STAT 3690 Lecture 26

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Apr 01, 2022

Factor scores

- Weighted least square (WLS) method
  - Given $\bar{\mathbf{Y}}$, $\widehat{\mathbf{L}}$, and $\widehat{\boldsymbol{\Psi}}$
  - For the ith observation $\mathbf{Y}_i$, to minimize $(\mathbf{Y}_i - \bar{\mathbf{Y}} - \widehat{\mathbf{L}}\mathbf{F})^\top \widehat{\boldsymbol{\Psi}}^{-1}(\mathbf{Y}_i - \bar{\mathbf{Y}} - \widehat{\mathbf{L}}\mathbf{F})$ with respect to $\mathbf{F}$
  - $\widehat{\mathbf{F}}_i = (\widehat{\mathbf{L}}^\top \widehat{\boldsymbol{\Psi}}^{-1}\widehat{\mathbf{L}})^{-1}\widehat{\mathbf{L}}^\top \widehat{\boldsymbol{\Psi}}^{-1}(\mathbf{Y}_i - \bar{\mathbf{Y}})$

- Regression method
  - Under normality $\mathbf{F} \sim MVN_p(\mathbf{0}, \mathbf{I})$ and $\mathbf{E} \sim MVN_p(\mathbf{0}, \boldsymbol{\Psi})$
    - $[\mathbf{Y}^\top - \boldsymbol{\mu}^\top, \mathbf{F}^\top]^\top$ is of zero mean and normally distributed with covariance matrix

$$\begin{bmatrix} \mathbf{L}\mathbf{L}^\top + \Psi & \mathbf{L} \\ \mathbf{L}^\top & \mathbf{I} \end{bmatrix}$$

    - Hence $\mathbf{F} \mid \mathbf{Y}$ is normally distributed with mean $\mathbf{L}^\top(\mathbf{L}\mathbf{L}^\top + \boldsymbol{\Psi})^{-1}(\mathbf{Y} - \boldsymbol{\mu})$ and covariance matrix $\mathbf{I} - \mathbf{L}^\top(\mathbf{L}\mathbf{L}^\top + \boldsymbol{\Psi})^{-1}\mathbf{L}$
  - Given $\bar{\mathbf{Y}}$, $\widehat{\mathbf{L}}$, and $\widehat{\boldsymbol{\Psi}}$,

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

    - Sometimes replace $\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top + \widehat{\boldsymbol{\Psi}}$ with $\mathbf{S}$

- 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