Hybrid course (onsite for IMR staff and online for external participants)

Data Exploration, Regression, GLM & GAM with introduction to R

Provided by: Highland Statistics Ltd

Organised by: The IMR Academy, Institute of Marine Research, Bergen, Norway

This is an onsite course for IMR staff and students, with the option for non-IMR participants to join online via Zoom.

This (live) online course consists of 5 modules representing a total of approximately 40 hours of work. Each module consists of live teaching, followed by exercises using real data sets. All the material is also available as on-demand video files. A discussion board allows for daily interaction between instructors and participants. The course includes a 1-hour face-to-face video chat with the instructors.

Course content

We begin with a short introduction to R and provide a protocol for data exploration to avoid common statistical problems. We will discuss how to detect outliers, deal with collinearity and transformations.

An important statistical tool is multiple linear regression. Various basic linear regression topics will be explained from a biological point of view. We will discuss potential problems and show how generalised linear models (GLM) can be used to analyse count data, continuous data, presence-absence data and proportional data.

Sometimes, parametric models (linear regression, GLM) do not quite fit the data and in such cases generalised additive models (GAM; a smoothing technique) can be used.

Date & Venue

Dates & times:
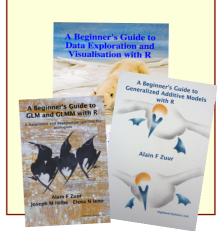- 17-21 November 2025
- 08.30-15.30 (Norwegian time)

Price for non-IMR participants: £500

Venue: IMR, Bergen, Norway

Instructors: Dr. Alain Zuur
　　　　　 Dr. Elena Ieno
Authors of 10 books and providers of over 300 courses

Included: 1 hour free face-to-face video chat about

KEYWORDS

Introduction to R. Outliers. Transformations. Collinearity (correlation between covariates). Multiple linear regression. Model selection. Visualising results. Poisson, Generalised Poisson, Negative binomial, Bernoulli, Gamma, Tweedie, beta and binomial GLMs.

COURSE CONTENT

Day 1:
- General introduction.
- Introduction to R.
- Theory presentation on data exploration (outliers, collinearity, relationships, interactions, etc.).
- One exercise.

Day 2:
- Theory presentation on linear regression.
    - Different strategies for model selection.
    - Interactions.
    - Dealing with categorical covariates.
    - Visualising covariate effects.
- Two exercises.

Days 3 and 4:
- Theory presentations on Poisson, negative binomial, generalised Poisson, Tweedie, Gamma, Bernoulli, binomial and beta distributions.
- Exercise 1: Poisson/generalised Poisson/negative binomial GLM for count data.
- Exercise 2: Bernoulli GLM for absence/presence data.
- Exercise 3: Gamma or Tweedie GLM for continuous (e.g. biomass) data.
- Time allowing: Exercises on beta GLM (for proportional data) and binomial GLM (proportional data). On-demand video files are available.

Day 5:
- Theory presentation on GAM.
    - Two exercises using Gaussian GAM and Poisson and negative binomial GAMs.
    - Based on various chapters in Zuur (2012).

FREE 1-HOUR FACE-TO-FACE MEETING
The course includes a 1-hour face-to-face meeting with one or both instructors. The meeting needs to take place within 12 months after the last live zoom meeting. You can discuss your own data, but we strongly advice that the statistical topics are within the content of the course. The 1-hour needs to be consumed in one session, and will take place at a mutual convenient time.

GENERAL INFORMATION

COURSE TIMES:
- 08.30am to 15.30pm including 1 hour lunch break and a 20 minutes break both morning and afternoon.

COURSE MATERIAL:
- Pdf files of all presentations are provided
- The powerpoint files are based on various chapters from:
  - Chapters 4 - 5 from Zuur, Ieno, Smith (2007). *Analysing Ecological Data*
  - *A Beginner's Guide to GLM and GLMM using MCMC with R.* (2013).
  - A *Beginner's Guide to Zero Inflated Models with R.* (2016)
  - Chapter 3 in *Beginner's Guide to GAM with R.* Zuur (2013).
  - Books are not included. The course can be followed without purchasing these books.

PRE-REQUIRED KNOWLEDGE:
Basic statistics (e.g. mean, variance, normality). No R knowledge is required. You will learn R 'on the fly'. This is a non-technical course.

RECOMMEND LITERATURE:
- Zuur, Hilbe, Ieno (2013). *A Beginner's Guide to GLM and GLMM with R.*
- Ieno, Zuur (2015) A *Beginner's Guide to Data Exploration and Visualisation with R.*
- Zuur (2013). *A Beginner's Guide to GAM with R.*
- These books are available from www.highstat.com

GENERAL
- Please ensure that you have system administration rights to install R and R packages on your computer.
- Instructions what to install will be provided before the start of the course.
- Course participants will be given access to the course website with all the videos, data sets, R solution code and course material one week before the start of the course.
- This is a hybrid (onsite and online) course. We will use Zoom. You will receive a link for the Zoom meeting and a password to enter the meeting.
- All video files are on-demand and can be watched online, as often as you want, at any time of the day, within a 12 month period.

INFORMATION ON COURSE CONTENT
Dr Alain F Zuur
highstat@highstat.com
www.highstat.com