ANÁLISE FATORIAL

$\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, vetor p-dim

- IMAGINE que existem DUAS variáveis F_1 e F_2 QUE NÃO SÃO DIRETAMENTE OBSERVÁVEIS
- DIZEMOS QUE SÃO LATENTES
- F_1 e F_2 SÃO INDEP.
- CADA UMA DAS X_i EM $\underset{\sim}{X}$ É APROXIMADAMENTE UMA COMBINAÇÃO LINEAR DE F_1 e F_2

ISTO É,

$$X_1 \approx \mu_1 + l_{11} F_1 + l_{12} F_2$$

$$X_2 \approx \mu_2 + l_{21} F_1 + l_{22} F_2$$

$$\vdots$$

$$X_p \approx \mu_p + l_{p1} F_1 + l_{p2} F_2$$

⊕ l_{ij} = CONSTANTES DESCONHECIDAS

= CARGAS DOS FATORES

FACTOR LOADING.

EXEMPLO $\underset{\sim}{X}$ = VETOR COM NOTAS DE UM ALUNO em 15 ASSUNTOS

X_1 = GRAMÁTICA

X_2 = Literatura

X_3 = Redação

X_4 = História

X_5 = Geografia

X_6 = Filosofia

X_7 = Álgebra

X_8 = Geometria

X_9 = Lógica

X_{10} = Química

X_{11} = Física

X_{12} = Biologia

X_{13} = Inglês

X_{14} = Sociologia

X_{15} = Filosofia

- Desempenho individual nestes 15 assuntos é muito variado:
  - alguns vão bem em todas as disciplinas
  - alguns vão bem apenas em algumas e têm um desempenho médio nas outras
  - alguns vão muito bem em algumas e muito mal em outras
  - alguns vão mal em todas.
- Como entender esta diversidade?

Psicólogos se perguntaram se esta variabilidade na capacidade cognitiva não poderia ser explicada pela existência de uns poucos traços latentes, não observados. (8)

Por exemplo,

FV = Fator refletindo habilidade verbal

FQ = Fator refletindo habilidade lógico-quantitativo

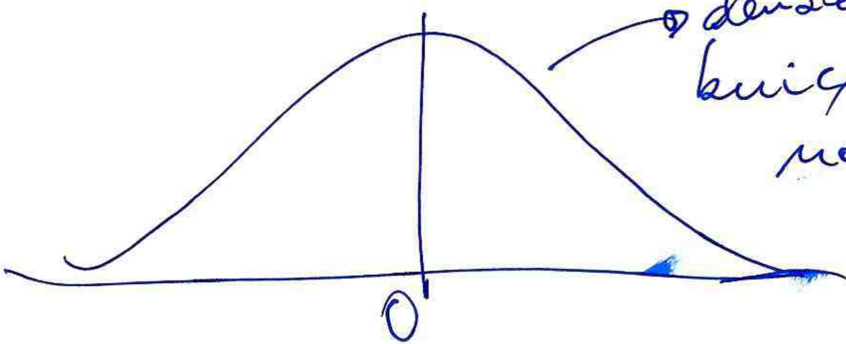
* Os fatores têm uma escala centrada em zero

Se $F=0 \Rightarrow$ indivíduo tem habilidade média no fator (6)

$F>0 \Rightarrow$ habilidade acima da média

$F \gg 0 \Rightarrow$ habilidade muito acima da média

$F<0 \Rightarrow$ abaixo da média



→ densidade da distribuição de um fator na população.

⊕ Cada indivíduo recebe uma "dose" de FV (7)
e uma "dose" de ~~[illegible]~~ FQ

⊕ Doses de FV e FQ são <u>independentes</u>

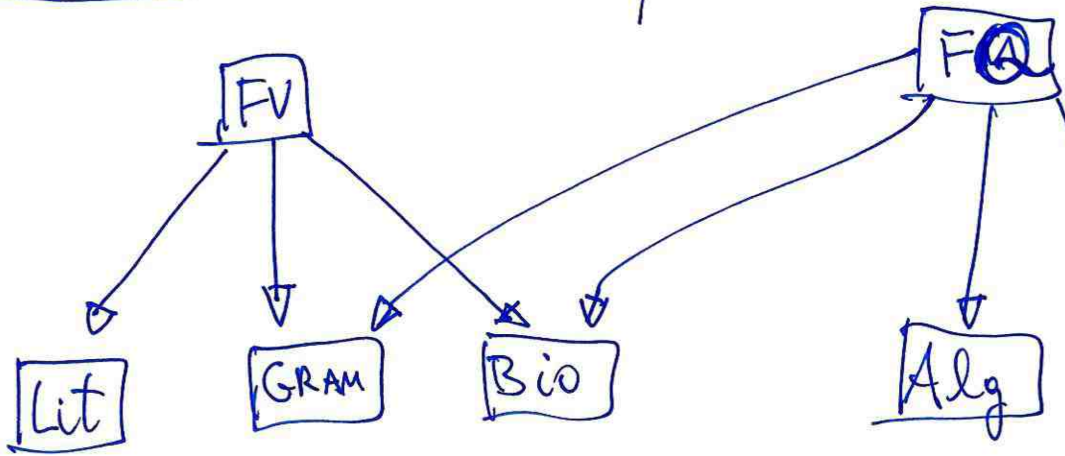
{ Por exemplo, alguns recebem muito de FV
Dentre estes, { metade recebe FQ +
metade recebe FQ −

⊕ As notas nos 15 assuntos são <u>reflexos e combinações</u> desses dois fatores, além de um ruído causado por outros fatores que não levamos em conta.

- Modelo hierárquico para a geração das 15 notas de um mesmo indivíduo

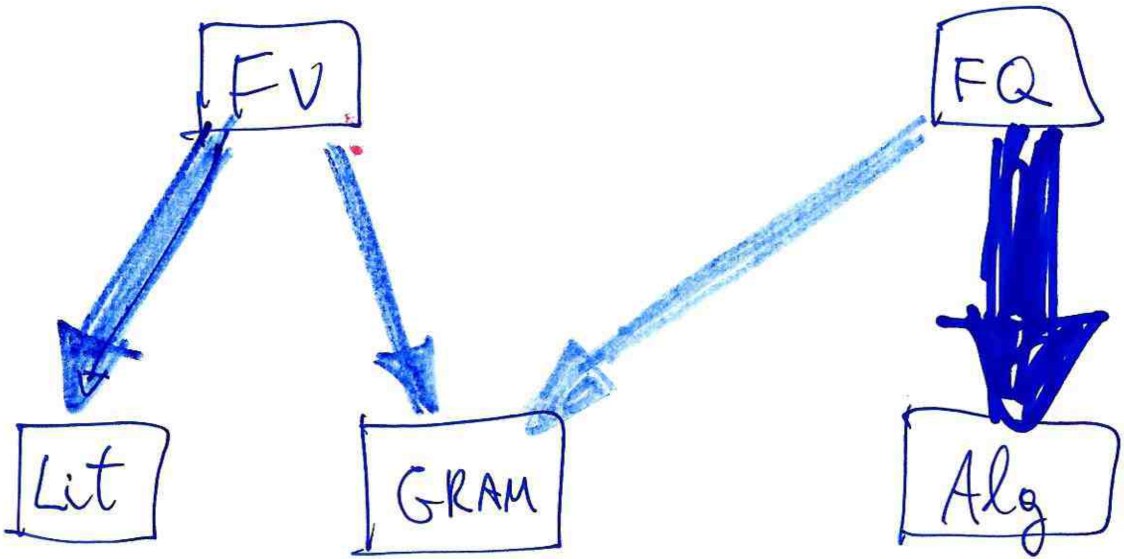
- **Passo 1:** Recebe doses indep de FV e FQ

- **Passo 2:** Estas doses afetam as notas dos 15 assuntos



AUSÊNCIA DE ARESTA ⇒ sem influência do fator

⊕ Peso ou importância da aresta representada por grossura da aresta



Peso da aresta é a carga do fator (factor loading).

Representação algébrica

$$\text{Gramática} = X_1 \approx \mu_{GRAM} + l_{GV} * FV + l_{GQ} * FQ$$

$$\text{Literatura} = X_2 \approx \mu_{Lit} + l_{LV} * FV + l_{LQ} * FQ$$

$\vdots$

CARGA DOS FATORES

FATOR VERBAL DO INDIVÍDUO, O MESMO PARA TODOS OS ASSUNTOS

MAIS FORMAL

$$X_1 = \mu_{Gr} + l_{GV} * FV + l_{GQ} * FQ + \varepsilon_{Gran}$$

$$X_2 = \mu_{Lit} + l_{LV} * FV + l_{LQ} * FQ + \varepsilon_{Lit}$$

$$\vdots \qquad\qquad\qquad\qquad\qquad \vdots$$

$$X_{15} = \mu_{Fil} + l_{FV} * FV + l_{FQ} * FQ + \varepsilon_{Fil}$$

Erros ou fatores não-observados

Um exemplo esquemático:

$$X_1 = \mu_1 + (1.0) * FV + (0.1) FQ + \varepsilon_1$$

$$X_2 = \mu_2 + (0.8) * FV + (0.01) FQ + \varepsilon_2$$

ASSUNTOS MUITO ASSOCIADOS COM <u>FV</u>

$$X_3 = \mu_3 + (0.1) FV + (0.1) FQ + \varepsilon_3$$

NENHUMA DAS DUAS HABILIDADES É MUITO RELEVANTE

$$X_4 = \mu_4 + (0.3) FV + (0.9) FQ + \varepsilon_4$$

PRECISA SER BOM EM FQ <u>E</u> MEIO RUIM EM FV

$$X_5 = \mu_5 + (0.1) FV + (1.2) FQ + \varepsilon_5$$

$$\vdots$$

BOM EM FQ E POUCO RELEVANTE EM FV.

Notação matricial para <u>um</u> indivíduo

$$\underset{\sim}{X} = \underset{\sim}{\mu} + \underset{p\times m}{L} \cdot \underset{m\times 1}{\underset{\sim}{F}} + \underset{p\times 1}{\underset{\sim}{\varepsilon}}$$

L = matriz das cargas dos fatores (loadings)

F = vetor dos fatores comuns (às p variáveis)

$\underset{\sim}{\varepsilon}$ = vetor dos erros ou fatores <u>específicos</u> (de cada variável).

Vamos imaginar dois indivíduos com seus dois vetores instanciados:

$$\underset{\sim}{X}^{(1)} = \begin{pmatrix} X_1^{(1)} \\ \vdots \\ X_{15}^{(1)} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_{15} \end{pmatrix} + \begin{bmatrix} l_{11} & l_{12} \\ \vdots & \vdots \\ l_{15,1} & l_{15,2} \end{bmatrix} \begin{bmatrix} FV^{(1)} \\ FQ^{(1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1^{(1)} \\ \vdots \\ \varepsilon_{15}^{(1)} \end{bmatrix} = \underset{\sim}{\mu} + L \cdot \underset{\sim}{F}^{(1)} + \underset{\sim}{\varepsilon}^{(1)}$$

são iguais

ESPECÍFICOS DO INDIV.

$$\underset{\sim}{X}^{(2)} = \begin{pmatrix} X_1^{(2)} \\ \vdots \\ X_{15}^{(2)} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_{15} \end{pmatrix} + \begin{bmatrix} l_{11} & l_{12} \\ \vdots & \vdots \\ l_{15,1} & l_{15,2} \end{bmatrix} \begin{bmatrix} FV^{(2)} \\ FQ^{(2)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1^{(2)} \\ \vdots \\ \varepsilon_{15}^{(2)} \end{bmatrix} = \underset{\sim}{\mu} + L\, \underset{\sim}{F}^{(2)} + \underset{\sim}{\varepsilon}^{(2)}$$

$\underset{\sim}{\mu}$ e L $\Longrightarrow$ iguais para todos os indivíduos

$\underset{\sim}{F}^{(i)} = \begin{bmatrix} FV^{(i)} \\ FQ^{(i)} \end{bmatrix}$ = doses dos fatores recebidos pelo indivíduo (i)

$\underset{\sim}{\varepsilon}^{(i)}$ = erros do indivíduo (i)

- Só observamos o vetor $\underset{\sim}{X}$ em vários indivíduos.
- Queremos entender como $\underset{\sim}{X}$ varia
- Basta entender como os fatores FV e FQ variam de indivíduo para indivíduo.
- As notas dos 15 assuntos apenas refletem e contaminam estes fatores através da matriz L
- Um pequeno ruído ε_j para cada disciplina é adicionado para levar em conta os demais fatores que estamos ignorando.
- Como podemos inferir L a partir dos dados?
- E os ~~valores~~ escores dos fatores $\begin{bmatrix} FV^{(i)} \\ FQ^{(i)} \end{bmatrix}$ de cada indivíduo, como obtê-los?

Suposições do modelo fatorial

Indivíduo (i) com $m=2$ fatores

$$\underset{p\times 1}{\underset{\sim}{X}^{(i)}} = \underset{p\times 1}{\underset{\sim}{\mu}} + \underset{p\times 2}{L}\ \underset{2\times 1}{\underset{\sim}{F}^{(i)}} + \underset{p\times 1}{\underset{\sim}{\varepsilon}^{(i)}}$$

- $\underset{\sim}{\mu}$ e L são comuns a todos os indivíduos, não-aleatórios
- $\underset{\sim}{F}^{(i)}$ e $\underset{\sim}{\varepsilon}^{(i)}$ variam de indivíduo para indivíduo, são aleatórios
- $\underset{2\times 1}{\mathbb{E}}\left(\underset{\sim}{F}^{(i)}\right) = \mathbb{E}\begin{pmatrix} FV^{(i)} \\ FQ^{(i)} \end{pmatrix} = \begin{pmatrix} \mathbb{E}(FV^{(i)}) \\ \mathbb{E}(FQ^{(i)}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \underset{\sim}{0}$
- $\underset{2\times 2}{Cov}\left(\underset{\sim}{F}^{(i)}\right) = \begin{bmatrix} Var(FV^{(i)}) & Cov(FV^{(i)}, FQ^{(i)}) \\ Cov(FV^{(i)}, FQ^{(i)}), & Var(FQ^{(i)}) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

- A opção de tomar a variância de cada fator igual a 1 (e portanto, tomar o <u>DP</u> de cada fator = 1), esta opção é baseada no seguinte argumento:
    - cada fator (FV ou FQ) terá a "mesma escala" indo de -2 a $+2$ aproximadamente ao variar dos menos habilidosos aos mais habilidosos.
    - Se o fator F_k afetar muito uma nota X_j isto será refletido numa carga l_{jk} muito positiva (ou muito negativa)
    - Mas a escala de todos os fatores é a mesma (em DP's): vai de -2 a $+2$, a<u>proximadamente</u>
- A covariância $Cov(FV^{(i)}, FQ^{(i)})=0$ pois estamos supondo fatores <u>independentes</u>

- Mais suposições, agora sobre $\underset{\sim}{\xi}^{(i)}$

$$\underset{15\times 1}{\mathbb{E}\left(\underset{\sim}{\xi}^{(i)}\right)} = \mathbb{E}\begin{pmatrix} \xi_1^{(i)} \\ \vdots \\ \xi_{15}^{(i)} \end{pmatrix} = \begin{pmatrix} \mathbb{E}(\xi_1^{(i)}) \\ \vdots \\ \mathbb{E}(\xi_{15}^{(i)}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \underset{\sim}{0}$$

$$\underset{15\times 15}{Cov\left(\underset{\sim}{\xi}\right)} = \begin{bmatrix} Var(\xi_1^{(i)}) & Cov(\xi_1^{(i)}, \xi_2^{(i)}) & \cdots & Cov(\xi_1^{(i)}, \xi_{15}^{(i)}) \\ Cov(\xi_2^{(i)}, \xi_1^{(i)}) & Var(\xi_2^{(i)}) & \cdots & Cov(\xi_2^{(i)}, \xi_{15}^{(i)}) \\ \vdots & & & \\ Cov(\xi_{15}^{(i)}, \xi_1^{(i)}) & Cov(\xi_{15}^{(i)}, \xi_2^{(i)}) & \cdots & Var(\xi_{15}^{(i)}) \end{bmatrix} = \begin{bmatrix} \psi_1 & & & \\ & \psi_2 & & \\ & & \ddots & \\ & & & \psi_{15} \end{bmatrix} =$$

$$= diag(\psi_1, \ldots, \psi_{15}) = \Psi$$

- É razoável deixar que ~~Var~~ $\cancel{\xi^{(i)}_{j}}$ $Var(\xi^{(i)}_{\ell j})$ varie
- Podemos ter $\psi_j \neq \psi_k$
- A razão é que a nota de Redação, digamos, pode ter muito mais variabilidade que a nota de Matemática devido a fatores não relacionados com FV ou FQ.
- A subjetividade do corretor da redação, a variação da qualidade da redação como fruto do conhecimento do aluno sobre o tema da redação, entre outras causas, pode gerar mais variação na nota de redação do que a variação induzida pela diversidade de FV e FQ.

- A covariância entre os erros de assuntos distintos, $Cov(\varepsilon_j^{(i)}, \varepsilon_k^{(i)})$, provavelmente não é zero mas deve ser pequena e por isto forçamos todas iguais a zero no modelo.
- Assim adotamos $Cov(\underset{\sim}{\varepsilon}) = \begin{pmatrix} \psi_1 & & \underset{\sim}{0} \\ & \ddots & \\ \underset{\sim}{0} & & \psi_{15} \end{pmatrix}$ = diagonal.
- Uma última suposição: $\underset{\sim}{F}$ e $\underset{\sim}{\varepsilon}$ são independentes.
- Isto implica que $\longrightarrow$

$$\underset{15\times 2}{\mathrm{Cov}(\underset{\sim}{\varepsilon}, \underset{\sim}{F})} = \begin{bmatrix} \mathrm{Cov}(\varepsilon_1, F_1) & \mathrm{Cov}(\varepsilon_1, F_2) \\ \mathrm{Cov}(\varepsilon_2, F_1) & \mathrm{Cov}(\varepsilon_2, F_2) \\ \vdots & \\ \mathrm{Cov}(\varepsilon_{15}, F_1) & \mathrm{Cov}(\varepsilon_{15}, F_2) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}$$

- Note que, como $\underset{15\times 1}{\mathbb{E}(\underset{\sim}{\varepsilon})} = \underset{\sim}{0}$ e $\underset{2\times 1}{\mathbb{E}(\underset{\sim}{F})} = \underset{\sim}{0}$, temos

$$\mathrm{Cov}(\underset{\sim}{\varepsilon}, \underset{\sim}{F}) = \mathbb{E}\left(\underset{\sim}{\varepsilon}\,\underset{\sim}{F}'\right) = \underset{\sim}{0} \quad \text{e tb.} \quad \mathbb{E}(\underset{\sim}{F}\,\underset{\sim}{\varepsilon}') = \underset{\sim}{0}$$

- Podemos agora obter a estrutura de covariância das observações $\underset{\sim}{X}$.

- $\mathbb{E}(\underset{\sim}{X}) = \mathbb{E}(\underset{\sim}{\mu} + L\underset{\sim}{F} + \underset{\sim}{\varepsilon})$

$$= \underset{\sim}{\mu} + \mathbb{E}(L\underset{\sim}{F}) + \mathbb{E}(\underset{\sim}{\varepsilon})$$

$$= \underset{\sim}{\mu} + L\,\underbrace{\mathbb{E}(\underset{\sim}{F})}_{=\underset{\sim}{0}} + \underbrace{\mathbb{E}(\underset{\sim}{\varepsilon})}_{=\underset{\sim}{0}}$$

$$= \underset{\sim}{\mu}$$

- É isto mesmo, o valor esperado das notas é o vetor $\underset{\sim}{\mu}$ que representa a média da população de interesse.

- $\underset{15\times 15}{\mathrm{Cov}(\underset{\sim}{X})} = \Sigma = \mathbb{E}\left((\underset{\sim}{X}-\underset{\sim}{\mu})(\underset{\sim}{X}-\underset{\sim}{\mu})'\right) = \mathbb{E}\left((L\underset{\sim}{F}+\underset{\sim}{\varepsilon})(L\underset{\sim}{F}+\underset{\sim}{\varepsilon})'\right)$

$$= \mathbb{E}\left((L\underset{\sim}{F})(L\underset{\sim}{F})' + \underset{\sim}{\varepsilon}(L\underset{\sim}{F})' + (L\underset{\sim}{F})\underset{\sim}{\varepsilon}' + \underset{\sim}{\varepsilon}\underset{\sim}{\varepsilon}'\right)$$

$$= \mathbb{E}(L\underset{\sim}{F}\,\underset{\sim}{F}'L') + \mathbb{E}(\underset{\sim}{\varepsilon}\,(L\underset{\sim}{F})') + \mathbb{E}((L\underset{\sim}{F})\,\underset{\sim}{\varepsilon}') + \mathbb{E}(\underset{\sim}{\varepsilon}\,\underset{\sim}{\varepsilon}')$$

$$= L \underbrace{\mathbb{E}(\underset{\sim}{F}\underset{\sim}{F}')}_{= I_2} L' + \underset{\sim}{0} + \underset{\sim}{0} + \psi$$

$$= L L' + \psi$$

Isto é, $\mathrm{Var}(\underset{\sim}{X}) = \Sigma = LL' + \psi$

No nosso exemplo com 15 notas e dois fatores:

$$\underset{15\times 5}{\Sigma} = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \\ \vdots & \\ \vdots & \\ l_{15,1} & l_{15,2} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \cdots & l_{15,1} \\ l_{12} & l_{22} & \cdots & l_{15,2} \end{bmatrix} + \begin{bmatrix} \psi_1 & & & \\ & \psi_2 & & \\ & & \ddots & \\ & & & \psi_{15} \end{bmatrix}$$

$$
\Sigma = \left[\begin{array}{c|c|c|c|c}
l_{11}^2 + l_{12}^2 + \psi_1 & l_{11}l_{12} + l_{12}l_{22} & l_{11}l_{31} + l_{12}l_{32} & \cdots & l_{11}l_{15,1} + l_{12}l_{15,2} \\
l_{11}l_{21} + l_{12}l_{22} & l_{21}^2 + l_{22}^2 + \psi_2 & l_{21}l_{31} + l_{22}l_{32} & \cdots & l_{21}l_{15,1} + l_{22}l_{15,2} \\
 & & & & \vdots \\
\hline
 & & & & l_{15,1}^2 + l_{15,2}^2 + \psi_{15}
\end{array}\right]
$$

Outra maneira de pensar sobre Σ é perceber que, se olharmos a matriz $\underset{15\times 2}{L}$ como um conjunto de 15 vetores-cargas,

$$
L = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \\ \vdots & \vdots \\ l_{15,1} & l_{15,2} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{l}_1' \\ \underset{\sim}{l}_2 \\ \vdots \\ \underset{\sim}{l}_{15} \end{bmatrix}
$$

onde $\underset{\sim}{l}_j$ = cargas da disciplina j

entoS

$$\underset{15\times 15}{\Sigma} = \begin{bmatrix} \langle \underset{\sim}{l}_1, \underset{\sim}{l}_1\rangle + \psi_1 & \langle \underset{\sim}{l}_2, \underset{\sim}{l}_1\rangle & \langle \underset{\sim}{l}_3, \underset{\sim}{l}_1\rangle & \cdots & \langle \underset{\sim}{l}_{15}, \underset{\sim}{l}_1\rangle \\ \langle \underset{\sim}{l}_1, \underset{\sim}{l}_2\rangle & \langle \underset{\sim}{l}_2, \underset{\sim}{l}_2\rangle + \psi_2 & \langle \underset{\sim}{l}_3, \underset{\sim}{l}_2\rangle & \cdots & \langle \underset{\sim}{l}_{15}, \underset{\sim}{l}_2\rangle \\ \langle \underset{\sim}{l}_1, \underset{\sim}{l}_3\rangle & \langle \underset{\sim}{l}_2, \underset{\sim}{l}_3\rangle & \langle \underset{\sim}{l}_3, \underset{\sim}{l}_3\rangle + \psi_3 & \cdots & \langle \underset{\sim}{l}_{15}, \underset{\sim}{l}_3\rangle \\ \vdots & \vdots & \vdots & & \vdots \\ \langle \underset{\sim}{l}_1, \underset{\sim}{l}_{15}\rangle & \langle \underset{\sim}{l}_2, \underset{\sim}{l}_{15}\rangle & \langle \underset{\sim}{l}_3, \underset{\sim}{l}_{15}\rangle & \cdots & \langle \underset{\sim}{l}_{15}, \underset{\sim}{l}_{15}\rangle + \psi_{15} \end{bmatrix}$$

$$= \begin{bmatrix} \|\underset{\sim}{l}_1\|^2 + \psi_1 & \langle \underset{\sim}{l}_2, \underset{\sim}{l}_1\rangle & \cdots & \langle \underset{\sim}{l}_{15}, \underset{\sim}{l}_1\rangle \\ \langle \underset{\sim}{l}_2, \underset{\sim}{l}_1\rangle & \|\underset{\sim}{l}_2\|^2 + \psi_2 & \cdots & \langle \underset{\sim}{l}_{15}, \underset{\sim}{l}_2\rangle \\ \vdots & \vdots & & \vdots \\ \langle \underset{\sim}{l}_{15}, \underset{\sim}{l}_1\rangle & \langle \underset{\sim}{l}_{15}, \underset{\sim}{l}_2\rangle & \cdots & \|\underset{\sim}{l}_{15}\|^2 + \psi_{15} \end{bmatrix}$$

Assim,

$$\mathbb{V}ar(X_i) = \Sigma_{ii} = \underbrace{\|\underset{\sim}{l}_i\|^2}_{\substack{\shortparallel \\ \text{COMUNALIDADE}}} + \underbrace{\psi_i}_{\text{variância específica}} = l_{i1}^2 + l_{i2}^2 + \psi_i$$

⊕ Se os dois fatores latentes não possuem impacto na disciplina i (por exemplo, se disciplina for educação física) então $l_{i1}^2 + l_{i2}^2 \simeq 0$ e toda a variância da nota é devida aos fatores específicos ε_i e diferentes dos fatores latentes.

⊕ Suponhamos que a disciplina X_i tenha uma carga grande do fator verbal ($l_{i1}^2 >> 0$) mas uma carga pequena do fator quantitativo ($l_{i2}^2 \simeq 0$). Então

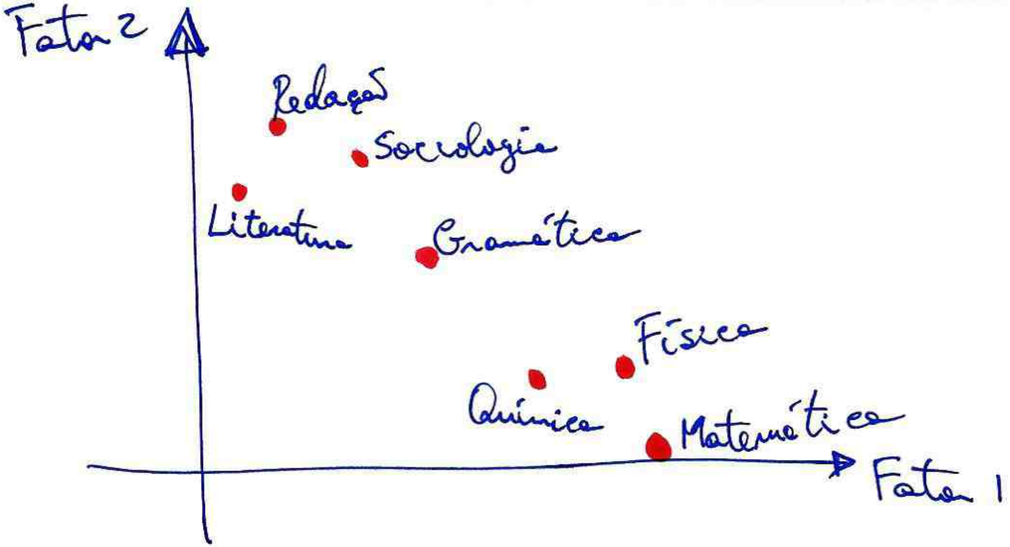
$$\mathrm{Var}(X_i) = l_{i1}^2 + l_{i2}^2 + \psi_i \approx l_{i1}^2 + \psi_i$$

- Toda a variabilidade das notas entre os alunos é devida às diferenças do fator verbal.
- Alunos ~~com~~ apenas o fator Quantitativo FQ muito diferente não terão notas muito distintas.

- ~~Suposição~~ Interpretando as cargas dos fatores

$$L = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \\ \vdots & \\ l_{15,1} & l_{15,2} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{l}_1 \\ \underset{\sim}{l}_2 \\ \vdots \\ \underset{\sim}{l}_{15} \end{bmatrix}$$

- Podemos plotar as linhas de L num gráfico planar.
- A primeira coordenada (fator 1) no eixo horizontal e a segunda coordenada no eixo vertical (fator 2)



- Esta representação mostra que as disciplinas Física, Química, Matemática ⇒ possuem cargas altas no Fator 1 e cargas baixas no Fator 2

- Redação, Literatura e Sociologia ⇒ pouca carga do Fator 1 e muita carga do Fator 2.

⊕ Isto indica que, nesta disposição de colunas da matriz L, a 1ª coluna (ou 1ª coordenada das linhas) representa o fator quantitativo

⊕ A segunda coluna de L representa o fator verbal.

⊕ Observe que "Gramática" ficou a meio caminho, ~~representado~~ com cargas medianas nos dois fatores. Para ter nota alta em "Gramática" é preciso ter "doses" razoáveis dos dois fatores <u>OU</u> uma "dose" bem grande de um dos fatores, qualquer um deles.

MÉTODOS DE ESTIMAÇÃO

1) MÁXIMA VEROSSIMILHANÇA (VEREMOS + TARDE NO CURSO)

2) COMPONENTES PRINCIPAIS

- Veremos apenas o segundo método.
- Pelo teorema espectral, $\Sigma = P \Lambda P$ onde $P = \left[\underset{\sim}{v}_1 \,\middle|\, - \,\middle|\, \underset{\sim}{v}_p \right]$ são os autovetores de Σ e $\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix}$ é matriz diagonal com os autovalores
- Manipulação matricial permite escrever

$$\Sigma = L^*(L^*)' = \left[\sqrt{\lambda_1}\, \vec{\underset{\sim}{v}}_1 \,\middle|\, \cdots \,\middle|\, \sqrt{\lambda_p}\, \vec{\underset{\sim}{v}}_p \right] \begin{bmatrix} \sqrt{\lambda_1}\, \underset{\sim}{v}_1' \\ \hline \vdots \\ \hline \sqrt{\lambda_p}\, \underset{\sim}{v}_p' \end{bmatrix}$$

- Suponha que os ~~xxxxx~~ últimos autovalores sejam ≈ 0.

- Isto implica que as últimas colunas de L^* são $\approx$ nulas e podem ser ignoradas.
- Mais formalmente, suponha que a soma dos k primeiros autovalores seja praticamente igual à soma de todos os p autovalores:

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_k + \lambda_{k+1} + \cdots + \lambda_p} \approx 1$$

- Ignorando as últimas colunas da matriz $\underset{p \times p}{L^*}$ ficamos com uma matriz $L_{p \times k}$:

$$\Sigma \approx L L' = \left[\sqrt{\lambda_1}\, \underset{\sim}{v}_1 \middle| \cdots \middle| \sqrt{\lambda_k}\, \underset{\sim}{v}_k \right] \begin{bmatrix} \sqrt{\lambda_1}\, \underset{\sim}{v}_1' \\ \hline \vdots \\ \hline \sqrt{\lambda_k}\, \underset{\sim}{v}_k' \end{bmatrix}$$

Para completar o modelo fatorial, estimamos a matriz diagonal

$$\Psi = \begin{bmatrix} \psi_1 & & \\ & \ddots & \\ & & \psi_p \end{bmatrix} = \mathrm{diag}\left[\Sigma - LL'\right]$$

Isto é, $\psi_i = \Sigma_{ii} - (LL')_{ii}$

RESUMO PRÁTICO: • Matriz de dados $\underset{n \times p}{X}$

• Obtenha $S = \mathrm{cov}(X)$

• Alternativamente, se os s_{ii} forem muito distintos, usamos a matriz $R = \mathrm{cor}(x)$, a matriz de correlaç.

• Obtenhos os autovalores ordenados e os autovetores de S

- Calcule a soma acumulada

$$\frac{\lambda_1 + \dots + \lambda_k}{\lambda_1 + \dots + \lambda_k + \lambda_{k+1} + \dots + \lambda_p} = a_k$$

- Se $a_k \approx 1$ com k pequeno então o modelo fatorial pode ser usado pois vai simplificar a estrutura dos dados
- Use os primeiros k autovetores (tal que $a_k \approx 1$) para criar a matriz de cargas

$$L = \left[\sqrt{\lambda_1}\, \underset{\sim}{v}_1 \,\middle|\, \dots \,\middle|\, \sqrt{\lambda_k}\, \underset{\sim}{v}_k \right] \quad e \quad \psi = \begin{bmatrix} \psi_1 & & \\ & \ddots & \\ & & \psi_p \end{bmatrix} \text{ onde } \psi_i = S_{ii} - (LL')_{ii}$$

- Um bom critério de escolha de k é verificar que a soma das entradas ao quadrado da matriz $\left(S - (LL' + \psi)\right) \leq \lambda_{k+1}^2 + \dots + \lambda_p^2$

Assim, se $\lambda_{k+1}^2 + \dots + \lambda_p^2 \approx 0 \Rightarrow S \approx LL' + \psi$ e o modelo fatorial é um bom ajuste

Exemplo 9.3 de Johnson & Wichern: Em um estudo sobre preferência de consumidores, uma amostra de clientes foram questionados sobre um novo produto. (34) As respostas numa escala de 7-pontos foram ~~differential scale~~ were tabulated and the attribute correlation matrix constructed. The correlation matrix is presented next.

$$
\begin{array}{lc}
\textit{Attribute (Variable)} & \\
\text{Taste} & 1 \\
\text{Good buy for money} & 2 \\
\text{Flavor} & 3 \\
\text{Suitable for snack} & 4 \\
\text{Provides lots of energy} & 5
\end{array}
\begin{array}{c}
\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{bmatrix}
1.00 & .02 & .96 & .42 & .01 \\
.02 & 1.00 & .13 & .71 & .85 \\
.96 & .13 & 1.00 & .50 & .11 \\
.42 & .71 & .50 & 1.00 & .79 \\
.01 & .85 & .11 & .79 & 1.00
\end{bmatrix}
\end{array}
$$

It is clear from the circled entries in the correlation matrix that variables 1 and 3 and variables 2 and 5 form groups. Variable 4 is "closer" to the (2, 5) group than the (1, 3) group. Given these results and the small number of variables, we might expect that the apparent linear relationships between the variables can be explained in terms of, at most, two or three common factors.

The first two eigenvalues $\hat{\lambda}_1 = 2.85$ and $\hat{\lambda}_2 = 1.81$ of **R** are the only eigenvalues greater than unity. Moreover, $m = 2$ common factors will account for a cumulative proportion

$$\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{p} = \frac{2.85 + 1.81}{5} = .93$$

of the total (standardized) sample variance. The estimated factor loadings, communalities, and specific variances, obtained using (9-15), (9-16), and (9-17), are given in Table 9.1.

TABLE 9.1

Variable	Estimated factor loadings $\tilde{\ell}_{ij} = \sqrt{\hat{\lambda}_i}\,\hat{e}_{ij}$ F_1	F_2	Communalities $\tilde{h}_i^2$	Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. Taste	.56	.82	.98	.02
2. Good buy for money	.78	−.53	.88	.12
3. Flavor	.65	.75	.98	.02
4. Suitable for snack	.94	−.11	.89	.11
5. Provides lots of energy	.80	−.54	.93	.07
Eigenvalues	2.85	1.81		
Cumulative proportion of total (standardized) sample variance	.571	.932		

(34)

Now

$$\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\boldsymbol{\Psi}} = \begin{bmatrix} .56 & .82 \\ .78 & -.53 \\ .65 & .75 \\ .94 & -.10 \\ .80 & -.54 \end{bmatrix} \begin{bmatrix} .56 & .78 & .65 & .94 & .80 \\ .82 & -.53 & .75 & -.10 & -.54 \end{bmatrix}$$

$$+ \begin{bmatrix} .02 & 0 & 0 & 0 & 0 \\ 0 & .12 & 0 & 0 & 0 \\ 0 & 0 & .02 & 0 & 0 \\ 0 & 0 & 0 & .11 & 0 \\ 0 & 0 & 0 & 0 & .07 \end{bmatrix} = \begin{bmatrix} 1.00 & .01 & .97 & .44 & .00 \\ & 1.00 & .11 & .79 & .91 \\ & & 1.00 & .53 & .11 \\ & & & 1.00 & .81 \\ & & & & 1.00 \end{bmatrix}$$

nearly reproduces the correlation matrix **R**. Thus on a purely descriptive basis, we would judge a two-factor model with the factor loadings displayed above as providing a good fit to the data. The communalities (.98, .88, .98, .89, .93) indicate that the two factors account for a large percentage of the sample variance of each variable.

We shall not interpret the factors at this point. As we noted in Section 9.2, the factors (and loadings) are unique up to an orthogonal rotation. A rotation of the factors often reveals a simple structure and aids interpretation. We shall consider this example again (Example 9.9 and Panel 9.1) after factor rotation has been discussed. ■

Example 9.4

Stock-price data consisting of $n = 100$ weekly rates of return on $p = 5$ stocks were introduced in Example 8.5. In that example the first two sample principal components were obtained from **R**. Taking $m = 1$ and $m = 2$, principal component solutions to the orthogonal factor model can be easily obtained. Specifically, the estimated factor loadings are the sample principal component coefficients (eigenvectors of **R**) scaled by the square root of the corresponding eigenvalues. The estimated factor loadings, communalities, specific variances, and proportion of total (standardized) sample variance explained by each factor for the $m = 1$ and $m = 2$ factor solutions are displayed in Table 9.2. The communalities are given by (9-17). So, for example, with $m = 2$, $\tilde{h}_1^2 = \tilde{\ell}_{11}^2 + \tilde{\ell}_{12}^2 = (.783)^2 + (-.217)^2 = .66$.

The residual matrix corresponding to the solution for $m = 2$ factors is

$$\mathbf{R} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}' - \tilde{\boldsymbol{\Psi}} = \begin{bmatrix} 0 & -.127 & -.164 & -.069 & .017 \\ -.127 & 0 & -.122 & .055 & .012 \\ -.164 & -.122 & 0 & -.019 & -.017 \\ -.069 & .055 & -.019 & 0 & -.232 \\ .017 & .012 & -.017 & -.232 & 0 \end{bmatrix}$$

TABLE 9.2

Variable	One-factor solution: Estimated factor loadings F_1	One-factor solution: Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$	Two-factor solution: Estimated factor loadings F_1	Two-factor solution: Estimated factor loadings F_2	Two-factor solution: Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. Allied Chemical	.783	.39	.783	−.217	.34
2. du Pont	.773	.40	.773	−.458	.19
3. Union Carbide	.794	.37	.794	−.234	.31
4. Exxon	.713	.49	.713	.472	.27
5. Texaco	.712	.49	.712	.524	.22
Cumulative proportion of total (standardized) sample variance explained	.571		.571	.733	

The proportion of total variance explained by the two-factor solution is appreciably larger than that for the one-factor solution. However, for $m = 2$, $\tilde{\mathbf{L}}\tilde{\mathbf{L}}'$ produces numbers that are, in general, larger than the sample correlations. This is particularly true for r_{45}.

It seems fairly clear that the first factor, F_1, represents general economic conditions and might be called a *market factor*. All of the stocks load highly on this factor and the loadings are about equal. The second factor contrasts the chemical stocks with the oil stocks (the chemicals have relatively large negative loadings and the oils have large positive loadings on the factor). Thus F_2 seems to differentiate stocks in different industries and might be called an *industry factor*. To summarize, rates of return appear to be determined by general market conditions and activities that are unique to the different industries, as well as a residual or firm specific factor. This is essentially the conclusion reached by an examination of the sample principal components in Example 8.5. ■

A Modified Approach—The Principal Factor Solution

A modification of the principal component approach is sometimes considered. We describe the reasoning in terms of a factor analysis of $\mathbf{R}$, although the procedure is also appropriate for $\mathbf{S}$. If the factor model $\boldsymbol{\rho} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ is correctly specified, the m common factors should account for the *off-diagonal* elements of $\boldsymbol{\rho}$, as well as the *communality portions* of the diagonal elements

$$\rho_{ii} = 1 = h_i^2 + \psi_i$$

If the specific factor contribution ψ_i is removed from the diagonal or, equivalently, the 1 replaced by h_i^2, the resulting matrix is $\boldsymbol{\rho} - \boldsymbol{\Psi} = \mathbf{L}\mathbf{L}'$.

⊕ Existe um problema de identificabilidade na determinação do modelo fatorial I_1

⊕ O problema é que a matriz de cargas L só pode ser conhecida a menos de uma rotação.

⊕ Seja $\underset{2\times 2}{T}$ uma matriz ortogonal

Isto é, $TT' = T'T = I_2$ = identidade $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

⊕ Da álgebra de matrizes, sabemos que matrizes ortogonais correspondem a uma rotação rígida dos eixos coordenados.

⊕ Isto significa que uma matriz $\underset{2\times 2}{T}$ tal que $TT' = T'T = I$ tem de ser da seguinte forma:

$$T = \begin{bmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{bmatrix} \qquad \text{ou} \qquad T = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}$$

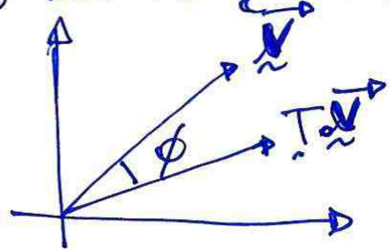
rotação clockwise — rotação counter-clockwise

para $\phi \in [0, 2\pi]$

⊕ Estas matrizes correspondem a rotações no plano.

⊕ Seja $\underset{2\times1}{\vec{v}} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ e $\underset{2\times2}{T} = \begin{bmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{bmatrix}$

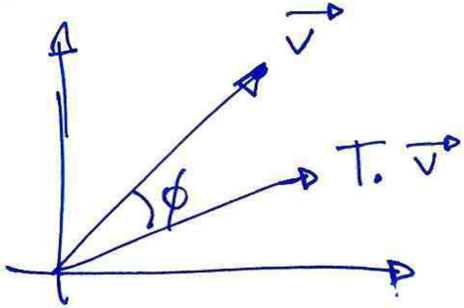
⊕ Então $\underset{2\times2}{T} \cdot \underset{2\times1}{\vec{v}}$ é um novo ponto no $\mathbb{R}^2$ obtido rotacionando $\vec{v}$ pelo ângulo ϕ na direção do relógio:



Considerando <u>vetores-linha</u>:

$$(T.\vec{v})' = \vec{v}'.T' = (x \;\; y)\begin{bmatrix}\cos(\phi) & -\mathrm{sen}(\phi)\\ \sin(\phi) & \cos(\phi)\end{bmatrix}$$

~~Considere~~



A situação geométrica continua a mesma de antes, representar o vetor como linha ou coluna não altera o ~~sentido~~ o significado

⊕ Vamos trabalhar com as <u>linhas</u> da matriz L

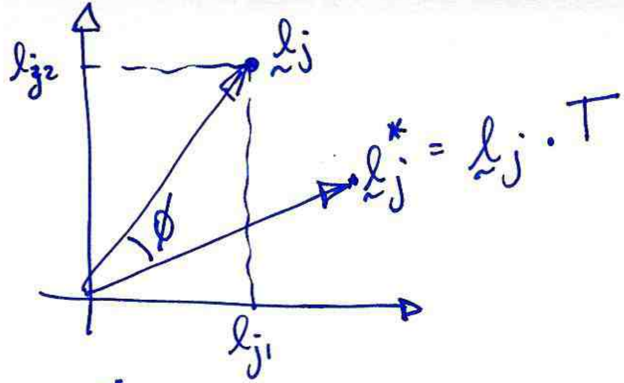
Considere a matriz L das cargas

$$L = \begin{bmatrix} \underset{\sim}{\ell}_1' \\ \vdots \\ \underset{\sim}{\ell}_{15}' \end{bmatrix} = \begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \\ \vdots & \\ \ell_{15,1} & \ell_{15,2} \end{bmatrix}$$ e T = matriz ortogonal

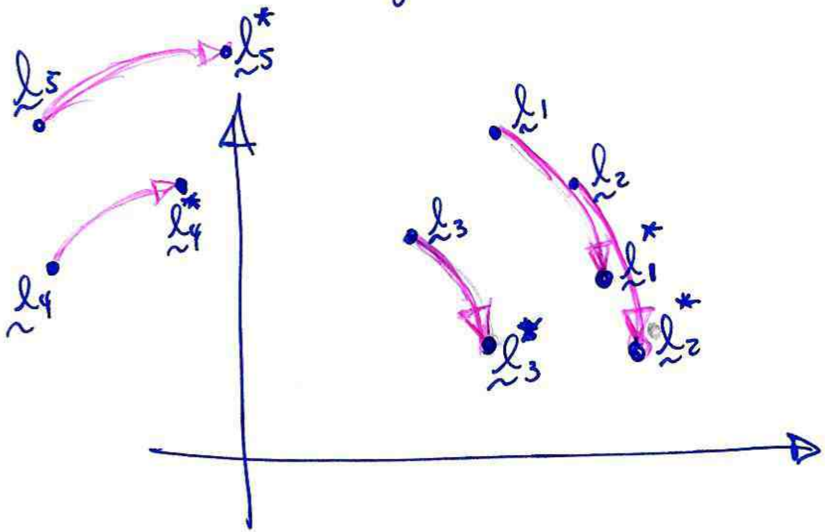
O produto $\underbrace{\underset{15\times 2}{L} \cdot \underset{2\times 2}{T}}_{15\times 2}$ pode ser pensado linha-a-linha

$$\underset{15\times 2}{L} \cdot \underset{2\times 2}{T} = \begin{bmatrix} \underset{\sim}{\ell}_1' \\ \underset{\sim}{\ell}_2' \\ \vdots \\ \underset{\sim}{\ell}_{15}' \end{bmatrix} = \begin{bmatrix} \underbrace{\underset{\sim}{\ell}_1'}_{1\times 2} \cdot \underbrace{T}_{2\times 2} \\ \underset{\sim}{\ell}_2' \cdot T \\ \vdots \\ \underset{\sim}{\ell}_{15}' \cdot T \end{bmatrix} = L^* \Rightarrow$$

As linhas de L^* são as linhas de L rotacionadas de certo ângulo ϕ associado à matriz T.



Rodando uma linha de L



Rodando cinco linhas de L

OK, o que tudo isto quer dizer?

⊕ Suponha que o modelo fatorial é correto e que realmente podemos (escrever ou decompor) a matriz de covariância de $\underset{\sim}{X}$ como:

$$\text{Var}(\underset{\sim}{X}) = \underset{15\times 5}{\Sigma} = \underset{15\times 2}{L}\,\underset{2\times 15}{L'} + \underset{15\times 15}{\Psi} \longrightarrow \text{diagonal}$$

Seja $\underset{2\times 2}{T}$ **qualquer** matriz ortogonal (de rotação no plano).

Então podemos escrever

$$\Sigma = LL' + \Psi = L\,\overbrace{TT'}^{I_2}\,L' + \Psi = \underbrace{(LT)}_{\underset{15\times 2}{L^*}}(LT)' + \Psi$$

$$= \underset{15\times 2}{L^*}\,\underset{2\times 15}{(L^*)'} + \Psi$$

Isto significa que, se <u>tivermos apenas Σ</u>,

teremos
$$LL' + \psi = \Sigma = L^*(L^*)' + \psi$$

onde $L^* = LT$ é diferente de L

~~As colunas de~~

- As linhas de L^* são as linhas de L rotacionadas de um ~~vetor ϕ~~ ângulo ϕ.
- Como T é arbitrária (pode ser <u>qualquer</u> T) isto significa que podemos rodar L à vontade, com qualquer ângulo ϕ, que <u>sempre</u> teremos uma representação de Σ da forma $\Sigma = L^*(L^*)' + \psi$.

- Mas então como interpretar os números que aparecem em L ?
- Qual a L "correta" ?
- Não é possível determinar uma única L tal que $\Sigma = L L' + \Psi$
- Existem infinitos L com esta propriedade.
- Qualquer $L^* = L.T$ (isto é, L rotacionada) terá a mesma propriedade.
- Todas as matrizes de carga L^* obtidas a partir de uma matriz L inicial terão a mesma capacidade de reproduzir a matriz de covariância Σ

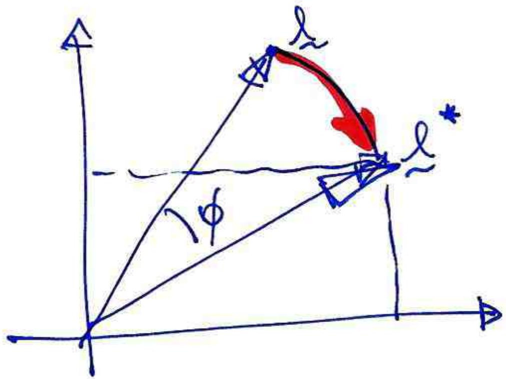
- Ao invés disso se tornar um problema, transformamos os limões numa limonada.
- ~~Como uma~~ Caso uma matriz de cargas L inicialmente obtida por algum método de estimação não fornecer uma boa interpretação para os fatores, nós procuramos uma versão rotacionada $L^* = L.T$ tal que as novas cargas sejam mais interpretáveis
- É comum sermos capazes de terminar com uma estrutura + simples que a matriz L inicial
- Qual é esta estrutura mais simples?

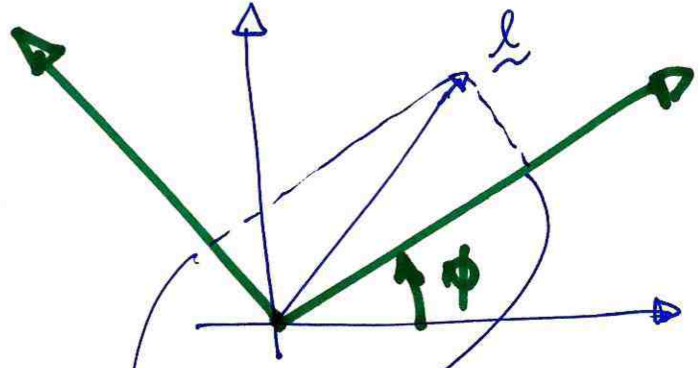
I10

- Idealmente, nós gostaríamos de ver um padrão em que cada variável tenha uma carga alta num dos fatores e uma carga $\simeq 0$ nos demais.

- O objetivo é procurar uma rotação dos eixos de forma que as novas cargas fiquem o mais próximo possível deste ideal.

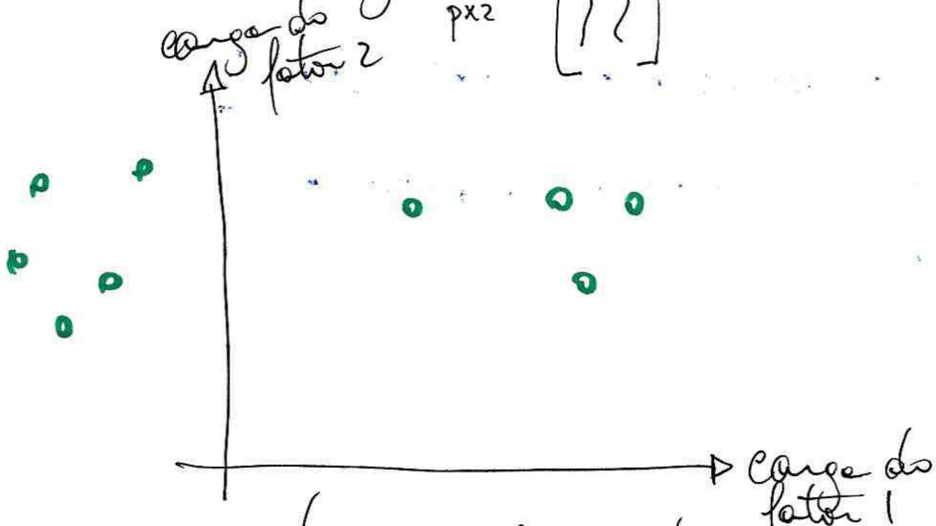
- $\boxed{\text{OBS}}$: Se temos 15 pontos no plano (as cargas $\underset{\sim}{\ell}_j$) e rodamos todos eles de um ângulo ϕ, isto é o mesmo que rodar os dois eixos do plano de $-\phi$ e deixar os "pontos intactos"



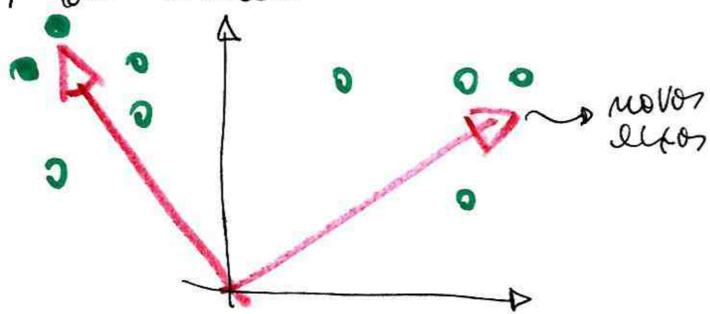


As coordenadas nos novos eixos são as mesmas de $\underset{\sim}{l}^*$ nos eixos antigos

Se nossas cargas $L_{p\times 2} = \begin{bmatrix} \vdots & \vdots \end{bmatrix}$ são assim:



procuramos rodar os eixos até que as cargas sejam próximas do ideal



Nos novos eixos, as cargas são ≈ 0 exceto em um único fator.

- Procedimento VARIMAX
- Defina $\tilde{l}_{ij}^{*} = \dfrac{l_{ij}^{*}}{h_i} = \dfrac{l_{ij}^{*}}{l_{i1}^{*} + l_{i2}^{*} + \dots + l_{ip}^{*}} \in [0,1]$

as cargas dos fatores rotacionados e ~~pes~~ normalizados

- Busque a rotaç T tal que maximize

$$V = \frac{1}{p} \sum_{j=1}^{m} \left[\sum_{i=1}^{p} \tilde{l}_{ij}^{*4} - \left(\sum_{i=1}^{p} \tilde{l}_{ij}^{*2} \right)^2 \Big/ p \right]$$

$\propto \sum_{j=1}^{m}$ (Variância das $(\text{cargas})^2$ normalizados do fator j)

Maximizar V significa espalhar as $(\text{cargas})^2$ o máximo possível, com valores altos em alguns fatores e valores ≈ 0 em outros

⊕ Tendo obtido a matriz de cargas, podemos estimar o valor <u>dos fatores</u> de cada indivíduo da amostra.

⊕ Suponha que o i-ésimo indivíduo tenha o vetor $\underset{\sim}{X}_i$ e que tenhamos estimado $\underset{\sim}{\mu}$ (a média das variáveis sobre a amostra) e tenhamos também a matriz de cargas L (talvez rotacionada).

⊕ O vetor $\underset{\sim}{F}_i$ <u>deste indivíduo</u> é estimado pela minimização da ~~comparação~~ diferença entre $\underset{\sim}{X}_i$ e $\underset{\sim}{\mu} + L\underset{\sim}{F}$.

Isto é, procuramos um vetor $\underset{\sim}{F}_i$ tal que ele minimize o comprimento

$$\| \underset{\sim}{X}_i - \underset{\sim}{\mu} - L \underset{\sim}{F}_i \|^2$$

⊕ Veja a lista de exercícios (example) para um exemplo.