Getting to fit models in **R**

Alexandre Courtiol

Leibniz Institute of Zoo and Wildlife Research

June 2018

Getting started with **R**

What is a linear model?

A statistical model represents, often in considerably idealized form, the data-generating process (`https://en.wikipedia.org/wiki/Statistical_model`).

What is a linear model?

A statistical model represents, often in considerably idealized form, the data-generating process (`https://en.wikipedia.org/wiki/Statistical_model`).

In a linear model, the data-generating process is assumed to be a linear function: it is constructed from a set of terms by multiplying each term by a constant (a model parameter) and adding the results.

What is a linear model?

A statistical model represents, often in considerably idealized form, the data-generating process (`https://en.wikipedia.org/wiki/Statistical_model`).

In a linear model, the data-generating process is assumed to be a linear function: it is constructed from a set of terms by multiplying each term by a constant (a model parameter) and adding the results.

R allows to fit efficiently and easily all main kinds of linear models:

- classical linear models (t-test, correlation, linear regression, ANOVA, ANCOVA): LM
- generalized linear models (logistic regression, Poisson regression...): GLM
- linear mixed-effects models: LMM
- generalized linear mixed-effects models: GLMM
- general additive models & general additive mixed models: GAM & GAMM

Linear models in **R**

R is very rich in terms of capabilities to fit linear models due to an increasing number of dedicated packages!

For now, no other software seems to be remotely as good (prognostic: only `Julia` or `Python` may change that within a decade but I find it unlikely).

Linear models in **R**

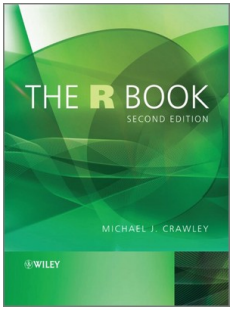
R is very rich in terms of capabilities to fit linear models due to an increasing number of dedicated packages!

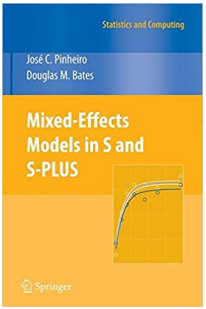
For now, no other software seems to be remotely as good (prognostic: only `Julia` or `Python` may change that within a decade but I find it unlikely).

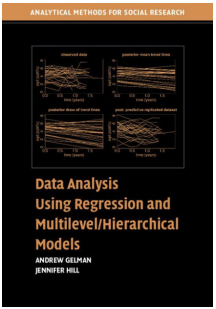
Models	Packages for fitting	Helper packages
LM	none; `spaMM`	`car`; `visreg`
GLM	none; `spaMM`	`car`; `DHARMa`; `visreg`
LMM	`lme4`; `spaMM`; `glmmTMB`	`DHARMa`; `pbkrtest`
GLMM	`lme4`; `spaMM`; `glmmTMB`	`DHARMa`; `pbkrtest`
GAM	`mgcv`	`visreg`
GAMM	`mgcv`	

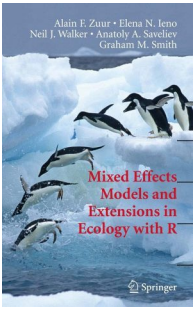
Note: those are my personal favorite ones but they are plenty more out there.

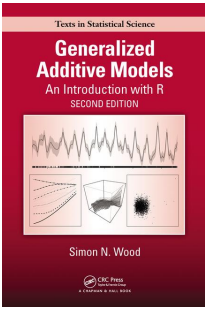
Useful books dealing with linear models in **R**











Note: it is also useful to look at books not focussed on **R**!

Preparing data for (G)LM(M) & GA(M)

To maximize the chances of success prepare your data as follow:

- one row = one observation (if repeated measures, use several rows!)
- qualitative variables of class `factor` (check the levels, drop unused ones, set the reference properly)
- no `NA` (models can somewhat deal with them but it is a major source of headackes)

Getting started with **R**

Getting started with **R**

The linear model

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 \times x_{1,i} + \hat{\beta}_2 \times x_{2,i} + \cdots + \hat{\beta}_p \times x_{p,i} + \varepsilon_i$$

- y_i = the observations to explain / response variable / dependent variable
- $x_{j,i}$ = constants derived from the predictors / explanatory variables / independent variables
- $\hat{\beta}_j$ = the (model parameter / regression coefficient) estimates
- ε_i = the residuals (i.e. the estimates for the errors)

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**

Getting started with **R**