Academic Year 2024/2025, period P1

Syllabus:
Multivariate Data Analysis 3613 (8ECTS, on-line, synchronized course)

Instructor: Agnieszka $\overbrace{\text{Jach}}^{surname}$, agnieszka.jach@hanken.fi

Office hours: Thursday 16-17h in the virtual room via Teams (link on Moodle)

Textbooks:
James, G., Witten D., Hastie, T., Tibshirani, R. (2017). *An Introduction to Statistical Learning.* 7th ed. New York: Springer. (textbook-based slides, videos, Rcode available on the Internet)
Hair, J.F, Black, W.C, Babin, B.J, Anderson, R.E (2018). *Multivariate data analysis: a global perspective.* 8th ed. or earlier. Upper Saddle River (N.J.): Prentice Hall.

Teaching materials: Lecture notes with code&text (Rmarkdown file, .Rmd), instructions, links, etc are available on Moodle. Self-enrolment key for Moodle: please, check Sisu about one week before the first lecture.

Schedule: 21 sessions in total (see Table 1 below for details), all in Teams' virtual room; access to the room via 'Join a team with a code', the code is available on Moodle:

- $18 = 9 \times 2$ (nine double-slot sessions): computer lab with the theory mixed-in

- $3 = 3 \times 1$ (three single-slot sessions): help-desk

Times: see Table 1 below and Schema.

Software: `R` (computations and graphics), `markdown` (generation of documents) used within `rstudio` (=IDE for `R`). Please, install the software on your personal laptop, following instructions on Moodle, and have your machine ready at every session.

Assessment: 40% (exam) + 60% (HW assignments). You do the exam and the assignments on your personal computer (both are open-Internet, open-book). You need to attempt most of the HW assignments and you need to attempt the final exam to be considered for passing the course. At least 50% (in total) is needed to pass the course. Group-work is involved (deadline for signing up for groups is after the 2nd class-meeting). **For details on the assessment and group-work see a separate document on Moodle.**

- Deadlines for HW assignments: see Table 1.

-

- Late submissions are not allowed/accepted.

- HW- and material-related questions can be posted on a specially designed Moodle forum and ideally should not be consulted via e-mail

Contents:

A.Preliminary: data types, graphical and numerical summaries (also for grouped data), outliers, missing observations, transformations, re-coding variables and subsetting

B.Preliminary: data standarization, centering, distance; optional statistical appendix

1. Analysis of variance (ANOVA)

2. Multivariate analysis of variance (MANOVA)

3. Multiple regression (REGR)

4. Logistic Regression (LOGIT)

5. Linear Discriminant Analysis (LDA)

6. Principal Component Analysis (PCA), Factor Analysis (FA)

7. Clustering (CLUSTER)

Week	Dates (Mon-Fri)	Day of week Tue	Wed	Fri	**HW due on Mon 18:00h**
36	02-06.09	A.Prelim 16:00-19:15h	B.Prelim 16:00-19:15h	Help-desk 16:00-17:30h	
37	09-13.09	1.ANOVA 16:00-19:15h	2.MANOVA 16:00-19:15h	— —	HW1 due
38	16-20.09	3.REGR 16:00-17:30h	— —	Help-desk 16:00-17:30h	HW2 due
39	23-27.09	4.LOGIT 16:00-19:15h	5.LDA 16:00-19:15h	— —	HW3 due
40	30.09-04.10	6.PCA, FA 16:00-19:15h	7.CLUSTER 16:00-19:15h	Help-desk 16:00-17:30h	HW4 due
41	07-11.10	—	—	—	HW5 due

Table 1: Detailed class schedule for 3613. All sessions take place in Teams' virtual room (no physical-room sessions).