Module 19 Multivariate Analysis for Genetic data
Session 01: Introduction

Jan Graffelman[1,2]

[1]Department of Statistics and Operations Research
Universitat Politècnica de Catalunya
Barcelona, Spain

[2]Department of Biostatistics
University of Washington
Seattle, WA, USA

28th Summer Institute in Statistical Genetics (SISG 2023)

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
UPC BARCELONATECH

BIOSTAT
W

Instructor and TA

Jan Graffelman
`jan.graffelman@upc.edu`
Associate professor in Statistics at UPC, Barcelona
Visiting associate professor at UW Biostatistics, Seattle
Accredited as full professor (ANECA, Spain)
PI of a project on Count and Compositional data
Research areas: statistical genetics; multivariate analysis;
compositional data analysis;

Antonio Olivas
`aolivas@uw.edu`
Fourth year PhD student in Biostatistics
Dpt. of BioStatistics, UW, Seattle
Teaching Assistent for Module 19 at SISG 2023

Course program

Wednesday 26th of July 2023 (Pacific time)
- 13.30-15.00pm S01: Introduction; Matrix algebra
- 15.00-15.30pm Coffee/tea break
- 15.30-17.00pm S02: Matrix decompositions & Biplots

Thursday 27th of July 2023
- 08.30-10.00am S03 Principal component analysis
- 10.00-10.30am Coffee/tea break
- 10.30-12.00am S04 Logratio principal component analysis
- 12.00-13.30pm Lunch
- 13.30-15.00pm S05 Multidimensional scaling
- 15.00-15.30pm Coffee/tea break
- 15.30-17.00pm S06 Correspondence analysis

Friday 28th of July 2023
- 08.30-10.00am S07 Canonical corrrelation analysis
- 10.00-10.30am Coffee/tea break
- 10.30-12.00am S08 Cluster analysis
- 12.00-13.30pm Lunch
- 13.30-15.00pm S09 Discriminant analysis
- 15.00-15.30pm Coffee/tea break
- 15.30-17.00pm S10 Multivariate normal & multivariate inference

Materials: slides, data and software

- Slides available in PDF format at the module's website
- Data sets
  - SNPs and STRs.
  - Genetic data sets from public repositories
  - Data sets from scientific articles
- Software
  - R (we use version 4.3.1) and R studio
  - .R scripts (provided)
  - R packages

Didactic approach

- Students can look at the slides prior to the session.
- In each session we:
  - summarise key concepts
  - provide hands-on training for analysis in the R environment
  - raise and answer questions

Bibliography

- Manly, B.F.J. (1989) *Multivariate statistical methods: a primer*. 3rd edition. Chapman and Hall, London.
- Johnson, R.A. & Wichern, D.W. (2002) *Applied Multivariate Statistical Analysis*, 5th edition, Prentice Hall.
- Mardia, K.V. et al. (1979) *Multivariate Analysis*. Academic press.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *An Introduction to Statistical Learning*. Springer, New York.