STAT 3690 Lecture Note

Part VIII: Factor analysis

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca, zhiyanggeezhou.github.io)

2023/Mar/12 20:18:01

Factor analysis

Latent variable model

- latent/unobservable variables give rise to observed data through a specific model, i.e., a regression model with unobservable covariates
- Factor analysis model is a special kind of latent variable model

Population version

- Model

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

  - $\boldsymbol{Y} = [Y_1, \ldots, Y_p]^\top \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$: random & observable p-vector
  - $\mathbf{L} = [\ell_{ij}]_{p\times q}$: fixed & unknown, a matrix of factor loadings
  - $\boldsymbol{F} \sim (\mathbf{0}, \mathbf{I}_q)$: random & unobservable, q-vector of latent/common factors
  - $\boldsymbol{E} \sim (\mathbf{0}, \boldsymbol{\Psi})$: random & unobservable, p-vector of error/specific factors, with $\boldsymbol{\Psi} = \operatorname{diag}(\psi_1, \ldots, \psi_p)$ and $\operatorname{cov}(\boldsymbol{F}, \boldsymbol{E}) = \mathbf{0}$
- Covariance structure
  - $\operatorname{var}(\boldsymbol{Y}) = \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^\top + \boldsymbol{\Psi}$
    * I.e., $\operatorname{var}(Y_i) = \sum_{j=1}^q \ell_{ij}^2 + \psi_i$
  - $\operatorname{cov}(\boldsymbol{Y}, \boldsymbol{F}) = \mathbf{L}$
  - $\sum_{i=1}^p \ell_{ij}^2$: the variability contributed by the jth latent factor

Sample version

- Model

$$\boldsymbol{Y}_i - \boldsymbol{\mu} = \mathbf{L}\boldsymbol{F}_i + \boldsymbol{E}_i, \quad i = 1, \ldots, n$$

  - $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \overset{\text{iid}}{\sim} \boldsymbol{Y}$
  - $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_n \overset{\text{iid}}{\sim} \boldsymbol{F}$
  - $\boldsymbol{E}_1, \ldots, \boldsymbol{E}_n \overset{\text{iid}}{\sim} \boldsymbol{E}$

Estimating $\mathbf{L}$ and $\boldsymbol{\Psi}$

- Selection of q, i.e., the number of latent factors, with one of the following rules
  - PCA stopping rule
  - Taking q such that $\sum_{i,j} \ell_{ij}^2/\operatorname{tr}(\mathbf{S})$ is over a preset percentage, where $\mathbf{S} = (n-1)^{-1}\sum_{i=1}^n (\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})^\top$

    * $\sum_{i=1}^q \ell_{ij}^2/\mathrm{tr}(\mathbf{S})$: the proportion of variation explained by the jth latent factor
  - Taking q as the number of positive eigenvalues of $\mathbf{S}$
  - Taking q as the number of eigenvalues of $\mathbf{S}$ that are above average
  - Taking q as the number of eigenvalues of correlation matrix greater than one
  - According to domain-knowledge expertise
- PC method
  1. Pick up q largest eigenvalues $\lambda_1, \dots, \lambda_q$ of $\mathbf{S}$ and corresponding eigenvectors $w_1, \dots, w_q$
  2. $\widehat{\mathbf{L}} = [\sqrt{\lambda_1}w_1, \dots, \sqrt{\lambda_q}w_q]_{p\times q}$ and $\widehat{\mathbf{\Psi}} = \mathrm{diag}(\mathbf{S} - \widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)$

- Exercise 8.1: `psych::bfi` involves 2800 subjects, covering their 25 personality assessments, gender, education and age.
- ML method
  - Assuming
    * $\boldsymbol{F} \sim \mathrm{MVN}_q(\mathbf{0}, \mathbf{I})$
    * $\boldsymbol{E} \sim \mathrm{MVN}_p(\mathbf{0}, \mathbf{\Psi})$
    * Diagonal $\mathbf{L}^\top\mathbf{\Psi}^{-1}\mathbf{L}$
  - Resorting to `R` functions `factanal` or `psych::fa`

- Comments on the estimation of $\mathbf{L}$ and $\mathbf{\Psi}$
  - Other methods
  - Different statistical softwares may apply different methods
    * Have to look into help manuals to figure out what is going on for different softwares/packages
  - Compare the outputs of multiple estimation methods
    * For a good fit, similar answers would be reached regardless of the method

---

Factor rotation

- $\mathbf{L}$ is not uniquely defined: if $\boldsymbol{Y} - \boldsymbol{\mu} = \mathbf{L}\boldsymbol{F} + \boldsymbol{E}$, then $\boldsymbol{Y} - \boldsymbol{\mu} = \widetilde{\mathbf{L}}\widetilde{\boldsymbol{F}} + \boldsymbol{E}$, where
  - $\widetilde{\mathbf{L}} = \mathbf{LP}$ and $\widetilde{\boldsymbol{F}} = \mathbf{P}^\top\boldsymbol{F}$ with $\mathbf{P}$ a $q\times q$ orthogonal matrix ($\mathbf{P}^{-1} = \mathbf{P}^\top$)
- A blessing to improve interpretation: pick up a $\mathbf{P}$ such that $\widetilde{\boldsymbol{F}}$ is more interpretable; to ease the interpretation, we want:
  - Each entry of $\boldsymbol{Y}$ to have large loadings for merely one latent factor and negligible loadings for remaining ones
- varimax: find $\mathbf{P}$ to maximize the sum of variance of squared (scaled) loadings over all the latent factors

$$\sum_{j=1}^q \left\{ \frac{1}{p}\sum_{i=1}^p \tilde{\ell}_{ij}^{*4} - \left(\frac{1}{p}\sum_{i=1}^p \tilde{\ell}_{ij}^{*2} \right)^2 \right\}$$

  - $\tilde{\ell}_{ij}^* = \tilde{\ell}_{ij}/\sqrt{\sum_{j=1}^q \tilde{\ell}_{ij}^2}$ with $\tilde{\ell}_{ij}$ the (i,j)-th entry of $\widetilde{\mathbf{L}} = \mathbf{LP}$

- Comments on factor rotation
  - Especially useful with loadings obtained through ML
  - Sometimes used even for PCA loadings

Factor scores

- Weighted least square (WLS) method
  - Given $\bar{\boldsymbol{Y}}$, $\widehat{\mathbf{L}}$, and $\widehat{\boldsymbol{\Psi}}$, then, for the ith observation $\boldsymbol{Y}_i$,

$$\widehat{\boldsymbol{F}}_i = (\widehat{\mathbf{L}}^\top \widehat{\boldsymbol{\Psi}}^{-1} \widehat{\mathbf{L}})^{-1} \widehat{\mathbf{L}}^\top \widehat{\boldsymbol{\Psi}}^{-1} (\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})$$

    * I.e., the minimizer of $(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}} - \widehat{\mathbf{L}}\boldsymbol{F})^\top \widehat{\boldsymbol{\Psi}}^{-1} (\boldsymbol{Y}_i - \bar{\boldsymbol{Y}} - \widehat{\mathbf{L}}\boldsymbol{F})$ with respect to $\boldsymbol{F}$

- Regression method
  - Assuming $\boldsymbol{F} \sim \text{MVN}_q(\mathbf{0}, \mathbf{I})$ and $\boldsymbol{E} \sim \text{MVN}_p(\mathbf{0}, \boldsymbol{\Psi})$,

$$\begin{bmatrix} \boldsymbol{Y} - \boldsymbol{\mu} \\ \boldsymbol{F} \end{bmatrix} \sim \text{MVN}_{p+q} \left(\mathbf{0}, \begin{bmatrix} \mathbf{L}\mathbf{L}^\top + \Psi & \mathbf{L} \\ \mathbf{L}^\top & \mathbf{I} \end{bmatrix} \right)$$

    and hence

$$\boldsymbol{F} \mid \boldsymbol{Y} \sim \text{MVN}_p(\mathbf{L}^\top (\mathbf{L}\mathbf{L}^\top + \boldsymbol{\Psi})^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}), \mathbf{I} - \mathbf{L}^\top (\mathbf{L}\mathbf{L}^\top + \boldsymbol{\Psi})^{-1} \mathbf{L})$$

  - Given $\bar{\boldsymbol{Y}}$, $\widehat{\mathbf{L}}$, and $\widehat{\boldsymbol{\Psi}}$, esitmate $\boldsymbol{F}_i$ by

$$\widehat{\boldsymbol{F}}_i = \widehat{\mathbf{L}}^\top (\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top + \widehat{\boldsymbol{\Psi}})^{-1} (\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})$$

    OR

$$\widehat{\boldsymbol{F}}_i = \widehat{\mathbf{L}}^\top \mathbf{S}^{-1} (\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})$$

- Comments on factor scores
  - More methods available
  - No uniformly superior way

Summary on factor analysis

- What we discussed is "exploratory" factor analysis
  - "Confirmatory" factor analysis would make stronger assumptions about the nature of the latent factors and perform statistical inference.
  - There are choices to make at every stage of factor analysis: estimation method, number of factors, factor rotation, and score estimation.
    * Too flexiable to be tracked
    * Close to an "art"
- General strategy for factor analysis
  1. Perform a PC factor analysis
     - It may help you identify potential outliers
  2. Perform an ML factor analysis.
     - Try a varimax rotation to see if it makes sense
  3. Compare the solutions of both methods to see if they generally agree.
  4. Repeat for different number of common factors q and check if adding more factors may improve the interpretation
  5. For large datasets, you can split your data, run the same model on both subsets, and compare the loadings to see if they generally agree

An example of factor analysis

- `state.x77` contains general information about all 50 US states
  - Population
  - Income per capita
  - Illiteracy rate
  - Life expectancy
  - Murder rate
  - High-school graduation rate
  - Average number of freezing degree days (with the temperature lower than 0 °C)
  - Total area