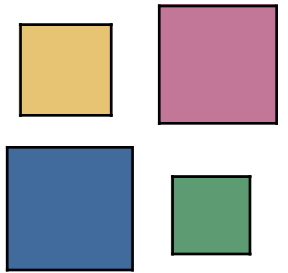


## Faktoren Analyse

- Oft kann eine interessierende Größe (Faktor) nicht direkt gemessen werden, sondern nur Indikatoren für diese Größe.  
Beispiele:
  - Schulnoten in Englisch und Deutsch als Indikator für die sprachliche Begabung, sowie Noten für Mathematik und Physik für die naturwissenschaftliche Begabung
  - Intelligenz ...
- Frage:  
Kann man diese (unbeobachtbaren) Faktoren durch eine kleine Zahl der möglichen Indikatoren zusammenfassen?
- Anwendungen finden sich oft in Studien die eine Vielzahl von Umfragedaten erheben, aus denen nur wenige interessante Ergebnisse extrahiert werden können.



## Faktoren Analyse: Modell

Faktorenanalyse ist in gewisser Weise mit der Hauptkomponentenanalyse verwandt, da es um die Beschreibung der Kovarianzen bzw. Correlationen zwischen den beobachteten Variablen geht, und wie diese durch eine kleine Zahl von unbeobachteten  $f_1, f_2, \dots, f_k$  latenten Variablen beschrieben werden können  $k < p$ .

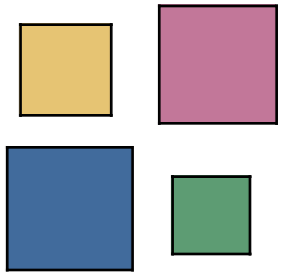
Als Modell ergibt sich:

$$X_j - \mu_j = l_{j1}F_1 + l_{j2}F_2 + \dots + l_{jk}F_k + \epsilon_j, \quad j = 1, \dots, p \quad \text{d.h.}$$

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\epsilon}$$

mit

$$\mathbf{L} = \begin{bmatrix} l_{11} & \dots & l_{1k} \\ \vdots & & \vdots \\ l_{p1} & \dots & l_{pk} \end{bmatrix}_{p \times k}, \quad \mathbf{F} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_k \end{bmatrix}_{k \times 1}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{bmatrix}_{p \times 1}$$



...

Die  $l_{ij}$  werden als **Ladungen** bezeichnet, **L** als Ladungsmatrix.

Weiter wird angenommen:

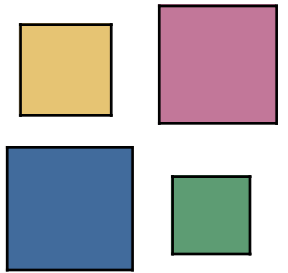
$$E(\mathbf{F}) = \mathbf{0}, \quad \text{Cov}(\mathbf{F}) = I_k; \quad E(\boldsymbol{\varepsilon}) = \mathbf{0},$$
$$\text{Cov}(\boldsymbol{\varepsilon}) = \Psi = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & & \\ \vdots & & \ddots & \\ 0 & & & \psi_p \end{bmatrix}$$

Woraus folgt:

$$\text{Var}(\mathbf{X}_i - \boldsymbol{\mu}_i) = \sigma_i^2 = \sum_{j=1}^k l_{ij}^2 + \psi_i$$

wobei  $\psi_i$  die Variance der  $\varepsilon_i$  ist.

Weiter ist,  $\text{Cov}(\mathbf{X}_i, \mathbf{X}_j)$  durch  $\sum_{h=1}^k l_{ih}l_{jh}$  gegeben.



## Varianzzerlegung

Es gilt:

- $\text{Cov}(\mathbf{X}) = \mathbf{L}\mathbf{L}' + \Psi$   
bzw.

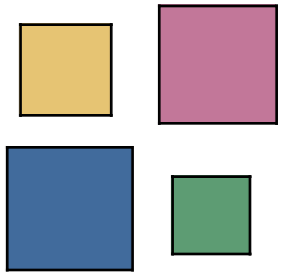
$$\begin{aligned}\text{Var}(X_i) &= l_{i1}^2 + \dots + l_{ik}^2 + \psi_i \\ \text{Cov}(\mathbf{X}_i, \mathbf{X}_j) &= l_{i1}l_{j1} + \dots + l_{ik}l_{jk} = \sum_{h=1}^k l_{ih}l_{jh}\end{aligned}$$

- $\text{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$   
bzw.

$$\text{Cov}(X_i, F_j) = l_{ij}$$

Man verteilt dann die Varianz in

$$\begin{aligned}\underbrace{\sigma_i}_{\text{Var}(X_i)} &= \underbrace{l_{i1}^2 + l_{i2}^2 + \dots + l_{ik}^2}_{\text{Kommunalität} =: h_i^2} + \underbrace{\psi_i}_{\text{Spezifität}} \\ \text{Var}(X_i) &= \text{Kommunalität} =: h_i^2 + \text{Spezifität}\end{aligned}$$



...

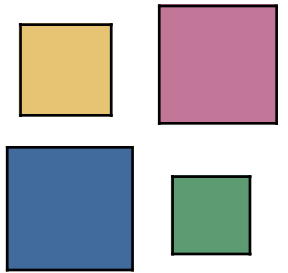
Die  $h_i^2$  wird die **Kommunalität** genannt, und ist die Varianz, die die Variable  $i$  über die gemeinsamen Faktoren mit den anderen Variablen teilt.

Der zweite Teil ist die **Spezifität**, d.h. der Teil der Varianz, der alleine aus dieser Variable stammt.

Somit lässt sich die Varianz Covarianz Matrix umschreiben zu

$$\mathbf{S} = \mathbf{LL}' + \Psi$$

Faktorenanalyse bricht im wesentlichen die Varianz einer Stichprobe auf in die allgemeine Varianz und die spezifische Varianz.



## Schätzung von $L$ (PCA Methode)

Die klassische Methode um  $L$  zu schätzen ist über die Eigenwertzerlegung der PCA.

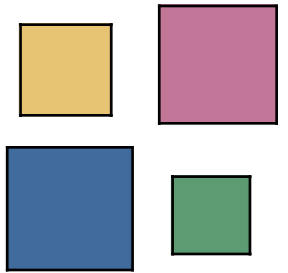
Die Matrix der geschätzten Ladungen ergibt sich aus:

$$\hat{\mathbf{L}} = (\sqrt{\lambda_1} \mathbf{e}_1 \quad \sqrt{\lambda_2} \mathbf{e}_2 \quad \dots \quad \sqrt{\lambda_k} \mathbf{e}_k)$$

Die spezifischen Varianzen können über  $\mathbf{S} - \hat{\mathbf{L}}\hat{\mathbf{L}}'$  berechnet werden:

$$\hat{\Psi} = \begin{bmatrix} s_{11} - \sum_i^k \hat{l}_{1i}^2 & 0 & \dots & 0 \\ 0 & s_{22} - \sum_i^k \hat{l}_{2i}^2 & & \\ \vdots & & \ddots & \\ 0 & & & s_{pp} - \sum_i^k \hat{l}_{pi}^2 \end{bmatrix}$$

Die Kommunalitäten werden durch  $\hat{h}_i^2 = \sum_j^k \hat{l}_{ik}^2$  geschätzt.



...

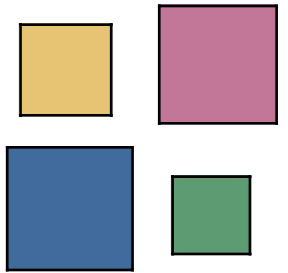
## Bemerkungen

- Entscheidend ist, dass die Zahl der Eigenvektoren verringert wird ( $k < p$ ), da sonst nur  $\Sigma = \mathbf{L}\mathbf{L}' + \mathbf{0}_{p \times p}$  berechnet würde.
- Die Ladungen  $\mathbf{L}$  sind nur bis auf eine orthogonale Matrix  $\mathbf{T}$  eindeutig, da für  $\mathbf{L}^* = \mathbf{T}\mathbf{L}$  gilt:

$$\text{Cov}(\mathbf{X}) = \mathbf{L}\mathbf{L}' + \Psi = \mathbf{L}\mathbf{T}\mathbf{T}'\mathbf{L}' + \Psi = (\mathbf{L}^*)(\mathbf{L}^*)' + \Psi$$

- Da die Diagonalelemente von  $\mathbf{L}\mathbf{L}'$  durch die Rotation nicht verändert werden, können die Faktoren **rotiert** werden.
- Durch die Faktorrotation wird versucht  $\hat{\mathbf{L}}$  eindeutig zu machen:
  - Rotiere so, dass möglichst viele Koeffizienten der Projektionsmatrix  $\hat{\mathbf{L}}$  gleich Null sind
  - Mittels des Varimax Prinzips:

$$\sum_{i,j} (d_{ij} - \bar{d}_{.j})^2 \rightarrow \max \quad \text{mit} \quad d_{ij} = l_{ij}^2 / h_i^2$$



# Schätzung über Maximum Likelihood

Vorraussetzung:

- Faktoren  $\mathbf{f}$ , und Fehler  $\varepsilon$  sind  $N(\boldsymbol{\Sigma}, 0)$
- Um  $\mathbf{L}$  eindeutig zu machen wird  $\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L}$  als Diagonal Matrix gefordert.

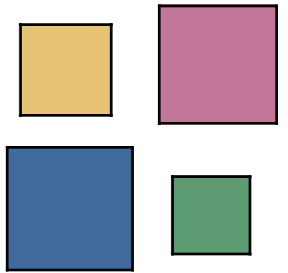
Maximiert wird dann

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{(n-1)p}{2}} |\boldsymbol{\Sigma}|^{-\frac{(n-1)}{2}} \exp\left\{\frac{-1}{2} \text{tr} \left( \boldsymbol{\Sigma}^{-1} \left( \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' \right) \right)\right\} \\ \times (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{\frac{-n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})\right\}$$

mit  $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ .  $\boldsymbol{\mu}$  wird mit dem Stichprobenmittel, und die Komunalitäten über

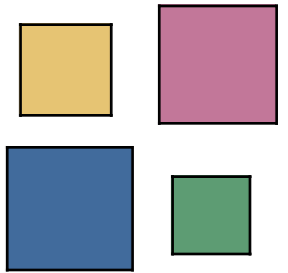
$$\hat{h}_i^2 = \sum_{j=1}^k l_{ij}, \quad i = 1, \dots, p$$

geschätzt.



## Bemerkungen

- Die Lösung über Hauptkomponenten ist eher historisch, weil sie auch ohne Computer relativ leicht zu berechnen war.  
Haupt Nachteil ist die Skalierungs-Invarianz
- Die Interpretation der Faktor Ladungen ist ohne Rotation meist nicht sinnvoll möglich. Ein – oft gewünschter – gemeinsamer Faktor aller Variablen verschwindet jedoch meist durch die Rotation.
- Je nach gewähltem Kriterium für die Rotation fällt die Interpretation unter Umständen anders aus.
- Kann man nicht davon ausgehen, dass die Faktoren unabhängig sind, können auch nicht-orthogonale Rotationen angebracht sein.



## Bartlett-Test auf Anzahl der Faktoren

Getestet wird

$$H_0 : \Sigma = \mathbf{L}\mathbf{L}' + \Psi \quad \text{vs.} \quad H_1 : \Sigma \text{ andere pos. def. Matrix}$$

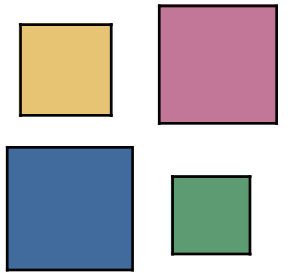
Da  $\hat{\Sigma} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}$  der maximum likelihood Schätzer von  $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$  ist, gilt:

$$-2 \ln \Lambda = n \ln \left( \frac{|\hat{\Sigma}|}{|\mathbf{S}_n|} \right) \sim \chi^2_{[(p-k)^2 - p - k]/2}(\alpha)$$

kann als Teststatistik verwendet werden.

Eine bessere Approximation der  $\chi^2$ -Verteilung erhält man, wenn man  $n$  durch  $(n - 1 - (2p + 4k + 5)/6)$  benutzt, d.h.  $H_0$  wird verworfen, falls gilt:

$$(n - 1 - (2p + 4k + 5)/6) \ln \left( \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}|}{|\mathbf{S}_n|} \right) > \chi^2_{[(p-k)^2 - p - k]/2}(\alpha)$$



## Beispiel: Börsendaten (via PCA)

- Der wöchentliche Return von 5 Firmen gemessen über 100 Wochen.

### 1 Faktor Modell

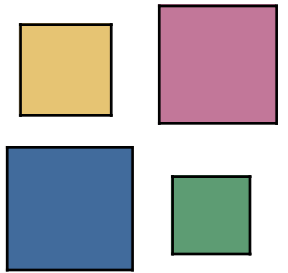
$$L_1 = \begin{pmatrix} 0.78 \\ 0.77 \\ 0.79 \\ 0.71 \\ 0.71 \end{pmatrix} \quad \hat{\Psi} = \begin{pmatrix} 0.39 \\ 0.40 \\ 0.37 \\ 0.49 \\ 0.49 \end{pmatrix}$$

### 2 Faktor Modell

$$\mathbf{L} = \begin{pmatrix} 0.78 & -0.22 \\ 0.77 & -0.46 \\ 0.79 & -0.23 \\ 0.71 & 0.47 \\ 0.71 & 0.52 \end{pmatrix} \quad \hat{\Psi} = \begin{pmatrix} 0.34 \\ 0.19 \\ 0.31 \\ 0.27 \\ 0.22 \end{pmatrix}$$

### Residual Matrix (2 Faktoren)

$$\varepsilon = \mathbf{R} - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\Psi} = \begin{pmatrix} 0.000 & -0.127 & -0.164 & -0.069 & 0.018 \\ -0.127 & 0.000 & -0.123 & 0.055 & 0.012 \\ -0.164 & -0.123 & 0.000 & -0.019 & -0.017 \\ -0.069 & 0.055 & -0.019 & 0.000 & -0.231 \\ 0.018 & 0.012 & -0.017 & -0.231 & 0.000 \end{pmatrix}$$



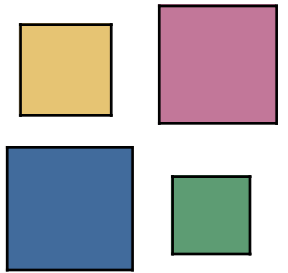
## Beispiel: Börsendaten (via MLE)

### 2 Faktor Modell (PCA)

$$\mathbf{L} = \begin{pmatrix} 0.78 & -0.22 \\ 0.77 & -0.46 \\ 0.79 & -0.23 \\ 0.71 & 0.47 \\ 0.71 & 0.52 \end{pmatrix} \quad \hat{\Psi} = \begin{pmatrix} 0.34 \\ 0.19 \\ 0.31 \\ 0.27 \\ 0.22 \end{pmatrix}$$

### 2 Faktor Modell (MLE)

$$\mathbf{L} = \begin{pmatrix} 0.68 & 0.19 \\ 0.69 & 0.52 \\ 0.68 & 0.25 \\ 0.62 & -0.07 \\ 0.79 & -0.44 \end{pmatrix} \quad \hat{\Psi} = \begin{pmatrix} 0.50 \\ 0.25 \\ 0.47 \\ 0.61 \\ 0.18 \end{pmatrix} \quad \text{Residual Matrix } \varepsilon = \mathbf{R} - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\Psi} = \begin{pmatrix} 0.00 & 0.01 & -0.00 & -0.02 & 0.00 \\ 0.01 & 0.00 & -0.00 & -0.00 & 0.00 \\ -0.00 & -0.00 & 0.00 & 0.03 & -0.00 \\ -0.02 & -0.00 & 0.03 & 0.00 & 0.00 \\ 0.00 & 0.00 & -0.00 & 0.00 & 0.00 \end{pmatrix}$$

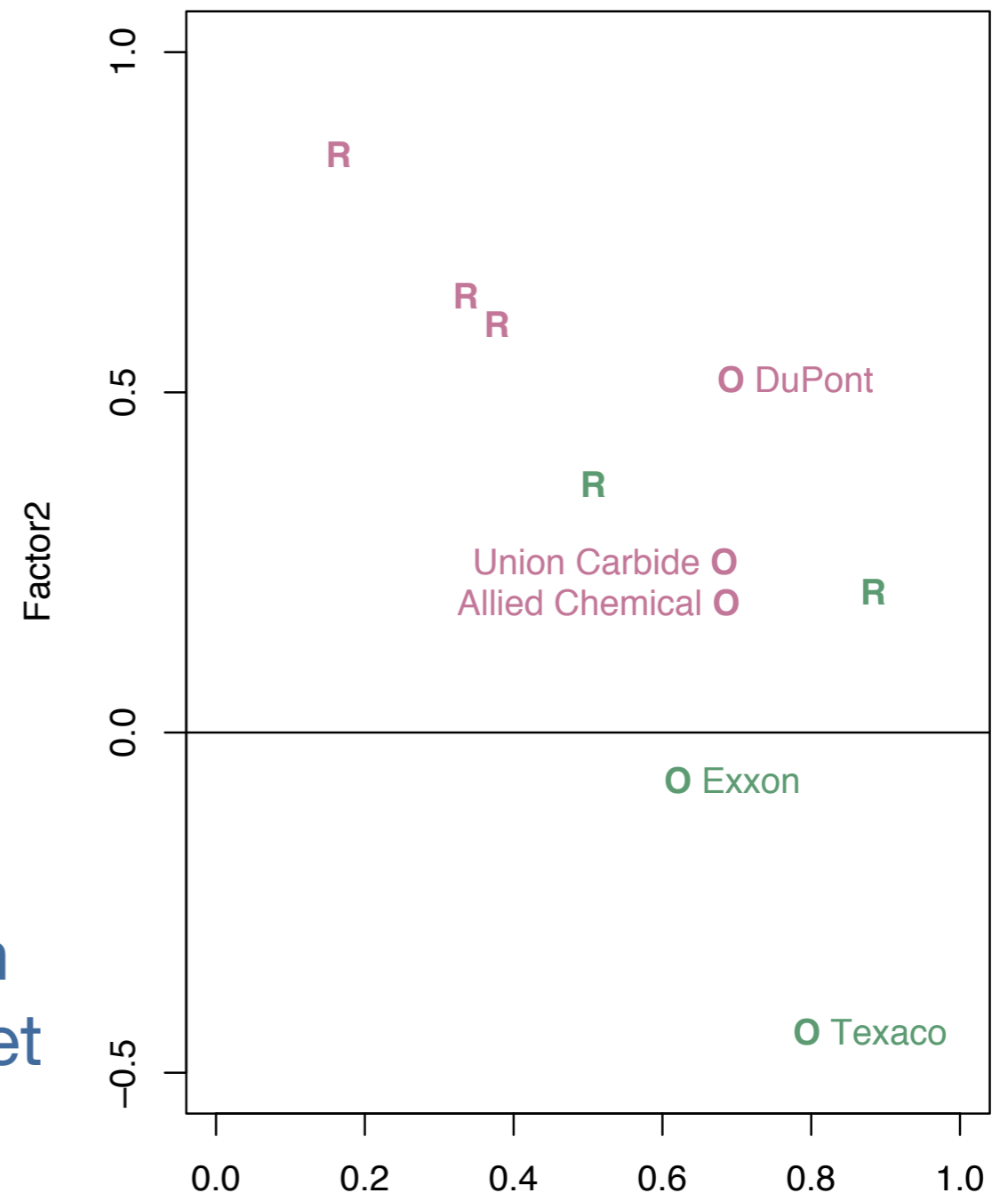


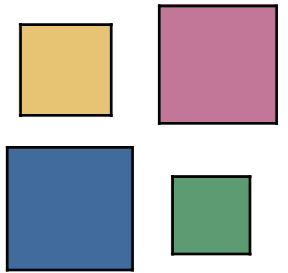
## Beispiel: Börsendaten – Faktorrotation

### 2 Faktor Modell (MLE) mit Rotation

$$\mathbf{L} = \begin{pmatrix} 0.60 & 0.38 \\ 0.85 & 0.17 \\ 0.64 & 0.34 \\ 0.37 & 0.51 \\ 0.22 & 0.88 \end{pmatrix} \quad \hat{\Psi} = \begin{pmatrix} 0.50 \\ 0.25 \\ 0.47 \\ 0.61 \\ 0.18 \end{pmatrix}$$

- Der erste Faktor gruppiert die Aktien der Chemie Industrie
- Der zweite Faktor die Ölindustrie
- Die Rotation hat den gemeinsamen Faktor der alle Aktien hoch bewertet eliminiert.





## Schätzung der Faktoren

Auch wenn die Ladungen  $\mathbf{L}$  meist von größerem Interesse sind, so werden die Faktoren selber (auch factor scores) auch benötigt. Mit der gewichteten kleinsten Quadrate Methode minimiert man

$$\sum_{i=1}^p \frac{\varepsilon_1^2}{\psi_i} = \boldsymbol{\varepsilon}' \boldsymbol{\Psi}^{-1} \boldsymbol{\varepsilon} = (\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f})' \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f})$$

welches Lösung

$$\hat{\mathbf{f}} = (\mathbf{L}' \boldsymbol{\Psi}^{-1} \mathbf{L})^{-1} \mathbf{L}' \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

hat.

Alternativ wird  $\mathbf{f}$  auch über die Regressionsmethode als

$$\hat{f}_j = \hat{\mathbf{L}}' \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}), \quad j = 1, \dots, n$$

geschätzt.