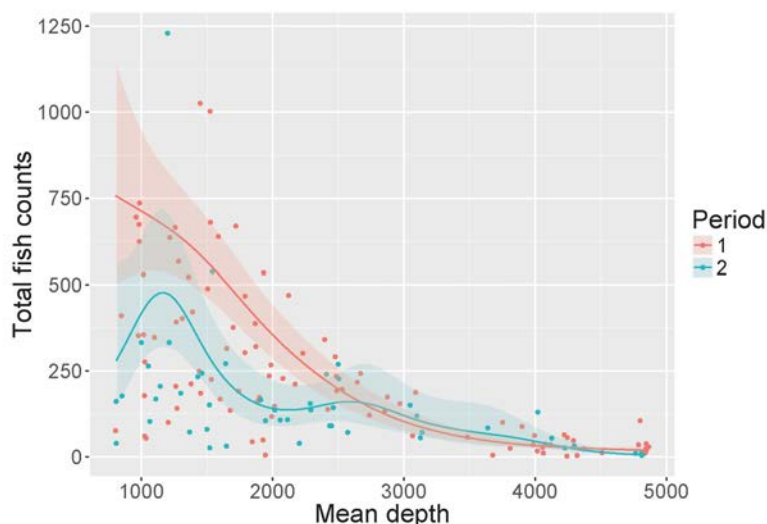# Data Exploration, Regression, GLM & GAM with introduction to R

Provided by: Highland Statistics Ltd

In cooperation with:

cE3c - CCIAM, Faculty of Sciences, University of Lisbon, Portugal
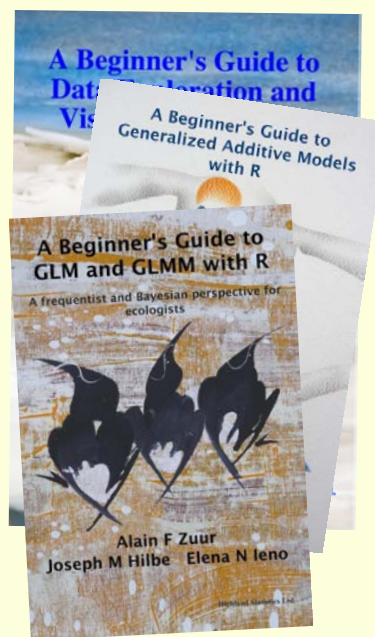

Date & Venue

Date: 18 - 22 February 2019

Venue: PT Meeting Centre, R. Bojador 47, Parque das Nações, Lisbon, Portugal

Price: £500

Instructors: Dr. Alain Zuur
Dr. Elena Ieno

Authors of 8 books and

We begin with an 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, 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.

During the course several case studies are presented, in which the statistical theory is integrated with applied analyses in a clear and understandable manner.

KEYWORDS

Introduction to R. Outliers. Transformations. Collinearity (correlation between covariates). Multiple linear regression. Model selection. Visualising results. Poisson GLM. Overdispersion. Negative binomial GLM. Binary and proportional data. ggplot2. Logistic regression.

COURSE CONTENT

Monday:
- General introduction.
- Introduction to R.
- Theory presentation on data exploration (outliers, collinearity, transformations, relationships, interactions).
  - Based on Zuur et al. (2010) and Ieno and Zuur (2015).
- Two exercises.

Tuesday & Wednesday morning:
- Theory presentation on linear regression.
  - Different strategies for model selection.
  - Interactions.
  - Dealing with categorical covariates.
  - Sketching model fit.
- Two exercises.

Wednesday afternoon and Thursday:
- Theory presentation on Poisson, negative binomial, Bernoulli and binomial distributions.
  - Based on Chapter 1 in Zuur et al. (2013).
- Theory presentation on GLM.
  - Poisson GLM.
  - Negative binomial GLM.
  - Bernoulli GLM.
  - Based on Chapter 1 in Zuur et al. (2013).
  - How to deal with overdispersion.
- Three exercises.

Friday:
- Theory presentation on GAM.
  - Two exercises using Gaussian GAM and Poisson and negative binomial GAMs.
  - Based on various chapters in Zuur (2012).
  - What to present in a paper

Time allowing:
  - Short theory presentation on binomial GLM and gamma GLM.
  - Two exercises for binomial and gamma GLM.
    - *Video solution files are available.*

Cited literature:
  - Zuur, Ieno, Elphick. (2010). A protocol for data exploration to avoiding common statistical problems. *Methods in Ecology and Evolution,* **1***: 3-14.*
  - Zuur (2013) Beginner's Guide to GAM with R.
  - Zuur, Hilbe, Ieno (2013). Beginner's Guide to GLM and GLMM with R.
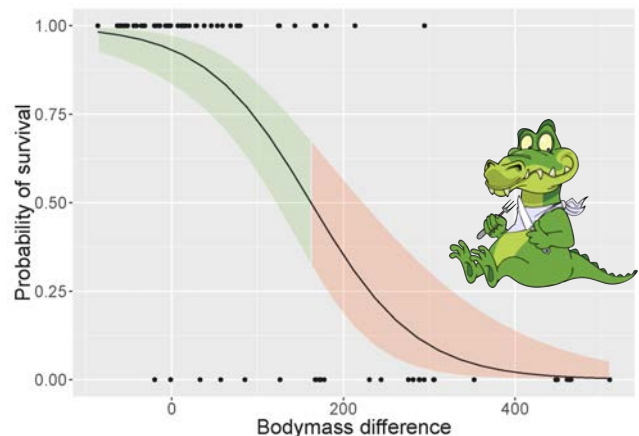
GENERAL INFORMATION

COURSE FEE: £500
- Credit card payments are charged in GBP currency. UK participants are subject to 20% VAT. EU participants (but non-UK) are not subject to UK VAT, but need to provide their institutional VAT number. Non-EU participants are not subject to VAT.
- The course fee excludes refreshments and lunch.

COURSE TIMES:
- 09.00am to 16.00pm including 1 hour lunch break and a 20 minutes break both morning and afternoon

COURSE MATERIAL:
- Pdf files of all powerpoint 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 in the course fee. 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.

CANCELLATION POLICY:
What if you are not able to participate? Once participants are given access to course exercises with R solution codes, pdf files of certain book chapters, pdf files of powerpoint files and video solution files, all course fees are <u>non-refundable</u>. However, we will offer you the option to attend a future course or you can authorise a colleague to attend this course. Terms and conditions see: http://highstat.com/index.php/sign-up2

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
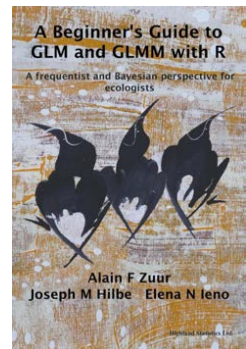
GENERAL
- You need to bring your own laptop.
  - Please ensure that you have system administration rights to install R and R packages on your computer.
  - Instructions what to install will be provided before the start of the course.
- You will be given access to a course website with around 25 - 30 fully worked out R exercises. These are all based on published papers and real data sets.
- *You will also get access to 11 fully worked out exercises with video solution files.*

REGISTRATION

www.highstat.com

Dr Alain F Zuur
highstat@highstat.com
www.highstat.com
Payment via credit card or bank transfer