Each eigenvalue has an eigenvector telling us which mix of variables defines that factor.

Eigenvector matrix V

	PC1	PC2	PC3	PC4
V1	−0.506	−0.498	+0.125	−0.693
V2	−0.524	−0.470	−0.067	+0.708
V3	−0.502	+0.490	−0.704	−0.113
V4	−0.467	+0.540	+0.696	+0.079
λ	2.137	1.467	0.218	0.179

How to read PC1

PC1 weights: $(-0.51, -0.52, -0.50, -0.47)$

All four variables contribute about equally with the same sign.

$\Rightarrow$ PC1 is the "overall level" direction.

How to read PC2

PC2 weights: $(-0.50, -0.47, +0.49, +0.54)$

V1, V2 negative; V3, V4 positive.

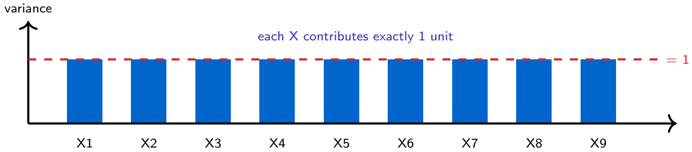
$\Rightarrow$ PC2 contrasts "V1+V2" vs "V3+V4".

- Eigenvectors are unit vectors: $\sum_i v_{ij}^2 = 1$ for each column.
- Sign is arbitrary — multiply column by -1 and PC stays valid.

Eigenvalues say "how much"; eigenvectors say "which direction".

Why the threshold is exactly $\lambda = 1$ (intuition)

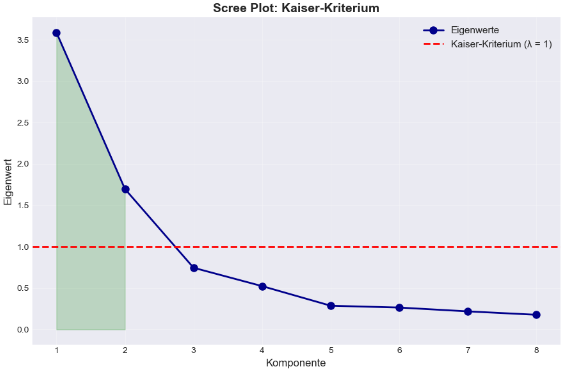
Each standardized variable carries variance $= 1$.



- 9 variables $\rightarrow$ 9 total units of variance.
- A factor that captures < 1 is less useful than ONE original variable.

A factor should do AT LEAST as much as a single variable.

Keep every factor with eigenvalue > 1.



- Bars above the red line $\rightarrow$ keep.
- Bars below $\rightarrow$ drop.

Kaiser (1960): $\lambda > 1$ keeps factors worth more than one variable.

Worked example (1 of 2) — the correlation matrix

4 student test items, 2 visible clusters.

	V1	V2	M1	M2
V1	1.00	0.85	0.15	0.10
V2	0.85	1.00	0.12	0.18
M1	0.15	0.12	1.00	0.90
M2	0.10	0.18	0.90	1.00

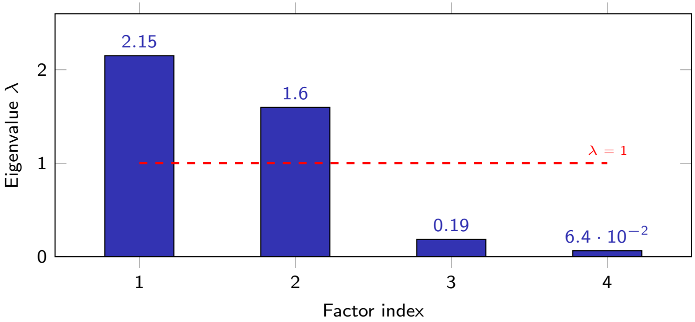
Verbal block: $r \approx 0.85$ Math block: $r \approx 0.90$

- Verbal block: $r \approx 0.85$.
- Math block: $r \approx 0.90$.

We expect 2 factors — let's verify with Kaiser.

Worked example (2 of 2) — eigenvalues and verdict

Eigenvalues: 2.151, 1.599, 0.185, 0.064 $\Rightarrow$ keep 2 factors.



- $\lambda_1 = 2.151 > 1$ ✓ $\lambda_2 = 1.599 > 1$ ✓
- $\lambda_3 = 0.185$, $\lambda_4 = 0.064$ × (drop)

Two factors, exactly as the correlation blocks suggested.

Worked example — the eigenvectors of R

The two "surviving" eigenvectors tell us what the 2 factors ARE.

Eigenvectors (top 2)

	PC1	PC2
V1	−0.470	−0.531
V2	−0.483	−0.514
M1	−0.521	+0.478
M2	−0.524	+0.475
λ	2.151	1.599

Reading the structure

PC1 (general factor): all 4 weights ≈ -0.5, same sign.
→ "overall test ability"

PC2 (contrast factor): V's negative, M's positive.
→ "verbal vs. math"

Worked PC2 score for student with z-scores $(1.0,\ 0.8,\ -0.5,\ -0.4)$:

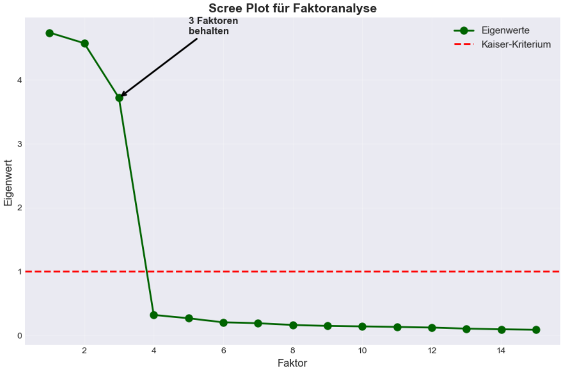
$$\begin{aligned} \text{PC2} &= 1.0(-0.531) + 0.8(-0.514) \\ &\quad + (-0.5)(0.478) + (-0.4)(0.475) \\ &= -0.531 - 0.411 - 0.239 - 0.190 \\ &= -\mathbf{1.371} \end{aligned}$$

Strongly verbal-leaning student.

- Block-correlated $R \Rightarrow$ block-structured eigenvectors.
- Sign of each eigenvector is arbitrary — structure is what matters.

Same matrix, two views: λ = how much, eigenvector = which mix.

The same verdict on a scree plot.

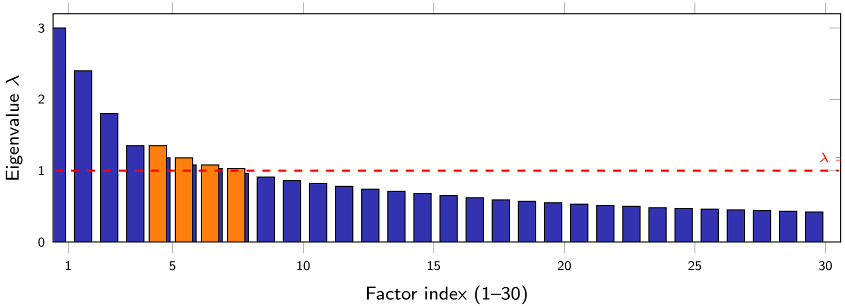


- Two bars above the line.
- Elbow also around the 3rd eigenvalue.

Kaiser and the elbow agree here.

A caveat — Kaiser over-picks for many variables

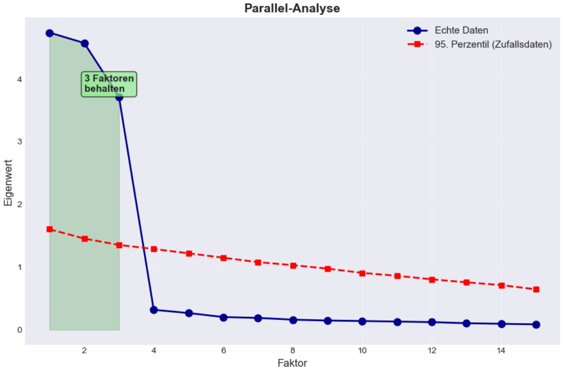
With many variables, several eigenvalues cluster just above 1.



- Kaiser can keep “borderline” factors (orange bars).
- These may be noise.

Kaiser is a rule of thumb, not a law.

Compare your eigenvalues to those from random data.



- Keep factors where real > random.
- More conservative than Kaiser.

Parallel analysis is generally preferred for factor counts.

Kaiser — summary card

Four-step recipe

Four steps, one threshold.

Thank you

Questions on Varimax or Kaiser? Let's discuss.