Factor Analysis

Max Turgeon

STAT 4690–Applied Multivariate Analysis

Latent variable models

- With PCA, we saw how we could reduce the dimension of data using the eigenvectors of the sample covariance matrix.
- Conversely, we could construe PCA has a generative model, where the principal components give rise to the observed data.
- **Latent Variable Models** formalise this idea:
    - Latent (i.e. unobserved) variables $\mathbf{F}$ give rise to observed data $\mathbf{Y}$ through a *specified* model.

Factor Analysis i

- **Factor Analysis** is a special kind of latent variable model.
- Let $\mathbf{Y}$ be a p-dimensional vector with mean μ and covariance matrix Σ.
- Let $\mathbf{F}$ be a m-dimensional *latent* vector.
- The *orthogonal factor model* is given by

$$\mathbf{Y} - \mu = L\mathbf{F} + \mathbf{E},$$

where L is a $p \times m$ *matrix of factor loadings*, and $\mathbf{E}$ is a p-dimensional vector of *errors*.

Factor Analysis ii

- $\mathbf{F}$ are also called *common factors*; $\mathbf{E}$ are also called *specific factors*.
- **Note**: This is essentially a multivariate regression model, but where the covariates are unobserved.

Assumptions i

- The model above is generally not identifiable, since there are too many parameters.
- We therefore need to impose further restrictions:
  - $E(\mathbf{F}) = 0$
  - $\mathrm{Cov}(\mathbf{Y}) = I$
  - $E(\mathbf{E}) = 0$
  - $\mathrm{Cov}(\mathbf{Y}) = \Psi = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{pmatrix}$
  - $\mathrm{Cov}(\mathbf{F}, \mathbf{E}) = 0$

Assumptions ii

- In other words:
- Both common and specific factors have mean zero;
- They are uncorrelated;
- The common factors are mutually uncorrelated and standardised;
- The specific factors each affect only one observed variable.

Structured Covariance i

- As a consequence of these assumptions, we can derive an assumption on the structure of $\Sigma = \mathrm{Cov}(\mathbf{Y})$:

$$
\begin{aligned}
\Sigma &= E\left((\mathbf{Y} - \mu)(\mathbf{Y} - \mu)^T\right) \\
&= E\left((L\mathbf{F} + \mathbf{E})(L\mathbf{F} + \mathbf{E})^T\right) \\
&= LE(\mathbf{F}\mathbf{F}^T)L + E(\mathbf{E}\mathbf{F}^T)L^T + LE(\mathbf{F}\mathbf{E}^T) + E(\mathbf{E}\mathbf{E}^T) \\
&= LIL^T + 0L^T + L0 + \Psi \\
&= LL^T + \Psi.
\end{aligned}
$$

Structured Covariance ii

- Similarly, we can show that

$$\mathrm{Cov}(\mathbf{Y}, \mathbf{F}) = L.$$

- If we write ℓ_{ij} for the (i, j)-th element of L, we see that

$$\mathrm{Var}(Y_i) = \sum_{k=1}^{m} \ell_{ik}^2 + \psi_i.$$

- Crucially, these assumptions are **testable**. In other words, we can check whether they are reasonable for our data.

Example i

- Let's look at an example where there is no solution.
- Assume $p = 3$, $m = 1$, with

$$\Sigma = \begin{pmatrix} 1 & 0.9 & 0.7 \\ 0.9 & 1 & 0.4 \\ 0.7 & 0.4 & 1 \end{pmatrix}.$$

- From our assumptions on the covariance structure, we derive a system of equations

$$\begin{matrix} 1 = \ell_{11}^2 + \psi_1 & 0.9 = \ell_{11}\ell_{21} & 0.7 = \ell_{11}\ell_{31} \\ & 1 = \ell_{22}^2 + \psi_2 & 0.4 = \ell_{21}\ell_{31} \\ & & 1 = \ell_{33}^2 + \psi_3 \end{matrix}$$

Example ii

- From $0.7 = \ell_{11}\ell_{31}$ and $0.4 = \ell_{21}\ell_{31}$, we get

$$\ell_{21} = \frac{0.4}{0.7}\ell_{11}.$$

- But since $0.9 = \ell_{11}\ell_{21}$, we can conclude that

$$\ell_{11} = \pm 1.255.$$

- However, since the first component Y_1 has unit variance, $\ell_{11} = \mathrm{Corr}(Y_1, F_1)$, and therefore the correlation is out of bounds.
- Similarly, we get

$$\psi_1 = 1 - \ell_{11}^2 = 1 - 1.575 = -0.575.$$

Example iii

- But since ψ_1 is the variance of the first error term, we once again get a non-sensical solution.

Factor Rotation i

- Even with our assumptions above, our model is still not uniquely identified.
- Let T be an $m \times m$ orthogonal matrix. We have

$$\begin{aligned}\mathbf{Y} - \mu &= L\mathbf{F} + \mathbf{E}\\ &= LTT^T\mathbf{F} + \mathbf{E}\\ &= \tilde{L}\tilde{\mathbf{F}} + \mathbf{E},\end{aligned}$$

where $\tilde{L} = LT$ and $\tilde{\mathbf{F}} = T^T\mathbf{F}$.

Factor Rotation ii

- Both models lead to the same covariance matrix:

$$\Sigma = LL^T + \Psi = LTT^T L^T + \Psi = \tilde{L}\tilde{L}^T + \Psi.$$

- As we will see, this will turn out to be a blessing in disguise:
    - We will impose a uniqueness condition to get one solution.
    - Then we will rotate our solution using T to improve interpretation.

Estimation–Principal Component Method i

- Recall the spectral decomposition of the covariance matrix:
$$\Sigma = \sum_{i=1}^p \lambda_i w_i w_i^T,$$
with $\lambda_1 \geq \cdots \geq \lambda_p$.
- If we let W be the matrix whose i-th column is $\sqrt{\lambda_i} w_i$, we can rewrite the spectral decomposition as
$$\Sigma = WW^T.$$
- In other words, if we let $m = p$ and $\Psi = 0$, we see that we recover the orthogonal factor model with $L = W$.

Estimation–Principal Component Method ii

- Of course, this is not very satisfactory, as the dimension of the common factors is the same as that of the original data.
- Instead, we select $m < p$ using one of the methods we discussed with PCA and we approximate

$$\Sigma \approx \sum_{i=1}^{m} \lambda_i w_i w_i^T.$$

- If we let L be the $p \times m$ matrix whose i-th column is $\sqrt{\lambda_i} w_i$, we can estimate Ψ as follows:

$$\psi_i = \sigma_{ii} - \sum_{j=1}^{m} \ell_{ij}^2.$$

Estimation–Principal Component Method iii

Algorithm

1. Let $\hat{\lambda}_1 \cdots > \hat{\lambda}_p$ and $\hat{w}_1, \hat{w}_p$ be the eigenvalues and eigenvectors of the covariance matrix S_n.
2. Select m using one of the PCA criteria.
3. Estimate $\hat{L}$ with the $p \times m$ matrix whose i-th column is $\sqrt{\hat{\lambda}_i}\hat{w}_i$.
4. Estimate $\hat{\Psi}$ with the diagonal elements of $S_n - \hat{L}\hat{L}^T$.

Example i

```r
library(psych)

dim(bfi)
```

```
## [1] 2800   28
```

```r
names(bfi)
```

Example ii

```
##  [1] "A1"        "A2"        "A3"        "A4"
##  [6] "C1"        "C2"        "C3"        "C4"
## [11] "E1"        "E2"        "E3"        "E4"
## [16] "N1"        "N2"        "N3"        "N4"
## [21] "O1"        "O2"        "O3"        "O4"
## [26] "gender"    "education" "age"
```

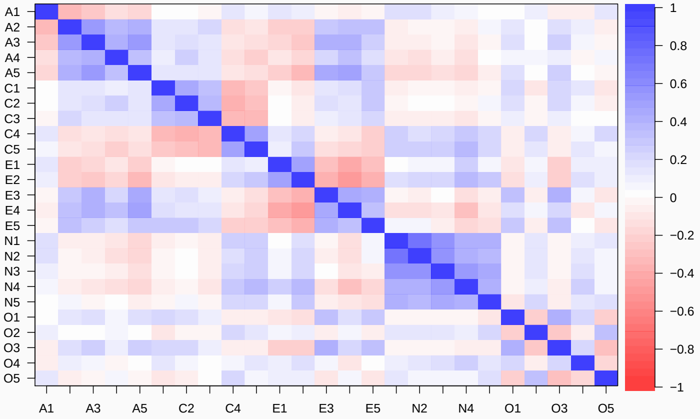
```
library(tidyverse)

data <- bfi %>%
  select(-gender, -education, -age) %>%
  filter(complete.cases(.))
```

Example iii

```
cor.plot(cor(data))
```

Example iv


A1
A2
A3
A4
A5
C1
C2
C3
C4
C5
E1
E2
E3
E4
E5
N1
N2
N3
N4
N5
O1
O2
O3
O4
O5
A1
A3
A5
C2
C4
E1
E3
E5
N2
N4
O1
O3
O5
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

Example v

```r
decomp <- prcomp(data)
summary(decomp)$importance[,1:3]
```

```
##                             PC1      PC2      PC3
## Standard deviation     3.291635 2.451538 2.030393
## Proportion of Variance 0.215650 0.119620 0.082050
## Cumulative Proportion  0.215650 0.335270 0.417320
```

Example vi

```r
cum_prop <- decomp %>%
  summary %>%
  .[["importance"]] %>%
  .["Cumulative Proportion",]

which(cum_prop > 0.8)
```

```
## PC14 PC15 PC16 PC17 PC18 PC19 PC20 PC21 PC22 PC23 PC
##   14   15   16   17   18   19   20   21   22   23
```

Example vii

```
Lhat <- decomp$rotation[,1:14] %*%
  diag(decomp$sdev[1:14])
Psi_hat <- diag(cov(data) - tcrossprod(Lhat))

sum((cov(data) - tcrossprod(Lhat) - diag(Psi_hat))^2)
```

```
## [1] 3.645694
```

```
sum(diag(cov(data)))
```

```
## [1] 50.24287
```