MS-E2112 Multivariate Statistical Analysis – 2021
Online Exam
Answer to all the questions.
In problem 1, you do not have to justify your answer. In all the other problems, justify your solutions and write down all your calculations.

---

1. True or False (4 p.)

   Determine whether the statement is true or false. In this problem, you do not have to justify your answers. Simply state wether the statement is true or false.(Every correct answer +1 p., every wrong answer -1 p., no answer 0 p.)

   (a) If the influence function of a functional Q is bounded (with respect to L_2 norm), then the asymptotical breakdown point of Q can not be 0.

   (b) Assume that we have two groups of variables and that we analyse the relationship between the groups of variables by applying canonical correlation analysis. Assume that in the first group, we have 6 variables, and in the second group, we have 4 variables. We now obtain max 4 pairs of canonical variables.

   (c) Fisher's linear discriminant analysis is based on maximizing the ratio of between groups dispersions and within group dispersions.

   (d) According to Zuo and Serfling, depth functions should be invariant under affine transformations.

2. Multiple Correspondence Analysis (6 p.)

A survey was administered to 201 first year engineering students. The goal of the survey was to understand what are the factors influencing the success of a student in her/his first year as a university student. Variables considered are given below:

- Attendance to class: *low, average, high*
- Average time spent studying for the exams: *less than 4h., between 4h. and 8h., more than 8h.*
- Activity in class: *low, high*
- Participation to additional guidance: *low, high*

Use the picture and the eigenvalues (next page) to justify your answers.

(a) What is the total variance of the variables?

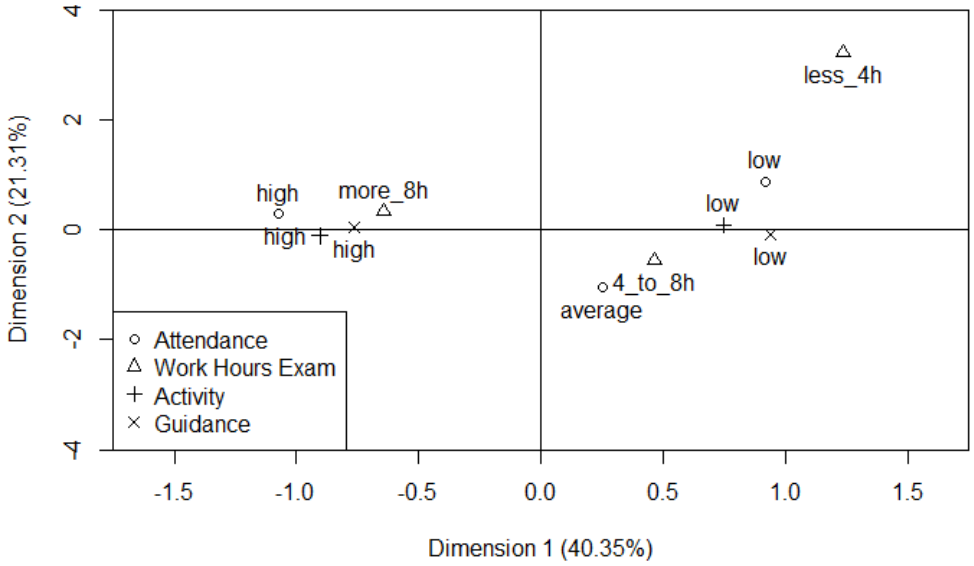
(b) How much of the total variance do the first two MCA components explain?

(c) Interpret Dimension 1 using the picture.

(d) Based on the picture, does it seem that students with average attendance study more than 8 hours to the exam? Justify!

Table 1: Eigenvalues (rounded) associated with the MCA transformation:

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
0.61	0.32	0.21	0.17	0.11	0.08



3. Multivariate Location and Scatter (6 p.)

   (a) Let x denote a p-variate random vector. Let μ_x denote the population mean vector, and Σ_x denote the population covariance matrix of x. Assume that μ_x and Σ_x exist as finite quantities. Let $A \in \mathbb{R}^{p \times p}$ be nonsingular, and let $b \in \mathbb{R}^p$. Let $y = Ax + b$. Let μ_y denote the population mean vector, and Σ_y denote the population covariance matrix of y. Show that $\mu_y = A\mu_x + b$ and that $\Sigma_y = A\Sigma_x A^T$. (3 p.)

   (b) Let x denote a p-variate random vector. Let F_x denote the cumulative distribution function of x. Assume that x is symmetrically distributed about θ. Assume that T ia an affine equivariant location functional and assume that $T(F_x)$ exists as a finite quantity. Show that $T(F_x) = \theta$. (3 p.)

4. Clustering (4 p.)

   Consider the following bivariate sample:

   $$A = (2, 1), B = (-3, 0), C = (1, -2), D = (1, 1), E = (0, 2).$$

   (a) Draw a scatter plot of the data. (1 p.)

   (b) Perform agglomerative hierarchical clustering on the data. Use Euclidian distance as the distance measure and in clustering, measure the distance between the groups by applying maximum distance. Draw a corresponding classification tree. If you choose the number of the final clusters to be two, what are the two clusters? (3 p.)

5. Principal component analysis (4 p.)

   (a) Describe the idea behind principal component transformation. (2 p.)

   (b) Explain how and under what conditions you can robustify principal component analysis. (2 p.)

BONUS QUESTION (2 p.):

Consider the following bivariate sample:

$$S = \{(1.5, 1.5), (-1.5, -2.5), (2.5, -1.5), (2.0, 1.0), (-0.5, 1.5), (0.0, 4.5)\}.$$

What is the half-space depth of the point $(-1.0, -1.0)$ with respect to the sample S?