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 $\widehat{\mathbf{\Psi}}$
  - For the ith observation $\mathbf{Y}_i$, to minimize $(\mathbf{Y}_i - \bar{\mathbf{Y}} - \widehat{\mathbf{L}}\mathbf{F})^\top \widehat{\mathbf{\Psi}}^{-1}(\mathbf{Y}_i - \bar{\mathbf{Y}} - \widehat{\mathbf{L}}\mathbf{F})$ with respect to $\mathbf{F}$
  - $\widehat{\mathbf{F}}_i = (\widehat{\mathbf{L}}^\top \widehat{\mathbf{\Psi}}^{-1}\widehat{\mathbf{L}})^{-1}\widehat{\mathbf{L}}^\top \widehat{\mathbf{\Psi}}^{-1}(\mathbf{Y}_i - \bar{\mathbf{Y}})$

```
install.packages(c('psych'))
library(psych)
library(tidyverse)
options(digits = 4)
head(psych::bfi)
data = bfi %>%
  select(-gender, -education, -age) %>%
  filter(complete.cases(.)) # Remove demographic variable and keep complete data

decomp = psych::fa(r=data, covar=T, nfactors=q, rotate="varimax", fm="ml", scores='Bartlett')

# Follow formula
L_ml = decomp$loadings
Psi_ml = diag(decomp$uniquenesses)
Psi_ml_inv = diag(decomp$uniquenesses^-1)
Weight_mat = solve(t(L_ml) %*% Psi_ml_inv %*% L_ml) %*% t(L_ml) %*% Psi_ml_inv
scores_wls1 = scale(data, center = T, scale = F) %*% t(Weight_mat)
head(scores_wls1)

# P.S. the `scores` component of `psych::fa` is correct only when
# factoring correlation matrix/standardized data
scores_wls2 = decomp$scores
head(scores_wls2)
```

- Regression method
  - Under normality $\mathbf{F} \sim MVN_p(\mathbf{0}, \mathbf{I})$ and $\mathbf{E} \sim MVN_p(\mathbf{0}, \mathbf{\Psi})$
    - $[\mathbf{Y}^\top - \boldsymbol{\mu}^\top, \mathbf{F}^\top]^\top$ 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 $\mathbf{S}$

```
decomp = psych::fa(r=data, covar=T, nfactors=q, rotate="varimax", fm ="ml", scores='regression')

# Follow formula
L_ml = decomp$loadings
Psi_ml = diag(decomp$uniquenesses)
Weight_mat = t(L_ml) %*% solve(cov(data))
scores_reg1 = scale(data, center = T, scale = F) %*% t(Weight_mat)
head(scores_reg1)

# P.S. the `scores` component of `psych::fa` is correct only when
# factoring correlation matrix/standardized data
scores_reg2 = decomp$scores
head(scores_reg2)
```

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