



Basic Statistical Analysis

Rodrigo Labouriau

Applied Statistics Laboratory
Department of Mathematics
Aarhus University

Preliminaries - 2024

1

¹Copyright © 2024 by Rodrigo Labouriau.

This material is only for internal use in the course. Please, do not circulate and do not record.





General Remark

This material is only for internal use in the course.

Please, do not circulate and **do not record**.





Preliminaries

- Practical matters
- Course overview
- Round presentation





PhD Courses at the Applied Statistics Laboratory

- **Introduction to R:**
Learning rudiments of R, 1 ECTS
- **Basic Statistic Analysis:**
Understanding the construction and use of simple statistical models, 4 ECTS
- **Mixed models:**
Some models for situations where the observations are not independent or arrive from sampling, 3 ECTS
- **Planning of Experiments:**
Designing simple comparative experiments and simple observational studies, 3 ECTS (under implementation)
- **Laboratory of Statistics:**
Discussing and performing statistical analyses in collaboration with statisticians, 5 ECTS (under implementation)





Practical Matters

- Course material posted on the page:

<https://tildeweb.au.dk/au33031/Courses/BasicStatisticalAnalysisWorkPage>

Username: BSA

Password: Abrir





Practical matters

- **Aim of the course:** to introduce the student to basic notions of statistical modelling and give an idea of a typical statistical modelling process
- The student should become able to:
 - Identify the key assumptions and critically analyse some chosen (simple) statistical model
 - Perform inference and draw conclusions using those models
 - Present/report the results of those analyses
- The aim of the course is to introduce the student to basic notions of statistical analyses and give an idea of a typical statistical modelling process. The course does not intend to cover systematically key statistical models nor to supply a large applicable statistical tool box in the research area of the student. Instead, the idea is build up a solid basis on general statistical principles, which will allow the student to understand and apply more complex statistical models used in their research area or in other statistical courses. The examples used are based on relatively simple real cases occurring in life and environmental sciences strategically chosen for pedagogical purposes.
- We will use R as a tool, **but this is not a course on R!**





Practical Matters:

Language and software:

- The course will be in English
- Course requirement:
Basic notions of \mathbb{R}
- I will assume from this point that you have this requirement.





Practical Matters: Tentative time schedule - 5 modules, 8 lecture days

- **Module 1 - Basic probability and statistics (first week)**
 - Lecture 1 - Tuesday: Overview and basic probability theory
 - Lecture 2 - Wednesday: Some more basic probability theory, parametric statistical models and basic statistical inference techniques
- **Module 2 - Binomial models (second week)**
 - Lecture 3 - Tuesday: Binomial models with discrete explanatory variables, binomial models with one- and two-way classification structures
 - Lecture 4 - Wednesday: Binomial models with continuous explanatory variables
- **Module 3 - Poisson models (third week)**
 - Lecture 5 - Tuesday: The Poisson distribution, Poisson models with continuous explanatory variable
 - Lecture 6 - Wednesday: Poisson models with discrete and discrete explanatory variables
- **Module 4 - Normal models (fourth week)**
 - Lecture 7 - Tuesday: Normal distribution, normal models with discrete and continuous explanatory variables
- **Module 5 - Concluding (fourth week)**
 - Lecture 8 - Wednesday: Case studies and techniques for model control, overview and concluding remarks
- **Oral exam: 18th or 19th of April 2024** (probably online via Zoom)





Practical Matters: Office hours

- I would like to reserve 1 h a week for consultations on matters related to the course (not in the first week)

Location: Rodrigo's office

Time: To be arranged in class.

Suggestion: Wednesday's lecture from 12:30 to 13:30
(not in the first week)

- **Instructions to go to Rodrigo's office:** Take the lift at the building 1530 (the main building) to the fourth floor, from the lift, take to the right and then take the corridor to the right, the office is located close at the end of this corridor.





Practical Matters:

The rules of the game:

- Each student will be evaluated in the final examination consisting of a written report and an oral examination based on a set of analyses proposed by the instructor
- A decision on whether the student is approved for the course or not will be taken based on this examination (together with an internal censor)
- A student is allowed to make a final examination provided that:
 - She/he has attended at least 5 of the 8 lectures
 - She/he delivers the final report on the assignments before the deadline (15th of April 2024, 11:59 pm!)





Practical Matters: Examination

- Six questions will be placed in the work-page of the course (a draft of the exam is already posted for discussion, final version, after discussion intended to be ready in the second week of the course)
- Each question involves producing one statistical analysis and it is supplied:
 - A basic description of the problem
 - A dataset
 - A R-program performing some analyses
- You have to choose three of those questions (among the 6 questions)
 - Each question involves one of three distributions: binomial, Poisson or normal
 - Each question involves one of two types of explanatory variables: Continuous or discrete

Your choice should be such that you have one analysis involving each of the distributions and the three analyses should not involve the same type of explanatory variable





Practical Matters: Examination

- You should prepare a **short** report explaining the key points of the analyses
- You should present and discuss orally the analyses you choose:
(to me and an internal observer)
30-45 minutes scheduled the 18th or 19th of March 2024
Location: To be arranged (online via Zoom)
- You might (if necessary, and you are very welcome to) discuss parts of the examination during the course in the period before the delivery of the examination
- You might present a draft of the written report and get feedback from me, **if** the draft is delivered before the 2nd of April 2024 (11:58pm)
- **The deadline for delivering the reports is the 15th of April 2024, 11:59 pm!**





The aStatLab

- The Applied Statistics Laboratory (aStatLab) at the Department of Mathematics maintains a statistical support service for all PhD Students, Post Doc Fellows and Researchers **from the Faculty of Natural Sciences (NAT)** at Aarhus University.
- The statistical support is composed of the following activities:
 - A system of statistical consultation (simple, daily questions)
 - Short ad hoc internal projects
(simple, well-defined questions requiring that a statistician from the aStaLab performs some specific analyses, short-duration max. one week of workload)
 - Statistical support for PhD and Master at NAT
(follow-up of statistical analyses related to PhD and Master theses).
- These activities do not involve costs for the clients, but the aStaLab reserves the right to limit the total workload of the activities.





Short Presentation

- **Rodrigo Labouriau**
Head of the Applied Statistics Laboratory, MATH, AU
- **Research interest:** Statistical and Mathematical Biology
Construction of new statistical methods inspired in applications
Complex mathematical/statistical models with applications in life and environmental sciences
Some areas of theoretical statistics, fundaments of statistics
- **Education:**
Agronomy/Biology (1982)
Master: Mathematics with specialisation in statistics (1988, IMPA)
Thesis: Robustness in exponential families
PhD: Mathematical Statistics
Thesis: Semiparametric models; Minor: Functional analysis, probability theory and spectral theory
Post-doc: Mathematical Statistics semiparametric models and mathematical foundation of Statistics (Utrecht, The Netherlands)
- Around 30 years experience in applied statistics in a variety of subjects:
Agricultural Sciences, Epidemiology, Social Sciences, Genetics and Evolution, Molecular Biology etc
Co-authored with more than 450 different researchers.
- **Hobbies:** MTB, Zwift, swim, play cello, play guitar (bossa-nova and jazz), visit labs and field experiments, work on foundations of statistics





Preliminaries

Round presentation

Please tell us:

- Your name
- The department (or group) you come from
- A few words on your PhD/research project

