

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) Multiple correspondence analysis (MCA) is based on applying bi-variate correspondence analysis on the so called complete disjunctive table.
- (d) Whereas PCA relies on euclidian distances, MCA relies on chi-square distances.
- (e) Discriminant analysis is a method for splitting a set of individuals into unknown homogenous groups.
- (f) The first step in agglomerative clustering is choosing k initial points randomly.

Statement	a	b	c	d	e	f
True						
False						

2. Clustering (6 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? (5 p.)

3. Robustness (6 p.)

- (a) Derive the finite sample breakdown point and the asymptotic breakdown point of the sample median. (3 p.)
- (b) Derive the empirical influence function of the sample mean. (3 p.)

4. Depth functions (6 p.)

- (a) 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. (4 p.)
- (b) Consider the following bivariate sample:

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

What is the half-space depth of the data point $(1.0, 1.0)$? (2 p.)

BONUS QUESTION (2 p.):

Show that, under the assumption of central symmetry, all affine equivariant location functionals measure the same population quantity.