Name: J. Hediger

HOMEWORK 3: Model Selection and Mixed Models

WSCI 6390 – 002: Population Parameter Estimation

Due 11:59 PM Tuesday, February 20

Let’s practice (1) building models and using AIC, AICc, and BIC to compare competing models and (2) building mixed effect models and interpreting that output.

INSTRUCTIONS:

We will be using two datasets for this assignment:

1. A dataset “caterpillar.csv” on growth of caterpillar larvae exposed to different levels of an herbicide in a greenhouse experiment. Caterpillars are housed in small cages in the greenhouses, with forage plants that have been treated. The exposure levels were designed to correspond with field-level treatments of 20, 30, or 40 lb/ha. In addition, the temperatures in the caterpillar cages were monitored, as temperature can affect growth, and you have average temperatures (C). Finally, growth is measured in mg over the course of the experiment.
2. A dataset “bats.csv” We have 300 measurements of little brown bat (*Myotis lucifugus*) body temperatures during hibernation, and we want to know how temperature is related to ambient temperature in the hibernation roost. We put RFID tags on 7 bats in each of 6 different colony roosts. We visited the colony roosts on 10 different days each winter and sampled all the bats we could find. There are 42 bats in the dataset, with between 5 and 10 temperature measures on each (i.e., each bat was located between 5 and 10 times). The data frame includes day, scaled ambient temperature, colony, individual within the colony, unique individual id, and bat temperature.

ASSIGNMENT:

1. First, read in the caterpillar dataset using

caterpillar <- read.csv(YOUR\_DIRECTORY/caterpillar.csv”, header=T)

Use

head(caterpillar)

to see what the top six rows of the dataset look like.

Now let’s build four different models with growth (mg) as a response variable:

#model with intercept only

mod1 <- lm(Growth ~ 1, data=caterpillar)

#model with growth as a function of herbicide exposure

mod2 <- lm(Growth ~ Exposure, data=caterpillar)

#model with growth as a function of temperature

mod3 <- lm(Growth ~ Temperature, data=caterpillar)

#model with growth as a function of herbicide exposure and temperature

mod4 <- lm(Growth ~ Exposure + Temperature, data=caterpillar)

Use

summary(mod4)

to look at the outputs of the model with both covariates.

**What can we say about the effect that herbicide exposure and temperature have on caterpillar growth?**

For every 1 unit of herbicide exposure, caterpillar growth increases 0.12 mg

For every 1 unit of temperature increase, caterpillar growth increases 2.69 mg

1. Let’s ensure that the AICcmodavg library is installed and loaded with

install.packages("AICcmodavg")

library(AICcmodavg)

Use these lines to put the model names in a list, give the elements in the list a more obvious name, and then calculate AIC:

#define list of models

models <- list(mod1, mod2, mod3, mod4)

#specify model names

mod.names <- c("intercept.only", "exposure", "temperature", "exposure.temperature")

#calculate AIC of each model

#a note that second.ord = T will calculate AICc, which comes next

aictab(cand.set = models, modnames = mod.names, second.ord = F)

**Please paste the model selection table below:**

Model selection based on AIC:

K AIC Delta\_AIC AICWt Cum.Wt LL

exposure.temperature 4 733.02 0.00 0.82 0.82 -362.51

temperature 3 736.27 3.25 0.16 0.99 -365.13

exposure 3 741.83 8.81 0.01 1.00 -367.92

intercept.only 2 744.72 11.70 0.00 1.00 -370.36

1. A rule of thumb is that any models within 2 of the top model have greatest support. **What model(s) have greatest support in this exercise?**

Exposure.temperature (i.e., mod 4)

1. Let’s rank the models using AICc:

aictab(cand.set = models, modnames = mod.names, second.ord = T)

**Please paste the model selection table below:**

Model selection based on AICc:

K AICc Delta\_AICc AICcWt Cum.Wt LL

exposure.temperature 4 733.37 0.00 0.81 0.81 -362.51

temperature 3 736.48 3.10 0.17 0.99 -365.13

exposure 3 742.04 8.67 0.01 1.00 -367.92

intercept.only 2 744.82 11.45 0.00 1.00 -370.36

1. **Using a rule of thumb of 2 having greatest support, what model(s) have greatest support in this exercise?**

Exposure.temperature

1. Let’s rank the models using BIC now using

bictab(cand.set = models, modnames = mod.names)

**Please paste the model selection table below:**

Model selection based on BIC:

K BIC Delta\_BIC BICWt Cum.Wt LL

exposure.temperature 4 744.17 0.00 0.53 0.53 -362.51

temperature 3 744.63 0.46 0.42 0.95 -365.13

exposure 3 750.19 6.02 0.03 0.98 -367.92

intercept.only 2 750.30 6.12 0.02 1.00 -370.36

1. **Using a rule of thumb of 2 having greatest support, what model(s) have greatest support in this exercise?**

Exposure.temperature (mod 4)

Temperature (mod 3)

1. **Why did AIC, AICc and BIC suggest different top model sets?**

AICc controls for small sample size

BIC favors simpler models

1. **Can we use AIC et al. to say that our top model is actually a good model? Why or why not?**

**No because it is only comparing the other models in the dataset**

1. Let’s pivot to our second dataset, “bats.csv” Please read in the dataset using the same code as before, but call it “bats.” Also, use the line

header(bats)

To see what the top six rows of data look like.

Let’s also install and load the package lme4:

install.packages("AICcmodavg")

library(AICcmodavg)

Let’s build a linear mixed model for these data relating bat temperature to ambient temperature. Because we already have an idea of how our data might be clustered, we are going to include random effects for day, colony, and individual nested within colony:

#model with three random intercepts

mod1 <- lmer(bat.temp ~ ambient + (1|day) + (1|colony) + (1|colony:ind.col), data=bats, REML=TRUE)

**Please paste the summary table after running this line:**

summary(mod1)

Linear mixed model fit by REML ['lmerMod']

Formula:

bat.temp ~ ambient + (1 | day) + (1 | colony) + (1 | colony:ind.col)

Data: bats

REML criterion at convergence: -58.9

Scaled residuals:

Min 1Q Median 3Q Max

-2.2181 -0.6875 -0.0672 0.6107 2.7999

Random effects:

Groups Name Variance Std.Dev.

colony:ind.col (Intercept) 0.005970 0.07727

day (Intercept) 0.009069 0.09523

colony (Intercept) 0.054734 0.23395

Residual 0.037026 0.19242

Number of obs: 312, groups: colony:ind.col, 42; day, 10; colony, 6

Fixed effects:

Estimate Std. Error t value

(Intercept) 2.044824 0.102195 20.01

ambient 1.205375 0.003015 399.83

Correlation of Fixed Effects:

(Intr)

ambient -0.118

1. **Looking at the value of the beta parameter for the fixed effect (ambient), what can we say about the effect of ambient temperature on bat temperature?**

For every 1 unit increase in ambient temperature, bat temperature increases 1.2 C

1. Let’s look at the random effects now. Basically, the higher the variance explained by the random effect (i.e., the values under “Variance” in the output table), the better the random effect covers unexplained variance (i.e., variation in the predictor after account for our fixed effect). Knowing that, **what random effect is the “best” in explaining the variance that is “left over” after using our fixed effect of ambient temperature?**

**Colony**

1. We now want to know whether we even want ambient temperature as a fixed effect. Let’s run models with and without ambient temperature.

Before we do that, note that we are now checking for fixed effect structure. **Do we want to use REML = T or REML = F, and why?**

If we do not have a fixed effect, REML would be False because there is no fixed

effect to estimate

1. OK. Now run both models and compare using AICc using these lines:

#model with ambient temp as a fixed effect

mod1 <- lmer(bat.temp ~ ambient + (1|day) + (1|colony) + (1|colony:ind.col), data=bats, REML=FALSE)

#model without ambient temp as a fixed effect

#ignore the warning here; it is telling us we may be building a model that is too complex, as the random effects are really small

mod2 <- lmer(bat.temp ~ 1 + (1|day) + (1|colony) + (1|colony:ind.col), data=bats, REML=FALSE)

#create list of models

models <- list(mod1, mod2)

#specify model names

mod.names <- c("with.ambient", "without.ambient")

#create AICc table

aictab(cand.set = models, modnames = mod.names, second.ord=T)

**Paste the AICc table here.**

Model selection based on AICc:

K AICc Delta\_AICc AICcWt Cum.Wt LL

with.ambient 6 -59.23 0.00 1 1 35.75

without.ambient 5 1847.36 1906.59 0 1 -918.58

1. **Do we have convincing support that ambient should remain a fixed effect in the model? What are the lines of evidence?**

Yes it should remain a fixed effect because delta AIC of the second performing model is not within 2.