

Academic Year 2021/2022, period P1

# Syllabus: Multivariate Data Analysis 3613 and 3613-V (8ECTS)

**Instructor:** Agnieszka <sup>*surname*</sup> Jach , agnieszka.jach@hanken.fi

**Office hours:** Thursday 14-15h in the virtual room via Teams (link on Moodle)

## Textbooks:

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.

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)

**Teaching materials:** Lecture notes with code&text (Rmarkdown file, .Rmd), instructions, links, etc are available on Moodle. Self-enrolment key for Moodle: please, check your e-mail.

**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 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.
- **Exam date(s): 21.10.2021 (Thu), 20.11.2021 (Sat), both from 14h-18h.**
- 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)	Tue	Day of week Wed	Fri	HW due on Mon 18:00h
35	30.08-03.09	A.Prelim 16:00-19:15h	B.Prelim 14:15-17:30h	Help-desk 14:15-15:45h	exception: Mon-Tue teaching! HW1 due HW2 due, help-desk on Wed! HW3 due HW4 due
36	06-10.09	<del>Tue</del> →Mon 1. ANOVA 14:15-17:30h	<del>Wed</del> →Tue 2. MANOVA 16:00-19:15h	—	
37	13-17.09	3. REGR 14:15-17:30h	Help-desk 16:00-17:30h	—	
38	20-24.09	4. LOGIT 14:15-17:30h	5. LDA 16:00-19:15h	—	
39	27.09-01.10	6. PCA, FA 14:15-17:30h	7. CLUSTER 16:00-19:15h	Help-desk 14:15-15:45h	
40	04-08.10	—	—	—	HW5 due

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