Module 18 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

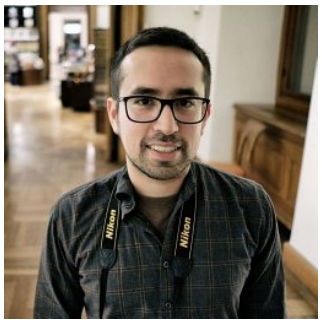
[2]Department of Biostatistics
University of Washington
Seattle, WA, USA

26th Summer Institute in Statistical Genetics (SISG 2021)

Human resources

Jan Graffelman
jan.graffelman@upc.edu
Associate professor in Statistics at UPC, Barcelona
Affiliate associate professor at UW Biostatistics, Seattle
Accredited as full professor (ANECA, Spain)

Iván Galván Femenía
ivan.galvan.femenia@gmail.com
Postdoc in Genome Data Science
Institute for Research in Biomedicine, Barcelona, Spain
Teaching Assistent for Module 18 at SISG 2021



Manuel Muñoz Aguirre
manuel.munoz@crg.eu
Postdoc at Computational Biology of RNA Processing Lab
Centre for Genomic Regulation, Barcelona, Spain
Teaching Assistent for Module 18 at SISG 2021

Course program

Wednesday 21th of July 2021 (Pacific time)

- 11.30-12.20pm S01: Introduction; Matrix algebra
- 12.30-13.30pm Lunch
- 13.30-14.20pm S02: Matrix decompositions

Thursday 22nd of July 2021

- 08.00-08.50am S03 Biplots
- 09.00-09.50am S04 Principal component analysis
- 10.00-10.50am S05 Logratio principal component analysis
- 11.00-11.50am S06 Multidimensional scaling
- 12.00-12.30pm Lunch
- 12.30-13.20pm S07 Correspondence analysis
- 13.30-14.20pm S08 Canonical corrrelation analysis

Friday 23rd of July 2021

- 08.00-08.50am S09 Cluster analysis I
- 09.00-09.50am S10 Cluster analysis II
- 10.00-10.50am S11 Discriminant analysis I
- 11.00-11.50am S12 Discriminant analysis II
- 12.00-12.30pm Lunch
- 12.30-13.20pm S13 Multivariate normal distribution
- 13.30-14.20pm S14 Multivariate inference

Materials: slides, data and software

- Slides available in PDF format at the module's website
- Recorded lectures available via 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 4.0.5) and R studio
  - .R scripts
  - PLINK

Didactic approach

- Students look at the slides prior to the online session.
- At the online session we:
  - summarise key concepts
  - provide hands-on training for analysis in the R environment
  - raise and answer questions
- Please take the module's pre-survey

Online interaction

- Post your question in the Zoom chat window.
- Expect an answer in the chat window.
- If it is common issue, we may unmute you and ask you to speak.
- You will receive a few polls. Please answer to give us feedback.
- We will use breakout rooms, dividing the class in small groups for resolving exercises.

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.