Exercises for linear mixed effect models, part 3

Timothée Bonnet

April 15, 2019

Contents

1 Logistic regression basics

* Exercise 1 Fit with glm()

Load the dataset `survivalsize.csv`. It contains fake data of individual-based measurements of body size and of survival from the time of measurement to the next year. Look at a summary of the data and plot them. Do you think size affects survival? Use the function glm() to fit a logistic regression. What should the `family` argument be? What is the direction of the effect of size on survival?

* Exercise 2 Model assumptions

In R some model assumptions of linear models are routinely checked using plot(lm()): residual normality, independance and homogeneous variance, and legerage. If you know about these diagnostics (and what the plots should ideally look like) check them for your glm. Should you worry?

** Exercise 3 Making sense of model estimates

Consider the model estimates for the intercept and slope. How to recover the mean survival from these two numbers? You will need to use the back-transformation inverse logit: probability $= \frac{1}{1+\exp(-y)}$ where y is a model prediction on the logit-scale. You can also use the function `predict()`.

How much does the ratio of surviving/dying probabilities changes for an individual of size 6 compared to an individual of size 5?

** Exercise 4 Visualize results

Use the function `predict()` to make a graph of the relationship between survival and size, with a confidence interval.

2 Classification

** Exercise 5 What populations are at risk?

Researchers are currently trying to predict animal population collapses in advance, by finding metrics that changed before historic populations collapsed. Let's imagine we have measured the time-autocorrelation in phenotypes (what that means does not matter) in 200 populations, some of which have collapsed in the past decades. We want to konw whether autocorrelation is related to collapse, and whether it could be used to predict which population will collapse in the future.

Load the dataset `decline.csv`, it contains fake data of past collapses and past autocorrelations in populations, as well as future collapses (yes, I am a fortune teller). Fit a logistic regression of `popcollapse` as a function of `atcor`. Does `atcor` increase the probability of population collapse? Visualize the relationship. Does it make sense to define a threshold of autocorrelation and allocate conservation resources to those populations that are at risk based on our metric? For instance, what do you think about conserving all populations above an autocorrelation of 0.5? How many future at-risk populations would be missed? How many populations that are not at risk would receive unecessary resources? How could you improve the prediction?

3 Repeatability

** Exercise 6 Repeatability, on what scale?

Load the dataset `runaway.csv`. It contains fake data of a behavioural experiment: you played the trumpet to an animal and see whether it an animal and see whether they ran away. 500 individuals were tested twice a year during 5 years. You measured distance between you and the animal. You would like to know whether individuals behave consistently that is, have a personality and are repeatable. We start by fitting a linear mixed model using the package lme4 (accounting for differences among years, and for the effect of distance):

```
roodat <- read.csv("runaway.csv")
lmm0 <- lmer(RunAway ~ 1 + distance +(1|individual) + (1|year),
             data = roodat)
VarCorr(lmm0)

##  Groups     Name        Std.Dev.
##  individual (Intercept) 0.0663
##  year       (Intercept) 0.0159
##  Residual               0.4139
```

This model suggests a repeatability of:

```
(0.0663^2)/(0.0663^2+0.0159^2+0.4139^2)

## [1] 0.02498
```

That is not a lot, and we wonder whether the small number is due to the lack of fit of the model. Fortunately it is possible to fit a logistic mixed model by changing `lmer` to `glmer` and specifying `family='binomial'`. If you do so, what repeatibility do you find? Where does the difference come from? What is right?