STAT 3690 Homework 3

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Due at Apr 11 11:59 pm (Central Time)

Answers must be submitted electronically via Crowdmark. Please enclose your R code trunks (if applicable) as well.

1. Figure out the covariance matrix for random vector $\mathbf{Y} = [Y_1, Y_2, Y_3]^\top$, when $\mathbf{Y}$ satisfies the following one-factor model:
   - $Y_1 = .6F + \epsilon_1$, $Y_2 = .8F + \epsilon_2$, and $Y_3 = .5F + \epsilon_3$;
   - Factors F, ϵ_1, ϵ_2, and ϵ_3 are uncorrelated with each other;
   - $\mathrm{var}(F) = 1$, $\mathrm{var}(\epsilon_1) = .64$, $\mathrm{var}(\epsilon_2) = .36$, and $\mathrm{var}(\epsilon_1) = .75$.

2. In the factor analysis, we assume the covariance matrix of error vector to be diagonal. Actually, we can formally test whether a covariance matrix is diagonal via likelihood ratio test (LRT). Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\mathbf{X}_i = [X_{i1}, \ldots, X_{ip}]^\top$, $i = 1, \ldots, n$.

   a. When $\boldsymbol{\Sigma}$ is diagonal, i.e.,

   $$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_p^2 \end{bmatrix},$$

   please point out WITHOUT proof the maximum likelihood estimators for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.

   b. Use the result of part a to derive the LRT statistic λ for hypotheses $H_0 : \boldsymbol{\Sigma}$ is diagonal vs. H_1 : otherwise. Specifically,

   $$\lambda = \frac{\text{maximum likelihood when } (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \boldsymbol{\Theta}_0}{\text{maximum likelihood when } (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \boldsymbol{\Theta}},$$

   where $\boldsymbol{\Theta}_0 = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \mid \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ is diagonal and positive-definite$\}$ and $\boldsymbol{\Theta} = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \mid \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ is positive-definite$\}$.

3. Typically, for the LRT in Q2b, the null distribution of $-2\ln\lambda$ is approximated by $\chi^2(p(p-1)/2)$.

   a. If $p = 3$, what is the rejection region for λ at level $\alpha = .01$?

   b. A simulation study may be used to investigate the Type I error rate of the LRT in Q2b. Generate a dataset of size $n = 100$ with $\boldsymbol{\mu} = [1, 2, 3]^\top$ and $\boldsymbol{\Sigma} = \mathbf{I}_3$ and then calculate λ for this dataset. Repeat $B = 1000$ times (i.e. you will have B realizations of λ). With each realization of λ, test hypotheses $H_0 : \boldsymbol{\Sigma}$ is diagonal vs. H_1 : otherwise. Get a conclusion on the goodness of χ^2-approximation by comparing $\alpha = .01$ with the Type I error rate (i.e., the rejection proportion in the B tests).

4. Suppose

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & .48 & .3 \\ .48 & 1 & .4 \\ .3 & .4 & 1 \end{bmatrix}.$$

Carry on a factor analysis with one common factor via the principal component method without any rotation. Compare your results with the model posited in Q1.

5. We are going to use dataset `OTE::Body` on 21 body measurements.

    a. Fit a factor model with $q = 3$ common factors for the 21 body measurements of female subjects only, using the maximum likelihood method. What are the estimates for the varimax loading matrix and score matrix (via weighted least squares), respectively?
    b. Illustrate and explain the association between each factor and Age, Weight, and Height.
    c. Refit the model for male observations only and compare the resulting loadings to those obtained in part a. Given these results, do you think the factor analysis is capturing the same unobserved structure for both genders?

```r
# data(OTE::Body)
dataset = OTE::Body[Body$Gender==0,!names(Body) %in% c('Gender','Age','Weight','Height')]
names(dataset)
```

```
##  [1] "Biacrom"   "Biiliac"   "Bitro"     "ChestDp"   "ChestD"    "ElbowD"
##  [7] "WristD"    "KneeD"     "AnkleD"    "ShoulderG" "ChestG"    "WaistG"
## [13] "AbdG"      "HipG"      "ThighG"    "BicepG"    "ForearmG"  "KneeG"
## [19] "CalfG"     "AnkleG"    "WristG"
```