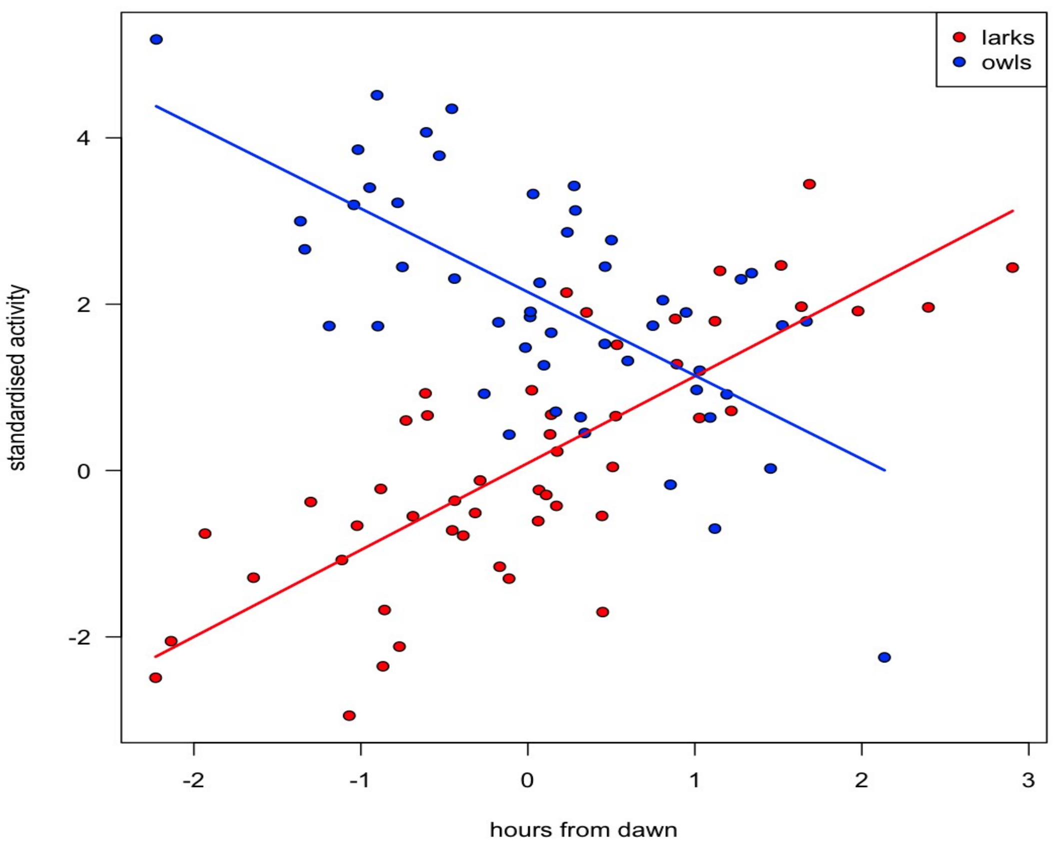
6F7Z1012
Statistics and Research Design

Week 11: Linear models and extensions

Dr Ed Harris



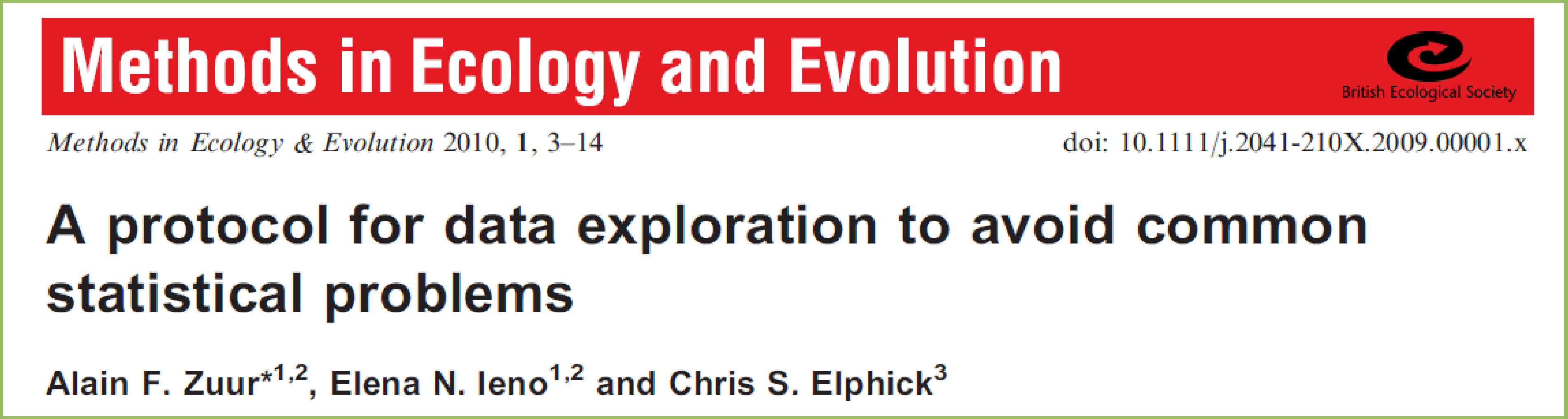
Announcements

-Syllabus

-Lecture: linear model extensions

-R lab: linear model extensions

-Questions / queries?



Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology & Evolution 2010, **1**, 3–14 doi: 10.1111/j.2041-210X.2009.00001.x

A protocol for data exploration to avoid common statistical problems

Alain F. Zuur*[1,2], Elena N. Ieno[1,2] and Chris S. Elphick[3]

The basic statistical toolbox:

-correlation

-regression

-t-test

-1- and 2-way ANOVA

-1-way ANOVA with a covariate (ANCOVA)

The basic statistical toolbox:

-correlation

-regression

-t-test

-1- and 2-way ANOVA

-1-way ANOVA with a covariate (ANCOVA)

All are specific instances of the “general linear model”

All are specific instances of the “general linear model”

General form:

$$Y = \text{alpha} + B1*X1 + \text{error}$$

Here we assume the error follows a normal distribution

$Y = alpha + B1*X1 + error$

Actually though, there can be many B and X terms

$Y = alpha + B1*X1 + B2*X2 + \ldots + Bn*Xn + error$

We have already looked at this

Y = alpha + B1*X1 + error

Today we will talk about 2 specific extensions to the general linear model

1) The Mixed Model

2) The GeneraliZED linear model (GLM)

1) The Mixed Model

The bad news:

The mixed model differs from the standard linear model in
A few ways

-all observations might not be independent

-there might be repeated measures on the same individuals

-variables might be “nested in one another”

1) The Mixed Model

The good news:

You can (must) control for these causes of non-independence
In your data very easily

Add a term to account for it!

$$Y = alpha + B1*X1 + RandomEffect + error$$

Y = alpha + B1*X1 + RandomEffect + error

A variable that adds such error is called a “random effect”

Our regular old independent variables are called “fixed effect”

Examples:

A bunch of samples taken in several different plots. Here, samples in a **plot** might be correlated – hence **plot** could be treated as a random effect

Chicks in **nests**

Mice in **cages** etc. etc.

Y = alpha + B1*X1 + RandomEffect + error

A few different packages in R deal with random effects and Mixed models

One of the oldest is the lme() function in the {nlme} package

Another is the {lme4} package

Y = alpha + B1*X1 + RandomEffect + error

Examples:

```
library(nlme)

#Usage

#lme(fixed, data, random, correlation, weights, subset, method,
#    na.action, control, contrasts = NULL, keep.data = TRUE)

M1 <- lme(LSpobee ~ fInfection01 * BeesN,
         random =~ 1 | fHive,
         data = Bees, method = "REML")
```

Bees data

Honey bee infection rate (based on counts of fungal spores), hive, Bee number, x and y coordinates, and presence or absence of Infection

The dependent var is the spores/bee

```
M1 <- lme(LSpobee ~ fInfection01 * BeesN,
          random =~ 1 | fHive,
          data = Bees, method = "REML")
```

We'll look at this data tonight...

Bees data

```
> summary(M1)
Linear mixed-effects model fit by REML
 Data: Bees
       AIC      BIC    logLik
  175.0129 188.3299 -81.50643

Random effects:
 Formula: ~1 | fHive
        (Intercept)  Residual
StdDev:   0.9666873 0.3373335

Fixed effects: LSpobee ~ fInfection01 * BeesN
                        Value Std.Error DF    t-value p-value
(Intercept)          2.643551 0.9281957 48  2.8480532  0.0065
fInfection011        3.646261 1.5420564 20  2.3645448  0.0283
BeesN               -0.000012 0.0000127 20 -0.9829788  0.3374
fInfection011:BeesN -0.000016 0.0000234 20 -0.6790603  0.5049
```

2) The GeneraliZED linear model (GLM)

This is an extension of the regular linear model framework

Here, the assumption of normal error distribution can be relaxed

The good news:

Often times error is not normalized in “real world” data and this approach makes it possible to analyse such data

The bad news:

You must specify the non-normal distribution!

2) The GeneraliZED linear model (GLM)

This is an extension of the regular linear model framework

Here, the assumption of normal error distribution can be relaxed

The good news:

Often times error is not normalized in “real world” data and this approach makes it possible to analyse such data

The bad news:

You must specify the non-normal distribution!

Red Squirrel data

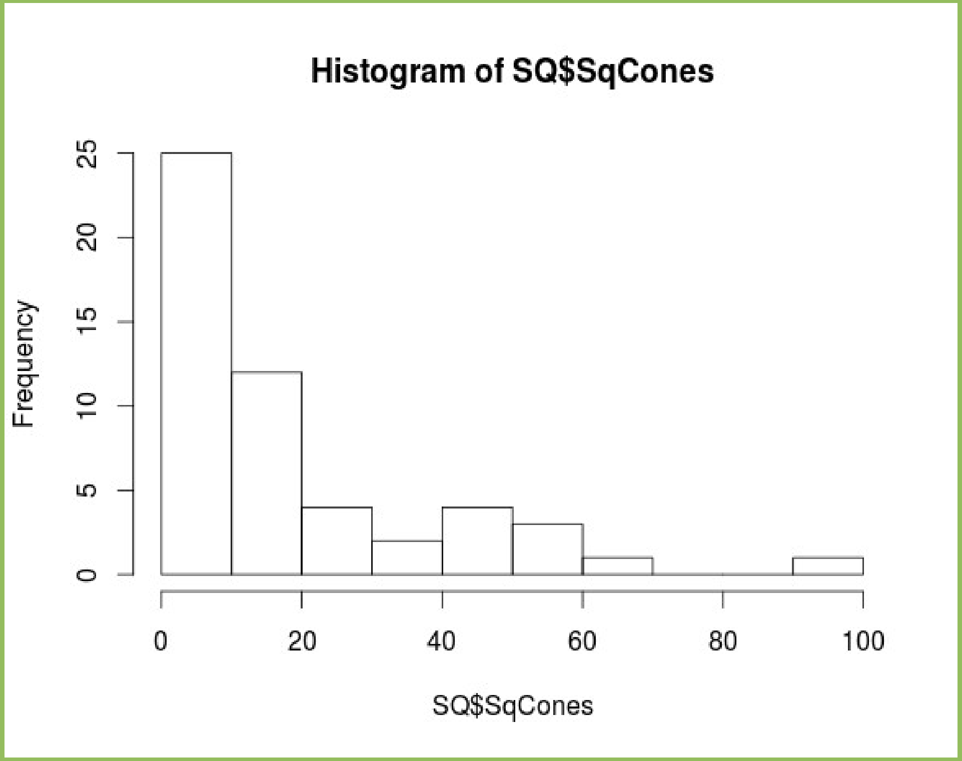
Simple data set looking at habitat characteristics that might affect the density of cones chewed on my squirrels

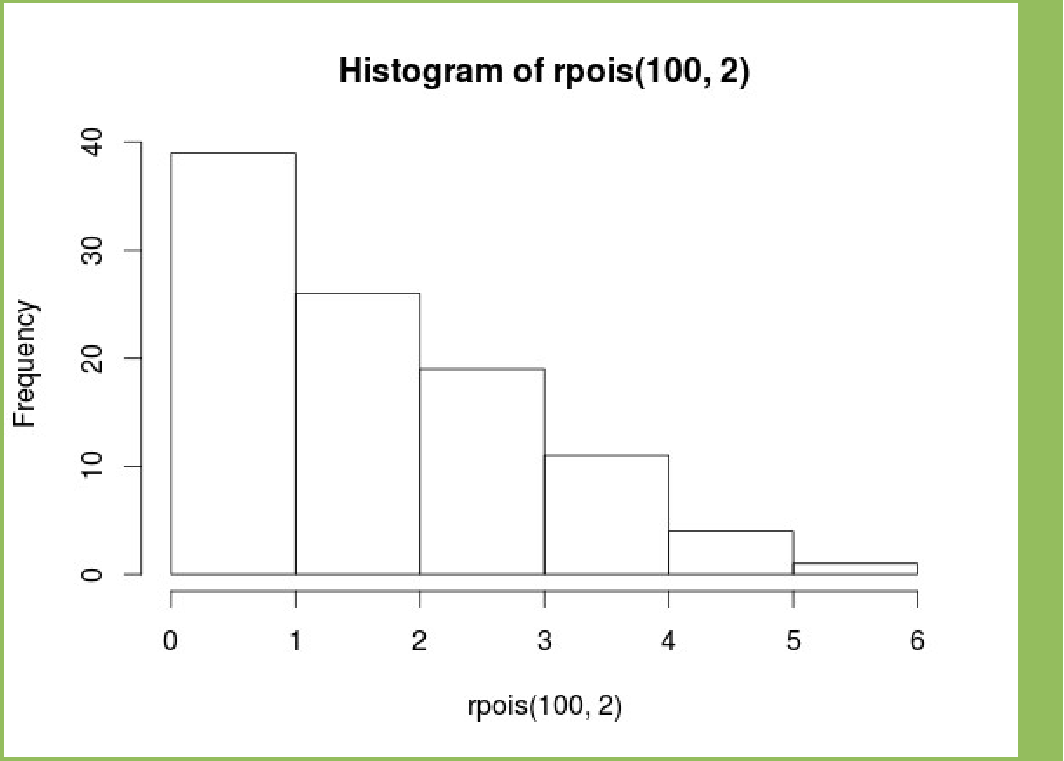
E.g., tree height, canopy cover, DBH, N trees.

The dependent variable is the count of chewed cones

Count data is often Poisson distributed

Red Squirrel data





Red Squirrel data

Here, we'll use the glm() function and we'll specify the “family” or shape of the error.

```
M1 <- glm(SqCones ~ Ntrees.std +  TreeHeight.std +
          CanopyCover.std,
          family = "poisson",
          data = SQ2)
```

Red Squirrel data

```
> summary(M1)

Call:
glm(formula = SqCones ~ Ntrees.std + TreeHeight.std + CanopyCover.std,
    family = "poisson", data = SQ2)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-6.581  -3.388  -1.385   1.488   7.324

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.62399    0.04252  61.718  < 2e-16 ***
Ntrees.std       0.27415    0.02889   9.490  < 2e-16 ***
TreeHeight.std   0.19669    0.04601   4.275 1.91e-05 ***
CanopyCover.std  0.52852    0.06602   8.006 1.19e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```