Multivariate Statistics in Ecology and Quantitative Genetics
Mixed-effects models

Dirk Metzler & Noémie Becker

http://evol.bio.lmu.de/_statgen

6. Juli 2014

Contents

Contents

We revisit the RIKZ dataset.

We revisit the RIKZ dataset.

Species abundance and many other variables were measured at 9 beaches.

We revisit the RIKZ dataset.

Species abundance and many other variables were measured at 9 beaches.

On every beach, 5 plots were sampled in the intertidal range.

We revisit the RIKZ dataset.

Species abundance and many other variables were measured at 9 beaches.

On every beach, 5 plots were sampled in the intertidal range.

Each plot was sampled only once. Thus, each line in the data table corresponds to one plot.

The random effect of the beach

```
> rikz <- read.csv("RIKZ_Tab4.csv")
> rikz$Beach <- factor(rikz$Beach)
> rikz$Week <- factor(rikz$Week)
> str(rikz)
'data.frame': 45 obs. of  15 variables:
 $ ShannonW      : num  0.76 0.72 0.85 0.53 0.74 0.13 0.4 0.29 1.02
 $ Richness      : int  11 10 13 11 10 8 9 8 19 17 ...
 $ angle1        : int  32 62 65 55 23 129 126 52 26 143 ...
 $ angle2        : int  96 96 96 96 96 89 89 89 89 89 ...
 $ exposure      : int  10 10 10 10 10 8 8 8 8 8 ...
 $ salinity      : num  29.4 29.4 29.4 29.4 29.4 29.6 29.6 29.6 29.6
 $ temperature   : num  17.5 17.5 17.5 17.5 17.5 20.8 20.8 20.8 20.8
 $ NAP           : num  0.045 -1.036 -1.336 0.616 -0.684 ...
 $ penetrability: num  254 227 237 249 252 ...
 $ grainsize     : num  222 200 194 221 202 ...
 $ humus         : num  0.05 0.3 0.1 0.15 0.05 0.1 0.1 0.1 0.15 0 ..
 $ chalk         : num  2.05 2.5 3.45 1.6 2.45 2.5 1.85 1.7 2.3 2.6
 $ sorting1      : num  69.8 59 59.2 67.8 57.8 ...
 $ Beach         : Factor w/ 9 levels "1","2","3","4",..: 1 1 1 1 1
 $ Week          : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 1 1 1
```

Today we use the following columns:

ShannonW	A biodiversity index based on the Shannon entropy.
Richness	The number of species.
Beach	Which beach.
NAP	The altitude of the plot compared to the average sea level.
exposure	An index composed of wave action, length of the surf zone, slope, grain size, depth of the anaerobic layer.
salinity	Salt contents.
humus	Fraction of organic material.

Today we use the following columns:

ShannonW	A biodiversity index based on the Shannon entropy.
Richness	The number of species.
Beach	Which beach.
NAP	The altitude of the plot compared to the average sea level.
exposure	An index composed of wave action, length of the surf zone, slope, grain size, depth of the anaerobic layer.
salinity	Salt contents.
humus	Fraction of organic material.

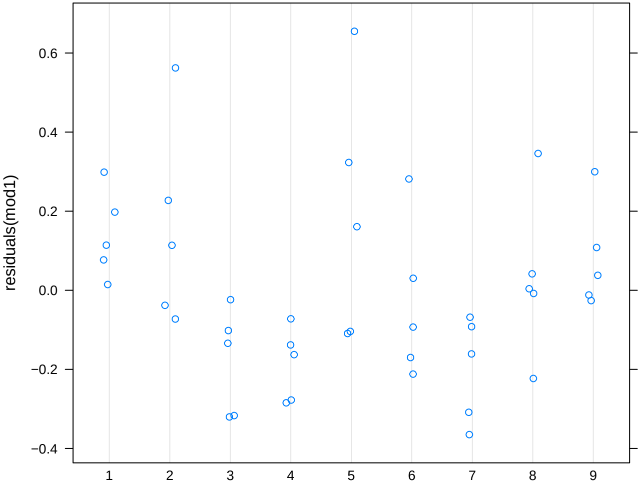
In an earlier analysis, NAP turned out to have an impact on species richness. Does ShannonW also depend on NAP?

```
> mod1 <- lm(ShannonW~NAP,data=rikz)
> summary(mod1)
[..]
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.47141    0.03627  12.996  < 2e-16 ***
NAP         -0.22583    0.03478  -6.493 7.01e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.2294 on 43 degrees of freedom
Multiple R-squared: 0.4951, Adjusted R-squared: 0.4833
F-statistic: 42.16 on 1 and 43 DF,  p-value: 7.01e-08
```

```
> mod1 <- lm(ShannonW~NAP,data=rikz)
> summary(mod1)
[..]
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.47141    0.03627  12.996  < 2e-16 ***
NAP         -0.22583    0.03478  -6.493 7.01e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.2294 on 43 degrees of freedom
Multiple R-squared: 0.4951, Adjusted R-squared: 0.4833
F-statistic: 42.16 on 1 and 43 DF,  p-value: 7.01e-08
```

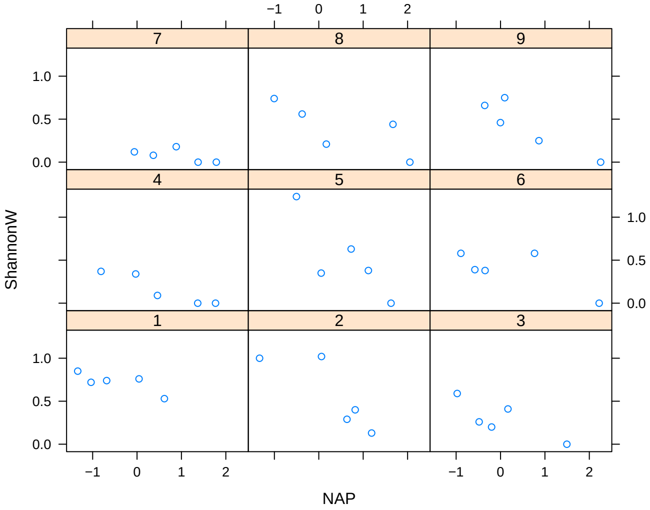
Can we really pool the data from the different beaches?

```
> library(lattice)
> dotplot(residuals(mod1)~rikz$Beach,jitter.x=TRUE,pch=1)
```



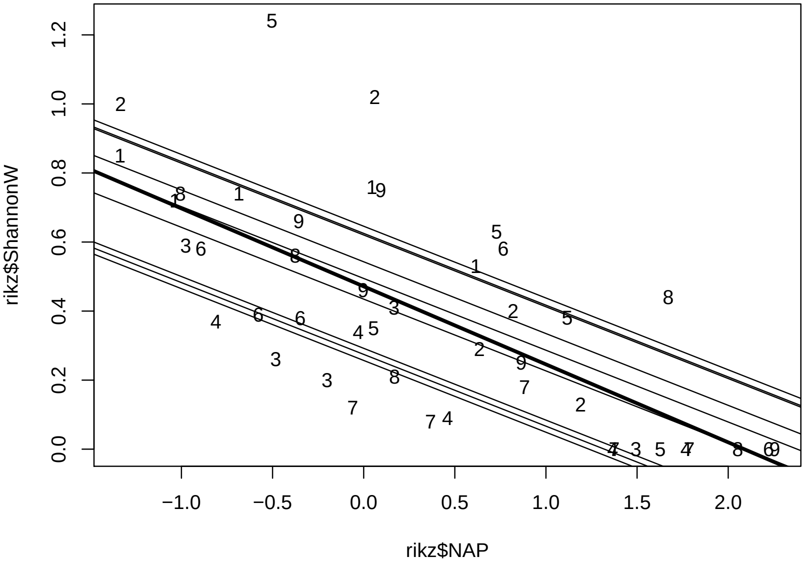
The random effect of the beach

```
> xyplot(ShannonW~NAP|Beach,data=rikz)
```



```
> mod2 <- lm(ShannonW~NAP+Beach,data=rikz)
> summary(mod2)
[..]
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.620395   0.086771   7.150 2.45e-08 ***
NAP         -0.207945   0.031366  -6.630 1.15e-07 ***
Beach2       0.004665   0.123148   0.038  0.96999
Beach3      -0.328311   0.121791  -2.696  0.01072 *
Beach4      -0.345859   0.125102  -2.765  0.00903 **
Beach5       0.025370   0.125550   0.202  0.84103
Beach6      -0.185777   0.122910  -1.511  0.13964
Beach7      -0.363441   0.128055  -2.838  0.00750 **
Beach8      -0.125882   0.124720  -1.009  0.31975
Beach9      -0.077617   0.125268  -0.620  0.53953
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

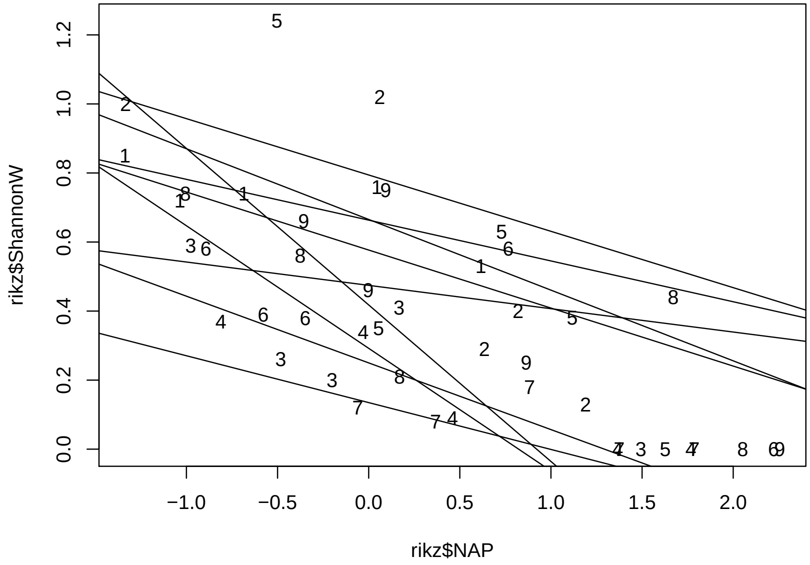
The random effect of the beach



The random effect of the beach

```
> mod3 <- lm(ShannonW~NAP*Beach,data=rikz)
> summary(mod3)
[..]
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.663388   0.097592   6.798 2.67e-07 ***
NAP         -0.118188   0.113333  -1.043  0.30627
Beach2       0.002042   0.129370   0.016  0.98752
Beach3      -0.371311   0.126891  -2.926  0.00688 **
Beach4      -0.413504   0.135632  -3.049  0.00510 **
Beach5       0.130807   0.142539   0.918  0.36691
Beach6      -0.245660   0.127976  -1.920  0.06554 .
Beach7      -0.528549   0.165405  -3.195  0.00354 **
Beach8      -0.188925   0.131537  -1.436  0.16241
Beach9      -0.086618   0.136291  -0.636  0.53043
NAP:Beach2  -0.236876   0.145877  -1.624  0.11604
NAP:Beach3  -0.075049   0.149266  -0.503  0.61919
NAP:Beach4  -0.044999   0.142857  -0.315  0.75519
NAP:Beach5  -0.335176   0.156110  -2.147  0.04093 *
NAP:Beach6  -0.017519   0.133795  -0.131  0.89679
NAP:Beach7   0.050573   0.166465   0.304  0.76361
NAP:Beach8  -0.049864   0.132654  -0.376  0.70993
```

The random effect of the beach



```
> anova(mod1,mod2,mod3)
[..]
Model 1: ShannonW ~ NAP
Model 2: ShannonW ~ NAP + Beach
Model 3: ShannonW ~ NAP * Beach
  Res.Df     RSS Df Sum of Sq      F   Pr(>F)
1     43 2.26293
2     35 1.27812  8   0.98481 3.7433 0.00458 **
3     27 0.88792  8   0.39020 1.4831 0.20951
[..]
> AIC(mod1,mod2,mod3)
     df         AIC
mod1  3  -0.8457103
mod2 11 -10.5528837
mod3 19 -10.9445763
```

- Model 2 (where beaches have an additive effect) is significantly better than model 1, and model 3 is not significantly better than model 2.

- Model 2 (where beaches have an additive effect) is significantly better than model 1, and model 3 is not significantly better than model 2.
- AIC favours (takes the lowest value for) model 3.

- Model 2 (where beaches have an additive effect) is significantly better than model 1, and model 3 is not significantly better than model 2.
- AIC favours (takes the lowest value for) model 3.
- How can AIC be negative?

- Model 2 (where beaches have an additive effect) is significantly better than model 1, and model 3 is not significantly better than model 2.
- AIC favours (takes the lowest value for) model 3.
- How can AIC be negative? It is still defined as $-2 \cdot \log(L) + 2 \cdot d$, whereas d is the number of parameters and L is the likelihood of the ML parameter estimate in the model. Since the response variable ShannonW has a continuos distribution (assumed to be normal), the likelihood is the probability *density* of the data given the parameter values. Densities can be larger than 1. If $L > 1$, then $\log(L) > 0$ and AIC can be negative.

- We are not interested in the precise effect of each beach

- We are not interested in the precise effect of each beach
- We do not want to estimate 8 extra paramters for the beaches

- We are not interested in the precise effect of each beach
- We do not want to estimate 8 extra paramters for the beaches
- Is there another way to take the difference between the beaches into account?

- We are not interested in the precise effect of each beach
- We do not want to estimate 8 extra paramters for the beaches
- Is there another way to take the difference between the beaches into account?
- Assume that the effect α_k of beach k is random. Do not estimate all α_k but only their standard deviation σ_α.

Mixed-effects model

Let S_i and N_i be the ShannonW and the NAP observed at plot i, which is on beach k.

$$S_i = a + b \cdot N_i + \alpha_k + \varepsilon_i$$

Mixed-effects model

Let S_i and N_i be the ShannonW and the NAP observed at plot i, which is on beach k.

$$S_i = a + b \cdot N_i + \alpha_k + \varepsilon_i$$

$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{45}$ are independently $\mathcal{N}(0, \sigma^2)$-distributed.

Mixed-effects model

Let S_i and N_i be the ShannonW and the NAP observed at plot i, which is on beach k.

$$S_i = a + b \cdot N_i + \alpha_k + \varepsilon_i$$

$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{45}$ are independently $\mathcal{N}(0, \sigma^2)$-distributed.
$\alpha_1, \alpha_2, \ldots, \alpha_9$ are independently $\mathcal{N}(0, \sigma_\alpha^2)$-distributed.

Mixed-effects model

Let S_i and N_i be the ShannonW and the NAP observed at plot i, which is on beach k.

$$S_i = a + b \cdot N_i + \alpha_k + \varepsilon_i$$

$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{45}$ are independently $\mathcal{N}(0, \sigma^2)$-distributed.
$\alpha_1, \alpha_2, \dots, \alpha_9$ are independently $\mathcal{N}(0, \sigma_\alpha^2)$-distributed.
Mixed-effects: *a* and *b* are *deterministic*, $\alpha_1, \alpha_2, \dots, \alpha_9$ are *random*.

Mixed-effects model

Let S_i and N_i be the ShannonW and the NAP observed at plot i, which is on beach k.

$$S_i = a + b \cdot N_i + \alpha_k + \varepsilon_i$$

$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{45}$ are independently $\mathcal{N}(0, \sigma^2)$-distributed.
$\alpha_1, \alpha_2, \ldots, \alpha_9$ are independently $\mathcal{N}(0, \sigma_\alpha^2)$-distributed.
Mixed-effects: *a* and *b* are *deterministic*, $\alpha_1, \alpha_2, \ldots, \alpha_9$ are *random*.

To be estimated: *a*,*b*, σ_α, σ.

```
> library(lme4)
> mmod0 <- lmer(ShannonW~1+NAP+(1|Beach),
+                               data=rikz,verbose=TRUE)
  0:     -3.0154190: 0.730297
  1:     -3.0321775: 0.694973
  2:     -3.0321844: 0.694215
  3:     -3.0321844: 0.694257
  4:     -3.0321844: 0.694257
```

```
> library(lme4)
> mmod0 <- lmer(ShannonW~1+NAP+(1|Beach),
+                           data=rikz,verbose=TRUE)
  0:     -3.0154190: 0.730297
  1:     -3.0321775: 0.694973
  2:     -3.0321844: 0.694215
  3:     -3.0321844: 0.694257
  4:     -3.0321844: 0.694257
```

It may be necessary to first install the lme4 package:

```
> install.packages("lme4")
```

This will only work if the R version installed is not too old.

The random effect of the beach

```
> summary(mmod0)
Linear mixed model fit by REML
Formula: ShannonW ~ 1 + NAP + (1 | Beach)
   Data: rikz
   AIC   BIC logLik deviance REMLdev
 4.968 12.19  1.516   -12.27  -3.032
Random effects:
 Groups   Name        Variance Std.Dev.
 Beach    (Intercept) 0.017595 0.13264
 Residual             0.036504 0.19106
Number of obs: 45, groups: Beach, 9

Fixed effects:
            Estimate Std. Error t value
(Intercept)  0.46722    0.05366   8.707
NAP         -0.21380    0.03060  -6.987

Correlation of Fixed Effects:
    (Intr)
NAP -0.198
```

The random effect of the beach

```
> summary(mmod0)
Linear mixed model fit by REML
Formula: ShannonW ~ 1 + NAP + (1 | Beach)
   Data: rikz
   AIC   BIC logLik deviance REMLdev
 4.968 12.19  1.516   -12.27  -3.032
Random effects:
 Groups   Name        Variance Std.Dev.
 Beach    (Intercept) 0.017595 0.13264
 Residual             0.036504 0.19106
Number of obs: 45, groups: Beach, 9

Fixed effects:
            Estimate Std. Error t value
(Intercept)  0.46722    0.05366   8.707
NAP         -0.21380    0.03060  -6.987

Correlation of Fixed Effects:
    (Intr)
NAP -0.198
```

What is REML?

The random effect of the beach

```
> summary(mmod0)
Linear mixed model fit by REML
Formula: ShannonW ~ 1 + NAP + (1 | Beach)
   Data: rikz
   AIC   BIC logLik deviance REMLdev
 4.968 12.19  1.516   -12.27  -3.032
Random effects:
 Groups   Name        Variance Std.Dev.
 Beach    (Intercept) 0.017595 0.13264
 Residual             0.036504 0.19106
Number of obs: 45, groups: Beach, 9

Fixed effects:
            Estimate Std. Error t value
(Intercept)  0.46722    0.05366   8.707
NAP         -0.21380    0.03060  -6.987

Correlation of Fixed Effects:
    (Intr)
NAP -0.198
```

What is REML?

Why are there *t*-values but no *p*-values?

REML vs. ML

- ML (Maximum Likelihood): estimate all parameters (here *a*, *b*, σ_α, σ) by maximizing their joint likelihood.

REML vs. ML

- ML (Maximum Likelihood): estimate all parameters (here *a*, *b*, σ_α, σ) by maximizing their joint likelihood.
- REML (Restricted Maximum Likelihood): first estimate variance parameters (here σ_α, σ) from the components of the response space that are orthogonal on components that can be explained by fixed effects. Using these estimates, the coefficients of the fixed effects (here *a* and *b*) are estimated with ML.

REML vs. ML

- ML (Maximum Likelihood): estimate all parameters (here *a*, *b*, σ_α, σ) by maximizing their joint likelihood.
- REML (Restricted Maximum Likelihood): first estimate variance parameters (here σ_α, σ) from the components of the response space that are orthogonal on components that can be explained by fixed effects. Using these estimates, the coefficients of the fixed effects (here *a* and *b*) are estimated with ML.
- Comparable to estimation of σ^2 from sample $X_1, \dots, X_n$ by $\frac{1}{n-1}\sum_i(\mu_X - X_i)^2$ instead of the biased ML estimator $\frac{1}{n}\sum_i(\mu_X - X_i)^2$

REML vs. ML

- ML (Maximum Likelihood): estimate all parameters (here a, b, σ_α, σ) by maximizing their joint likelihood.
- REML (Restricted Maximum Likelihood): first estimate variance parameters (here σ_α, σ) from the components of the response space that are orthogonal on components that can be explained by fixed effects. Using these estimates, the coefficients of the fixed effects (here a and b) are estimated with ML.
- Comparable to estimation of σ^2 from sample $X_1, \ldots, X_n$ by $\frac{1}{n-1}\sum_i(\mu_X - X_i)^2$ instead of the biased ML estimator $\frac{1}{n}\sum_i(\mu_X - X_i)^2$
- Also for fitting parameters of mixed-effects models, ML estimation is biased and REML is usually preferred.

REML vs. ML

- ML (Maximum Likelihood): estimate all parameters (here a, b, σ_α, σ) by maximizing their joint likelihood.
- REML (Restricted Maximum Likelihood): first estimate variance parameters (here σ_α, σ) from the components of the response space that are orthogonal on components that can be explained by fixed effects. Using these estimates, the coefficients of the fixed effects (here a and b) are estimated with ML.
- Comparable to estimation of σ^2 from sample $X_1, \ldots, X_n$ by $\frac{1}{n-1}\sum_i(\mu_X - X_i)^2$ instead of the biased ML estimator $\frac{1}{n}\sum_i(\mu_X - X_i)^2$
- Also for fitting parameters of mixed-effects models, ML estimation is biased and REML is usually preferred.
- ML estimation should be used when a likelihood ratio test shall be applied to models with different fixed effects and the same random effects.

Why no *p*-values for the *t*-values?

- The *t*-values computed like in the usual linear model, but in the case of mixed-effects models they are in general not *t*-distributed (under the null hypothesis). Thus, it is not clear how to get *p*-values from the *t*-values.

Why no *p*-values for the *t*-values?

- The *t*-values computed like in the usual linear model, but in the case of mixed-effects models they are in general not *t*-distributed (under the null hypothesis). Thus, it is not clear how to get *p*-values from the *t*-values.
- Some other programs give *p*-values which can be very imprecise.

Why no *p*-values for the *t*-values?

- The *t*-values computed like in the usual linear model, but in the case of mixed-effects models they are in general not *t*-distributed (under the null hypothesis). Thus, it is not clear how to get *p*-values from the *t*-values.
- Some other programs give *p*-values which can be very imprecise.
- Exception: small balanced datasets. Here, *t*-values are approximately *t*-distributed and $|t| > 2$ usually indicates significance on the 5% level.

One possibility to visualize the estimations for the parameter and to assess their significance is based on sampling parameter values from their posterior distribution by an MCMC method.

One possibility to visualize the estimations for the parameter and to assess their significance is based on sampling parameter values from their posterior distribution by an MCMC method.

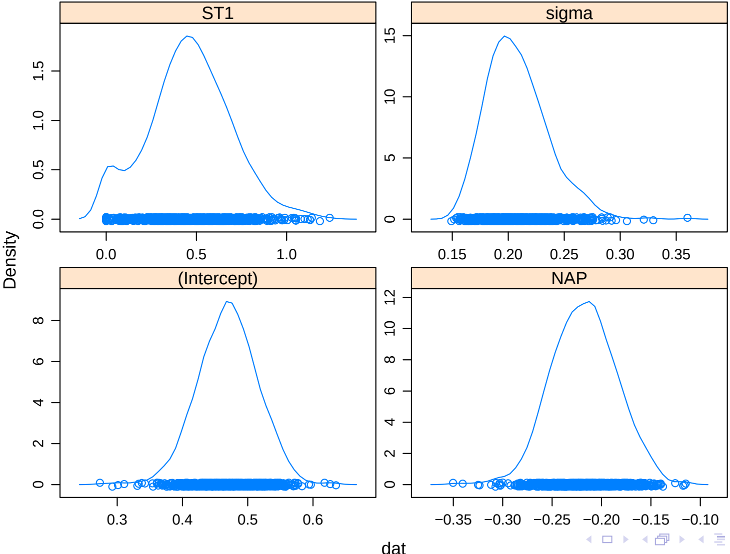
In contrast to most other methods discussed in this lecture, this is a Bayesian approach and thus needs prior distributions for the parameter values (or at least pseudo priors).

One possibility to visualize the estimations for the parameter and to assess their significance is based on sampling parameter values from their posterior distribution by an MCMC method.

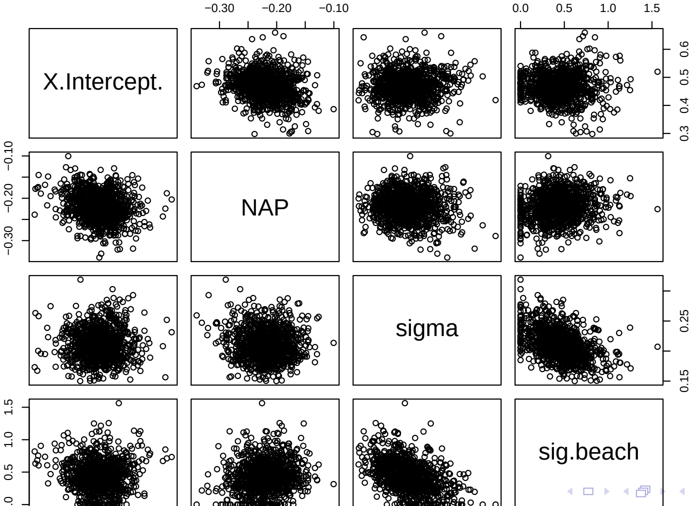
In contrast to most other methods discussed in this lecture, this is a Bayesian approach and thus needs prior distributions for the parameter values (or at least pseudo priors).

MCMC sampling is not supported anymore in the current version of lme4. We will come back to it in the context of the package MCMCglmm, which provides a Bayesian variant of (generalized) linear mixed models.

```
> samp <- mcmcsamp(mmod0,1000)  ## from old lme4 version
> densityplot(samp)
```



```
> sampdf <- data.frame(t(samp@fixef),
+              sigma=t(samp@sigma),sig.beach=t(samp@ST))
> plot(sampdf)
```



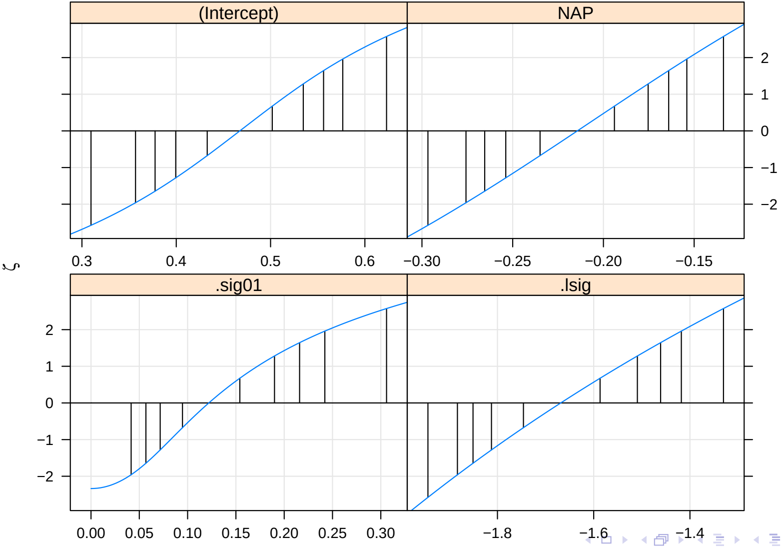
The computation of confidence ranges in lme4 is based on likelihood ratios. The profiles in the first plots are likelihood ratios.

The computation of confidence ranges in lme4 is based on likelihood ratios. The profiles in the first plots are likelihood ratios.

Vertical lines indicate 50%, 80%, 90%, 95%, and 99% confidence intervals.

The random effect of the beach

```
> pr0 <- profile(mmod0)
> xyplot(pr0)
```



Display confidence intervals (per default 95%).

```
> confint(pr0)
                 2.5 %     97.5 %
.sig01       0.0415353  0.2420610
.lsig       -1.8831563 -1.4178656
(Intercept)  0.3568131  0.5765609
NAP         -0.2757402 -0.1540307
```

Display confidence intervals (per default 95%).

```
> confint(pr0)
                 2.5 %     97.5 %
.sig01       0.0415353  0.2420610
.lsig       -1.8831563 -1.4178656
(Intercept)  0.3568131  0.5765609
NAP         -0.2757402 -0.1540307
```

The next plot shows pairwise confidence ranges (again 50%, 80%, 90%, 95%, and 99%).

Display confidence intervals (per default 95%).

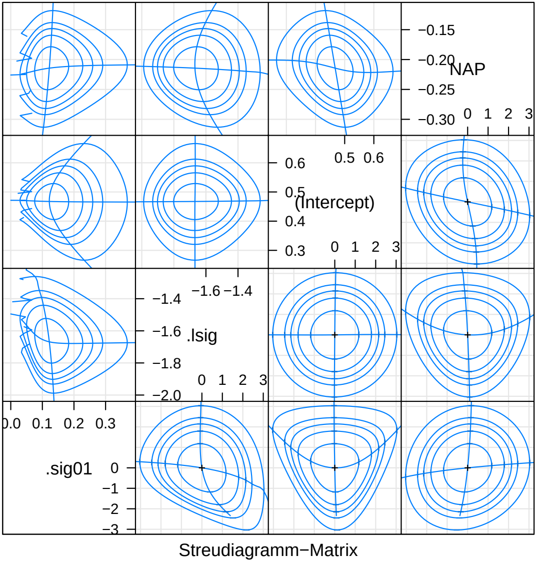
```
> confint(pr0)
                 2.5 %     97.5 %
.sig01       0.0415353  0.2420610
.lsig       -1.8831563 -1.4178656
(Intercept)  0.3568131  0.5765609
NAP         -0.2757402 -0.1540307
```

The next plot shows pairwise confidence ranges (again 50%, 80%, 90%, 95%, and 99%).

Panels above the diagonal are in the original scale of the parameters, panels below the diagonal are on the $\sqrt{\text{likelihoodratio}}$ scale.

The random effect of the beach

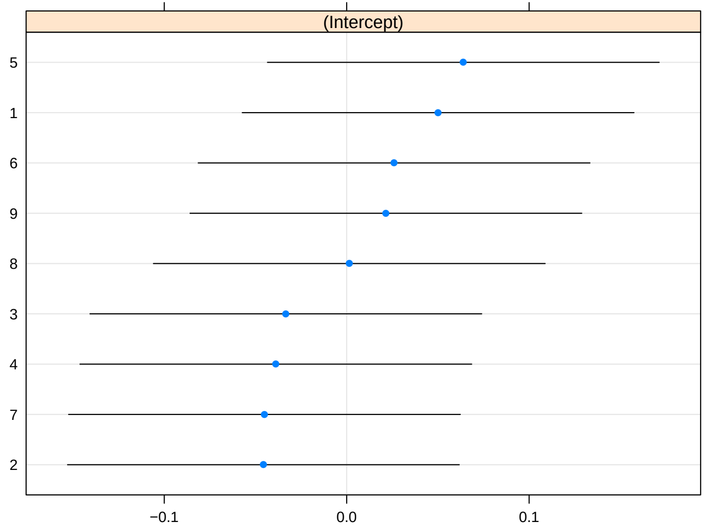
```
> splom(pr0)
```



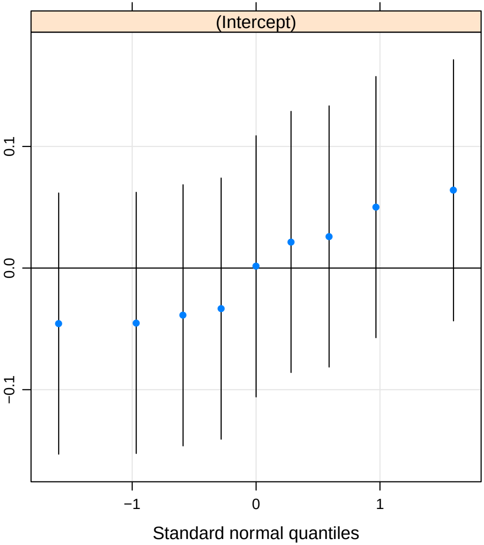
Back to what is possible with older version of lme4.

The next two plots visualize estimates for the random effects of the beaches.

```
> dotplot(ranef(B1,postVar=TRUE))
```



```
> qqmath(ranef(B1,postVar=TRUE))
```



Contents

General considerations

- What is the purpose of the model?
  1. Making predictions as precise as possible
  2. or to understand what the most influential paramters are?

General considerations

- What is the purpose of the model?
  1. Making predictions as precise as possible
  2. or to understand what the most influential paramters are?
- In the first case AIC may be appropriate.

General considerations

- What is the purpose of the model?
    1. Making predictions as precise as possible
    2. or to understand what the most influential paramters are?
- In the first case AIC may be appropriate.
- In the second case it may be better to use likelihood-ratio tests and remove all parameters which do not significantly improve the fit.

General considerations

- What is the purpose of the model?
  1. Making predictions as precise as possible
  2. or to understand what the most influential paramters are?
- In the first case AIC may be appropriate.
- In the second case it may be better to use likelihood-ratio tests and remove all parameters which do not significantly improve the fit.
- Variable selection should not only depend on statistics but also on the relevance of the parameter for the biological question.

We begin with a comparison of two nested models with the same random parameters

We begin with a comparison of two nested models with the same random parameters

```
> B0 <- lmer(ShannonW~1+exposure*NAP+(1|Beach),
+                                    data=rikz,verbose=TRUE)
  0:     -3.1079878: 0.730297
  1:     -3.9300741: 0.431067
  2:     -3.9406377: 0.471682
  3:     -3.9424344: 0.460627
  4:     -3.9424352: 0.460386
  5:     -3.9424352: 0.460388
```

```
> summary(B0)
Linear mixed model fit by REML
Formula: ShannonW ~ 1 + exposure * NAP + (1 | Beach)
   Data: rikz
   AIC   BIC logLik deviance REMLdev
 8.058 18.90  1.971   -23.54  -3.942
Random effects:
 Groups   Name        Variance  Std.Dev.
 Beach    (Intercept) 0.0069857 0.08358
 Residual             0.0329581 0.18154
Number of obs: 45, groups: Beach, 9

Fixed effects:
             Estimate Std. Error t value
(Intercept)   1.92603    0.44765   4.303
exposure     -0.14294    0.04368  -3.272
[..]
```

```
> summary(B0)
[..]
Fixed effects:
             Estimate Std. Error t value
(Intercept)   1.92603    0.44765   4.303
exposure     -0.14294    0.04368  -3.272
NAP          -0.89673    0.33620  -2.667
exposure:NAP  0.06661    0.03270   2.037

Correlation of Fixed Effects:
            (Intr) exposr NAP
exposure    -0.996
NAP         -0.233  0.235
exposur:NAP  0.235 -0.238 -0.996
```

```
> B1 <- update(B0,~.-exposure:NAP)
  0:     -3.5739227: 0.730297
  1:     -4.1112507:  0.00000
  2:     -4.1112507: 3.65618e-06
  3:     -4.1112511: 0.000144407
  4:     -4.1650501: 0.0555603
  5:     -4.7984144: 0.246592
  6:     -5.0104872: 0.374551
  7:     -5.0104967: 0.373651
  8:     -5.0104968: 0.373725
  9:     -5.0104968: 0.373722
```

```
> summary(B1)
Linear mixed model fit by REML
Formula: ShannonW ~ exposure + NAP + (1 | Beach)
   Data: rikz
  AIC   BIC logLik deviance REMLdev
 4.99 14.02  2.505   -19.62   -5.01
Random effects:
 Groups   Name        Variance  Std.Dev.
 Beach    (Intercept) 0.0050958 0.071385
 Residual             0.0364847 0.191010
Number of obs: 45, groups: Beach, 9

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.71143    0.41570   4.117
exposure    -0.12166    0.04053  -3.001
NAP         -0.21555    0.02990  -7.209

Correlation of Fixed Effects:
         (Intr) exposr
exposure -0.996
```

```
> anova(B0,B1)
Data: rikz
Models:
B1: ShannonW ~ exposure + NAP + (1 | Beach)
B0: ShannonW ~ 1 + exposure * NAP + (1 | Beach)
   Df      AIC      BIC  logLik  Chisq Chi Df Pr(>Chisq)
B1  5  -9.6222 -0.58886  9.8111
B0  6 -11.5446 -0.70461 11.7723 3.9224      1    0.04765
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> anova(B0,B1)
Data: rikz
Models:
B1: ShannonW ~ exposure + NAP + (1 | Beach)
B0: ShannonW ~ 1 + exposure * NAP + (1 | Beach)
   Df      AIC      BIC  logLik  Chisq Chi Df Pr(>Chisq)
B1  5  -9.6222 -0.58886  9.8111
B0  6 -11.5446 -0.70461 11.7723 3.9224      1    0.04765
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The *p*-value of 0.04765 may be imprecise because REML was used.

```
> anova(B0,B1)
Data: rikz
Models:
B1: ShannonW ~ exposure + NAP + (1 | Beach)
B0: ShannonW ~ 1 + exposure * NAP + (1 | Beach)
   Df      AIC      BIC  logLik  Chisq Chi Df Pr(>Chisq)
B1  5  -9.6222 -0.58886  9.8111
B0  6 -11.5446 -0.70461 11.7723 3.9224      1    0.04765
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The *p*-value of 0.04765 may be imprecise because REML was used.To get a slighly more reliable *p*-value we fit both models with ML.

```
> B1ML <- update(B1,REML=FALSE)
  0:     -16.875166: 0.730297
  1:     -19.518802:  0.00000
  2:     -19.518802: 3.18807e-06

> B0ML <- update(B0,REML=FALSE)
  0:     -21.820395: 0.730297
  1:     -22.739010:  0.00000
  2:     -22.739010: 7.52231e-06
  3:     -22.739011: 0.000259128
  4:     -22.880262: 0.0868904
  5:     -23.616152: 0.288666
  6:     -23.703097: 0.366825
  7:     -23.703443: 0.361997
  8:     -23.703444: 0.362261
  9:     -23.703444: 0.362262
```

```
> anova(B1ML,B0ML)
Data: rikz
Models:
B1ML: ShannonW ~ exposure + NAP + (1 | Beach)
B0ML: ShannonW ~ 1 + exposure * NAP + (1 | Beach)
     Df      AIC      BIC  logLik  Chisq Chi Df Pr(>Chisq)
B1ML  5  -9.5188 -0.48549  9.7594
B0ML  6 -11.7034 -0.86347 11.8517 4.1846      1    0.04079 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> anova(B1ML,B0ML)
Data: rikz
Models:
B1ML: ShannonW ~ exposure + NAP + (1 | Beach)
B0ML: ShannonW ~ 1 + exposure * NAP + (1 | Beach)
     Df      AIC      BIC  logLik  Chisq Chi Df Pr(>Chisq)
B1ML  5  -9.5188 -0.48549  9.7594
B0ML  6 -11.7034 -0.86347 11.8517 4.1846      1    0.04079 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The model B0 (or B0ML) with interaction between NAP and exposure fits significantly better.

Zuur *et al.* [1] recommend the following strategy for the selection of random and fixed parameters.

1. Start with a model that contains as many of the relevant parameters and interactions as possible.
2. First select random parameters. To decide between models which have different random parameters, fit models with REML and choose model of minimal AIC.
3. Now select fixed parameters. This can be done with the help of AIC or with likelihood ratio tests. If likelihood ratio tests are used, apply ML to fit the models to the data.
4. Never remove covariates that are still involved in interactions.
5. Fit the final model with REML.

---

[1] A.F. Zuur, E.N. Ieno, N.J. Walker, A.A. Saveliev, G.M. Smith (2009) *Mixed effects models and extensions in ecology with R.* Springer.

```
> B2 <- lmer(ShannonW~1+(exposure+salinity+NAP+humus)^2+(
+             .... [TRUNCATED]
  0:     -5.8942807: 0.730297
  1:     -5.8978355: 0.838579
  2:     -5.9174068: 0.789454
  3:     -5.9175376: 0.784845
  4:     -5.9175385: 0.785156
  5:     -5.9175385: 0.785156
```

```
> summary(B2)
Linear mixed model fit by REML
Formula: ShannonW ~ 1 + (exposure + salinity + NAP + humus)^2
                                                  + (1 | Beach)
   Data: rikz
   AIC   BIC logLik deviance REMLdev
 20.08 43.57  2.959   -37.55  -5.918
Random effects:
 Groups   Name        Variance Std.Dev.
 Beach    (Intercept) 0.015652 0.12511
 Residual             0.025390 0.15934
Number of obs: 45, groups: Beach, 9

Fixed effects:
                  Estimate Std. Error t value
(Intercept)       40.32851   21.45346   1.880
exposure          -4.16161    2.12543  -1.958
salinity          -1.38750    0.74706  -1.857
NAP                0.65692    0.92037   0.714
humus             40.36916   16.33692   2.471
exposure:salinity  0.14522    0.07425   1.956
```

```
[..]
Correlation of Fixed Effects:
            (Intr) exposr salnty NAP    humus  expsr:s ex:NAP expsr
exposure    -0.998
salinity    -0.999  0.998
NAP         -0.079  0.062  0.074
humus        0.117 -0.142 -0.129  0.152
expsr:slnty  0.996 -0.999 -0.998 -0.057  0.155
exposur:NAP  0.173 -0.165 -0.177 -0.661  0.067  0.167
exposur:hms -0.285  0.295  0.306 -0.008 -0.694 -0.315  -0.331
salinty:NAP  0.005  0.012  0.003 -0.920 -0.224 -0.020   0.316  0.18
salinty:hms  0.025  0.002 -0.022 -0.197 -0.888 -0.006   0.128  0.28
NAP:humus    0.149 -0.145 -0.158  0.109 -0.119  0.154   0.215 -0.14
            slnty:
exposure
salinity
NAP
humus
expsr:slnty
exposur:NAP
exposur:hms
```

```
> B3 <- lmer(ShannonW~1+(exposure+salinity+NAP+humus)^2
+          +(1+NAP|Beach), data=rikz,verbose=TRUE)
  0:     -1.5323306: 0.730297 0.700275  0.00000
  1:     -5.4597824:  1.16363  0.00000 -0.193541
  2:     -5.7892458:  1.04081 4.21403e-06 -0.0735358
  3:     -5.9531543: 0.913041 6.58205e-06 -0.188264
[..]
 12:     -6.0798268: 0.829521  0.00000 -0.109646
 13:     -6.0798268: 0.829519  0.00000 -0.109651
 14:     -6.0798268: 0.829519  0.00000 -0.109651
```

Next, we fit a model where there is not only a random intercept for every beach but also a random coefficient of NAP. Again, let S_i and N_i be the ShannonW and the NAP observed at plot i, which is on beach k. The model says

$$S_i = a + [\text{fixed effects terms}] + \alpha_k + \beta_k \cdot N_i + \varepsilon_i.$$

$\varepsilon_1, \dots, \varepsilon_{45}$ are independently $\mathcal{N}(0, \sigma^2)$-distributed,
$\alpha_1, \dots, \alpha_9$ are independently $\mathcal{N}(0, \sigma_\alpha^2)$-distributed,
$\beta_1, \dots, \beta_9$ are independently $\mathcal{N}(0, \sigma_\beta^2)$-distributed,

Next, we fit a model where there is not only a random intercept for every beach but also a random coefficient of NAP. Again, let S_i and N_i be the ShannonW and the NAP observed at plot i, which is on beach k. The model says

$$S_i = a + [\text{fixed effects terms}] + \alpha_k + \beta_k \cdot N_i + \varepsilon_i.$$

$\varepsilon_1, \dots, \varepsilon_{45}$ are independently $\mathcal{N}(0, \sigma^2)$-distributed,
$\alpha_1, \dots, \alpha_9$ are independently $\mathcal{N}(0, \sigma_\alpha^2)$-distributed,
$\beta_1, \dots, \beta_9$ are independently $\mathcal{N}(0, \sigma_\beta^2)$-distributed,

Besides the fixed-effects coefficients we have to estimate σ, σ_α and σ_β.

```
> summary(B3)
[..]
   AIC   BIC logLik deviance REMLdev
 23.92 51.02   3.04   -37.69   -6.08
Random effects:
 Groups   Name        Variance   Std.Dev. Corr
 Beach    (Intercept) 0.01730652 0.131554
          NAP         0.00020808 0.014425 -1.000
 Residual             0.02515108 0.158591
Number of obs: 45, groups: Beach, 9

Fixed effects:
                  Estimate Std. Error t value
(Intercept)       40.93936   21.72964   1.884
exposure          -4.22455    2.15239  -1.963
salinity          -1.40881    0.75753  -1.860
NAP                0.67896    0.93389   0.727
humus             40.42412   16.30099   2.480
exposure:salinity  0.14742    0.07528   1.958
exposure:NAP       0.08480    0.03663   2.315
exposure:humus    -2.10698    0.78447  -2.686
```

```
> B4 <- lmer(ShannonW˜1+(exposure+salinity+NAP+humus)ˆ2
+             +(1+NAP+exposure|Beach),data=rikz,verbose=TRUE)
  0:    -1.6536097: 0.730297 0.700275 0.0711568  0.00000  0.00000
  1:    -3.4551238: 0.700827 0.500322  0.00000 -0.0225207 -0.207903
  2:    -5.5427746: 0.630604 0.176807  0.00000 -0.00365199 -0.00670
  3:    -5.7754683: 0.635022 0.175143  0.00000 -0.00439670 0.027882
[...]
267:    -6.2309242: 0.0242437 1.71415e-05 7.45081e-05 -3.81706  3.1
268:    -6.2309242: 0.0242437 1.71415e-05 7.45081e-05 -3.81706  3.1
269:    -6.2309242: 0.0242437 1.71415e-05 7.45081e-05 -3.81706  3.1
```

```
> summary(B4)
[..]
   AIC   BIC logLik deviance REMLdev
 29.77 62.29  3.115   -38.41  -6.231
Random effects:
 Groups   Name        Variance   Std.Dev.  Corr
 Beach    (Intercept) 1.4754e-05 0.0038411
          NAP         2.1496e-04 0.0146616 -1.000
          exposure    1.4259e-04 0.0119410  1.000 -1.000
 Residual             2.5102e-02 0.1584364
Number of obs: 45, groups: Beach, 9

Fixed effects:
                  Estimate Std. Error t value
(Intercept)       42.77575   20.81776   2.055
exposure          -4.41472    2.06398  -2.139
salinity          -1.47464    0.72415  -2.036
NAP                0.67731    0.93297   0.726
humus             40.48334   16.27068   2.488
exposure:salinity  0.15425    0.07203   2.142
exposure:NAP       0.08472    0.03662   2.313
```

```
> B5 <- lmer(ShannonW~1+(exposure+salinity+NAP+humus)^2
+            +(1+NAP+exposure|Beach)+(1|Week),data=rikz,verbose=TRU
  0:    -1.9949020: 0.730297 0.700275 0.0711568  0.00000  0.00000
  1:    -2.5896994: 0.696448 0.572039  0.00000 -0.0146206 -0.256429
  2:    -5.3531758: 0.661521 0.525272  0.00000 -0.00132594 -0.10653
[..]
 62:    -9.1285182:  4.30055 1.60825e-07  0.00000 0.0323185 -0.0979
 63:    -9.1285182:  4.30055 1.49005e-08 5.79252e-07 0.0323201 -0.0
 64:    -9.1285182:  4.30055  0.00000  0.00000 0.0323202 -0.0979503
```

```
> summary(B5)
Linear mixed model fit by REML
Formula: ShannonW ~ 1 + (exposure + salinity + NAP + humus)^2
                 + (1 + NAP +      exposure | Beach) + (1 | Week)
   Data: rikz
   AIC  BIC logLik deviance REMLdev
 28.87 63.2  4.564   -40.07  -9.129
Random effects:
 Groups   Name        Variance   Std.Dev. Corr
 Beach    (Intercept) 0.44901920 0.670089
          NAP         0.00046904 0.021657  1.000
          exposure    0.00430800 0.065635 -1.000 -1.000
 Week     (Intercept) 0.02093924 0.144704
 Residual             0.02427827 0.155815
Number of obs: 45, groups: Beach, 9; Week, 4

Fixed effects:
                  Estimate Std. Error t value
(Intercept)       21.42334   15.49344   1.383
exposure          -2.34537    1.53851  -1.524
salinity          -0.74919    0.54787  -1.367
```

```
> anova(B2,B3,B4,B5)
Data: rikz
Models:
B2: ShannonW ~ 1 + (exposure + salinity + NAP + humus)^2
         + (1 | Beach)
B3: ShannonW ~ 1 + (exposure + salinity + NAP + humus)^2
         + (1 + NAP |  Beach)
B4: ShannonW ~ 1 + (exposure + salinity + NAP + humus)^2
         + (1 + NAP + exposure | Beach)
B5: ShannonW ~ 1 + (exposure + salinity + NAP + humus)^2
         + (1 + NAP + exposure | Beach) + (1 | Week)
   Df      AIC    BIC logLik  Chisq Chi Df Pr(>Chisq)
B2 13 -11.5528 11.934 18.776
B3 15  -7.6911 19.409 18.846 0.1383      2     0.9332
B4 18  -2.4079 30.112 19.204 0.7168      3     0.8693
B5 19  -2.0744 32.252 20.037 1.6665      1     0.1967
```

Don't trust the *p*-values on the previous slide! The problem is not only that the models were fitted with REML. The main problem ist that the null hypotheses (e.g. $\sigma_\beta = 0$ in the case of B2/B3) are on the boundary of the parameter space. σ_β can only be ≥ 0, and deviations from $\sigma_\beta = 0$ are thus only possible in one direction. The χ^2-approximation of likelihood ratio tests are only reliable when deviations from the expectation under the null hypothesis are possible in all directions, for example if the null hypothesis $\theta = 0$ is tested for some parameter θ, and estimates of θ can lead to positive as well as negative values.

Don’t trust the *p*-values on the previous slide! The problem is not only that the models were fitted with REML. The main problem ist that the null hypotheses (e.g. $\sigma_\beta = 0$ in the case of B2/B3) are on the boundary of the parameter space. σ_β can only be ≥ 0, and deviations from $\sigma_\beta = 0$ are thus only possible in one direction. The χ^2-approximation of likelihood ratio tests are only reliable when deviations from the expectation under the null hypothesis are possible in all directions, for example if the null hypothesis $\theta = 0$ is tested for some parameter θ, and estimates of θ can lead to positive as well as negative values.

Thus, we rather base our decision on the AIC values. This is, of course, also not stringent. However, in our case, all criteria favor model B2.

Now we keep the random effects “`(1 | Beach)`” and search for a a good selection of fixed effects. Here, our aim is to get a simple, understandable model. Therefore apply a strict criterion an remove all variables which do not significantly improve the model fit on the 5% level.

Now we keep the random effects “`(1 | Beach)`” and search for a a good selection of fixed effects. Here, our aim is to get a simple, understandable model. Therefore apply a strict criterion an remove all variables which do not significantly improve the model fit on the 5% level.

(If our aim was a model to make good predictions, we would rather take AIC as a criterion, which leaves more parameters in.)

```
> summary(B2)
Linear mixed model fit by REML
Formula: ShannonW ~ 1 + (exposure + salinity + NAP + humus)^2
                                             + (1 | Beach)
[..]
Fixed effects:
                  Estimate Std. Error t value
(Intercept)       40.32851   21.45346   1.880
exposure          -4.16161    2.12543  -1.958
salinity          -1.38750    0.74706  -1.857
NAP                0.65692    0.92037   0.714
humus             40.36916   16.33692   2.471
exposure:salinity  0.14522    0.07425   1.956
exposure:NAP       0.08357    0.03615   2.312
exposure:humus    -2.06931    0.78148  -2.648
salinity:NAP      -0.06337    0.02639  -2.401
salinity:humus    -0.67100    0.43169  -1.554
NAP:humus          1.01589    0.59247   1.715
[..]
```

```
> B6 <- update(B2,~.-salinity:humus)
  0:     -3.3840456: 0.730297
  1:     -3.3848655: 0.748586
  2:     -3.3851681: 0.742306
  3:     -3.3851681: 0.742296
  4:     -3.3851681: 0.742299
  5:     -3.3851681: 0.742299
```

```
> anova(B6,B2)
Data: rikz
Models:
B6: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B6:     exposure:salinity + exposure:NAP + exposure:humus + salinit
B6:     NAP:humus
B2: ShannonW ~ 1 + (exposure + salinity + NAP + humus)^2 + (1 | Bea
   Df     AIC    BIC logLik  Chisq Chi Df Pr(>Chisq)
B6 12 -10.720 10.960 17.360
B2 13 -11.553 11.934 18.776 2.8329      1    0.09235 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Remember that we should better use ML instead of REML if we want to apply likelihood ratio tests!

```
> B2ML <- update(B2,REML=FALSE)
  0:     -37.815253: 0.730297
  1:     -38.227458: 0.520323
  2:     -38.238951: 0.555035
  3:     -38.239535: 0.548840
  4:     -38.239535: 0.548745
  5:     -38.239535: 0.548714
  6:     -38.239535: 0.548724

> B6ML <- update(B6,REML=FALSE)
  0:     -34.786862: 0.730297
  1:     -35.481002: 0.455308
  2:     -35.502068: 0.503911
  3:     -35.503579: 0.494055
  4:     -35.503579: 0.493868
  5:     -35.503579: 0.493857
  6:     -35.503579: 0.493844
```

```
> anova(B6ML,B2ML)
Data: rikz
Models:
B6ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B6ML:     exposure:salinity + exposure:NAP + exposure:humus + salin
B6ML:     NAP:humus
B2ML: ShannonW ~ 1 + (exposure + salinity + NAP + humus)^2 + (1 | B
     Df     AIC    BIC logLik Chisq Chi Df Pr(>Chisq)
B6ML 12 -11.504 10.176 17.752
B2ML 13 -12.239 11.247 19.120 2.736      1    0.09811 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> summary(B6)
Linear mixed model fit by REML
Formula: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach)
     exposure:salinity + exposure:NAP + exposure:humus + salinity:N
                    NAP:humus
[...]
                  Estimate Std. Error t value
(Intercept)       40.98076   21.18993   1.934
exposure          -4.13572    2.10009  -1.969
salinity          -1.40592    0.73832  -1.904
NAP                0.37102    0.92444   0.401
humus             17.61688    7.69600   2.289
exposure:salinity  0.14382    0.07340   1.959
exposure:NAP       0.09008    0.03673   2.453
exposure:humus    -1.70126    0.76643  -2.220
salinity:NAP      -0.05584    0.02661  -2.098
NAP:humus          1.24688    0.58490   2.132
[..]
```

```
> B7 <- update(B6,~.-exposure:salinity)
  0:     -3.0026064: 0.730297
  1:     -3.2146723:  1.04756
  2:     -3.2845341: 0.957924
  3:     -3.2886569: 0.926746
  4:     -3.2887713: 0.931248
  5:     -3.2887715: 0.931100
  6:     -3.2887715: 0.931099

> B7ML <- update(B7,REML=FALSE)
  0:     -30.038157: 0.730297
  1:     -30.038162: 0.725726
  2:     -30.038210: 0.727949
  3:     -30.038210: 0.727947
```

```
> anova(B7ML,B6ML)
Data: rikz
Models:
B7ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B7ML:     exposure:NAP + exposure:humus + salinity:NAP + NAP:humus
B6ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B6ML:     exposure:salinity + exposure:NAP + exposure:humus + salin
B6ML:     NAP:humus
     Df      AIC    BIC logLik  Chisq Chi Df Pr(>Chisq)
B7ML 11  -8.0382 11.835 15.019
B6ML 12 -11.5036 10.176 17.752 5.4654      1    0.01940 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> B8 <- update(B6,~.-salinity:NAP)
  0:     -4.5729780: 0.730297
  1:     -4.6975945: 0.596090
  2:     -4.6982261: 0.606261
  3:     -4.6982498: 0.604628
  4:     -4.6982498: 0.604628
  5:     -4.6982498: 0.604628
```

```
> B8ML <- update(B8,REML=FALSE)
  0:     -29.489856: 0.730297
  1:     -30.710552:  0.00000
  2:     -30.710552: 2.50994e-05
  3:     -30.710558: 0.000649769
  4:     -31.015846: 0.162724
  5:     -31.356389: 0.335700
  6:     -31.356637: 0.345734
  7:     -31.356869: 0.341604
  8:     -31.356869: 0.341641
  9:     -31.356869: 0.341630
```

```
> anova(B8ML,B6ML)
Data: rikz
Models:
B8ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B8ML:     exposure:salinity + exposure:NAP + exposure:humus + NAP:h
B6ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B6ML:     exposure:salinity + exposure:NAP + exposure:humus + salin
B6ML:     NAP:humus
     Df      AIC    BIC logLik  Chisq Chi Df Pr(>Chisq)
B8ML 11  -9.3569 10.516 15.678
B6ML 12 -11.5036 10.176 17.752 4.1467      1    0.04172 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> summary(B6)
[...]
                   Estimate Std. Error t value
(Intercept)        40.98076   21.18993   1.934
exposure           -4.13572    2.10009  -1.969
salinity           -1.40592    0.73832  -1.904
NAP                 0.37102    0.92444   0.401
humus              17.61688    7.69600   2.289
exposure:salinity   0.14382    0.07340   1.959
exposure:NAP        0.09008    0.03673   2.453
exposure:humus     -1.70126    0.76643  -2.220
salinity:NAP       -0.05584    0.02661  -2.098
NAP:humus           1.24688    0.58490   2.132
[...]
```

```
> B9 <- update(B6,~.-NAP:humus)
  0:     1.7262210: 0.730297
  1:     1.4342459: 0.527474
  2:     1.4326078: 0.545056
  3:     1.4324476: 0.540992
  4:     1.4324476: 0.540993
  5:     1.4324476: 0.540993
  6:     1.4324476: 0.540993
```

```
> B9ML <- update(B9,REML=FALSE)
  0:     -29.340922: 0.730297
  1:     -31.445430:  0.00000
  2:     -31.445430: 1.57944e-05
  3:     -31.445431: 0.000268510
  4:     -31.471347: 0.0479773
  5:     -31.646917: 0.147977
  6:     -31.778206: 0.347977
  7:     -31.821753: 0.257110
  8:     -31.831799: 0.282988
  9:     -31.831884: 0.285733
 10:     -31.831884: 0.285584
 11:     -31.831884: 0.285537
 12:     -31.831884: 0.285545
```

```
> anova(B9ML,B6ML)
Data: rikz
Models:
B9ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B9ML:     exposure:salinity + exposure:NAP + exposure:humus + salin
B6ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B6ML:     exposure:salinity + exposure:NAP + exposure:humus + salin
B6ML:     NAP:humus
     Df      AIC    BIC logLik  Chisq Chi Df Pr(>Chisq)
B9ML 11  -9.8319 10.041 15.916
B6ML 12 -11.5036 10.176 17.752 3.6717      1    0.05534 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> summary(B9)
[..]
                  Estimate Std. Error t value
(Intercept)       33.46430   18.61198   1.798
exposure          -3.37260    1.84346  -1.829
salinity          -1.12647    0.64909  -1.735
NAP                0.02515    0.96985   0.026
humus             12.41735    7.87781   1.576
exposure:salinity  0.11548    0.06449   1.791
exposure:NAP       0.07247    0.03847   1.884
exposure:humus    -1.21902    0.78897  -1.545
salinity:NAP      -0.03552    0.02661  -1.335
[..]
```

```
> B10 <- update(B9,~.-salinity:NAP)
  0:     -1.7647565: 0.730297
  1:     -2.2271740: 0.471298
  2:     -2.2311842: 0.500273
  3:     -2.2317909: 0.492691
  4:     -2.2317912: 0.492525
  5:     -2.2317912: 0.492501
  6:     -2.2317912: 0.492501

> B10ML <- update(B10,REML=FALSE)
  0:     -26.956276: 0.730297
  1:     -29.939179:  0.00000
  2:     -29.939179: 1.55125e-06
  3:     -29.939179: 1.55125e-06
```

```
> anova(B10ML,B9ML)
Data: rikz
Models:
B10ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B10ML:     exposure:salinity + exposure:NAP + exposure:humus
B9ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B9ML:     exposure:salinity + exposure:NAP + exposure:humus
B9ML       + salinity:NAP
      Df     AIC     BIC logLik  Chisq Chi Df Pr(>Chisq)
B10ML 10 -9.9392  8.1274 14.970
B9ML  11 -9.8319 10.0414 15.916 1.8927      1     0.1689
```

```
> summary(B10)
[..]
(Intercept)        34.86718   18.02729   1.934
exposure           -3.47464    1.78723  -1.944
salinity           -1.17017    0.62943  -1.859
NAP                -1.16367    0.37065  -3.139
humus              11.46664    7.96861   1.439
exposure:salinity   0.11852    0.06259   1.894
exposure:NAP        0.09198    0.03577   2.571
exposure:humus     -1.12149    0.79790  -1.406
[..]
```

```
> B11 <- update(B10,~.-exposure:humus)
  0:      1.9746978: 0.730297
  1:      1.8085855:  0.00000
  2:      1.8085855:  0.00000

> B11ML <- update(B11,REML=FALSE)
  0:     -24.123141: 0.730297
  1:     -28.644278:  0.00000
  2:     -28.644278:  0.00000
```

```
> anova(B11ML,B10ML)
Data: rikz
Models:
B11ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B11ML:     exposure:salinity + exposure:NAP
B10ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B10ML:     exposure:salinity + exposure:NAP + exposure:humus
      Df      AIC    BIC logLik  Chisq Chi Df Pr(>Chisq)
B11ML  9 -10.6443 5.6157 14.322
B10ML 10  -9.9392 8.1274 14.970 1.2949      1     0.2551
```

```
> B11
Linear mixed model fit by REML
Formula: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach)
                              +      exposure:salinity + exposure:N
[..]
Fixed effects:
                  Estimate Std. Error t value
(Intercept)       25.44676   12.14435   2.095
exposure          -2.53976    1.20472  -2.108
salinity          -0.82116    0.41941  -1.958
NAP               -0.89459    0.35486  -2.521
humus              0.26450    0.51705   0.511
exposure:salinity  0.08385    0.04175   2.008
exposure:NAP       0.06663    0.03459   1.926
[...]
```

```
> B12 <- update(B11,~.-exposure:salinity)
  0:  -0.098587691: 0.730297
  1:   -0.66093519: 0.474219
  2:   -0.66484670: 0.500408
  3:   -0.66548356: 0.493367
  4:   -0.66548380: 0.493228
  5:   -0.66548380: 0.493229

> B12ML <- update(B12,REML=FALSE)
  0:    -22.593499: 0.730297
  1:    -24.104251:  0.00000
  2:    -24.104251: 6.18232e-06
  3:    -24.104252: 0.000170993
  4:    -24.139061: 0.0457497
  5:    -24.470138: 0.165738
  6:    -24.715631: 0.405714
  7:    -24.801367: 0.303438
```

```
> anova(B12ML,B11ML)
Data: rikz
Models:
B12ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B12ML:     exposure:NAP
B11ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B11ML:     exposure:salinity + exposure:NAP
      Df      AIC    BIC logLik  Chisq Chi Df Pr(>Chisq)
B12ML  8  -8.8116 5.6417 12.406
B11ML  9 -10.6443 5.6157 14.322 3.8327      1    0.05026 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> B12
Linear mixed model fit by REML
Formula: ShannonW ~ exposure + salinity + NAP + humus +
                  (1 | Beach) +   exposure:NAP
[..]
             Estimate Std. Error t value
(Intercept)   1.15024    1.46738   0.784
exposure     -0.12376    0.05269  -2.349
salinity      0.01993    0.04045   0.493
NAP          -0.93216    0.34402  -2.710
humus         0.35074    0.51842   0.677
exposure:NAP  0.07059    0.03354   2.105
[..]
```

```
> B13 <- update(B12,~.-salinity)
  0:     -4.0925532: 0.730297
  1:     -5.0279973: 0.406952
  2:     -5.0421919: 0.454465
  3:     -5.0447002: 0.441360
  4:     -5.0447011: 0.441102
  5:     -5.0447011: 0.441104

> B13ML <- update(B13,REML=FALSE)
  0:     -22.405729: 0.730297
  1:     -23.632872:  0.00000
  2:     -23.632872: 6.82081e-06
  3:     -23.632873: 0.000209714
  4:     -23.702115: 0.0625813
  5:     -24.233233: 0.221826
  6:     -24.453896: 0.338432
  7:     -24.454528: 0.346503
```

```
> anova(B13ML,B12ML)
Data: rikz
Models:
B13ML: ShannonW ~ exposure + NAP + humus + (1 | Beach) + exposure:N
B12ML: ShannonW ~ exposure + salinity + NAP + humus + (1 | Beach) +
B12ML:     exposure:NAP
      Df      AIC    BIC logLik Chisq Chi Df Pr(>Chisq)
B13ML  7 -10.4546 2.1921 12.227
B12ML  8  -8.8116 5.6417 12.406 0.357      1     0.5502
```

```
> B13
Linear mixed model fit by REML
Formula: ShannonW ~ exposure + NAP + humus +
                    (1 | Beach) + exposure:NAP
[..]
             Estimate Std. Error t value
(Intercept)   1.82886    0.45860   3.988
exposure     -0.13557    0.04412  -3.073
NAP          -0.93632    0.34287  -2.731
humus         0.40236    0.50705   0.794
exposure:NAP  0.07087    0.03343   2.120
[..]
```

```
> B14 <- update(B13,~.-humus)
  0:     -3.1079878: 0.730297
  1:     -3.9300741: 0.431067
  2:     -3.9406377: 0.471682
  3:     -3.9424344: 0.460627
  4:     -3.9424352: 0.460386
  5:     -3.9424352: 0.460388

> B14ML <- update(B14,REML=FALSE)
  0:     -21.820395: 0.730297
  1:     -22.739010:  0.00000
  2:     -22.739010: 7.52231e-06
  3:     -22.739011: 0.000259128
  4:     -22.880262: 0.0868904
  5:     -23.616152: 0.288666
  6:     -23.703097: 0.366825
  7:     -23.703443: 0.361997
```

```
> anova(B14ML,B13ML)
Data: rikz
Models:
B14ML: ShannonW ~ exposure + NAP + (1 | Beach) + exposure:NAP
B13ML: ShannonW ~ exposure + NAP + humus + (1 | Beach) + exposure:N
      Df     AIC      BIC logLik  Chisq Chi Df Pr(>Chisq)
B14ML  6 -11.703 -0.86347 11.852
B13ML  7 -10.455  2.19207 12.227 0.7511      1     0.3861
```

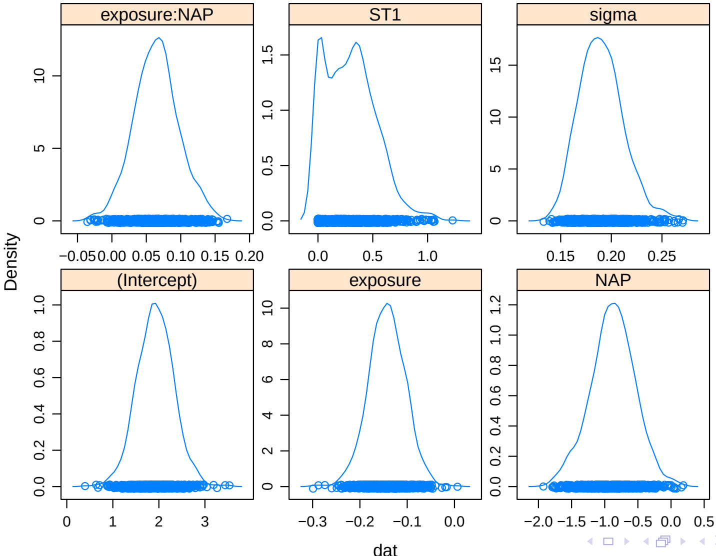
```
> B14
[..]
Fixed effects:
             Estimate Std. Error t value
(Intercept)   1.92603    0.44765   4.303
exposure     -0.14294    0.04368  -3.272
NAP          -0.89673    0.33620  -2.667
exposure:NAP  0.06661    0.03270   2.037
[..]
```

```
> B15 <- update(B14,~.-exposure:NAP)
  0:     -3.5739227: 0.730297
  1:     -4.1112507:  0.00000
  2:     -4.1112507: 3.65618e-06
  3:     -4.1112511: 0.000144407
  4:     -4.1650501: 0.0555603
  5:     -4.7984144: 0.246592
  6:     -5.0104872: 0.374551
  7:     -5.0104967: 0.373651
  8:     -5.0104968: 0.373725
  9:     -5.0104968: 0.373722

> B15ML <- update(B15,REML=FALSE)
  0:     -16.875166: 0.730297
  1:     -19.518802:  0.00000
  2:     -19.518802: 3.18807e-06
```

```
> anova(B15ML,B14ML)
Data: rikz
Models:
B15ML: ShannonW ~ exposure + NAP + (1 | Beach)
B14ML: ShannonW ~ exposure + NAP + (1 | Beach) + exposure:NAP
      Df      AIC      BIC  logLik  Chisq Chi Df Pr(>Chisq)
B15ML  5  -9.5188 -0.48549  9.7594
B14ML  6 -11.7034 -0.86347 11.8517 4.1846      1    0.04079 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> summary(B14)
[..]
Fixed effects:
             Estimate Std. Error t value
(Intercept)   1.92603    0.44765   4.303
exposure     -0.14294    0.04368  -3.272
NAP          -0.89673    0.33620  -2.667
exposure:NAP  0.06661    0.03270   2.037
[..]
```

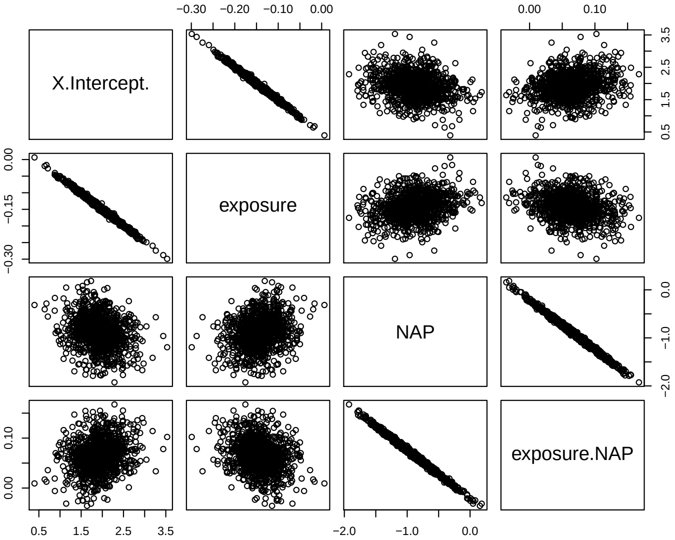
```
> samp14 <- mcmcsamp(B14,1000)
> densityplot(samp14)
```

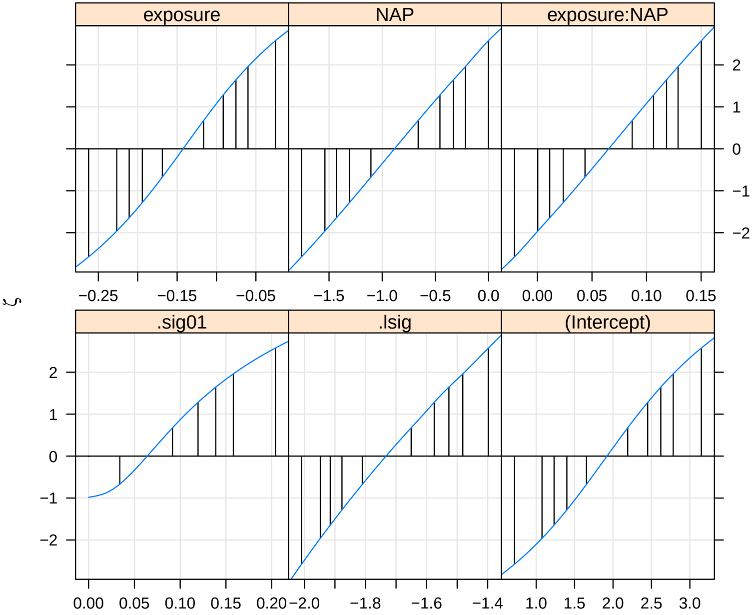


```
> sum(samp14@ST==0)/1000
[1] 0.153
```

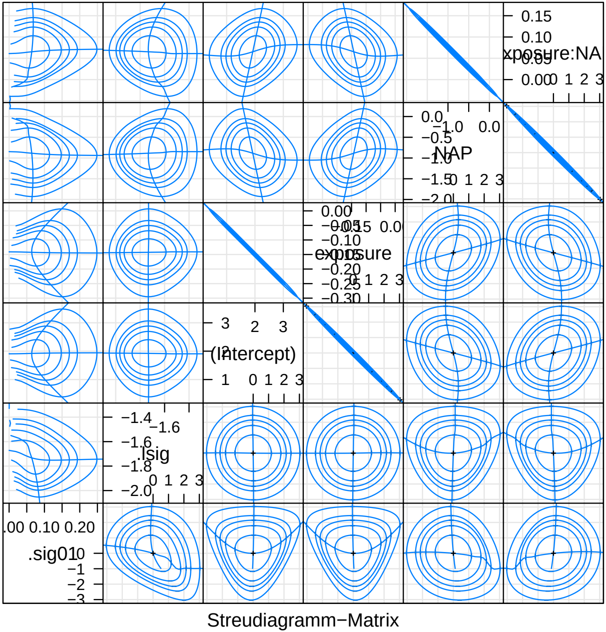
15.3% of the sampled parameter sets have $\sigma_\alpha = 0$. This means, with a posterior probability of 15.3% there is no beach effect.

```
> plot(data.frame(t(samp14@fixef)))
```



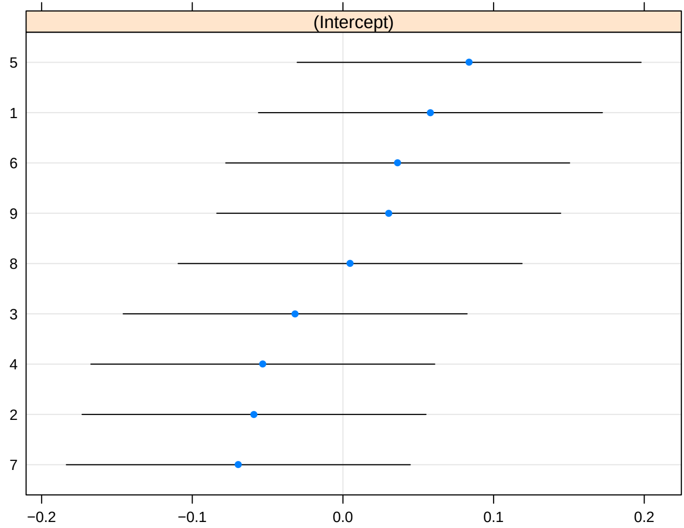

exposure
NAP
exposure:NAP
.sig01
.lsig
(Intercept)
ζ
−0.25
−0.15
−0.05
−1.5
−1.0
−0.5
0.0
0.00
0.05
0.10
0.15
0.00
0.05
0.10
0.15
0.20
−2.0
−1.8
−1.6
−1.4
1.0
1.5
2.0
2.5
3.0
2
1
0
−1
−2

```
> confint(pr14)
                      2.5 %       97.5 %
.sig01                   NA   0.15806902
.lsig         -1.9478665471 -1.48146028
(Intercept)    1.0759293994  2.78295226
exposure      -0.2265797054 -0.06004775
NAP           -1.5386892998 -0.21810411
exposure:NAP   0.0003071935  0.12889265
```

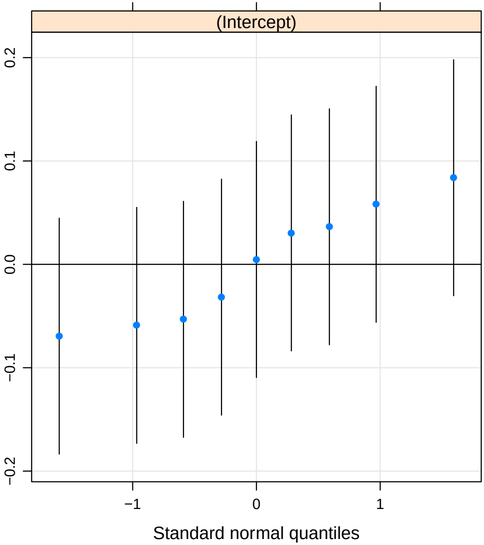


xposure:NA
0.15
0.10
0.05
0.00
0 1 2 3
NAP
0.0
−0.5
−1.0
−1.5
−2.0
−1.0 0.0
exposure
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.30
−0.15 0.0
(Intercept)
3
2
1
2 3
.lsig
−1.4
−1.6
−1.8
−2.0
−1.6
.sig01
0
−1
−2
−3
.00 0.10 0.20
Streudiagramm−Matrix

```
> dotplot(ranef(B14,postVar=TRUE))
```



```
> qqmath(ranef(B14,postVar=TRUE))
```



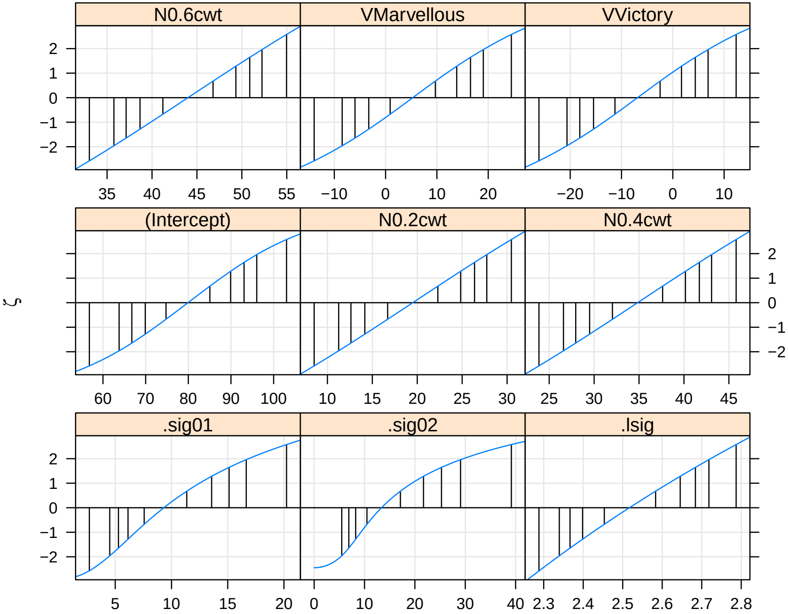
Contents

Nested Anova of a split-plot experiment.

```
> library(MASS)
> oats.aov <- aov(Y˜N+V+Error(B/V),data=oats,qr=T)
> model.tables(oats.aov,type="mean")

Grand mean
103.9722

N
0.0cwt 0.2cwt 0.4cwt 0.6cwt
 79.39  98.89 114.22 123.39

V
Golden.rain  Marvellous     Victory
     104.50      109.79       97.63
> oats.lmer <- lmer(Y˜N+V+(1 | B/V),data=oats)
> fixef(oats.lmer)
 (Interc) N0.2cwt N0.4cwt N0.6cwt VMarvellous VVictory
  79.917   19.50    34.83   44.0       5.292   -6.875
```

```
> oats.pr <- profile(oats.lmer)
> confint(oats.pr)
                  2.5 %    97.5 %
.sig01         4.511185 16.646920
.sig02         5.482973 29.072479
.lsig          2.339491  2.718178
(Intercept)   63.786567 96.046767
N0.2cwt       11.258823 27.741177
N0.4cwt       26.592156 43.074511
N0.6cwt       35.758823 52.241177
VMarvellous   -8.461640 19.045023
VVictory     -20.628307  6.878357
```

Nested Anova of a split-plot experiment.



Contents

- Generalized linear mixed-effects models can be fitted with the glmer command in the lme4 package.

- Generalized linear mixed-effects models can be fitted with the glmer command in the lme4 package.
- REML is not applied, more complex algorithms are applied to fit models.

- Generalized linear mixed-effects models can be fitted with the glmer command in the lme4 package.
- REML is not applied, more complex algorithms are applied to fit models.
- All *p*-values can be very imprecise, so do not trust them too much, especially if they are close to the significance level.

- Generalized linear mixed-effects models can be fitted with the glmer command in the lme4 package.
- REML is not applied, more complex algorithms are applied to fit models.
- All *p*-values can be very imprecise, so do not trust them too much, especially if they are close to the significance level.
- Likelihood profiles for GLMMs can only be generated with very recent versions of lme4.

- Generalized linear mixed-effects models can be fitted with the glmer command in the lme4 package.
- REML is not applied, more complex algorithms are applied to fit models.
- All *p*-values can be very imprecise, so do not trust them too much, especially if they are close to the significance level.
- Likelihood profiles for GLMMs can only be generated with very recent versions of lme4.
- As an example we fit an overdispersed Poisson model to the RIKZ data with Richness as the response variable. (Note “quasipoisson” is not allowed in the most recent versions of lme4)

The following analysis with step-wise parameter selection is performed with an older version of lme4. We will later see what's now possible with the upcoming versions.

```
> mmod11 <- glmer(Richness~1+(exposure+salinity+NAP+humus)^2
+                  +(1 | Beach),
+                 family=quasipoisson,data=rikz,verbose=TRUE)
  0:     49.833560: 0.730297  64.4576 -6.91148 -2.16330  2.32707[..
  1:     49.818337: 0.730288  64.4576 -6.91149 -2.16331  2.32707[..
  2:     49.813761: 0.730246  64.4576 -6.91148 -2.16330  2.32707[..
[...]
 52:     32.698606: 2.36530e-08  64.4582 -6.91155 -2.16332  2.32[..
 53:     32.698606: 1.27632e-08  64.4582 -6.91155 -2.16332  2.32[..
 54:     32.698606: 2.55421e-09  64.4582 -6.91155 -2.16332  2.32[..
```

```
> summary(mmod11)
Generalized linear mixed model fit by the Laplace approximation
Formula: Richness ~ 1 + (exposure + salinity + NAP + humus)^2
                                              + (1 | Beach)
   Data: rikz
  AIC   BIC logLik deviance
 58.7 82.19 -16.35     32.7
Random effects:
 Groups   Name        Variance   Std.Dev.
 Beach    (Intercept) 1.2605e-17 3.5504e-09
 Residual             1.9321e+00 1.3900e+00
Number of obs: 45, groups: Beach, 9

Fixed effects:
                  Estimate Std. Error t value
(Intercept)       64.45819   45.35670   1.421
exposure          -6.91155    4.44718  -1.554
salinity          -2.16332    1.56826  -1.379
NAP                2.32704    4.22804   0.550
humus             62.83613   49.51439   1.269
exposure:salinity  0.23884    0.15421   1.549
```

```
> mmod12 <- glmer(Richness~1+(exposure+salinity+NAP+humus)^2
+                  +(1+NAP | Beach),
+                  family=quasipoisson,data=rikz,verbose=TRUE)
  0:     59.544991: 0.730297 0.700275  0.00000  64.4576 -6.91148 -2
  1:     59.537494: 0.730285 0.700265 -5.54163e-08  64.4576 -6.9114
  2:     59.531152: 0.730266 0.700248 -2.01145e-07  64.4576 -6.9114
[...]
 59:     32.698606:  0.00000  0.00000 0.134189  64.4553 -6.91124 -2
 60:     32.698606:  0.00000  0.00000 0.134153  64.4577 -6.91148 -2
 61:     32.698606:  0.00000 7.78293e-09 0.134153  64.4577 -6.91148
 62:     32.698606:  0.00000 6.42409e-09 0.134153  64.4577 -6.91148
```

```
> summary(mmod12)
Generalized linear mixed model fit by the Laplace approximation
Formula: Richness ~ 1 + (exposure + salinity + NAP + humus)^2
                                       + (1 + NAP |      Beach)
   Data: rikz
  AIC   BIC logLik deviance
 62.7 89.8 -16.35     32.7
Random effects:
 Groups   Name        Variance   Std.Dev.   Corr
 Beach    (Intercept) 0.0000e+00 0.0000e+00
          NAP         7.9737e-17 8.9296e-09   NaN
 Residual             1.9321e+00 1.3900e+00
Number of obs: 45, groups: Beach, 9

Fixed effects:
                  Estimate Std. Error t value
(Intercept)       64.45774   45.35664   1.421
exposure          -6.91148    4.44718  -1.554
salinity          -2.16330    1.56826  -1.379
NAP                2.32709    4.22804   0.550
humus             62.83178   49.51448   1.269
```

```
> anova(mmod11,mmod12)
Data: rikz
Models:
mmod11: Richness ~ 1 + (exposure + salinity + NAP + humus)^2
                       + (1 | Beach)

mmod12: Richness ~ 1 + (exposure + salinity + NAP + humus)^2
                      + (1 + NAP | Beach)

       Df    AIC    BIC  logLik     Chisq Chi Df Pr(>Chisq)
mmod11 13 58.699 82.185 -16.349
mmod12 15 62.699 89.799 -16.349 4.148e-10      2          1
```

```
> summary(mmod11)
[...]
                  Estimate Std. Error t value
(Intercept)       64.45819   45.35670   1.421
exposure          -6.91155    4.44718  -1.554
salinity          -2.16332    1.56826  -1.379
NAP                2.32704    4.22804   0.550
humus             62.83613   49.51439   1.269
exposure:salinity  0.23884    0.15421   1.549
exposure:NAP      -0.01075    0.13433  -0.080
exposure:humus    -3.34134    2.07648  -1.609
salinity:NAP      -0.10649    0.12179  -0.874
salinity:humus    -0.93055    1.39050  -0.669
NAP:humus          4.47479    2.36577   1.892
[...]
```

```
> mmod13 <- update(mmod11,~.-exposure:NAP)
  0:     49.843351: 0.730297  64.9374 -6.96048 -2.18191  2.08793[..
  1:     49.827666: 0.730288  64.9374 -6.96048 -2.18192  2.08793[..
  2:     49.820544: 0.730200  64.9374 -6.96048 -2.18190  2.08794[..
[...]
 56:     32.710979: 3.62620e-05  64.9381 -6.96054 -2.18195  2.08291
 57:     32.710979:  0.00000  64.9381 -6.96056 -2.18193  2.08813[..
 58:     32.710979:  0.00000  64.9381 -6.96055 -2.18194  2.08799[..
 59:     32.710979:  0.00000  64.9381 -6.96056 -2.18193  2.08798[..
 60:     32.710979:  0.00000  64.9381 -6.96056 -2.18193  2.08794[..
```

```
> anova(mmod13,mmod11)
Data: rikz
Models:
mmod13: Richness ~ exposure + salinity + NAP + humus + (1 | Beach)
mmod13:     exposure:salinity + exposure:humus + salinity:NAP +
mmod13:     salinity:humus + NAP:humus
mmod11: Richness ~ 1 + (exposure + salinity + NAP + humus)^2 + (1 |
       Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
mmod13 12 56.711 78.391 -16.355
mmod11 13 58.699 82.185 -16.349 0.0124      1     0.9114
```

```
> summary(mmod13)
Generalized linear mixed model fit by the Laplace approximation
Formula: Richness ~ exposure + salinity + NAP + humus +
        (1 | Beach) +      exposure:salinity + exposure:humus +
               salinity:NAP + salinity:humus +      NAP:humus
[..]
Fixed effects:
                  Estimate Std. Error t value
(Intercept)        64.9381    45.1037   1.440
exposure           -6.9606     4.4192  -1.575
salinity           -2.1819     1.5560  -1.402
NAP                 2.0879     3.0016   0.696
humus              63.5644    48.7489   1.304
exposure:salinity   0.2407     0.1529   1.575
exposure:humus     -3.4145     1.8665  -1.829
salinity:NAP       -0.1018     0.1073  -0.949
salinity:humus     -0.9299     1.3925  -0.668
NAP:humus           4.4724     2.3660   1.890
[..]
```

```
> mmod14 <- update(mmod13,~.-salinity:humus)
  0:     50.977682: 0.730297  68.3906 -7.14795 -2.29828  2.41247  3.
  1:     50.963879: 0.730289  68.3906 -7.14795 -2.29829  2.41247  3.
  2:     50.963059: 0.730193  68.3906 -7.14795 -2.29828  2.41247  3.
  3:     50.007435: 0.636302  68.3910 -7.14725 -2.29102  2.41331  3.
  4:     47.892486: 0.542366  68.3913 -7.14654 -2.28372  2.41404  3.
  5:     39.995619: 0.166545  68.3925 -7.14406 -2.25563  2.41702  3.
  6:     39.769482: 0.166545  68.3925 -7.14406 -2.25565  2.41702  3.
  7:     39.647185: 0.166543  68.3925 -7.14406 -2.25566  2.41701  3.
0.242384 -3.00727 -0
[...]
 62:     33.567645: 1.55968e-07  68.3895 -7.14784 -2.29824  2.41238
 63:     33.567645:  0.00000  68.3904 -7.14793 -2.29827  2.41247  3.
 64:     33.567645: 1.19137e-09  68.3906 -7.14795 -2.29828  2.41247
```

```
> anova(mmod14,mmod13)
Data: rikz
Models:
mmod14: Richness ~ exposure + salinity + NAP + humus + (1 | Beach)
mmod14:     exposure:salinity + exposure:humus + salinity:NAP + NAP
mmod13: Richness ~ exposure + salinity + NAP + humus + (1 | Beach)
mmod13:     exposure:salinity + exposure:humus + salinity:NAP + sal
mmod13:     NAP:humus
       Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
mmod14 11 55.568 75.441 -16.784
mmod13 12 56.711 78.391 -16.355 0.8567      1     0.3547
```

```
> summary(mmod14)
[..]
                   Estimate Std. Error t value
(Intercept)         68.3906    45.2987   1.510
exposure            -7.1479     4.4665  -1.600
salinity            -2.2983     1.5638  -1.470
NAP                  2.4125     3.0405   0.794
humus               33.2900    17.8258   1.867
exposure:salinity    0.2468     0.1546   1.596
exposure:humus      -3.0208     1.7998  -1.678
salinity:NAP        -0.1145     0.1088  -1.052
NAP:humus            5.1862     2.1676   2.393
[..]
```

```
> mmod15 <- update(mmod14,~.-salinity:NAP)
  0:     54.119609: 0.730297  64.2333 -6.76223 -2.13153 -
  1:     54.105967: 0.730289  64.2333 -6.76224 -2.13154 -
  2:     54.105023: 0.730191  64.2333 -6.76223 -2.13154 -
[...]
 33:     35.825219:  0.00000  64.2334 -6.76224 -2.13154 -
 34:     35.825219:  0.00000  64.2333 -6.76223 -2.13153 -
```

```
> anova(mmod15,mmod14)
Data: rikz
Models:
mmod15: Richness ~ exposure + salinity + NAP + humus + (1 | Beach)
mmod15:     exposure:salinity + exposure:humus + NAP:humus
mmod14: Richness ~ exposure + salinity + NAP + humus + (1 | Beach)
mmod14:     exposure:salinity + exposure:humus + salinity:NAP + NAP
       Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
mmod15 10 55.825 73.892 -17.913
mmod14 11 55.568 75.441 -16.784 2.2576      1     0.1330
```

```
> summary(mmod15)
[...]
                  Estimate Std. Error t value
(Intercept)        64.2333    45.8648   1.400
exposure           -6.7622     4.5261  -1.494
salinity           -2.1315     1.5821  -1.347
NAP                -0.7919     0.1650  -4.800
humus              26.3211    17.2232   1.528
exposure:salinity   0.2313     0.1566   1.477
exposure:humus     -2.3704     1.7606  -1.346
NAP:humus           4.1208     1.9210   2.145
[...]
```

```
> mmod16 <- update(mmod15,~.-exposure:humus)
  0:     58.266620: 0.730297  46.8934 -5.01697 -1.48258 -
  1:     58.254732: 0.730288  46.8934 -5.01698 -1.48259 -
  2:     58.235790: 0.728826  46.8934 -5.01694 -1.48244 -
  3:     46.756746: 0.120192  46.8961 -5.00220 -1.42242 -
[...]
 28:     39.633451:  0.00000  46.8961 -5.01723 -1.48267 -
 29:     39.633451:  0.00000  46.8945 -5.01709 -1.48262 -
 30:     39.633451:  0.00000  46.8935 -5.01699 -1.48259 -
```

```
> anova(mmod16,mmod15)
Data: rikz
Models:
mmod16: Richness ~ exposure + salinity + NAP + humus + (1 | Beach)
mmod16:     exposure:salinity + NAP:humus
mmod15: Richness ~ exposure + salinity + NAP + humus + (1 | Beach)
mmod15:     exposure:salinity + exposure:humus + NAP:humus
       Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
mmod16  9 57.633 73.893 -19.817
mmod15 10 55.825 73.892 -17.913 3.8082      1      0.051 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> summary(mmod16)
[...]
                   Estimate Std. Error t value
(Intercept)         46.8935    47.4863   0.988
exposure            -5.0170     4.6770  -1.073
salinity            -1.4826     1.6240  -0.913
NAP                 -0.7177     0.1653  -4.341
humus                3.2326     1.7443   1.853
exposure:salinity    0.1658     0.1604   1.034
NAP:humus            4.0391     2.0136   2.006
[...]
```

```
> mmod17 <- update(mmod16,~.-exposure:salinity)
  0:     58.637737: 0.730297 -2.14684 -0.183246 0.196099
  1:     58.598479: 0.729260 -2.14690 -0.183880 0.194558
  2:     58.564423: 0.727583 -2.14686 -0.183644 0.195548
[...]
 33:     42.290178: 8.33188e-06 -2.14646 -0.183254 0.1960
 34:     42.290178: 3.01290e-06 -2.14671 -0.183249 0.1960
 35:     42.290178: 2.25278e-09 -2.14684 -0.183246 0.1960
 36:     42.290178: 1.36106e-11 -2.14684 -0.183246 0.1960
```

```
> anova(mmod17,mmod16)
Data: rikz
Models:
mmod17: Richness ~ exposure + salinity + NAP + humus + (1
mmod17:     NAP:humus
mmod16: Richness ~ exposure + salinity + NAP + humus + (1
mmod16:     exposure:salinity + NAP:humus
       Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
mmod17  8 58.290 72.743 -21.145
mmod16  9 57.633 73.893 -19.817 2.6567      1     0.1031
```

```
> summary(mmod17)
[...]
            Estimate Std. Error t value
(Intercept)  -2.1468     4.0782  -0.526
exposure     -0.1832     0.1281  -1.430
salinity      0.1961     0.1137   1.725
NAP          -0.7116     0.1780  -3.997
humus         3.4743     1.8951   1.833
NAP:humus     3.8253     2.1777   1.757
[...]
```

```
> mmod18 <- update(mmod17,~.-exposure)
  0:     59.363947: 0.730297 -6.50861 0.283750 -0.782548
  1:     59.309792: 0.729344 -6.50868 0.281666 -0.782629
  2:     59.249973: 0.724342 -6.50859 0.283825 -0.783080
  3:     49.314348: 0.249654 -6.50743 0.281775 -0.821932
  4:     49.028180: 0.0874159 -6.50621 0.286181 -0.768039
  5:     47.878333: 0.0874274 -6.50630 0.283516 -0.768044
  6:     47.829181: 0.111395 -6.50597 0.283569 -0.776400
  7:     47.811523: 0.108337 -6.48757 0.282663 -0.780176
  8:     47.782569: 0.107177 -6.38537 0.279448 -0.780133
  9:     47.771312: 0.120894 -6.28677 0.276050 -0.770252
 10:     47.770238: 0.117932 -6.26689 0.275277 -0.772068
 11:     47.769283: 0.116689 -6.25496 0.274892 -0.775140
 12:     47.768895: 0.118127 -6.25212 0.274784 -0.773427
 13:     47.768886: 0.118026 -6.24913 0.274674 -0.773555
 14:     47.768886: 0.118044 -6.24792 0.274631 -0.773578
 15:     47.768885: 0.118049 -6.24780 0.274626 -0.773576
```

```
> anova(mmod18,mmod17)
Data: rikz
Models:
mmod18: Richness ~ salinity + NAP + humus + (1 | Beach) + NAP:humus
mmod17: Richness ~ exposure + salinity + NAP + humus + (1 | Beach)
mmod17:     NAP:humus
       Df    AIC    BIC  logLik  Chisq Chi Df Pr(>Chisq)
mmod18  7 61.769 74.416 -23.884
mmod17  8 58.290 72.743 -21.145 5.4787      1    0.01925 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> summary(mmod17)
[...]
            Estimate Std. Error t value
(Intercept)  -2.1468     4.0782  -0.526
exposure     -0.1832     0.1281  -1.430
salinity      0.1961     0.1137   1.725
NAP          -0.7116     0.1780  -3.997
humus         3.4743     1.8951   1.833
NAP:humus     3.8253     2.1777   1.757
[..]
```

```
> mmod19 <- update(mmod17,~.-NAP:humus)
  0:     69.899155: 0.730297 -0.528882 -0.288714 0.180644 -0.498037
  1:     69.858054: 0.729229 -0.528938 -0.289361 0.179067 -0.498187
  2:     69.822146: 0.727528 -0.528902 -0.289110 0.180093 -0.498425
  3:     69.752883: 0.723874 -0.528954 -0.289891 0.178653 -0.498916
  4:     59.543657: 0.331011 -0.528734 -0.314567 0.189047 -0.545780
  5:     56.406034: 0.224112 -0.528678 -0.320623 0.190064 -0.550253
  6:     54.889391: 0.136913 -0.528511 -0.324521 0.195036 -0.558595
  7:     54.771875: 0.0497220 -0.528591 -0.329548 0.191390 -0.56709
  8:     53.704231: 0.00628396 -0.528348 -0.325059 0.194014 -0.5708
  9:     52.557547: 0.0391547 -0.528090 -0.298380 0.183467 -0.56589
 10:     52.326076:  0.00000 -0.527933 -0.291355 0.181607 -0.563221
 11:     52.106386:  0.00000 -0.528402 -0.324021 0.191978 -0.491460
 12:     52.008303: 3.46690e-15 -0.528380 -0.323749 0.192565 -0.491
 13:     51.758028: 4.00663e-13 -0.528270 -0.314098 0.188878 -0.492
 14:     51.716750:  0.00000 -0.528142 -0.305373 0.187209 -0.488564
 15:     51.457747:  0.00000 -0.527829 -0.297049 0.183507 -0.489009
 16:     51.427700:  0.00000 -0.527500 -0.294752 0.182725 -0.496743
 17:     51.426304: 0.00277884 -0.527461 -0.294440 0.182564 -0.4964
 18:     51.418913:  0.00000 -0.527550 -0.291428 0.181696 -0.495694
 19:     51.411862:  0.00000 -0.527743 -0.288447 0.180484 -0.495154
```

```
> mmod20 <- update(mmod17,~.-salinity)
  0:     59.931609: 0.730297  4.64070 -0.307879 -0.705878  3.87577
  1:     59.686506: 0.720204  4.63974 -0.318464 -0.705170  3.87486
  2:     59.438573: 0.708236  4.64065 -0.310064 -0.704618  3.87391
  3:     58.974741: 0.681684  4.63965 -0.322306 -0.703439  3.87166
  4:     53.658984:  0.00000  4.65615 -0.299096 -0.648086  3.71811
  5:     51.878726: 6.21175e-08  4.65523 -0.308198 -0.648525  3.718
  6:     51.089349:  0.00000  4.66017 -0.308044 -0.704827  3.70271
  7:     51.085319:  0.00000  4.65088 -0.309370 -0.704784  3.74770
  8:     51.005801:  0.00000  4.64691 -0.307597 -0.704035  3.77021
  9:     50.974766:  0.00000  4.64425 -0.307696 -0.711910  3.79163
 10:     50.607459: 0.0432925  4.62618 -0.307252 -0.698582  3.96917
 11:     50.157784: 0.0897194  4.60979 -0.305835 -0.686628  4.13325
 12:     50.124822: 0.0902944  4.60961 -0.307876 -0.686690  4.13318
 13:     50.106719: 0.0905979  4.60987 -0.306959 -0.686892  4.13179
 14:     50.087485: 0.0919311  4.61015 -0.307911 -0.687372  4.12889
 15:     50.068168: 0.0915007  4.61097 -0.307049 -0.687990  4.12263
 16:     50.043262: 0.0918963  4.61239 -0.308146 -0.689342  4.11006
 17:     50.022777: 0.0885245  4.61654 -0.307269 -0.691585  4.08530
 18:     49.988336: 0.0991556  4.61353 -0.308123 -0.687550  4.13588
 19:     49.914042: 0.0956507  4.62255 -0.307547 -0.692348  4.08647
```

```
> mmod17
Generalized linear mixed model fit by the Laplace approximation
Formula: Richness ~ exposure + salinity + NAP + humus + (1 | Beach)
   Data: rikz
   AIC   BIC logLik deviance
 58.29 72.74 -21.15    42.29
Random effects:
 Groups   Name        Variance   Std.Dev.
 Beach    (Intercept) 5.2533e-22 2.2920e-11
 Residual             2.8358e+00 1.6840e+00
Number of obs: 45, groups: Beach, 9

Fixed effects:
            Estimate Std. Error t value
(Intercept)  -2.1468     4.0782  -0.526
exposure     -0.1832     0.1281  -1.430
salinity      0.1961     0.1137   1.725
NAP          -0.7116     0.1780  -3.997
humus         3.4743     1.8951   1.833
NAP:humus     3.8253     2.1777   1.757
```

In newer versions of lme4 a few things are different:

- *p*-values for GLMMs
- no “quasipoisson” or “quasibinomial” in GLMMs
- profiles for GLMMs
- mcmcsamp has been removed
- parametric bootstrapping

```
> mmod1 <- glmer(Richness~1+exposure+salinity+NAP+humus
+                 +(1 | Beach),family=poisson,data=rikz)
> summary(mmod1)
[...]
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.52888    2.36277  -0.224  0.82288
exposure    -0.28871    0.06894  -4.188 2.81e-05 ***
salinity     0.18064    0.06731   2.684  0.00728 **
NAP         -0.49804    0.07027  -7.087 1.37e-12 ***
humus        1.23952    0.82561   1.501  0.13327
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
[...]
```

How to get overdispersion into GLMMs

One possibility is to add a random factor whose groups are all the single observation:

```
> rikz$plot <- factor(1:45)
> mmod2 <- glmer(Richness~1+exposure+salinity+NAP+humus
+                 +(1 | Beach) +(1 | plot) ,
+                 family=poisson,data=rikz)
> summary(mmod2)
[...]
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.08937    2.70527   0.033 0.973648
exposure    -0.31359    0.08581  -3.654 0.000258 ***
salinity     0.16629    0.07605   2.187 0.028767 *
NAP         -0.54076    0.08655  -6.248 4.17e-10 ***
humus        1.45350    1.06765   1.361 0.173386
[...]
```

Another possibility is to use the *negative binomial* family, which can be considered as an instance of oberdispersed Poisson. The results are very similar:

```
> mmod1nb <- glmer.nb(Richness~1+exposure+salinity+NAP+humus
+                     +(1 | Beach), data=rikz)
> summary(mmod1nb)
Generalized linear mixed model fit by maximum likelihood ['summary.
 Family: Negative Binomial(21.4896) ( log )
[...]
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.08463    2.69141  -0.031 0.974915
exposure    -0.30929    0.08321  -3.717 0.000202 ***
salinity     0.17194    0.07605   2.261 0.023771 *
NAP         -0.53971    0.08338  -6.473 9.61e-11 ***
humus        1.41223    1.02911   1.372 0.169975
[...]
```

Confidence intervals can be computed with profiles (but not for glmer.nb results):

```
> pr1 <- profile(mmod1) ## ...takes VERY long
> confint(pr1)
                  2.5 %     97.5 %
.sig01               NA         NA
(Intercept) -5.19621036  4.1480336
exposure    -0.44103410 -0.1523536
salinity     0.04935443  0.3137404
NAP         -0.63701933 -0.3613148
humus       -0.41958988  2.8211782
```

Confidence intervalls can also be computed with parameteric bootstrapping, where “parametric” means that data a simulated according to model assumptions using the estimated parameter values.

```
> mySumm <- function(.) {
+   c(beta=fixef(.),sigma=sigma(.),sig01=unlist(VarCorr(.)))
+ }
> boot1 <- bootMer(mmod1, mySumm, nsim = 1000) ## Takes VERY long!
> require("boot")
> boot.ci(boot1, index=4, type=c("norm", "basic", "perc"))  ## NAP
[...]
Level      Normal              Basic              Percentile
95%   (-0.6409, -0.3595 )   (-0.6439, -0.3648 )   (-0.6313, -0.3521 )
Calculations and Intervals on Original Scale
> boot.ci(boot1, index=5, type=c("norm", "basic", "perc"))  ## humus
[...]
Level      Normal              Basic              Percentile
95%   (-0.364,  2.914 )   (-0.415,  2.917 )   (-0.438,  2.894 )
Calculations and Intervals on Original Scale
> boot.ci(boot1, index=7, type=c("norm", "basic", "perc"))  ## sig01.Beach
[...]
Level      Normal              Basic              Percentile
95%   (-0.0120,  0.0102 )   (-0.0098,  0.0000 )   ( 0.0000,  0.0098 )
Calculations and Intervals on Original Scale
> ## Negative values for sig01 do not make sense. Here, only the
> ## percentile methods seems to give reasonable results.
```

Contents

So far we have taken the perspective of frequentistic statistics, where parameters are assumed to be unknown but non-random.

So far we have taken the perspective of frequentistic statistics, where parameters are assumed to be unknown but non-random.

A key concept of frequentistic statistics is the likelihood, which is the probability of the observed data D as a function of the parameter value θ:

$$L(\theta) = \Pr_{\theta}(D)$$

So far we have taken the perspective of frequentistic statistics, where parameters are assumed to be unknown but non-random.

A key concept of frequentistic statistics is the likelihood, which is the probability of the observed data D as a function of the parameter value θ:

$$L(\theta) = \mathrm{Pr}_\theta(D)$$

Bayesian statistics is an alternative approach, in which parameters are assumed to be random.

So far we have taken the perspective of frequentistic statistics, where parameters are assumed to be unknown but non-random.

A key concept of frequentistic statistics is the likelihood, which is the probability of the observed data D as a function of the parameter value θ:

$$L(\theta) = \Pr_\theta(D)$$

Bayesian statistics is an alternative approach, in which parameters are assumed to be random. Parameter estimation is then done according to the posterior distribution, that is, the probability distribution of the parameters, conditioned on the given the data:

$$\Pr(\theta|D) = \frac{\Pr(\theta, D)}{\Pr(D)} = \frac{\Pr(D|\theta) \cdot \Pr(\theta)}{\Pr(D)}$$

To compute the posterior probability with the Bayes-Formula

$$\Pr(\theta|D) = \frac{\Pr(D|\theta) \cdot \Pr(\theta)}{\Pr(D)}$$

we need:

- The probability of D, given θ, which is, in frequentistic terms, the likelihood:

$$\Pr(D|\theta) = \Pr_\theta(D) = L(D)$$

- $\Pr(\theta)$, the prior probability of θ, which does not depend on the data but on prior knowledge and/or on subjective belief (which is a reason for criticism).
- $\Pr(D)$, the probability of the data, not for a given parameter values, but averaged over all parameter values weighted by their prior probability:

$$\Pr(D) = \int \Pr(D|\theta) \cdot \Pr(\theta)\, d\theta$$

$$\Pr(D) = \int \Pr(D|\theta) \cdot \Pr(\theta)\, d\theta$$

is usually hard to compute, but it is not necessary to do this, if the posterior probabilities of two candidate parameter values are to be compared:

$$\frac{\Pr(\theta_1|D)}{\Pr(\theta_2|D)} = \frac{\frac{\Pr(D|\theta_1)\cdot\Pr(\theta_1)}{\Pr(D)}}{\frac{\Pr(D|\theta_2)\cdot\Pr(\theta_2)}{\Pr(D)}} = \frac{\Pr(D|\theta_1) \cdot \Pr(\theta_1)}{\Pr(D|\theta_2) \cdot \Pr(\theta_2)}$$

$$\Pr(D) = \int \Pr(D|\theta) \cdot \Pr(\theta) \, d\theta$$

is usually hard to compute, but it is not necessary to do this, if the posterior probabilities of two candidate parameter values are to be compared:

$$\frac{\Pr(\theta_1|D)}{\Pr(\theta_2|D)} = \frac{\frac{\Pr(D|\theta_1)\cdot\Pr(\theta_1)}{\Pr(D)}}{\frac{\Pr(D|\theta_2)\cdot\Pr(\theta_2)}{\Pr(D)}} = \frac{\Pr(D|\theta_1) \cdot \Pr(\theta_1)}{\Pr(D|\theta_2) \cdot \Pr(\theta_2)}$$

Markov-Chain Monte Carlo (MCMC) methods, e.g. based on the Metropolis-Hastings algorithm need only such pairwise comparisons and generate a sample of possible parameter values (approximately) according to the posterior probability distribution.

$$\Pr(D) = \int \Pr(D|\theta) \cdot \Pr(\theta)\, d\theta$$

is usually hard to compute, but it is not necessary to do this, if the posterior probabilities of two candidate parameter values are to be compared:

$$\frac{\Pr(\theta_1|D)}{\Pr(\theta_2|D)} = \frac{\frac{\Pr(D|\theta_1)\cdot\Pr(\theta_1)}{\Pr(D)}}{\frac{\Pr(D|\theta_2)\cdot\Pr(\theta_2)}{\Pr(D)}} = \frac{\Pr(D|\theta_1) \cdot \Pr(\theta_1)}{\Pr(D|\theta_2) \cdot \Pr(\theta_2)}$$

Markov-Chain Monte Carlo (MCMC) methods, e.g. based on the Metropolis-Hastings algorithm need only such pairwise comparisons and generate a sample of possible parameter values (approximately) according to the posterior probability distribution.(But there is no way to avoid specifying prior distributions for the parameters.)

Concepts from frequentistic statistics:

- statistical test
- likelihood
- confidence interval

Concepts from frequentistic statistics:

- statistical test
- likelihood
- confidence interval

Concepts from Bayesian statistics:

- prior and posterior probability of parameter values
- maximum a posteriory value
- expected value of posterior distribution
- Bayesian sampling
- credibility interval (also: credible interval)

The R package MCMCglmm provides a Bayesian approach of fitting (generalized) linear mixed models. Examples are given in file mcmcglmm.R.

The R package MCMCglmm provides a Bayesian approach of fitting (generalized) linear mixed models. Examples are given in file mcmcglmm.R.

If no priors are specified by the user, mcmcglmm assumes improper priors, that is, it assumes that the prior probability density is constant.

The R package MCMCglmm provides a Bayesian approach of fitting (generalized) linear mixed models. Examples are given in file mcmcglmm.R.

If no priors are specified by the user, mcmcglmm assumes improper priors, that is, it assumes that the prior probability density is constant.

This can be done in some Bayesian analyses, even though there is not really a probability distribution that is constant on all real numbers (or on all positive real numbers); that’s why it is called “improper”.

The R package MCMCglmm provides a Bayesian approach of fitting (generalized) linear mixed models. Examples are given in file mcmcglmm.R.

If no priors are specified by the user, mcmcglmm assumes improper priors, that is, it assumes that the prior probability density is constant.

This can be done in some Bayesian analyses, even though there is not really a probability distribution that is constant on all real numbers (or on all positive real numbers); that's why it is called "improper".

Note also that uniform priors for continuous random variables become non-uniform if the parameters are non-linearly rescaled, e.g. log transformed.

The R package MCMCglmm provides a Bayesian approach of fitting (generalized) linear mixed models. Examples are given in file mcmcglmm.R.

If no priors are specified by the user, mcmcglmm assumes improper priors, that is, it assumes that the prior probability density is constant.

This can be done in some Bayesian analyses, even though there is not really a probability distribution that is constant on all real numbers (or on all positive real numbers); that's why it is called "improper".

Note also that uniform priors for continuous random variables become non-uniform if the parameters are non-linearly rescaled, e.g. log transformed.

For this and other reasons, I find it misleading to call uniform priors "uniformative" or "uninformed" and do not recommend to do so.

If you want to apply Bayesian methods, you need to think about your priors. (And you may only be able to convince others if their priors are not to dissimilar to yours).

If you want to apply Bayesian methods, you need to think about your priors. (And you may only be able to convince others if their priors are not to dissimilar to yours).

The mcmcglmm packages allows you to use Inverse Wishart distributions as priors.

If you want to apply Bayesian methods, you need to think about your priors. (And you may only be able to convince others if their priors are not to dissimilar to yours).

The mcmcglmm packages allows you to use Inverse Wishart distributions as priors.

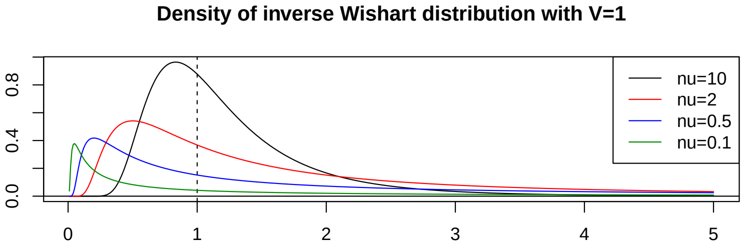
Inverse Wishart distributions have two parameters, their expectation value V and a dispersion parameter ν.

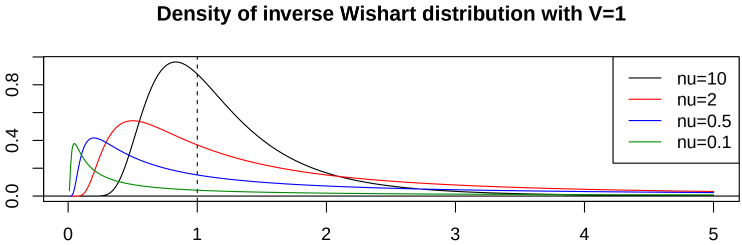
If you want to apply Bayesian methods, you need to think about your priors. (And you may only be able to convince others if their priors are not to dissimilar to yours).

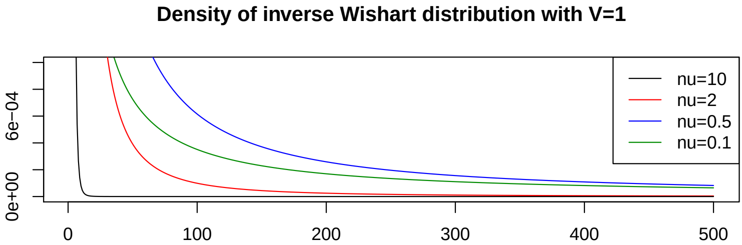
The mcmcglmm packages allows you to use Inverse Wishart distributions as priors.

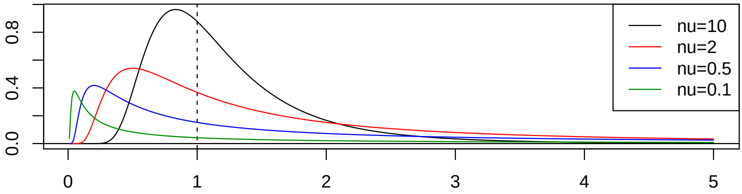
Inverse Wishart distributions have two parameters, their expectation value V and a dispersion parameter ν.

Note that, especially for small ν the mode of an Inverse Wishart distribution can be much smaller than its expectation value V, and this is compensated by a long right tail of the distribution.

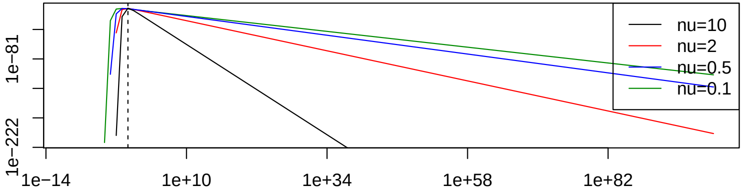

Density of inverse Wishart distribution with V=1
0.0
0.4
0.8
0
1
2
3
4
5
nu=10
nu=2
nu=0.5
nu=0.1


Density of inverse Wishart distribution with V=1
0.0
0.4
0.8
0
1
2
3
4
5
nu=10
nu=2
nu=0.5
nu=0.1

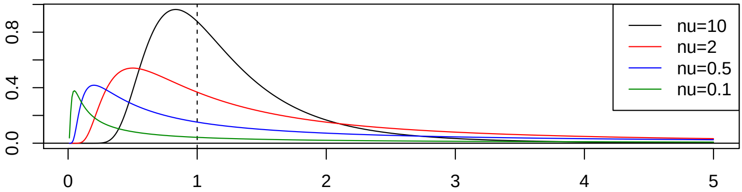

Density of inverse Wishart distribution with V=1
0e+00
6e−04
0
100
200
300
400
500
nu=10
nu=2
nu=0.5
nu=0.1



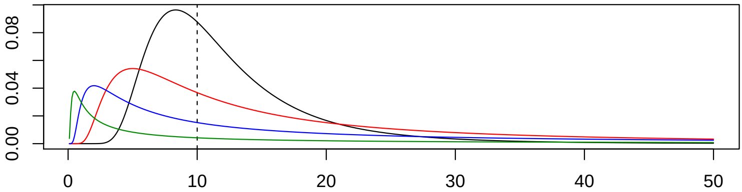
Density of inverse Wishart distribution with V=1



Density of inverse Wishart distribution with V=1



Density of inverse Wishart distribution with V=1



Density of inverse Wishart distribution with V=10