9   Correspondence Analysis

Exercise 9.1 (Full CA by Yourself)
This exercise exploits the file "studentStudy.txt" which contains a contingency table presenting the academic track selected by a student given the socio-economic category of his/her parent. The socio-economic category can take on the values $\{x_1, \ldots, x_N\}$ with $N = 5$ and the academic background can take on the values $\{y_1, \ldots, y_M\}$ with $M = 4$.

1. Load the file with the command

```
students <- read.table(file="studentStudy.txt",
                       row.names=1,header=T,sep="\t")
print(students)
```

2. What is the R class of the variable "students".

3. Compute the total number of students and compute the frequency contingency table.

4. Compute the frequency contingency table which occurs under the assumption that the two variables are independent.

5. Compute the standardized residuals Z_{ij} for all $i = 1, \ldots, N$ and $j = 1, \ldots, M$.

6. Compute Φ.

7. Compute the contribution of each row modality x_i to the test statistics χ^2.

8. Compute the matrix R of centered and shifted row profiles.

9. Compute the matrix V and the eigenvectors of the eigendecomposition for the row profile analysis.

10. Compute the percentage of variance explanation for each axis directed by an eigenvector.

11. Compute the row score vectors s_1 and s_2 associated to the two largest eigenvalues.

12. Compute the contribution of each row modality x_i on the construction of the axis directed by the kth eigenvector u_k for $k = 1, 2$.

13. Compute the quality of the representation, i.e., compute the squared cosine of angle between the row profile R_i and the eigenvector u_k for $k = 1, 2$.

14. Plot the 2D coordinates of the projected row profiles onto the plane spanned by u_1 and u_2, i.e., plot s_2 as a function of s_1. For each projected row profile, add a text label on the plot which shows in blue the modality x_i of the row.

15. Repeat steps 8 to 15 for the column profiles. You must replace the matrix R with the matrix Q. The modality y_j must appear in red on the figure showing the 2D coordinates of the projected column profiles.

16. Plot on the same figure the coordinates of the row profiles and the coordinates of the column profile after the dimension reduction.

17. Verify the numerical relationships between the row coordinates and the column coordinates.

18. Run the following commands :

```
1 # Loading FactoMineR
2 library(FactoMineR)
3 # Correspondence Analysis
4 res <- CA(students)
5 summary(res)
```

Explain the role and the output of each command. Verify the consistency of these results with your results computed in the previous questions. Comment any change.

19. How to interpret the origin of the plane ? What can we deduce about a student whose mother or father is a business leader ?

20. Can we say that a student whose mother or father is a laborer is interested by an Associate Degree ?

Exercise 9.2 (On the Theoretical Aspects of the CA Decomposition)
The following questions are based on the notations defined in the lecture.

1. Show that $V = Z^T Z$.

2. Show that $W = ZZ^T$.

3. Show that the non-zero eigenvalues of V and W are the same.

4. Show that if $u \in \mathbb{R}^M$ is an eigenvector of V, then $Zu \in \mathbb{R}^N$ is an eigenvector of W.

5. Show that $\Phi^2 = \mathrm{Tr}(V) = \mathrm{Tr}(W)$.