Academic Year 2025/2026, period P1

Syllabus:
Multivariate Data Analysis 3681
(Helsinki and Vaasa, 5ECTS)

Instructor: Agnieszka $\overbrace{\text{Jach}}^{surname}$, agnieszka.jach@hanken.fi, Teaching Assistant in Vaasa (Maryam Nasiri Kheshtmasjedi)

Office hours: Thursday 16:30-17:30h 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 class.

	Dates	Day of week and time		
Week	Mon-Fri	Tue 18h	Wed at 14:15h or 16h	Thu at 14:15h
36	01-05.09		16h A.Prelim 1x1.5h A210 Hfors, V325 Vasa	B.Prelim 1x1.5h A309 Hfors, V312 Vasa
37	08-12.09	HW-Quiz1 due	1.ANOVA 2x1.5h A309 Hfors, V325 Vasa	2.MANOVA 2x1.5h A210 Hfors, V312 Vasa
38	15-19.09	HW-Quiz2 due	— —	3.REGR 2x1.5h A210 Hfors, V312 Vasa
39	22-26.09	HW-Quiz3 due	4.PCA, FA 2x1.5h A309 Hfors, V325 Vasa	5.CLUSTER 2x1.5h A210 Hfors, V312 Vasa
40	29.09-03.10	HW-Quiz4 due, Article due	16h Articles 1x1.5h A210 Hfors, V325 Vasa	— —

Table 1: Detailed class schedule for 3681.

Schedule: 13 sessions in total (see Table 1 below for details). Campus-based teaching in Helsinki (with the examiner), campus-based teaching in Vasa (with the Teaching Assistant and a Teams's stream from Helsinki; access to the room via 'Join a team with a code', the code is available on Moodle):

- $2 = 2 \times 1$ (two single-slot sessions): computer lab with the theory mixed-in
- $10 = 5 \times 2$ (five double-slot sessions): computer lab with the theory mixed-in
- $1 = 1 \times 1$ (one single-slot session): articles (special class) - obligatory class participation

Times: see Table 1 and Sisu.

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 quiz) + 60% (hw quizzes and articles),
60%=4x12% (4 hw-quizzes)+2x6% (two-part assignment with articles).
You do the exam-quiz and the hw-quizzes on your personal computer (both are open-Internet, open-book). You need to attempt at least half of the hw-quizzes and you need to attempt the exam-quiz to be considered for passing the course. At least 50% (in total) is needed to pass the course. Group-work (the best option to participate in the course) is strongly recommended (deadline for signing up for groups is after the 2nd class-meeting). Articles: see detailed info on Moodle. **For more on quizzes, group-work, articles, etc see Moodle.**

- Deadlines for hw quizzes and articles: see Table 1.
- Exam date(s): 23.10.2025 (Thu), 22.11.2025 (Sat), both from 14h-18h.
- Late submissions are not allowed/accepted.
- hw-quiz- and material-related questions can be posted on a specially designed Moodle forum and ideally should not be consulted via e-mail

Use of AI: exam-quiz and hw-quizzes 1-4: red light; articles: individual submission - green light, class submission - red light.

Contents:

A.B. Preliminary: data types, graphical and numerical summaries (ungrouped/grouped data), transformations, outliers, missing observations, 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. Principal Component Analysis (PCA), Factor Analysis (FA)

5. Clustering (CLUSTER)

X. Articles (mandatory class attendance)