

STATISTICS FOR EES — EXERCISE SHEET 10

1. Ten bees from the same nest were marked with “A”, “B”, ..., “J” and their speed in km/h of their flights to a food source was measured. Measurements were repeated with different sugar concentrations in the food source, ranging from 0.2 to 0.8 mol/l. The data were analysed as follows:

```
> str(data.frame(speed, bee, sugar))
'data.frame': 200 obs. of 3 variables:
 $ speed: num 23.9 22.9 23.4 23.4 20.8 ...
 $ bee : Factor w/ 10 levels "A","B","C","D",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ sugar: num 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0.2 0.4 ...
> library(lme4)
> mod <- lmer(speed ~ sugar + (1|bee))
> summary(mod)
Linear mixed model fit by REML ['lmerMod']
Formula: speed ~ sugar + (1 | bee)
```

REML criterion at convergence: 784.5

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.36349	-0.66968	-0.08663	0.55515	3.06487

Random effects:

Groups	Name	Variance	Std.Dev.
bee	(Intercept)	1.029	1.015
	Residual	2.685	1.639

Number of obs: 200, groups: bee, 10

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	22.0211	0.4283	51.410
sugar	2.4479	0.5182	4.724

Correlation of Fixed Effects:

```
(Intr)
sugar -0.605
> ranef(mod)
$bee
(Intercept)
A -0.84175401
B -0.09292865
C 1.83066348
D -0.04737105
E 0.08399406
F 1.41047205
G 0.07426331
H -1.11244575
I -0.78071567
J -0.52417775
```

with conditional variances for \bee"

- (a) Ten more flights of bee "C" and a sugar concentration of 0.5 ml/l are to be observed. What does the model predict about the distribution of the velocities in km/h that will be measured in this new series of observations? Specify distribution family, mean value and standard deviation.
 - (b) Ten more flights with a sugar concentration of 0.5 ml/l are to be observed, this time each with a different random bee from the same nest. What does the model predict about the distribution of the velocities in km/h that will be measured in this new series of observations? Give all information that is needed to fully characterise the predicted distribution. (Neglect standard errors and correlations of fixed-effect values.)
 - (c) Propose how to extend the model specification in the `lmer` command to account for the possibility that the bees do not only differ in their random velocity but also in how the flight speeds depends on the sugar concentration of the food source. (Model this also as a random effect.)
2. Imagine that heart beat rates and metabolic rates have been measured repeatedly for 26 birds of prey of three different species (harpy eagle, osprey, golden eagle). Load the simulated data in R with the following command:

```
eagles <- read.csv("http://evol.bio.lmu.de/_statgen/StateES/eagles.csv")
```

Assume that the columns show the bird identity, the species, the weight of the bird in kg, the heart beat per minute (`hbpm`) and the metabolic rate (`mbr`) in $J/(g \cdot h)$, and that we want to analyse how the current metabolic rate in a bird can be predicted from the other variables. (Be aware that this data is simulated and the values and their correlations might be quite unrealistic.)

- (a) First assume that there is no bird effect. That is, analyse the data as if all measurements came from different birds of the respective species. Apply the R commands `lm` and `drop1` to analyse whether species and heart beat and additionally also the weight of the bird have a significant effect on the metabolic rate. Also check whether the effect of heart beat on the metabolic rate differs between the species.
 - (b) Account for possible differences in metabolic rate between the birds by adding the bird ID to the above analyses. Give mathematical specifications of the resulting models. What happens with the p-values of `weight` and `species` and why?
 - (c) Now analyse the data with the `lmer` command from the `lme4` package. Model the bird effect as a random factor with an additive intercept effect. Repeat the above tests with this approach (e.g. with test from the `pbkrtest` package). To what extent do you obtain different results than before and why?
 - (d) Use `lmer` to fit a model with bird as a random factor (with additive effect on the intercept) and species and heart beat as fixed effects.
 - i. What value of the metabolic rate does this model predict for bird g3 when its heart beat is 45 beats per minute?
 - ii. What standard deviation does this model predict for metabolic rate measurements for bird g3 when its heart beat is 45 beats per minute?
 - iii. What value of the metabolic rate does this model predict for an average golden eagle when its heart beat is 45 beats per minute?
 - iv. What standard deviation does this model predict for a series of metabolic rate measurements with different golden eagles taken when the heart beat rate of the birds was 45 bpm?
 - v. Give a full mathematical specification of this model.
 - vi. Visualize the model predictions and check the model assumptions graphically.
3. The daily food consumption in gram of 23 male and 26 female randomly sampled pet cats in Martinsried has been surveyed over 20 summer days, and the temperature T at 10 a.m. has been measured each day (in Celsius). The data has been analysed as follows:

```

> str(data.frame(food, catID, sex, T))
'data.frame': 980 obs. of 4 variables:
 $ food : num 165 168 204 178 170 ...
 $ catID: chr "M 1" "M 2" "M 3" "M 4" ...
 $ sex : chr "M" "M" "M" "M" ...
 $ T : num 21 21 21 21 21 21 21 21 21 21 ...
> library(lme4)
> mod <- lmer(food ~ T + sex + (1 + T | catID))
> summary(mod)
Linear mixed model fit by REML ['lmerMod']
Formula: food ~ T + sex + (1 + T | catID)

REML criterion at convergence: 4748

Scaled residuals:
    Min       1Q   Median       3Q      Max
-3.00131 -0.65462  0.02099  0.61029  2.80229

Random effects:
 Groups   Name      Variance Std.Dev. Corr
catID    (Intercept) 2187.800 46.774
         T           2.453  1.566  -0.95
Residual             4.045  2.011
Number of obs: 980, groups: catID, 49

Fixed effects:
              Estimate Std. Error t value
(Intercept) 233.2080     6.9694  33.462
T            -4.2431     0.2242 -18.922
sexM         33.3478     4.1697   7.998

Correlation of Fixed Effects:
      (Intr) T
T    -0.912
sexM -0.281  0.000
>

```

- (a) Give a full specification of the model (assumptions and parameter values) that has been fitted to the data. (Use standard maths formulas instead of R code.)
 - (b) Ten more random male cats from the Martinsried populations are fed on a day on which it was 20 degrees at 10 a.m. What does the model predict about the distribution of food intake among these cat on that day? (Neglect standard errors and correlations of fixed-effect values.)
4. Zygotes of lab flies of a certain species are treated with a mutagenic substance. The higher the concentration c of the substance (measured in mg/l), the larger the expectation value μ for the number X of mutations in a fly, and for c in the range from 0 to 50 this relationship is modeled by the formula

$$\log(\mu) = 0.8 + 0.3 \cdot \log(c + 1)$$

and the assumption that X is Poisson distributed with expectation value μ . (\log is the natural logarithm.)

- (a) For $c = 5$ calculate how the probabilities that the number of mutations in the fly is 0, 1 or 2.

- (b) Visualise how the expectation value and the variance of X depend on c .
- (c) Visualise how the probability that the fly has less than 3 mutations depends on c