MS-E2112 Multivariate Statistical Analysis – 2019
Exam
Answer to all the questions.

You are allowed to have pens and pencils, an eraser and a ruler, a calculator and one size A4 note (handwritten, text on one side only, name on the top right corner).

1. True or False (6 p.)

   Determine whether the statement is true or false. (Every correct answer +1 p., every wrong answer -1 p., no answer 0 p.)

   (a) PCA transformation is invariant under affine transformations.

   (b) PCA is sensitive to heterogenous scaling of the variables.

   (c) Asymptotical breakdown point of the sample median is $\frac{1}{2}$.

   (d) The empirical influence function of the sample median is bounded.

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

   (f) The initial K centers do not have an effect on the results of the moving centers clustering methods (K-means clustering methods).

Statement	a	b	c	d	e	f
True						
False						

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?

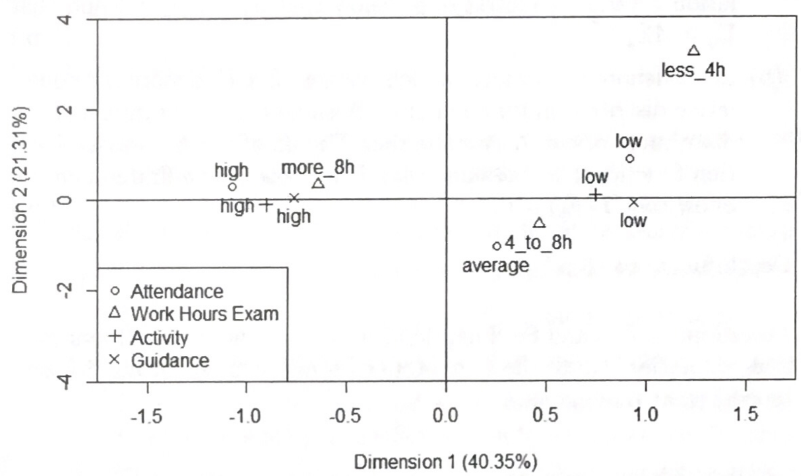
(b) How much of the total variance do the first three MCA components explain?

(c) Why do you think that the modality *less than 4h.* is far away from the center?

(d) Based on the picture, does it seem that students with average attendance to class are active in class? 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. Depth functions (6 p.)

   According to Zuo and Serfling, depth functions should fulfill four general properties. State the four properties and explain (using 2-3 sentences) what they mean.

BONUS QUESTION (2 p.):

Consider the following bivariate sample:

$$\{(2.5, -1.0), (-3.5, 1.5), (2.5, 1.5), (1.0, 1.0), (1.5, 1.5), (-2.0, 2.5)\}.$$

What is the half-space depth of the data point $(1.0, 1.0)$?