

# 2023 honours multivariate FA practical

## Question One (4)

Show that the covariance matrix

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & 0.63 & 0.45 \\ 0.63 & 1 & 0.35 \\ 0.45 & 0.35 & 1 \end{bmatrix}$$

for the  $p = 3$  standardised random variables  $Z_1$ ,  $Z_2$  and  $Z_3$  can be generated by the  $m = 1$  factor model

$$Z_1 = 0.9F_1 + \epsilon_1$$

$$Z_2 = 0.7F_1 + \epsilon_2$$

$$Z_3 = 0.5F_1 + \epsilon_3$$

where  $\text{Var}[F_1] = 1$ ,  $\text{Cov}[\boldsymbol{\epsilon}, F_1] = \mathbf{0}$  and

$$\boldsymbol{\Psi} = \text{Cov}[\boldsymbol{\epsilon}] = \begin{bmatrix} 0.19 & 0 & 0 \\ 0 & 0.51 & 0 \\ 0 & 0 & 0.75 \end{bmatrix}.$$

That is, write  $\boldsymbol{\rho}$  in the form  $\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ , where  $\mathbf{L}$  is the matrix of loadings and  $\boldsymbol{\Psi}$  is the diagonal matrix of specific variances.

## Question Two (13)

Use the air pollution data stored in `Air Pollution Data.csv`.

- Generate a sample correlation matrix. (1)
- Obtain the principal component solution to a factor model with  $m = 1$  and  $m = 2$  using the `eigen` function (as before, you must use the `eigen` function). (3)
- Obtain the maximum likelihood factor model for  $m = 2$  using `factanal`. (2)
- Compare the factorization obtained by the principal component and maximum likelihood methods. (2)
- Perform a varimax rotation of the principal component solution with  $m = 2$  factors. Interpret the results. (3)
- Calculate the factor scores from the  $m=2$  maximum likelihood approach using weighted least squares (i.e. using the matrix formula for the factor scores). (2)

## Question Three (4)

In FA we are concerned with accurately capturing the covariance of the variables. In this question we prove an upper bound for the sum of squared errors between our fitted covariance matrix and the sample covariance matrix, when we use the principal component method.

For  $\mathbf{S} : p \times p$  the sample covariance matrix,  $\hat{\mathbf{L}} : p \times m$  the matrix of loadings estimated using the principal component method and  $\hat{\mathbf{\Psi}}$  the diagonal matrix of specific variances, show that

$$\text{tr}[(\mathbf{S} - (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}))'(\mathbf{S} - (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}))] \leq \sum_{i=m+1}^p \lambda_i^2.$$

In other words, show that the sum of squared errors between the sample covariance matrix and the fitted covariance matrix is upper bounded by the sum of the squared eigenvalues of unused principal components.

- Note that  $\mathbf{S} - (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}})$  has zeroes on the diagonal, meaning that  $\text{tr}[(\mathbf{S} - (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}))'(\mathbf{S} - (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}))] \leq \text{tr}[(\mathbf{S} - \hat{\mathbf{L}}\hat{\mathbf{L}}')'(\mathbf{S} - \hat{\mathbf{L}}\hat{\mathbf{L}}')]$
- Also note that  $\mathbf{S} - \hat{\mathbf{L}}\hat{\mathbf{L}}' = \sum_{i=m+1}^p \lambda_i \mathbf{e}_i \mathbf{e}_i'$ , where  $\mathbf{e}_i$  is the eigenvector of  $\mathbf{S}$  associated with the  $i$ -th largest eigenvalue  $\lambda_i$ .
- The trace property that  $\text{tr}[\mathbf{ABC}] = \text{tr}[\mathbf{CAB}] = \text{tr}[\mathbf{BCA}]$  and the transpose property that  $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$  are useful here.