

Test Exercise 4 (R)

1. Read the data in **Acardamines_pop.txt**. The file contains data on 179 *Anthocharis cardamines* pupae. The pupae are from four populations (A = Ångermanland (Swe), B = Bournemouth (UK), D = Durham (UK), and S = Skåne (Swe)) and have experienced different winter durations (30, 60, 90 or 120 days). We suspect that the pupae require a certain winter length in order to break their winter resting stage (diapause) and your job is to now investigate whether this is true by analyzing if the winter duration affects the probability of a pupae hatching in the spring. You should also investigate if there is any difference in requires winter duration between the populations. For the sake of the exercise you should treat winter duration as a continuous variable.

You can get a hatch-indicating variable as follows:

```
dat$hatch <- as.numeric(dat$hatch_YN) - 1
```

This variable will be convenient for a logistic regression.

(a) Perform a logistic regression with population and winter duration in an interaction. Do the populations react differently in their probability of hatching to winter duration? If the interaction is not significant simplify the model. Finally, what is your final model and what are your conclusions? Is there a significant effect of winter duration on probability of hatching? Do populations differ in their requirement of cold in order to break diapause? Which population requires the most cold?

(b) Next, what is the fitted equation for the probability of hatching as a function of winter duration for a pupa from Bournemouth? (*Hint: see Lecture 11*)

(c) Finally, use the formula to estimate at which winter duration a pupa from Skåne is equally likely to hatch or not hatch. You can use the graph to verify your result.

2. Read the data in **Mcinxia3.txt**. You analysed these data using a mixed model in Exercise 5. Now you will perform the same kind of analysis but with a Bayesian approach. The data file contains *M. cinxia* female pupal weights. The females derive from two sites on Öland: Hildeborg and Littorinavallen. The individuals were reared in the lab from larvae collected in the wild. *M. cinxia* larvae are group-living (a larval group consists of one or several sibling groups) and those collected from a group in the wild were raised together in the lab. The variable `Lgroup` identifies the larval group (you need to make it into a factor). Your job is to determine if females from the two sites differ in pupal weight. **(a)** First, repeat one of the analyses you performed before, by loading the package `lme4` and fitting and summarizing the output from a mixed model.

Do you detect a statistically significant pupal weight difference between the two localities? Load the `MCMCglmm` package and **(b)** make a Bayesian analysis.

What is the Bayesian estimate and the confidence interval for the effect of site and what do you conclude from this interval. **(c)** Finally, give the Bayesian estimates and 95% confidence intervals of the variance components in the model. There are two variance components: the variance in pupal weight among larval groups within sites and the variance in pupal weight within larval groups.

3. Read the data in **Deer2.txt**. The data are from an experiment on fallow deer foraging, where 16 deer individually performed feeding trials. In a trial, a deer was exposed to two hazel bundles, each containing 8 hazel branches. The bundles were tied to two stakes that were placed 6 m apart. Each hazel branch had been dipped into either a high tannin solution or a low tannin solution (tannin is bitter tasting and one might expect the deer to prefer eating low tannin hazel). In one of the bundles (the “bad patch”), 7 of the branches were high tannin and one was low tannin. In the other bundle (the “good patch”), one branch was high tannin and 7 branches were low tannin. The data file contains the amount eaten per branch (g) for the bad and good patches and the high and low tannin branches in these patches. There are thus 4 amounts for each deer (this is then a so-called repeated measures design). Your job is to perform a Bayesian mixed-model analysis of the data, and to test if patch quality and/or tannin level had an effect of the consumption per branch.

First, use a weak prior and the default MCMC sampling parameters, like this:

```
pri1 <- list(R=list(V=1, nu=0.002), G = list(G1=list(V=1,
  nu=0.002), G2=list(V=1, nu=0.002), G3=list(V=1, nu=0.002)))

fm1 <- MCMCglmm(Eaten ~ PatchQ + Tannin,
  random = ~ Deer + Deer:PatchQ + Deer:Tannin, data=dat,
  prior=pri1, nitt=13000, thin=10, burnin=3000, verbose=FALSE)
```

Three random effects have been taken into account (in addition to the residual), and they correspond to differences between the deer in how much they tend to eat and how their appetite is influenced by patch quality and by the tannin level of a branch.

(a) From examining the output of `summary(fm1)`, and also from examining the autocorrelation plots, `autocorr.plot(mod1$VCV)` and `autocorr.plot(mod1$Sol)`, do you think this analysis is OK? Give a reason for your opinion about the analysis.

(b) Second, try using a stronger prior, for instance

```
pri2 <- list(R=list(V=1, nu=1), G = list(G1=list(V=1,
  nu=1), G2=list(V=1, nu=1), G3=list(V=1, nu=1)))
```

and/or higher spacing between MCMC iterations, for instance

```
fm2 <- MCMCglmm(Eaten ~ PatchQ + Tannin,
  random = ~ Deer + Deer:PatchQ + Deer:Tannin, data=dat,
  prior=pri2, nitt=100000, thin=100, burnin=10000,
  verbose=FALSE)
```

When you think you have an acceptable MCMC analysis, give 95% confidence intervals for the (fixed) effects of patch quality and tannin level and interpret them biologically. Are these effects statistically significant?

(c) Motivate briefly why you in the end decided on the model (priors, thinning etc.) you used.

We wish you good luck with statistics in your career!