# MAST90138 Assignment 2

**Instructions**:

- This assignment contains 3 problems with a total of 21 +1 marks which count towards 15% of the final mark for the subject. It is due by 5pm Wednesday 4 October, 2023.
- **[1 mark]** Your assignment should clearly show your name and student ID number, your tutor's name and the time and day of your tutorial class. Your answers must be clearly numbered and in the same order as the assignment questions. Your answers must be easy to read (marks may be deducted for illegible handwriting). Include all of your working out in your answers. All R outputs, including graphs and tables, must be accompanied by your concise and clearly written R code used to produce it. Any graph, table or R code must be accompanied by clear and concise comments. Uncommented R code is not acceptable.
- Use concise text explanations to support your answers. Comments should be brief and concise: marks will be awarded for clarity.
- No late submission is allowed.
- Your lecturer may not help you directly with assignment questions, but may provide some appropriate guidance.

**Problem 1** [8 marks]: Let $X$ denote a random $p$-vector with variance-covariance matrix $\Sigma$. Recall that in the orthogonal factor model, we assume that

$$X - E(X) = QF + U\,,$$

where $Q$ a $p \times q$ matrix of loadings, $F = (F_1, \ldots, F_q)^T \sim (0, I_q)$ a random vector of $q < p$ common factors and $U = (U_1, \ldots, U_p)^T \sim (0, \Psi)$ is a random vector of $p$ specific factors with $\Psi$ is a $p \times p$ diagonal matrix, and $\mathrm{cov}(F, U) = 0$. Under this model, the variance-covariance matrix $\Sigma$ of $X$ can be expressed as

$$\Sigma = QQ^T + \Psi\,. \tag{1}$$

(a) Prove by direct calculation (using no more than one single line of calculation), that if equation (1) is satisfied for $Q = Q_1$ and for some $\Psi$, then (1) is also satisfied for $Q = -Q_1$ and the same $\Psi$.

(b) We saw in class that if $Q$ satisfies equation (1) for some $\Psi$, then for any $q \times q$ orthogonal matrix $G$, $Q_G = QG$ satisfies

$$\Sigma = Q_G Q_G^T + \Psi \tag{2}$$

for the same $\Psi$. In the case where $q = 1$:

  (i) Prove that there are only two $q \times q$ orthogonal matrices $G_1$ and $G_2$ (and explicitly give $G_1$ and $G_2$).

  (ii) Explain how our answer at (b)(i) can be used to give an alternative proof for question (a) above in the particular case where $q = 1$.

(c) Consider equation (1) in the case where $p = 3$, $q = 1$ and

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} = \begin{pmatrix} 4 & -2 & 3 \\ -2 & 5 & -1 \\ 3 & -1 & 6 \end{pmatrix},$$

so that $Q = (q_1, q_2, q_3)^T$ and $\Psi = \mathrm{diag}(\psi_1, \psi_2, \psi_3)$, where $q_1, q_2, q_3$ and $\psi_1, \psi_2, \psi_3$ are unknowns that need to be determined. There is a unique $\Psi$ and there are two matrices $Q$ that satisfy (1). Your task: compute the unique $\Psi$ and the two matrices $Q$ that satisfy (1). In this case, is the unique $\Psi$ is interpretable? Explain why.

**Hint:** For $k = 1, 2, 3$, $q_k^2$ can be expressed in terms of $\sigma_{k1}, \sigma_{k2}, \sigma_{k3}$. Once you find the expression for each $q_k^2$ you deduce that for each $\psi_k$.

**Problem 2** [8 marks]: In this problem we will gain an understanding of the function `factanal` in `R` to fit, by maximum likelihood, a normal orthogonal factor model to a dataset. To achieve this, you will be redoing the analysis of the Boston Housing data carried out in section 12.4 of the reference book (Härdle and Simar, 2019). The Boston Housing data have been used and described in chapter 5 of the slides. The original data, which you will use here, have $n = 506$ observations on 14 variables:

```
crim: per capita crime rate by town
zn:  proportion of residential land zoned for lots over 25,000 sq.ft
indus:  proportion of non-retail business acres per town
chas:  Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
nox:  nitric oxides concentration (parts per 10 million)
rm:  average number of rooms per dwelling
age:  proportion of owner-occupied units built prior to 1940
dis:  weighted distances to five Boston employment centres
rad:  index of accessibility to radial highways
tax:  full-value property-tax rate per USD 10,000
ptratio:  pupil-teacher ratio by town
b:  1000(B - 0.63)^2 where B is the proportion of African American by town
lstat: percentage of lower status of the population
medv:  median value of owner-occupied homes in USD 1000's
```

As in section 12.4 of Härdle and Simar (2019), you will work with the following transformed version of those variables:

`X1=log(crim), X2=zn/10, X3=log(indus), X4=chas, X5=log(nox), X6=log(rm),`

`X7=age`$^{2.5}$`/10,000, X8=log(dis), X9=log(rad), X10=log(tax),`

`X11=exp(0.4*ptratio)/1000, X12=b/100, X13=sqrt(lstat), X14=log(medv)`

Table 12.3 below, taken from section 12.4 of Härdle and Simar (2019), summarises the numerical results of the factor analysis carried out in their section 12.4 for these transformed data, using maximum likelihood and varimax rotation.

**Table 12.3** Estimated factor loadings, communalities, and specific variances, MLM, varimax rotation MVAfacthous

| | Estimated factor loadings | | | Communalities | Specific variances |
|---|---|---|---|---|---|
| | $\widehat{q}_1$ | $\widehat{q}_2$ | $\widehat{q}_3$ | $\widehat{h}_j^2$ | $\widehat{\psi}_{jj} = 1 - \widehat{h}_j^2$ |
| 1. Crime | 0.7247 | −0.2705 | −0.5525 | 0.9036 | 0.0964 |
| 2. Large lots | −0.1570 | 0.2377 | 0.5858 | 0.4248 | 0.5752 |
| 3. Nonretail acres | 0.4195 | −0.3566 | −0.6287 | 0.6909 | 0.3091 |
| 5. Nitric oxides | 0.4141 | −0.2468 | −0.7896 | 0.8561 | 0.1439 |
| 6. Rooms | −0.0799 | 0.6691 | 0.1644 | 0.4812 | 0.5188 |
| 7. Prior 1940 | 0.2518 | −0.2934 | −0.7688 | 0.7406 | 0.2594 |
| 8. Empl. centers | −0.3164 | 0.1515 | 0.8709 | 0.8816 | 0.1184 |
| 9. Accessibility | 0.8932 | −0.1347 | −0.2736 | 0.8908 | 0.1092 |
| 10. Tax-rate | 0.7673 | −0.2772 | −0.3480 | 0.7867 | 0.2133 |
| 11. Pupil/Teacher | 0.3405 | −0.4065 | −0.1800 | 0.3135 | 0.6865 |
| 12. African American | −0.3917 | 0.2483 | 0.1813 | 0.2480 | 0.7520 |
| 13. Lower status | 0.2586 | −0.7752 | −0.4072 | 0.8337 | 0.1663 |
| 14. Value | −0.3043 | 0.8520 | 0.2111 | 0.8630 | 0.1370 |

(a) Extract the original Boston Housing data ($n = 506$ observations on 14 variables) from the mlbench R package (install it if needed), discard the binary variable `chas` and transform the other variables as described above. Store the resulting data in a $506 \times 13$ matrix called `XBoston`.

(b) Apply `factanal` to the scaled and centered version of `XBoston`, *with the varimax rotation*, to obtain the orthogonal factor model with $q = 3$ fitted by maximum likelihood and print the R output.

(c) Extract (and create), from the R outputs, a $3 \times 3$ diagonal matrix $\Psi$ that contains, on the diagonal, the 3 specific variances of the fitted factor model. Create a $13 \times 3$ matrix $Q$ that contains the loadings for the fitted factor model and ask R to print $Q$ and the diagonal of $\Psi$, displaying the values rounded to the first four 4 digits, in the sense that numbers like 4.90567819 and $-15.57031476$ are displayed as 4.9057 and $-15.5703$, respectively. Show your outputs.

(d) Give two distinct ways of obtaining, from the R output, the $p = 13$ communalities in the fitted model (and compute them in R using both ways and show your outputs, again rounding so as to display only the first 4 digits as in the previous question).

(e) Explain what are the differences between the loadings, communalities and specific variances of the model fitted in (b) above, with those from Table 12.3 of Härdle and Simar (2019), and explain the reason of those differences.

**Problem 3** [5 marks]: In this problem we will work with the wheat data available in `.txt` format on the LMS web page within the *Assignments* menu. The data come from three different varieties of wheat denoted by 1 to 3 in the dataset. Each row of the dataset corresponds to a different wheat kernel. Seven numerical characteristics were measured on the data: X1: area, X2: perimeter, X3: compactness, X4: length of kernel, X5: width of kernel, X6: asymmetry coefficient, X7: length of kernel groove, whereas the eighth variable X8 contains values 1, 2 or 3 which each code the variety of wheat the kernel comes from (there are three varieties).

Read the wheat data in R and create a data matrix `X` of size $n \times p$, where $n = 210$ and $p = 7$, which contains the seven attributes X1 to X7 described above from all $n$ kernels.

(a) Using the R function `factanal`, fit a normal factor model to the scaled and centered version of `X` by maximum likelihood, using $q = 3$ factors and the varimax rotation of the factors. Display the R output.

(b) In the fitted model, what is the percentage of variance of each of the seven attributes X1 to X7, explained by the 3 factors?

(c) Using the model fitted in (a), compute the $7 \times 3$ matrix `R` that contains the correlations between each of the seven attributes X1 to X7, and each of the three factors $F_1, F_2, F_3$. For each factor, give the variables that are the most correlated with the factor and use this to interpret what your factors may represent, explicitly discussing their connection with the attributes X1 to X7, or groups of them.