ST2304 - Statistical Modelling for Biologists/Biotechnologists

Bob O'Hara

bob.ohara@ntnu.no

This week we will...

Start the course

- admin
- try to motivate you
- overview of the course

Start learning R

- introduction
- hands-on work

Administration Matters

(we will deal with these in more detail later)

- ▶ Reference Group
- ▶ Blackboard
- ▶ web page: https://www.math.ntnu.no/emner/ST2304/2023v/

How the Course Will Run: Modules

One module a week

Modules on web pages

- ▶ text
- ▶ exercises (with hints)
- ▶ short recorded lectures

Active Learning, Group work, Problem solving

How the Course Will Run: Contact time

Any announcements at the start

In effect everything will be exercise sessions in groups

- ▶ "Lectures" mainly for the modules
- ▶ Exercise session will be mainly for the exercises

(but feel free to ignore these)

Work in groups, ask for help when you want it.

Assessment

Complete 8 exercise sets (of about 10)

- ▶ do in groups
- ▶ pass/fail
- ▶ first couple of weeks won't count
  - ▶ we will tell you when they start to count

Virtual or in Person?

https://www.menti.com/al6sim51dgx9

Teachers

Me

Bert van der Veen

TA: Kenneth Aase

Resources

Blackboard

- announcements
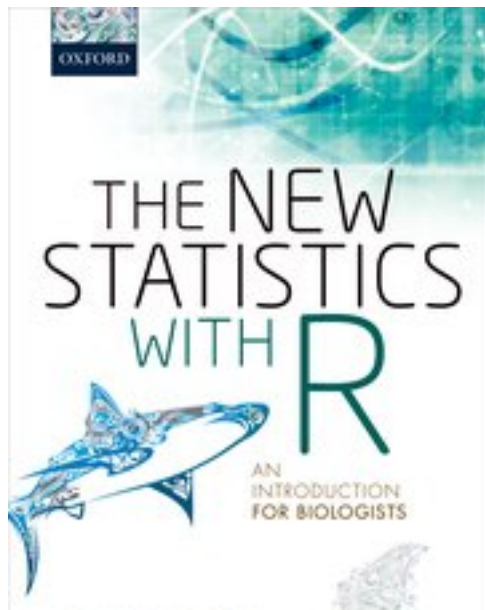- links to more material
- exercises

Discussion board on Discourse:
https://mattelab2023v.math.ntnu.no/c/st2304/52

Web page: https://wiki.math.ntnu.no/st2304/2023v/start (this includes other links from this presentation)

- modules

Text books

Text Books

New Statistics with R - Andy Hector

Text Books

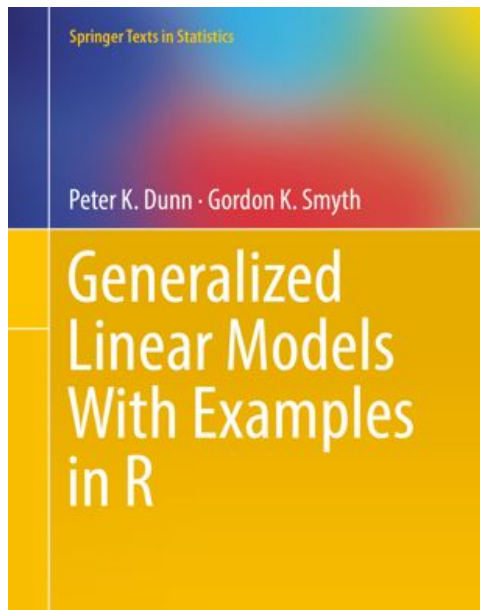The Analysis of Biological Data - Whitlock & Schluter

The Analysis of Biological Data

Text Books

Generalized Linear Models With Examples in R - Dunn & Smyth

Springer Texts in Statistics

Peter K. Dunn · Gordon K. Smyth

Generalized Linear Models With Examples in R

Other Resources

Blackboard

Discourse

Wiki

Google (yes, use it!)

Recap: why do we we use statistics in biology

What do you remember from ST0103?

Discuss in groups,

- ▶ Come up with 3 topics you learned about, and for each give an example where they are used in biology (or biotechnology!)
- ▶ Add the topics to TaskCard: https://ntnu.taskcards.app/#/board/448a3a68-62df-4a5a-9820-0d03428d06fb

.

What we are aiming for

We want you to be able to analyse your own data (and understand what you are doing!)

- ▶ fit the right models to data
- ▶ assess if the model is any good
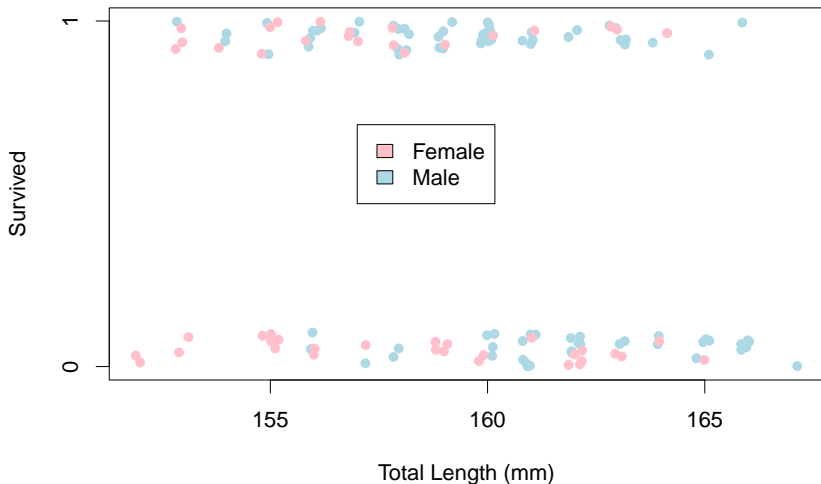- ▶ compare models and decide which is 'best'
- ▶ interpret the models

Types of model I: Linear models (regression, ANOVA)
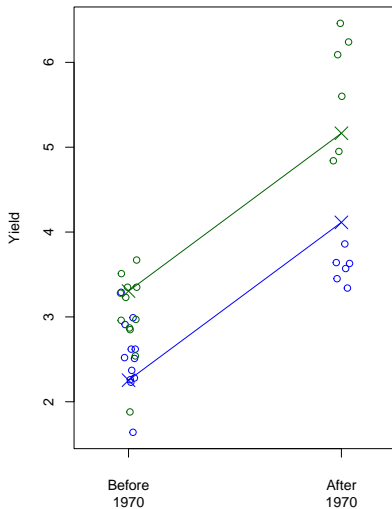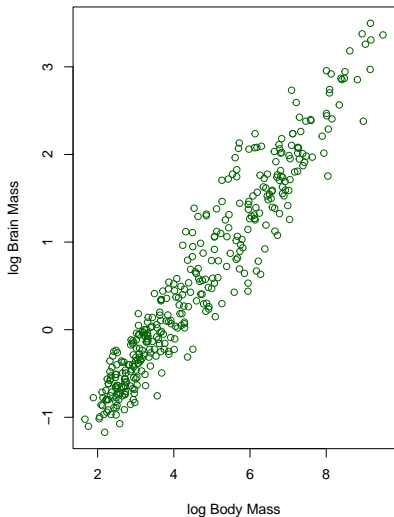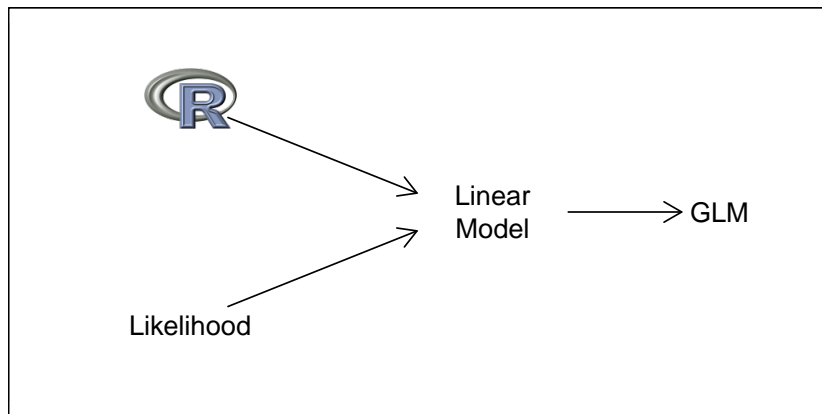
Types of model I: Generalised Linear models

When things aren't normal

- ▶ binary (e.g. survive/died)
- ▶ counts (e.g. how many sparrows are there?)

How do we get there?

Need some theory (likelihood) and to know how to write the models in the computer (R)

Then can start modelling. linear models is complicated regression

GLMs are complicated linear models

Likelihood

The statistical framework to do this

Likelihood = probability of the data

- ▶ means we can write everything as probabilities

R

The stats package we will use

- free, most commonly used
- more shortly

Course Structure

Weeks 1-3: Likelihood and R

- ▶ statistical theory, and programming
- ▶ the background you will need to understand what follows, and to do it

Weeks 4-10: Linear models (regression, ANOVA)

- ▶ fitting straight lines

Weeks 11-13: Generalised Linear models

- ▶ fitting straight lines to different types of data