

# Introduction to STA5069Z

Modern multivariate statistical techniques

Miguel Rodo

2024-02-12

# Welcome to STA5069Z

- ▶ This course follows on from the honours module in multivariate statistics. It is aimed at looking at the application of multivariate techniques to modern-day data problems, including big data problems. Techniques introduced at honours level will be reviewed and extended and new techniques will be introduced.
  - ▶ See [course website](#) for specific topics covered.
- ▶ **Lecturers**
  - ▶ Mr Miguel Rodo (PD Hahn 5.52)
  - ▶ Dr Sebnem Er (PD Hahn 5.55)

## Key references

- ▶ Primary: Izenman (2008)
  - ▶ Available (for purchase) at [Springer](#) and [Amazon](#)
- ▶ Also used: Hastie, Tibshirani, and Friedman (2009)
  - ▶ Available for free at [Hastie's own website](#)
- ▶ Paper references will be provided as we go

# Course structure: lectures

- ▶ The [course website](#) has all the information you need.

## Format

- ▶ Each lecture will cover a new set of multivariate techniques.
- ▶ There is a *lot* of content, which will not all be covered in lectures.
- ▶ As such, significant self study is required.
  - ▶ Lecture slides and associated recommended readings will be provided in advance.
- ▶ During lectures, we will cover key concepts, tackle student questions and review tutorial solutions.

## Venue and timing

- ▶ Lectures will be 4-5.45pm on Mondays and Wednesdays in PD Hahn 4.26 (the Quiet Room).

# Course structure: practice and testing

## Tutorials

- ▶ For each lecture, tutorials (class exercises) to implement the methods discussed in R will be provided.

## Assignment

- ▶ 10-page summary report on applying modern multivariate methods to particular problem areas, such as bioinformatics, finance, or marketing.

## Exam

- ▶ *Format*: Open-book take-home exam.
- ▶ *DP requirement*: lecture attendance and timely submission of tutorials.
- ▶ *Pass requirement*: Average of 50% across exam and assignment, with 40% sub-minimum for each.

# Assignment

- ▶ See [course website](#) for details.
- ▶ A template repository is available [here](#) for creating a GitHub repo with the desired structure for the project.

# Technical prerequisites

- ▶ Primarily, linear algebra. In particular:
  - ▶ Matrix norms
  - ▶ Eigendecomposition
  - ▶ SVD

- Hastie, Trevor., Robert. Tibshirani, and J. H. (Jerome H.) Friedman. 2009. *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. New York: Springer.
- Izenman, Alan J. 2008. *Modern Multivariate Statistical Techniques*. Springer Texts in Statistics. New York, NY: Springer. <https://doi.org/10.1007/978-0-387-78189-1>.