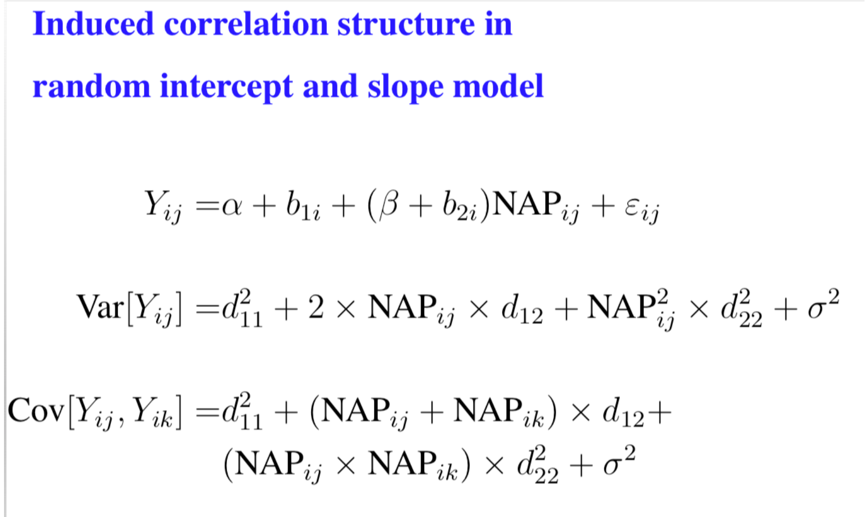
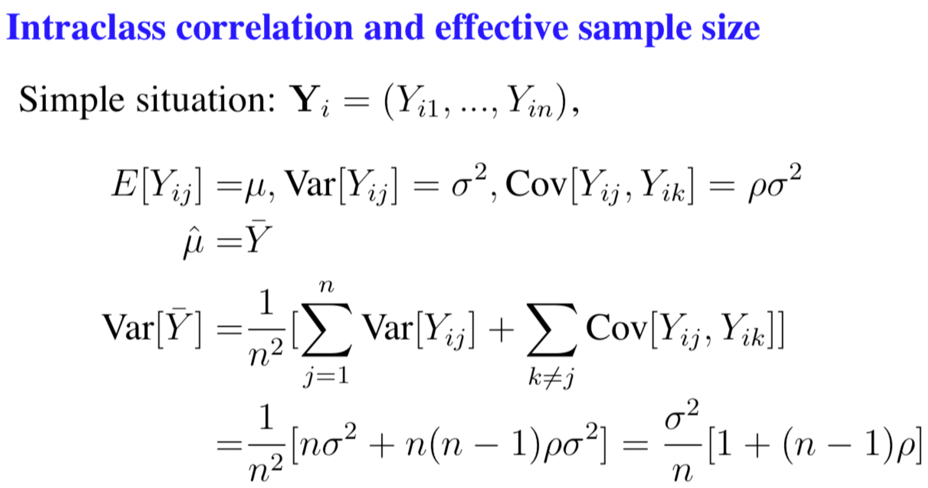
Mixed models can get very technical



Induced correlation structure in random intercept and slope model

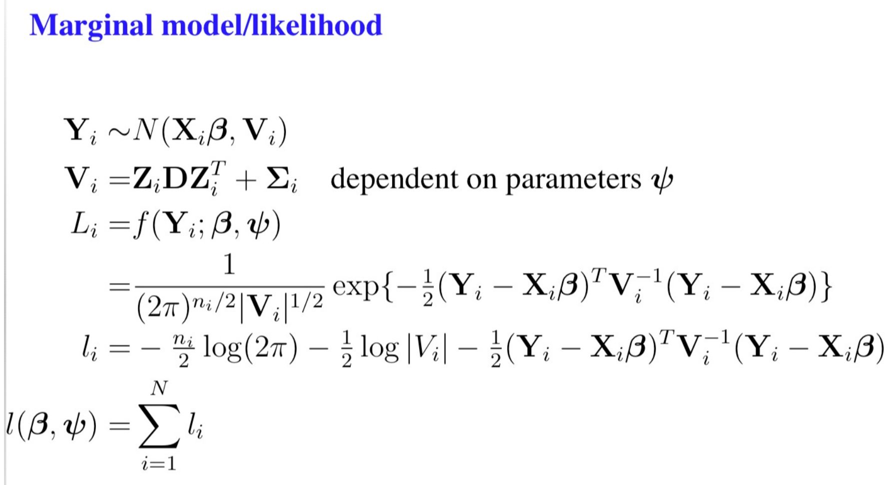
$$
\begin{aligned}
Y_{ij} &= \alpha + b_{1i} + (\beta + b_{2i})\text{NAP}_{ij} + \varepsilon_{ij} \\
\text{Var}[Y_{ij}] &= d_{11}^2 + 2 \times \text{NAP}_{ij} \times d_{12} + \text{NAP}_{ij}^2 \times d_{22}^2 + \sigma^2 \\
\text{Cov}[Y_{ij}, Y_{ik}] &= d_{11}^2 + (\text{NAP}_{ij} + \text{NAP}_{ik}) \times d_{12} + \\
&\qquad (\text{NAP}_{ij} \times \text{NAP}_{ik}) \times d_{22}^2 + \sigma^2
\end{aligned}
$$



Intraclass correlation and effective sample size

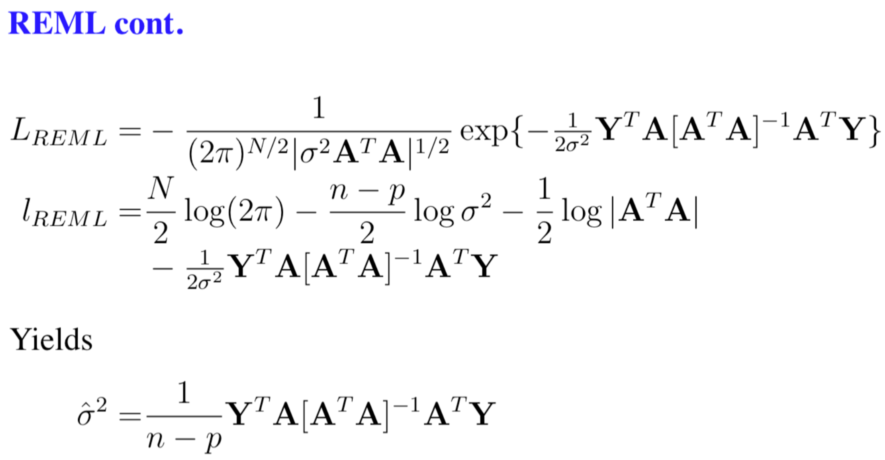
Simple situation: $\mathbf{Y}_i = (Y_{i1}, ..., Y_{in})$,

$$
\begin{aligned}
E[Y_{ij}] &= \mu, \text{Var}[Y_{ij}] = \sigma^2, \text{Cov}[Y_{ij}, Y_{ik}] = \rho\sigma^2 \\
\hat{\mu} &= \bar{Y} \\
\text{Var}[\bar{Y}] &= \frac{1}{n^2}\Big[\sum_{j=1}^{n} \text{Var}[Y_{ij}] + \sum_{k \neq j} \text{Cov}[Y_{ij}, Y_{ik}]\Big] \\
&= \frac{1}{n^2}[n\sigma^2 + n(n-1)\rho\sigma^2] = \frac{\sigma^2}{n}[1 + (n-1)\rho]
\end{aligned}
$$



Marginal model/likelihood

$$
\begin{aligned}
\mathbf{Y}_i &\sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i) \\
\mathbf{V}_i &= \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i \quad \text{dependent on parameters } \boldsymbol{\psi} \\
L_i &= f(\mathbf{Y}_i; \boldsymbol{\beta}, \boldsymbol{\psi}) \\
&= \frac{1}{(2\pi)^{n_i/2}|\mathbf{V}_i|^{1/2}} \exp\{-\tfrac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})\} \\
l_i &= -\tfrac{n_i}{2}\log(2\pi) - \tfrac{1}{2}\log|V_i| - \tfrac{1}{2}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) \\
l(\boldsymbol{\beta}, \boldsymbol{\psi}) &= \sum_{i=1}^{N} l_i
\end{aligned}
$$



REML cont.

$$
\begin{aligned}
L_{REML} &= -\frac{1}{(2\pi)^{N/2}|\sigma^2\mathbf{A}^T\mathbf{A}|^{1/2}} \exp\{-\tfrac{1}{2\sigma^2}\mathbf{Y}^T\mathbf{A}[\mathbf{A}^T\mathbf{A}]^{-1}\mathbf{A}^T\mathbf{Y}\} \\
l_{REML} &= \frac{N}{2}\log(2\pi) - \frac{n-p}{2}\log\sigma^2 - \frac{1}{2}\log|\mathbf{A}^T\mathbf{A}| \\
&\quad - \tfrac{1}{2\sigma^2}\mathbf{Y}^T\mathbf{A}[\mathbf{A}^T\mathbf{A}]^{-1}\mathbf{A}^T\mathbf{Y}
\end{aligned}
$$

Yields

$$
\hat{\sigma}^2 = \frac{1}{n-p}\mathbf{Y}^T\mathbf{A}[\mathbf{A}^T\mathbf{A}]^{-1}\mathbf{A}^T\mathbf{Y}
$$

https://www.uio.no/studier/emner/matnat/math/STK3100/h14/lectures/lecture8.pdf

RECAP: Zuur et al. (2007) used marine benthic data from **nine inter-tidal areas** along the Dutch coast collected by the RIKZ institute (summer of 2002).

In **each intertidal zone** (zone where ocean meets land; denoted by ‘beach’), five samples were taken, and the macro-fauna and abiotic variables were measured.

The **FINAL** goal is to model how species richness change as a function of **NAP** (Normal Amsterdam Level: the height of a sampling station compared to mean tidal level) and **Exposure** –- a nominal index for the entire beach (high/low) composed of the following elements: wave action, length of the surf zone, slope, grain size, and the depth of the anaerobic layer.

$$R_{ij} = b_0 + b_1 \times \underset{\downarrow}{NAP_{ij}} + b_2 \times \underset{\downarrow}{Exposure_j} + e_{ij}$$

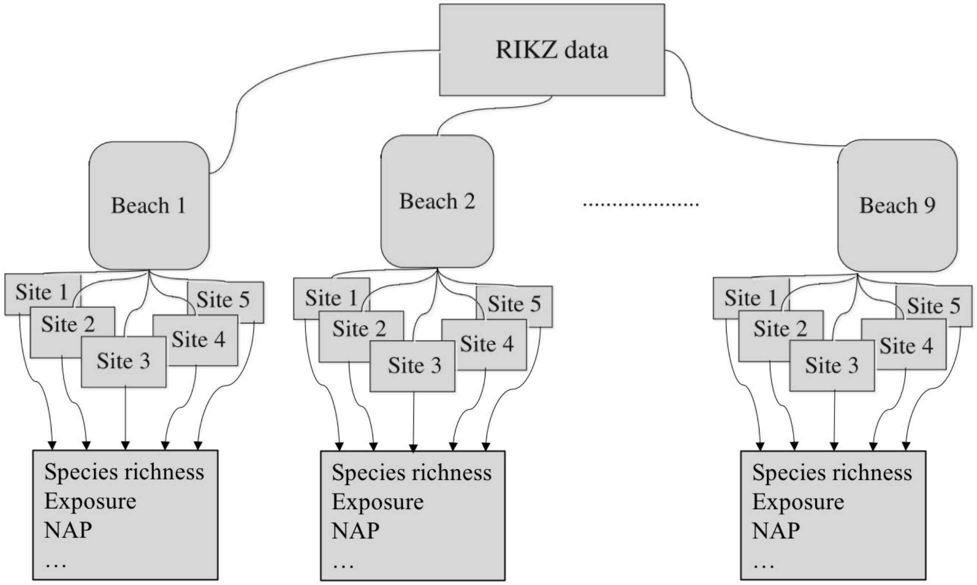
Each site for each beach has a NAP value

One value per beach

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$i = \text{sites};$$

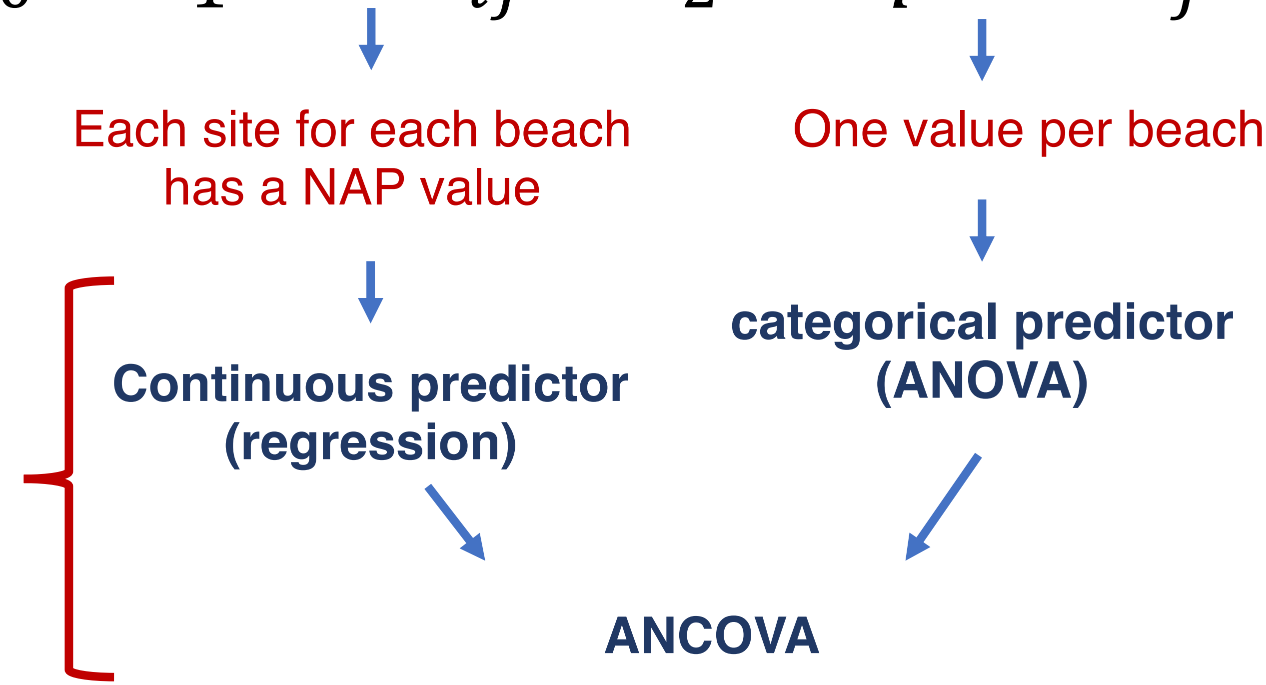
$$j = \text{beach}$$



Zuur AF, Ieno EN, Smith GM (2007) Analysing Ecological Data. Springer.

As we will see, once we consider the hierarchical nature of data, different models can be set and made compete to describe the same set of data

$$R_{ij} = b_0 + b_1 \times NAP_{ij} + b_2 \times Exposure_j + e_{ij}$$



Each site for each beach has a NAP value

One value per beach

Continuous predictor (regression)

categorical predictor (ANOVA)

Fixed versus Mixed effect models

ANCOVA

You may not be able to see it right now, but there are 10 or more possible linear models (covered in the tutorial) for these data; and one of them will best describe the data (i.e., best at predicting Richness)

Let’s concentrate on NAP for now particularly because it changes from beach to beach whereas Exposure does not (i.e., NAP may have a hierarchical dependence and exposure not)

$$R_{ij} = b_0 + b_1 \times \underset{\big\downarrow}{NAP_{ij}} + b_2 \times \underset{\big\downarrow}{Exposure_j} + e_{ij}$$

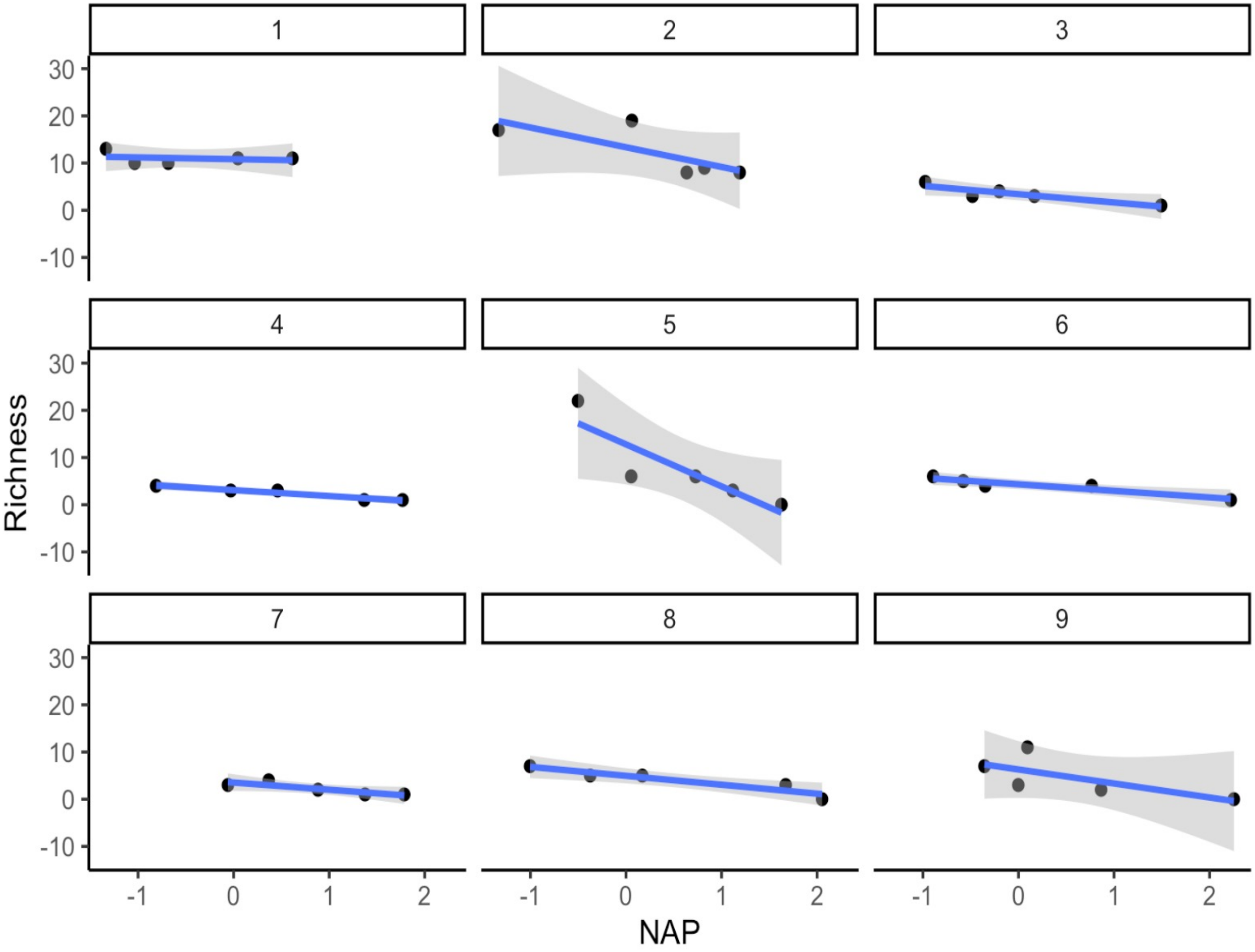
Each site for each beach has a NAP value

One value per beach

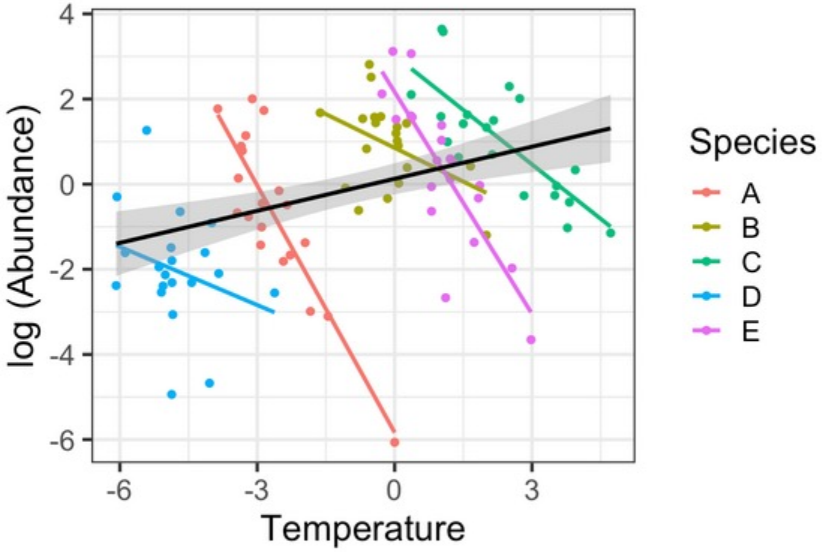
$$R_{ij} = b_0 + b_1 \times NAP_{ij} + e_{ij} \quad \Leftarrow$$

Regression lines (models) of Richness on NAP: they change from beach to beach (as we saw in the last lecture)

$$R_{ij} = b_0 + b_1 \times NAP_{ij} + e_{ij}$$

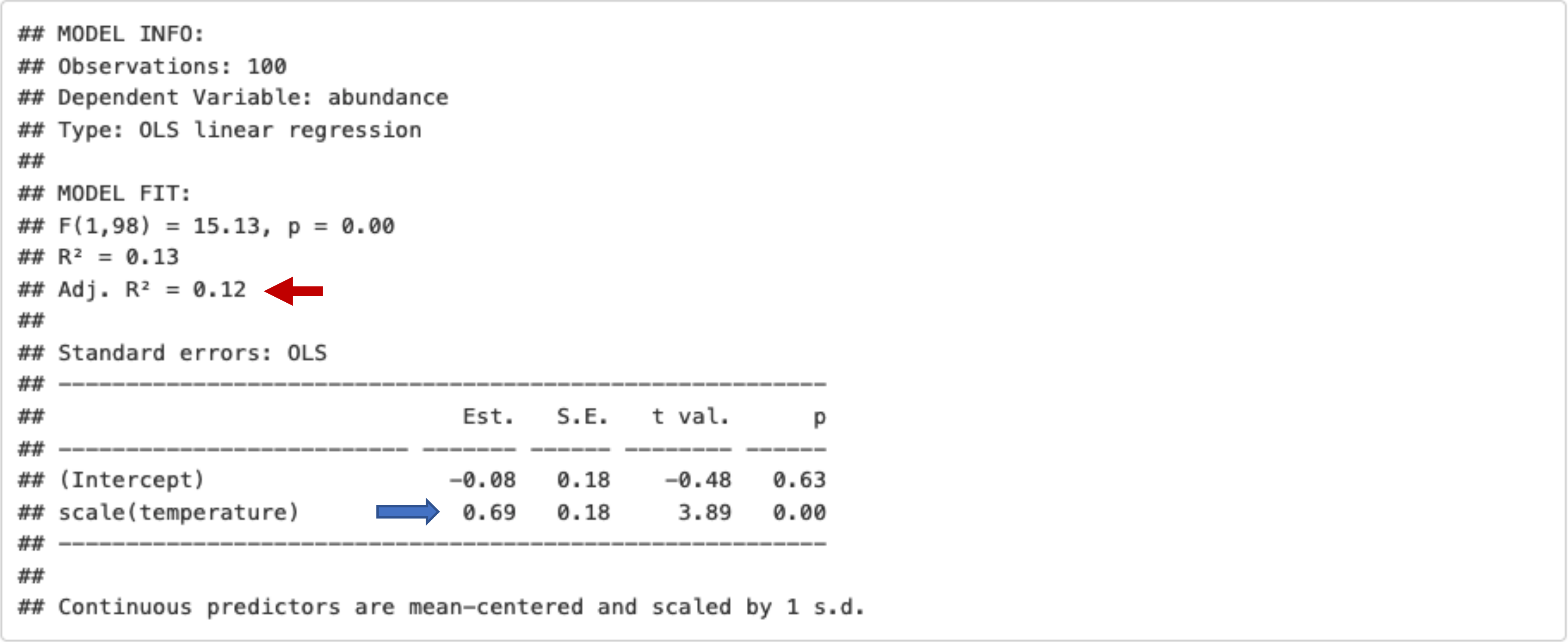


Interpretation of fixed *versus* mixed effect model

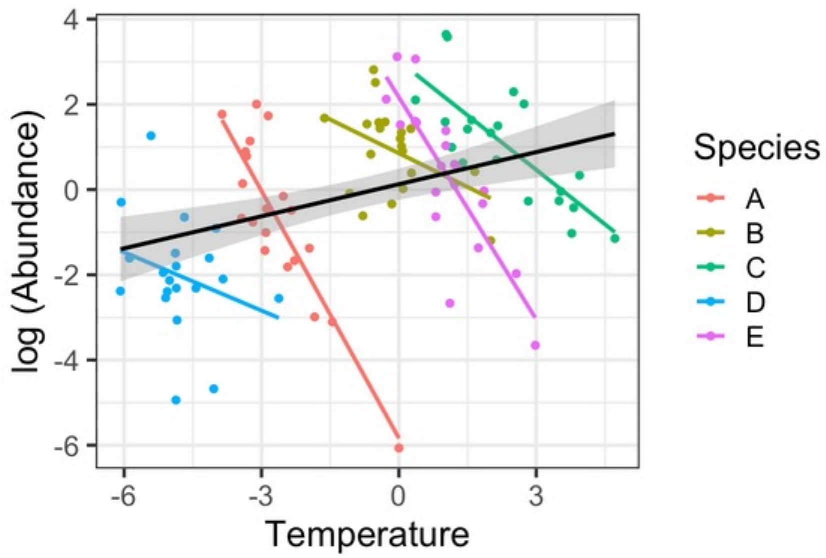


Fixed effect model

(results are LESS likely to apply to predict other species)



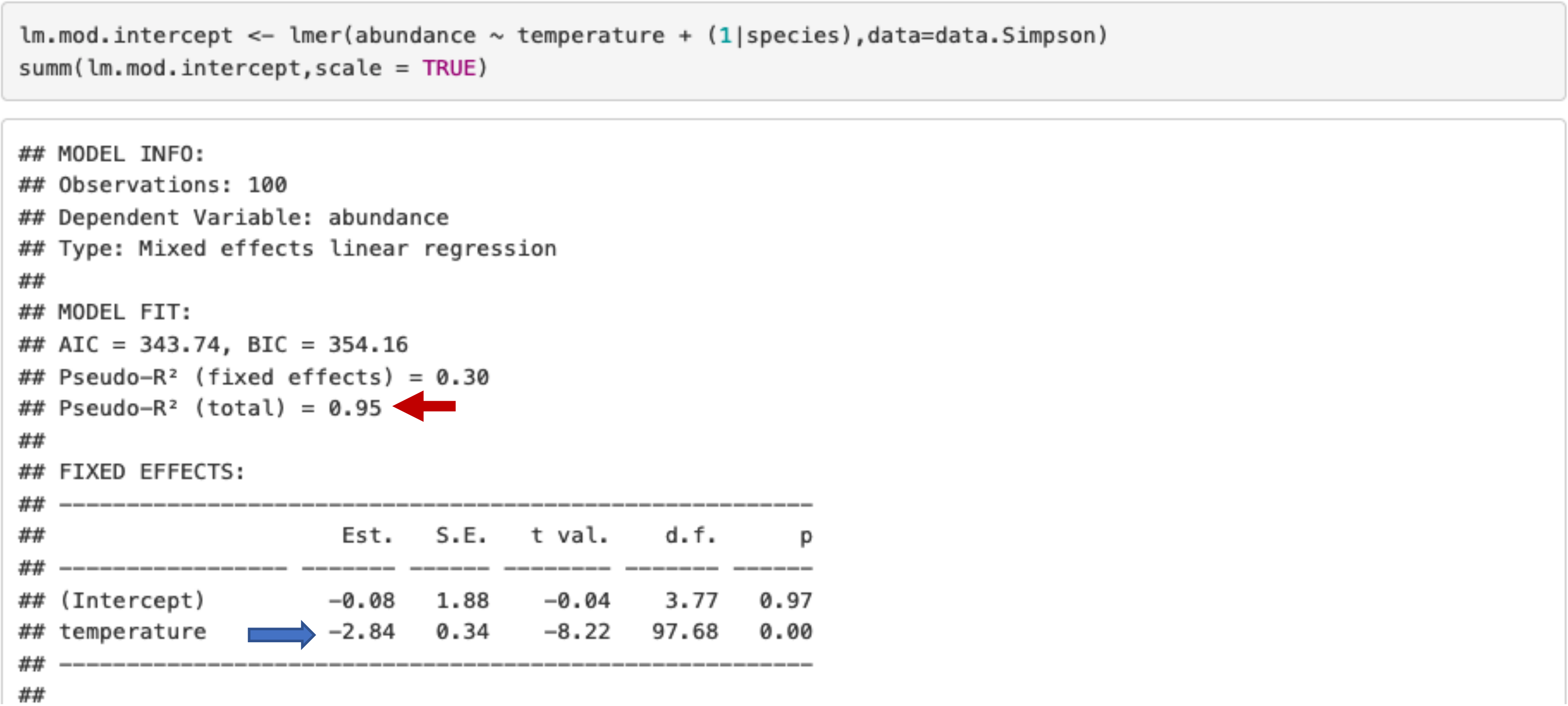
```
## MODEL INFO:
## Observations: 100
## Dependent Variable: abundance
## Type: OLS linear regression
##
## MODEL FIT:
## F(1,98) = 15.13, p = 0.00
## R² = 0.13
## Adj. R² = 0.12
##
## Standard errors: OLS
## ------------------------------------------------------------
##                               Est.   S.E.   t val.      p
## -------------------------- -------- ------ -------- ------
## (Intercept)                  -0.08   0.18    -0.48   0.63
## scale(temperature)            0.69   0.18     3.89   0.00
## ------------------------------------------------------------
##
## Continuous predictors are mean-centered and scaled by 1 s.d.
```



Interpretation of fixed *versus* mixed effect model

Mixed effect model

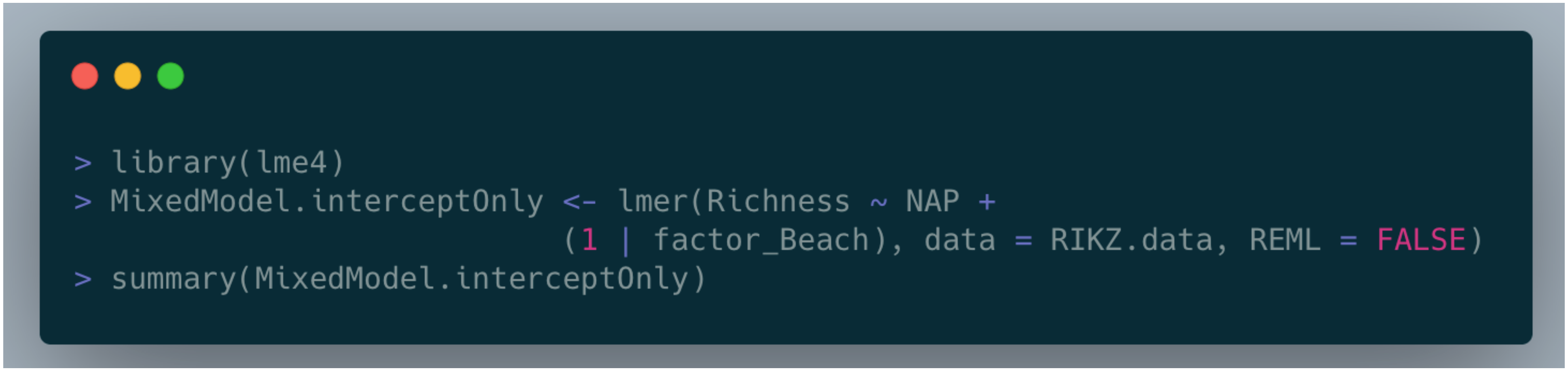
(results are MORE likely to apply to predict other species)



```
lm.mod.intercept <- lmer(abundance ~ temperature + (1|species),data=data.Simpson)
summ(lm.mod.intercept,scale = TRUE)
```

```
## MODEL INFO:
## Observations: 100
## Dependent Variable: abundance
## Type: Mixed effects linear regression
##
## MODEL FIT:
## AIC = 343.74, BIC = 354.16
## Pseudo-R² (fixed effects) = 0.30
## Pseudo-R² (total) = 0.95
##
## FIXED EFFECTS:
## --------------------------------------------------------
##                      Est.   S.E.   t val.    d.f.      p
## ----------------- ------- ------ -------- ------- ------
## (Intercept)         -0.08   1.88    -0.04    3.77   0.97
## temperature         -2.84   0.34    -8.22   97.68   0.00
## --------------------------------------------------------
##
```

RANDOM INTERCEPT MODEL: assumes a common slope and allow intercept to vary; IS THIS A GOOD MODEL for these data?

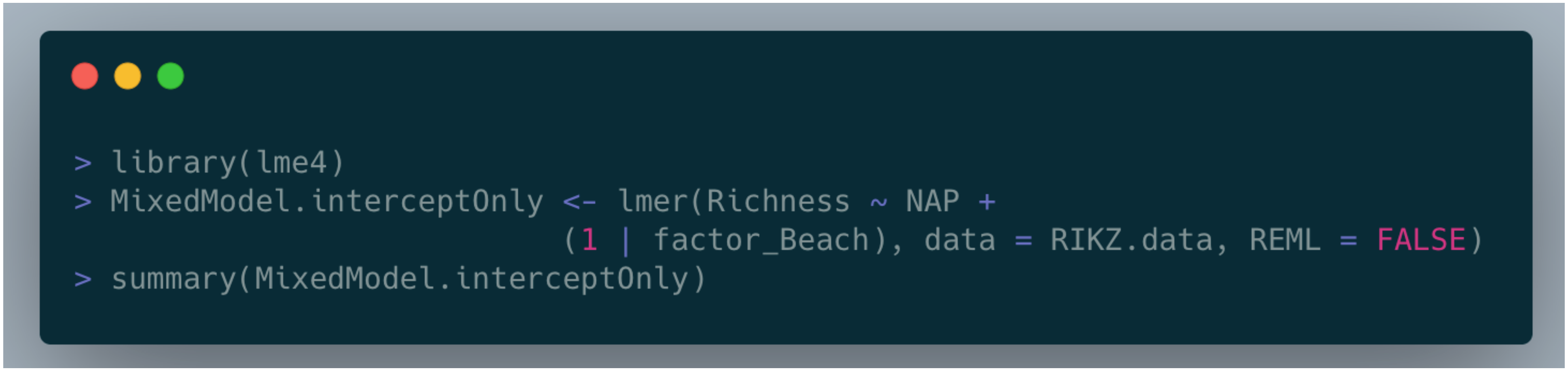


```
> library(lme4)
> MixedModel.interceptOnly <- lmer(Richness ~ NAP +
                           (1 | factor_Beach), data = RIKZ.data, REML = FALSE)
> summary(MixedModel.interceptOnly)
```

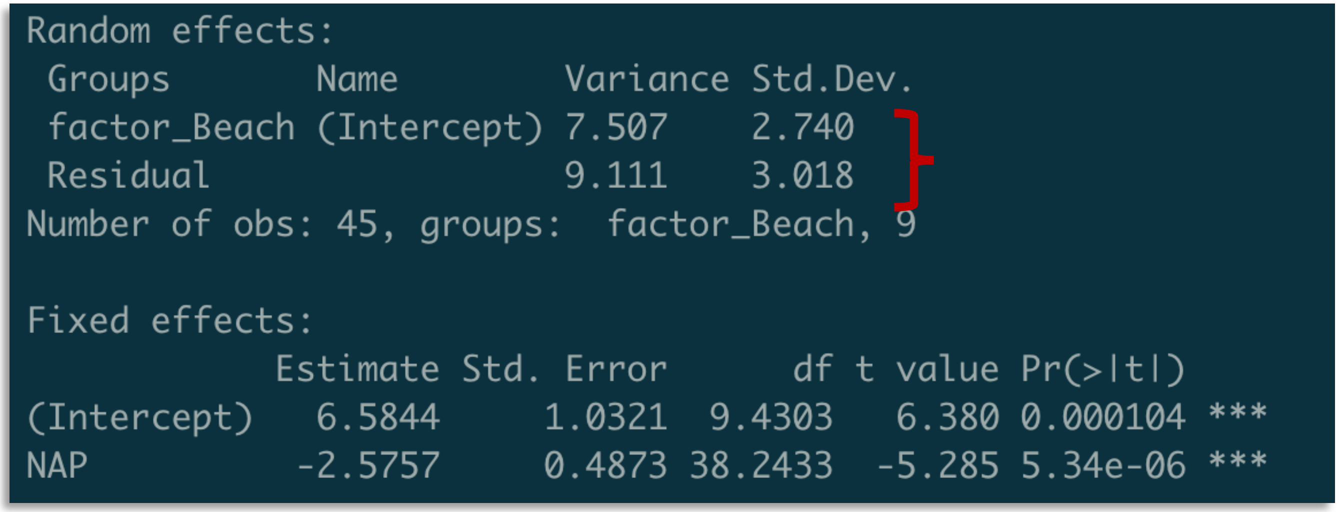
NAP is the fixed predictor of interest.

(1 | factor_Beach) is the random effect term, where the **1** denotes this is a random-intercept model and the term on the right of **|** is a factor to be used as the random effect. The factor here is beach, i.e., we are nesting sites within beaches to form the random effect.

RANDOM INTERCEPT MODEL: assumes a common slope and allow intercept to vary; IS THIS A GOOD MODEL for these data?



```
> library(lme4)
> MixedModel.interceptOnly <- lmer(Richness ~ NAP +
                         (1 | factor_Beach), data = RIKZ.data, REML = FALSE)
> summary(MixedModel.interceptOnly)
```

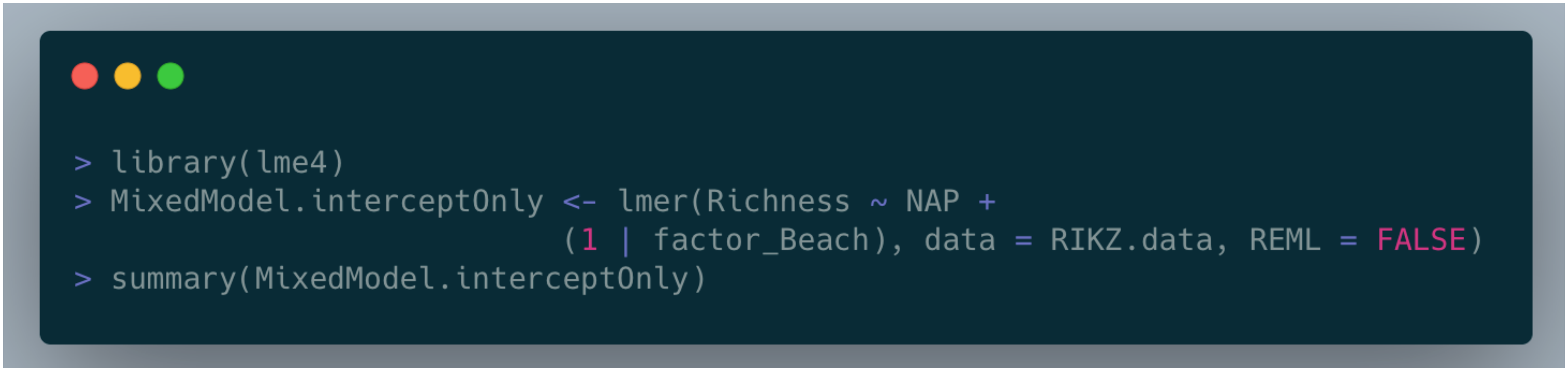


```
Random effects:
 Groups       Name        Variance Std.Dev.
 factor_Beach (Intercept) 7.507    2.740
 Residual                 9.111    3.018
Number of obs: 45, groups:  factor_Beach, 9

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   6.5844     1.0321  9.4303   6.380 0.000104 ***
NAP          -2.5757     0.4873 38.2433  -5.285 5.34e-06 ***
```

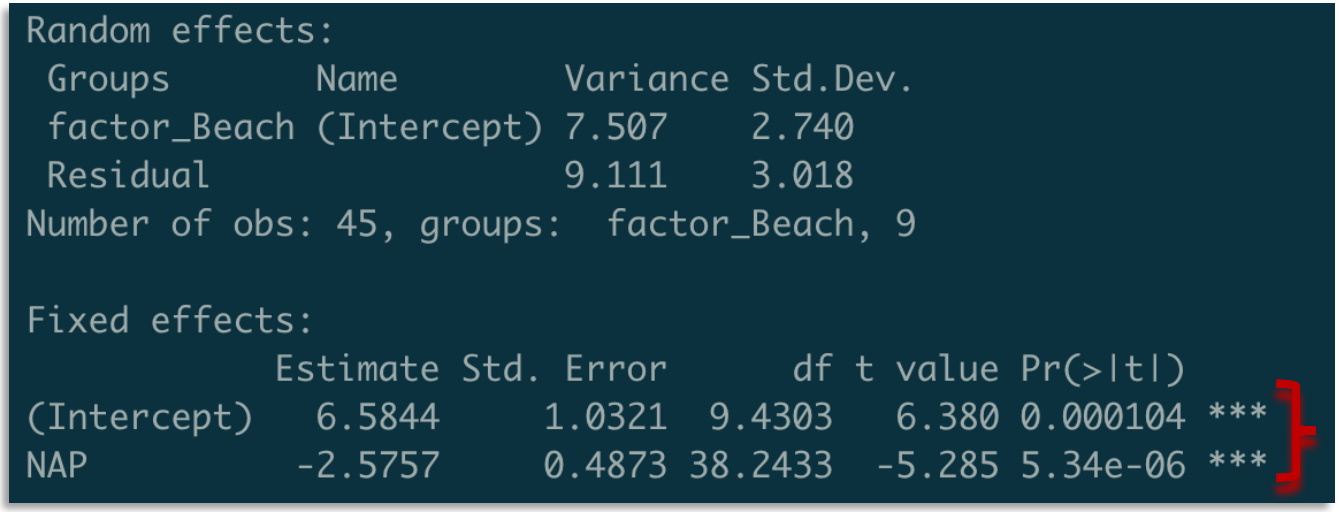
Random effect components: effect due to variation in **Intercepts** among beaches and **residuals** of the random component.

This mixed model have two sets of residuals (random and fixed). The parameters (variance) in these two sets are called hyperparameters (i.e., set by by the way that the study was designed and not likely generalizable by the model to other similar systems).

RANDOM INTERCEPT MODEL: assumes a common slope and allow intercept to vary; IS THIS A GOOD MODEL for these data?



```
> library(lme4)
> MixedModel.interceptOnly <- lmer(Richness ~ NAP +
                                (1 | factor_Beach), data = RIKZ.data, REML = FALSE)
> summary(MixedModel.interceptOnly)
```

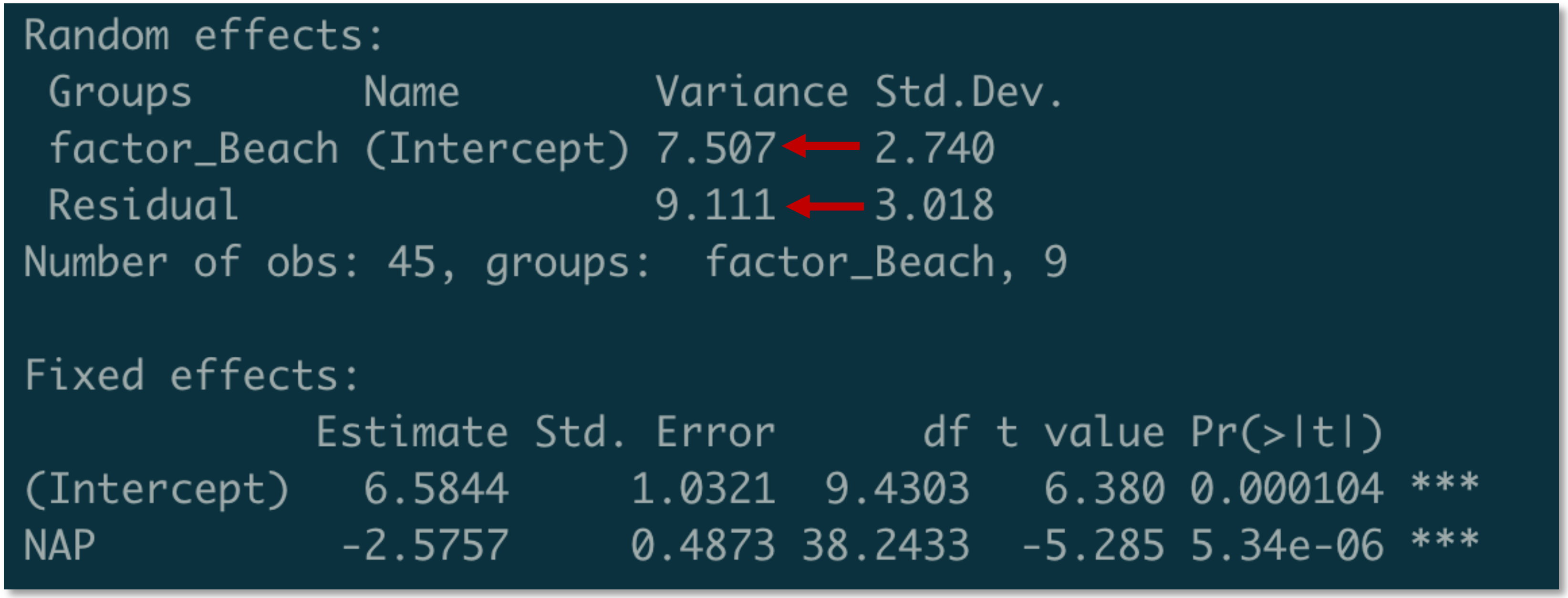


```
Random effects:
 Groups       Name        Variance Std.Dev.
 factor_Beach (Intercept) 7.507    2.740
 Residual                 9.111    3.018
Number of obs: 45, groups:  factor_Beach, 9

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   6.5844     1.0321  9.4303   6.380 0.000104 ***
NAP          -2.5757     0.4873 38.2433  -5.285 5.34e-06 ***
```

Fixed effect components: global (across all data) intercept and slope.

This part is more likely to be generalized to other systems as the design component (random effects) were used to estimate the fixed effect.

RANDOM INTERCEPT MODEL: assumes a common slope and allow intercept to vary; IS THIS A GOOD MODEL for these data?



```
Random effects:
 Groups       Name        Variance Std.Dev.
 factor_Beach (Intercept) 7.507 ←  2.740
 Residual                 9.111 ←  3.018
Number of obs: 45, groups:  factor_Beach, 9

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   6.5844     1.0321  9.4303   6.380 0.000104 ***
NAP          -2.5757     0.4873 38.2433  -5.285 5.34e-06 ***
```

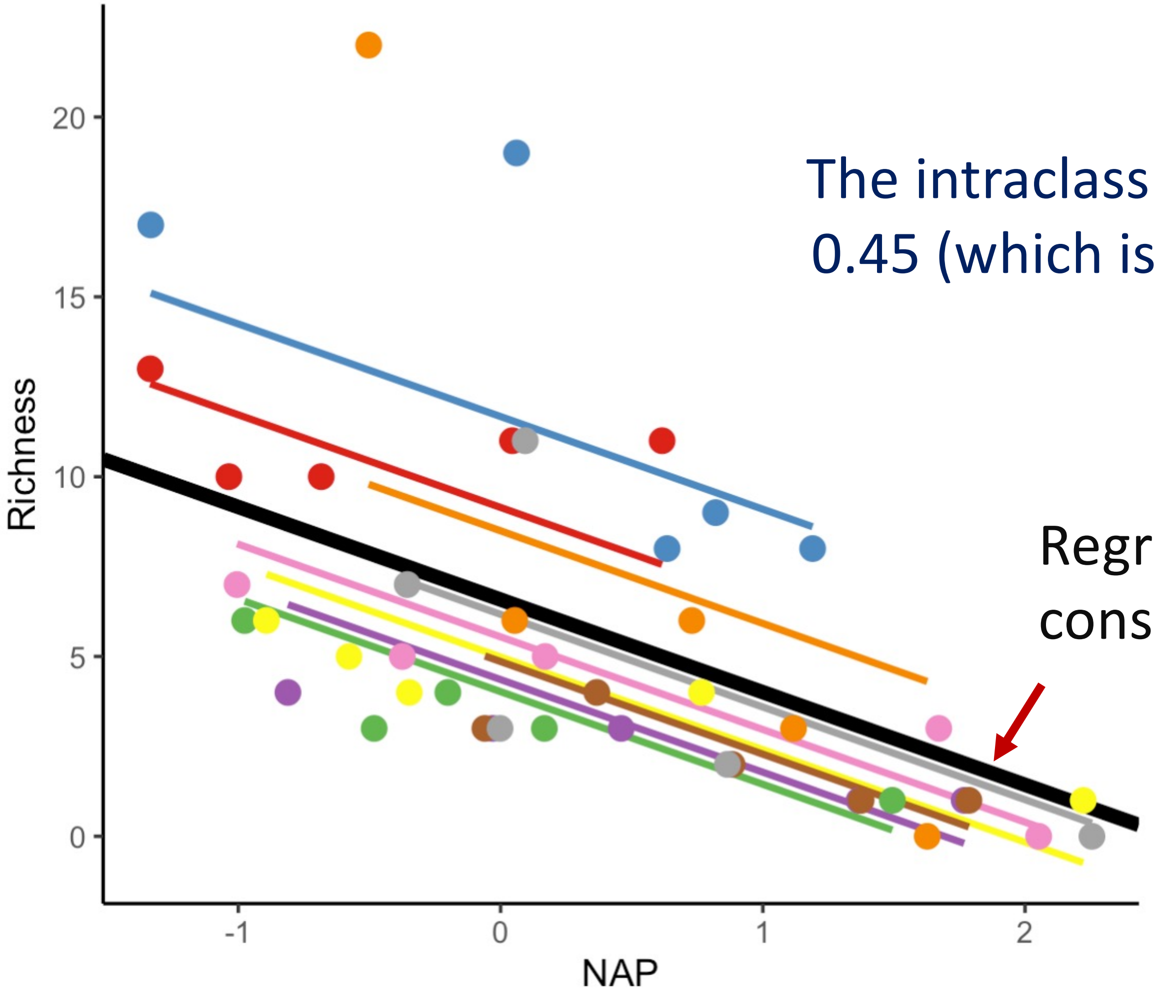
The intraclass correlation (ICC or Rho) here describes how strongly variation in predicted values within the same beach resemble each other.

ICC = 7.507 / (7.507 + 9.111) = 0.45 (which is pretty high, does indicating strong hierarchical structure in the data)

The RANDOM INTERCEPT MODEL

The intraclass correlation (ICC or Rho) here describes how strongly variation in predicted values within the same beach resemble each other

(note how predicted values are more similar within than among beaches).

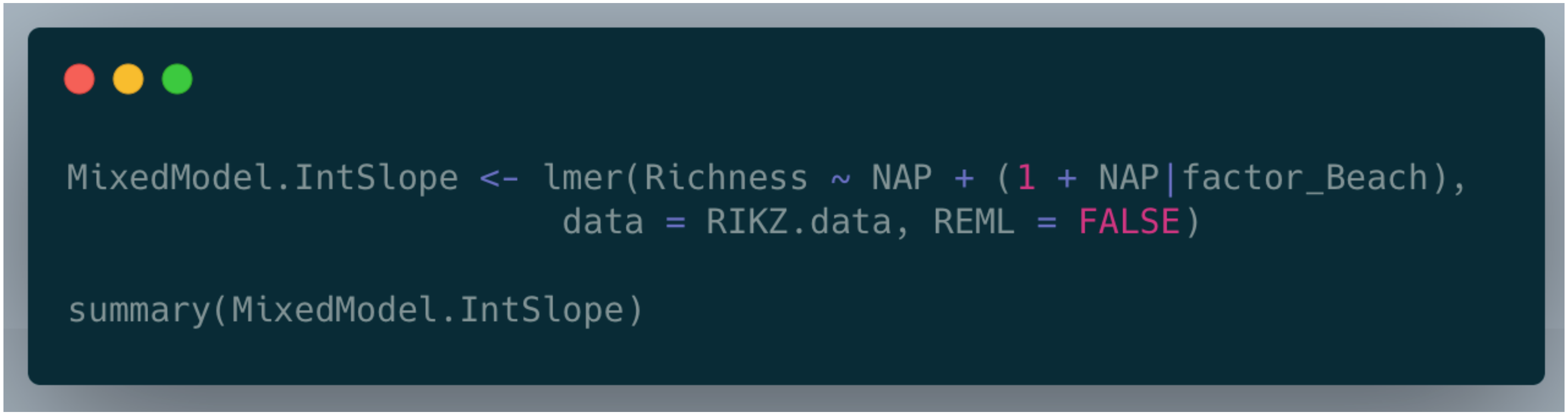


The intraclass correlation: 0.45 (which is pretty high)

Regression line considering all data

RANDOM INTERCEPT AND SLOPE MODEL: intercepts and slopes are allowed to vary

IS THIS A GOOD MODEL for these data?



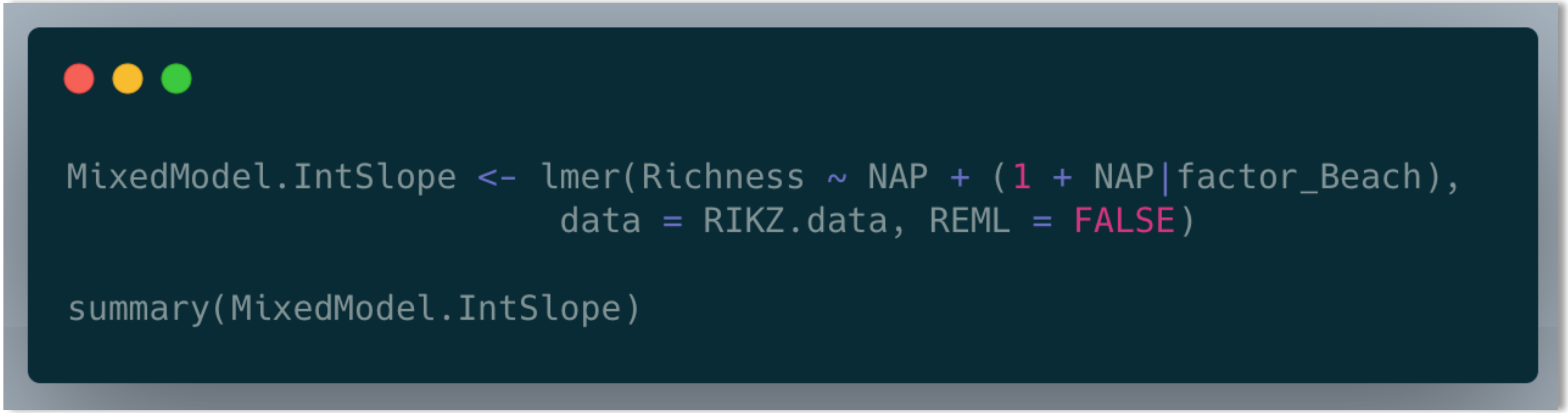
```r
MixedModel.IntSlope <- lmer(Richness ~ NAP + (1 + NAP|factor_Beach),
                        data = RIKZ.data, REML = FALSE)

summary(MixedModel.IntSlope)
```

NAP is the fixed predictor of interest.

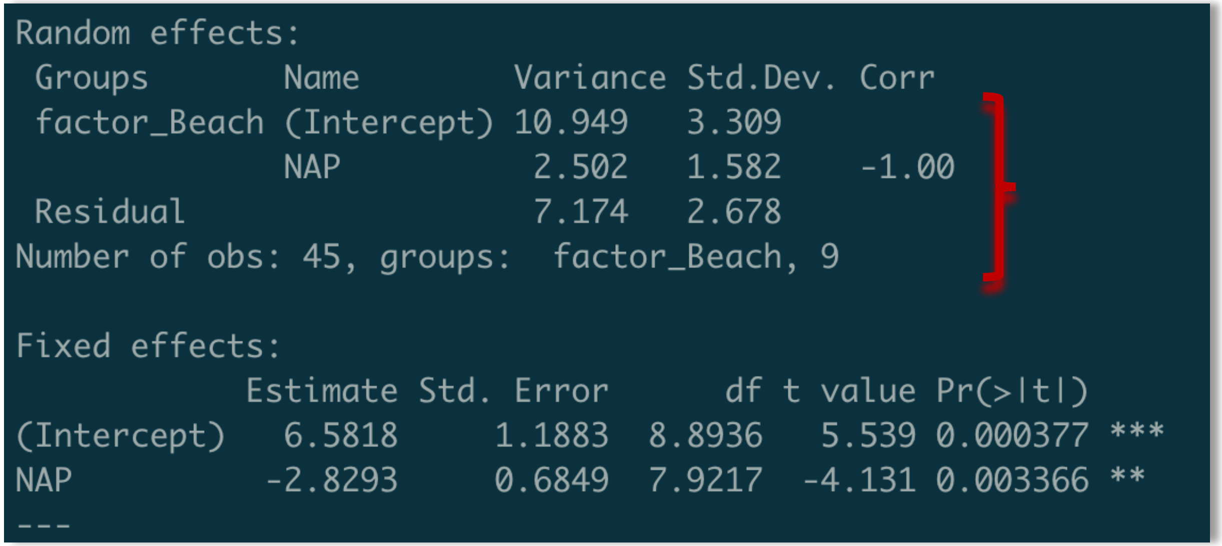
(1 + NAP | factor_Beach) is the random effect term, where the **1** denotes that we should consider variation in intercepts and also variation in slopes of NAP among beaches, i.e., **NAP | factor_Beach**, i.e., we are nesting sites within beaches to form the random effect.

RANDOM INTERCEPT AND SLOPE MODEL: intercepts and slopes are allowed to vary

IS THIS A GOOD MODEL for these data?



```
MixedModel.IntSlope <- lmer(Richness ~ NAP + (1 + NAP|factor_Beach),
                            data = RIKZ.data, REML = FALSE)

summary(MixedModel.IntSlope)
```

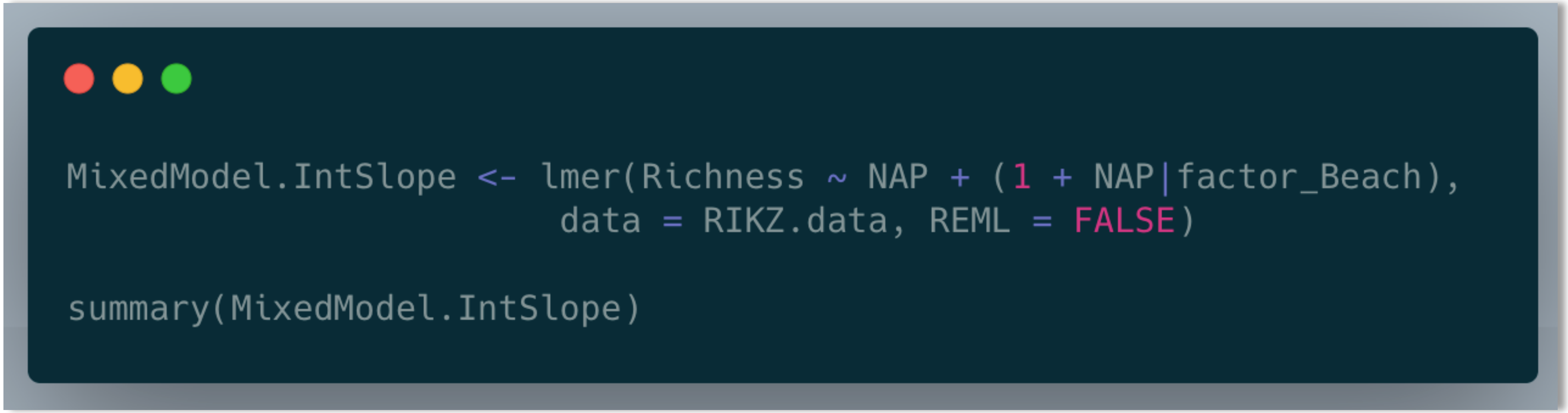


```
Random effects:
 Groups       Name        Variance Std.Dev. Corr
 factor_Beach (Intercept) 10.949   3.309
              NAP          2.502   1.582    -1.00
 Residual                  7.174   2.678
Number of obs: 45, groups:  factor_Beach, 9

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   6.5818     1.1883  8.8936   5.539 0.000377 ***
NAP          -2.8293     0.6849  7.9217  -4.131 0.003366 **
---
```

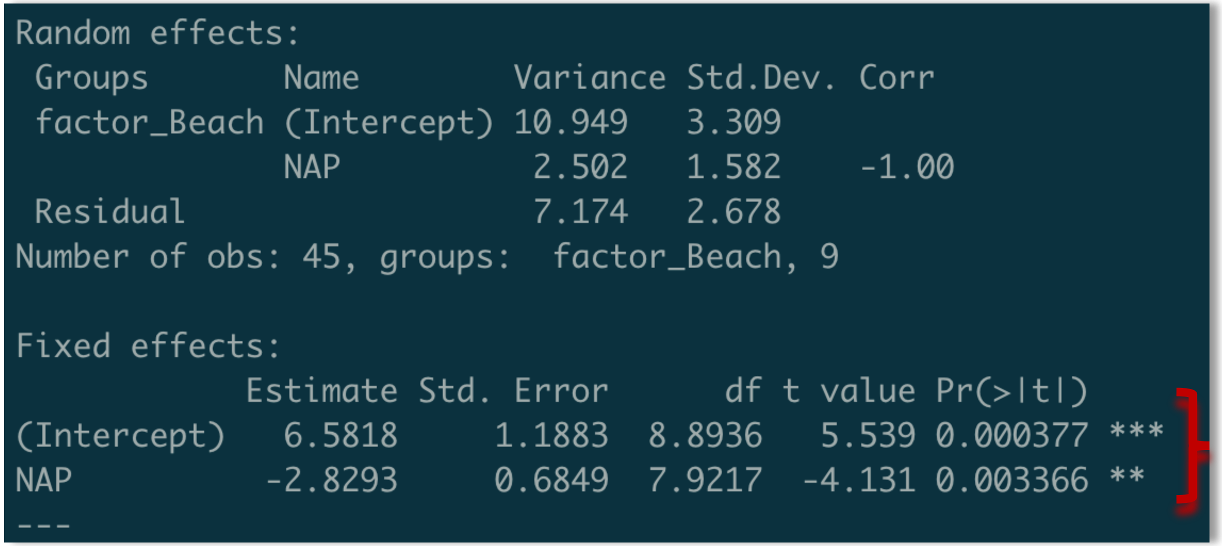
Random effect components: effect due to variation in Intercepts among beaches, variation in slopes among beaches and residuals of the random component.

Corr = correlation between slopes and intercepts of the separate models.

RANDOM INTERCEPT AND SLOPE MODEL: intercepts and slopes are allowed to vary
IS THIS A GOOD MODEL for these data?



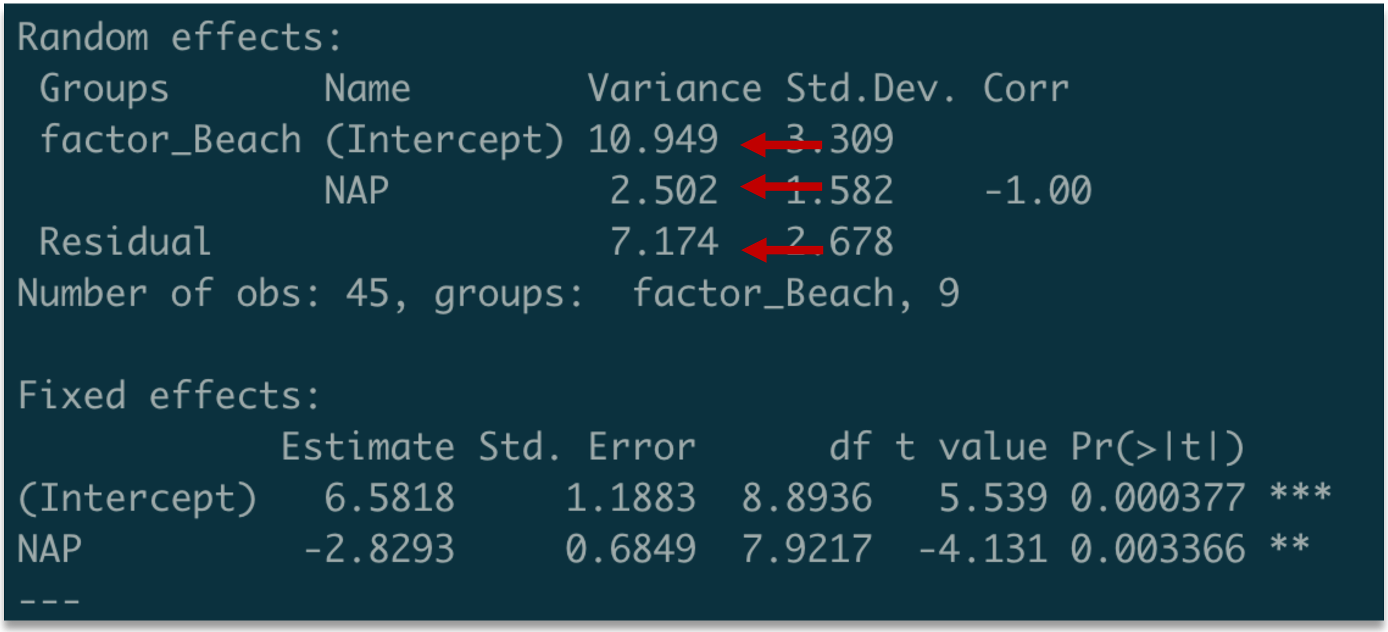
```
MixedModel.IntSlope <- lmer(Richness ~ NAP + (1 + NAP|factor_Beach),
                            data = RIKZ.data, REML = FALSE)

summary(MixedModel.IntSlope)
```



```
Random effects:
 Groups       Name        Variance Std.Dev. Corr
 factor_Beach (Intercept) 10.949   3.309
              NAP          2.502   1.582    -1.00
 Residual                  7.174   2.678
Number of obs: 45, groups:  factor_Beach, 9

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   6.5818     1.1883  8.8936   5.539 0.000377 ***
NAP          -2.8293     0.6849  7.9217  -4.131 0.003366 **
---
```

Fixed effect components: global (across all data) intercept and slope.

RANDOM INTERCEPT AND SLOPE MODEL: intercepts and slopes are allowed to vary
IS THIS A GOOD MODEL for these data?



```
Random effects:
 Groups       Name        Variance Std.Dev. Corr
 factor_Beach (Intercept) 10.949   3.309
              NAP          2.502   1.582    -1.00
 Residual                  7.174   2.678
Number of obs: 45, groups:  factor_Beach, 9

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   6.5818     1.1883  8.8936   5.539 0.000377 ***
NAP          -2.8293     0.6849  7.9217  -4.131 0.003366 **
---
```

The intraclass correlation (ICC or Rho) here describes how strongly variation in predicted values within the same beach resemble each other.

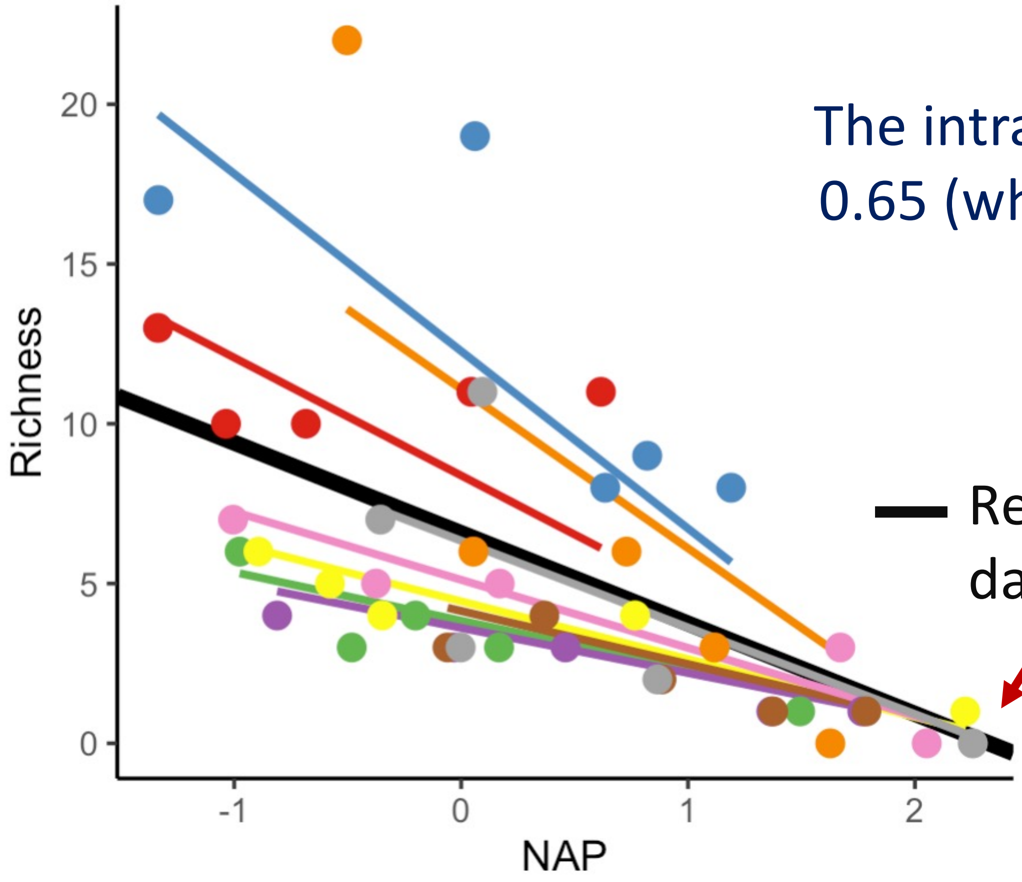
$$\text{ICC} = (10.949+2.502) / (10.949+2.502+7.174) = 0.65$$

(even higher than the previous random intercept model)

RANDOM INTERCEPT AND SLOPE MODEL

The intraclass correlation (ICC or Rho) here describes how strongly variation in predicted values within the same beach resemble each other

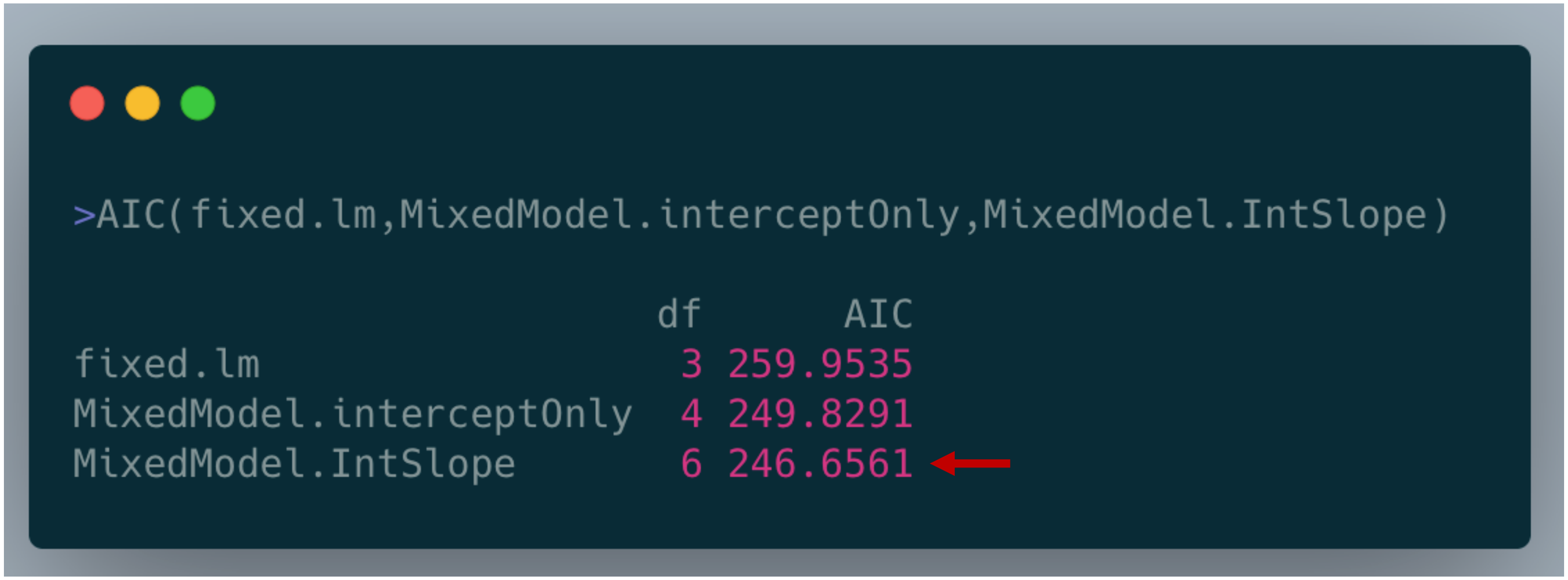
(note how predicted values are more similar within than among beaches).



The intraclass correlation:
0.65 (which is pretty high)

— Regression line considering all data versus different models

Which model to retain? The RANDOM INTERCEPT OR the RANDOM INTERCEPT AND SLOPE MODEL?



```
>AIC(fixed.lm,MixedModel.interceptOnly,MixedModel.IntSlope)

                         df      AIC
fixed.lm                  3 259.9535
MixedModel.interceptOnly  4 249.8291
MixedModel.IntSlope       6 246.6561
```

AIC is a widely used metric of goodness of fit and smaller AIC values indicate the model with the best fit.

$$AIC = 2k + n \, Log(RSS/n)$$

k = number of parameters in the model (intercept, slopes)
n = number of observations
RSS = Residual Sum-of-square

Next – “Go big or go home”: Going complex!

How do “competing” models compare with one another? Which model best fit the data?

Data

Model1

Model2

...

Model selection

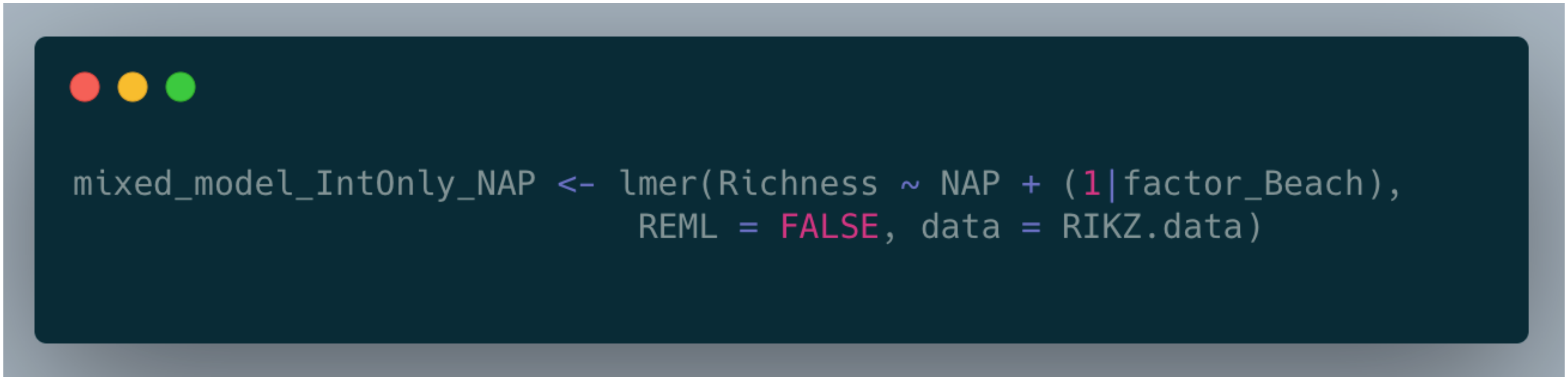
Decision-making

Figure source: https://wires.onlinelibrary.wiley.com/doi/full/10.1002/wics.1607

Let’s now consider different models, make them compete and select the one that best describe the same set of data (i.e., predict Richness).

$$R_{ij} = b_0 + b_1 \times NAP_{ij} + e_{ij}$$

MODEL 1: No interaction or main effect of exposure, i.e., just NAP under a random intercept model (as seen earlier):

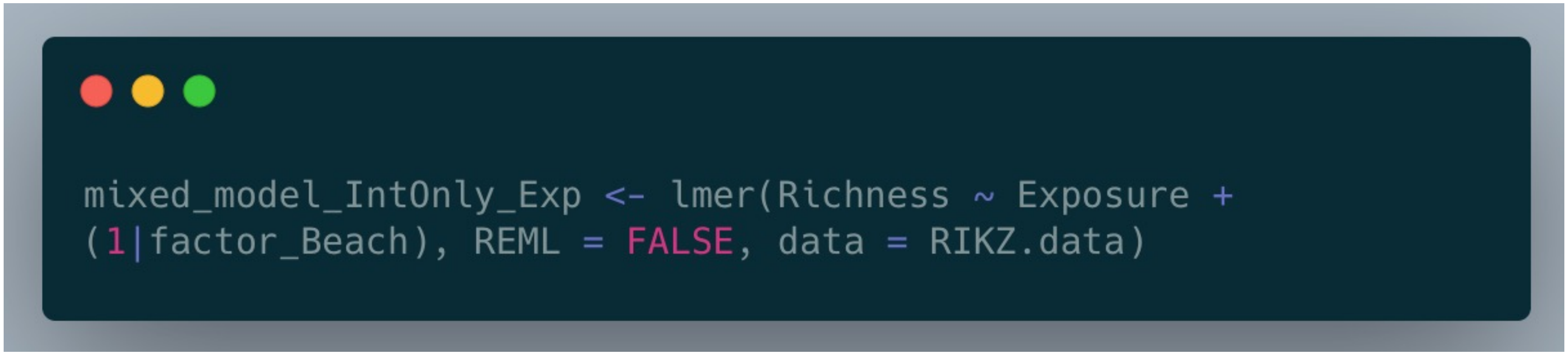


```
mixed_model_IntOnly_NAP <- lmer(Richness ~ NAP + (1|factor_Beach),
                            REML = FALSE, data = RIKZ.data)
```

Competing models

$$R_{ij} = b_0 + b_2 \times Exposure_j + e_{ij}$$

MODEL 2: No interaction or main effect of NAP, i.e., just EXPOSURE under a random intercept model:

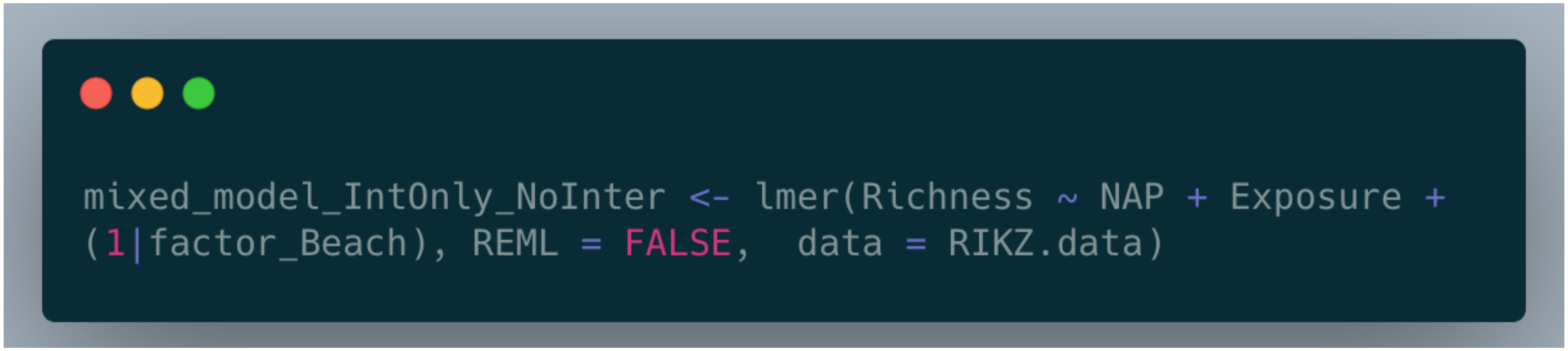


```
mixed_model_IntOnly_Exp <- lmer(Richness ~ Exposure +
(1|factor_Beach), REML = FALSE, data = RIKZ.data)
```

Competing models

$$R_{ij} = b_0 + b_1 \times NAP_{ij} + b_2 \times Exposure_j + e_{ij}$$

MODEL 3: Main effects (NAP and EXPOSURE) but no interaction under a random intercept model.

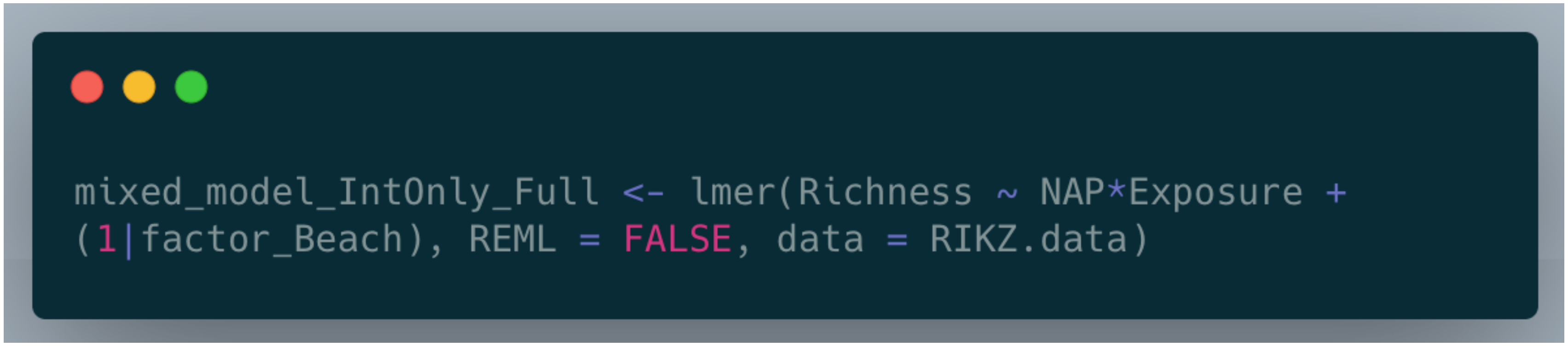


```
mixed_model_IntOnly_NoInter <- lmer(Richness ~ NAP + Exposure +
(1|factor_Beach), REML = FALSE,  data = RIKZ.data)
```

Competing models

$$R_{ij} = b_0 + b_1 \times NAP_{ij} + b_2 \times Exposure_j$$
$$b_3 \times (NAP_{ij} \times Exposure_j)\ + e_{ij}$$

MODEL 4: Main effects (NAP and EXPOSURE) and their interaction under a random intercept model.

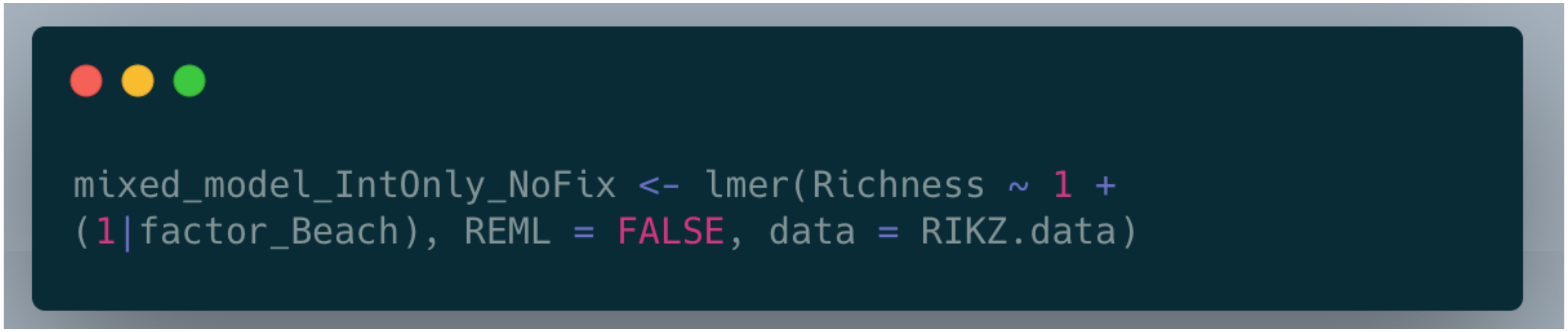


```
mixed_model_IntOnly_Full <- lmer(Richness ~ NAP*Exposure +
(1|factor_Beach), REML = FALSE, data = RIKZ.data)
```

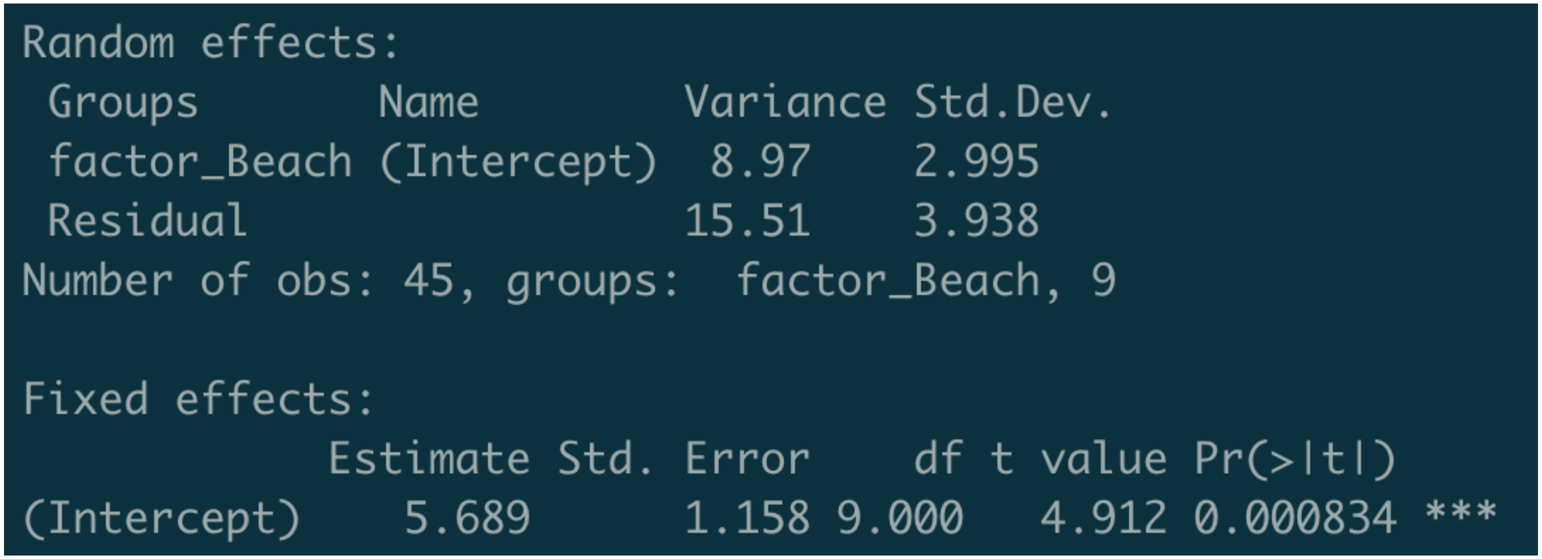
Competing models

$$R_{ij} = b_0 + e_{ij}$$

MODEL 5: Model with a fixed intercept and only random effects (i.e., the ”simplest” model)



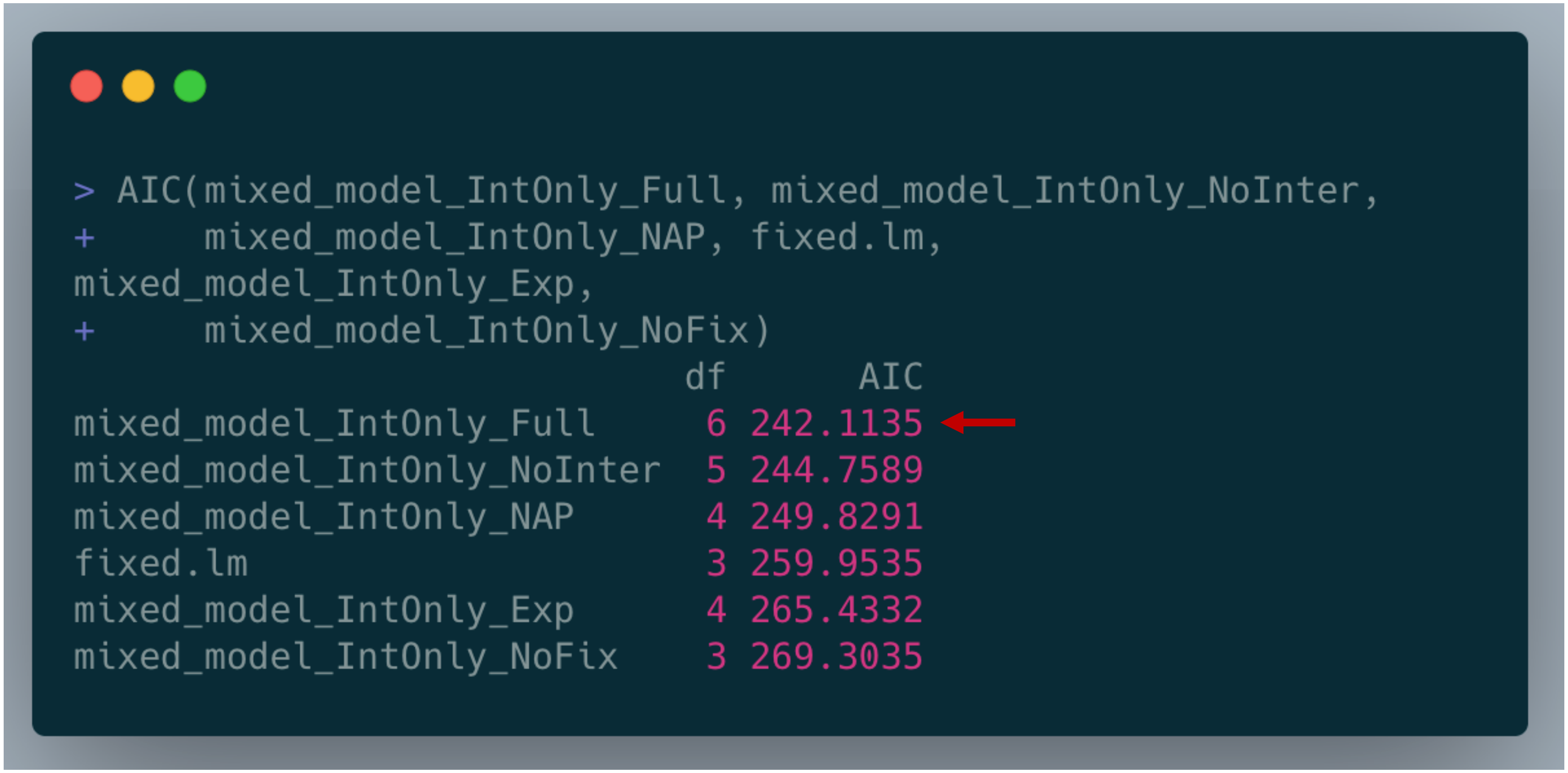
```r
mixed_model_IntOnly_NoFix <- lmer(Richness ~ 1 +
(1|factor_Beach), REML = FALSE, data = RIKZ.data)
```



```
Random effects:
 Groups       Name        Variance Std.Dev.
 factor_Beach (Intercept)  8.97    2.995
 Residual                 15.51    3.938
Number of obs: 45, groups:  factor_Beach, 9

Fixed effects:
            Estimate Std. Error    df t value Pr(>|t|)
(Intercept)    5.689      1.158 9.000   4.912 0.000834 ***
```

Competing models

$$R_{ij} = b_0 + b_1 \times NAP_{ij} + b_2 \times Exposure_j \\ b_3 \times (NAP_{ij} \times Exposure_j) \ + e_{ij}$$



```
> AIC(mixed_model_IntOnly_Full, mixed_model_IntOnly_NoInter,
+       mixed_model_IntOnly_NAP, fixed.lm,
mixed_model_IntOnly_Exp,
+       mixed_model_IntOnly_NoFix)
                            df      AIC
mixed_model_IntOnly_Full     6 242.1135
mixed_model_IntOnly_NoInter  5 244.7589
mixed_model_IntOnly_NAP      4 249.8291
fixed.lm                     3 259.9535
mixed_model_IntOnly_Exp      4 265.4332
mixed_model_IntOnly_NoFix    3 269.3035
```

NOTES:

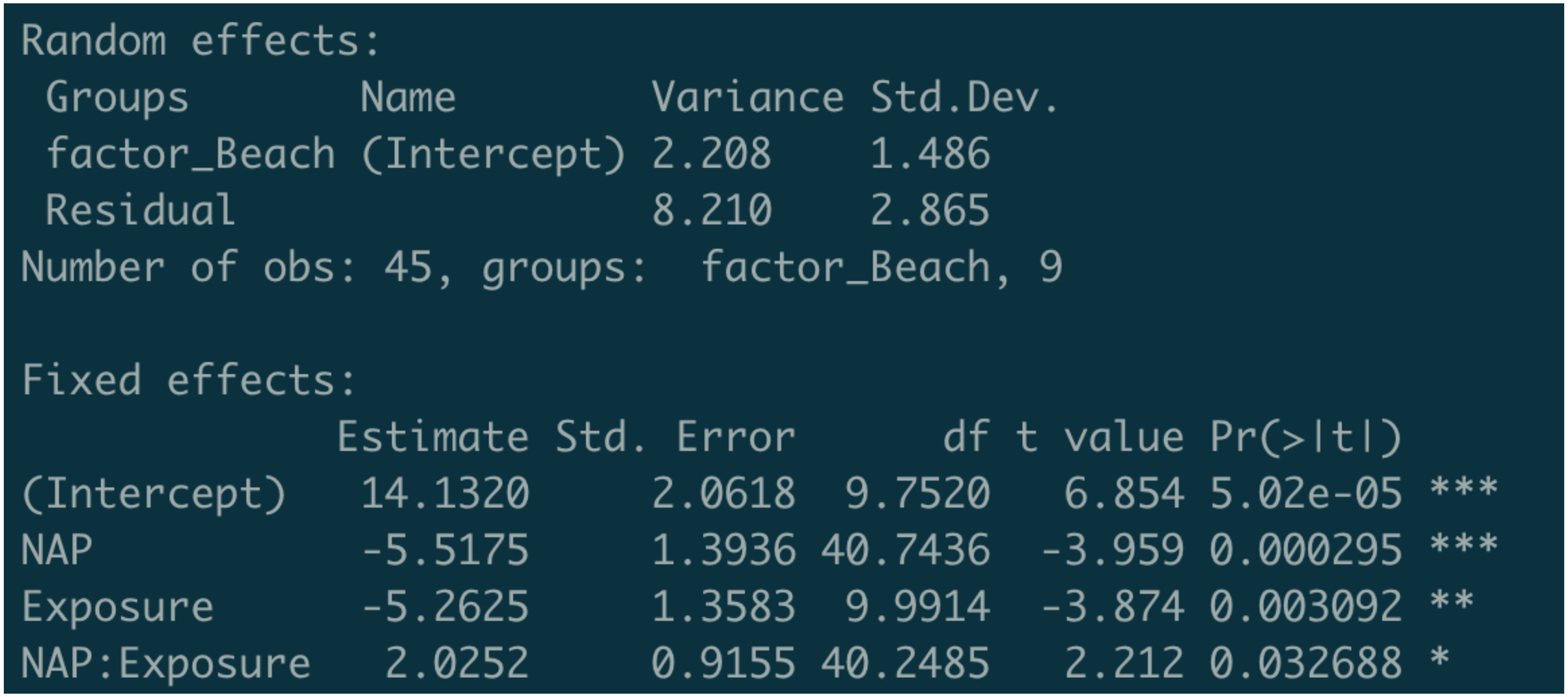
1) We only considered the intercept only model. We could have considered for each model the intercept and slope model (the Tutorial 10 does).

2) We could have also considered all the fixed effect only:
   Intercept only
   NAP only
   Exposure only
   NAP + Exposure
   NAP x Exposure (main effects + interaction).

3) And once all models are built, compare them using AIC.

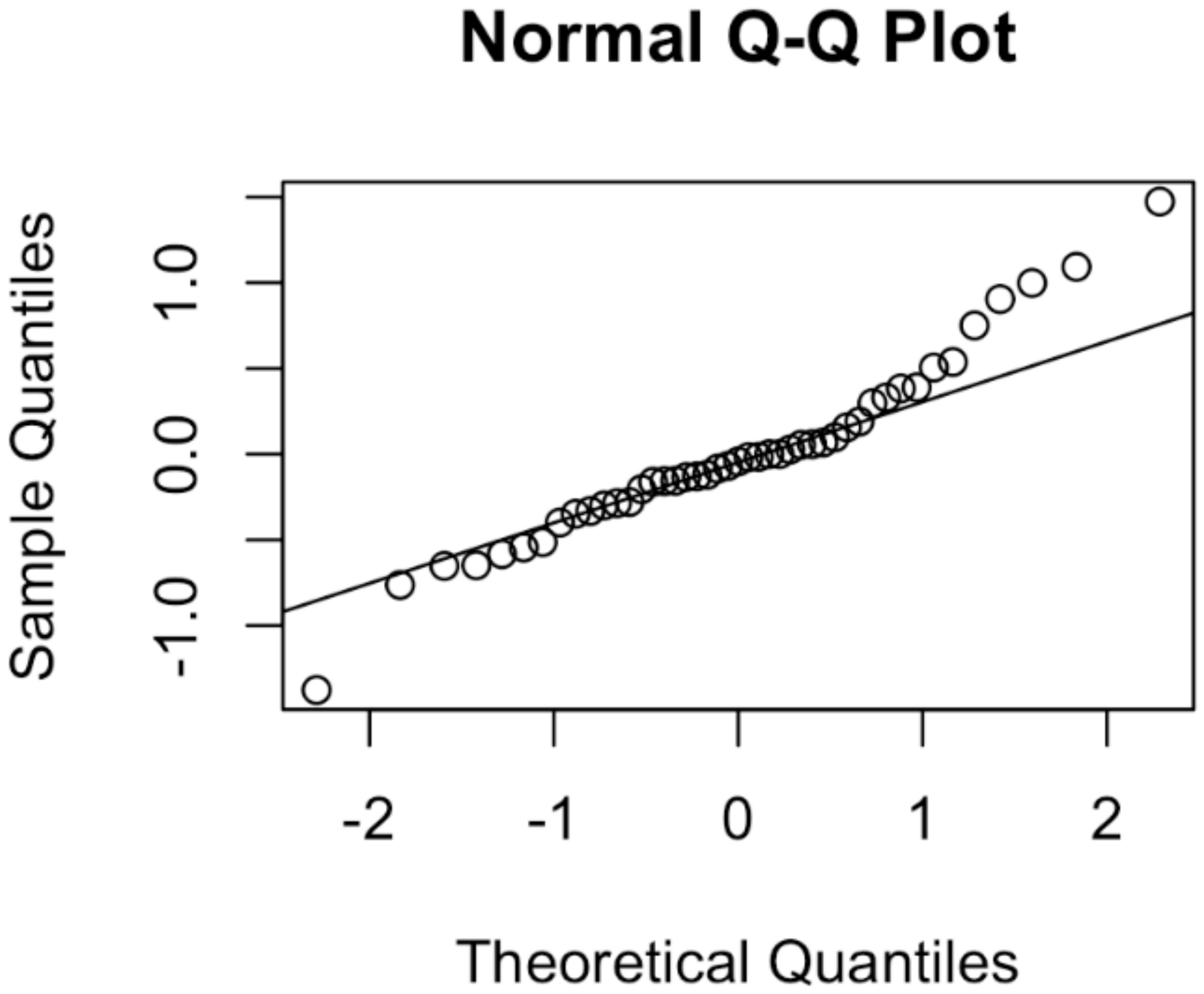
The best model (amongst the ones we compared in this lecture; but more models in the tutorial)!

$$R_{ij} = b_0 + b_1 \times NAP_{ij} + b_2 \times Exposure_j$$
$$b_3 \times (NAP_{ij} \times Exposure_j) \; + e_{ij}$$

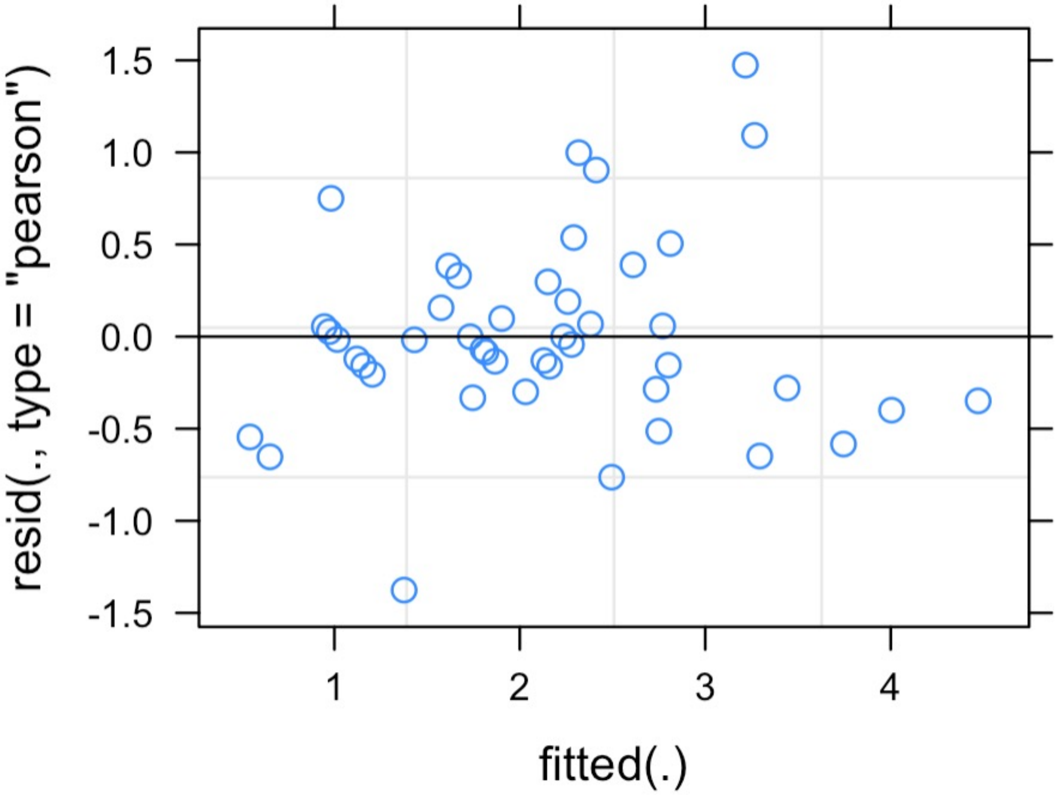


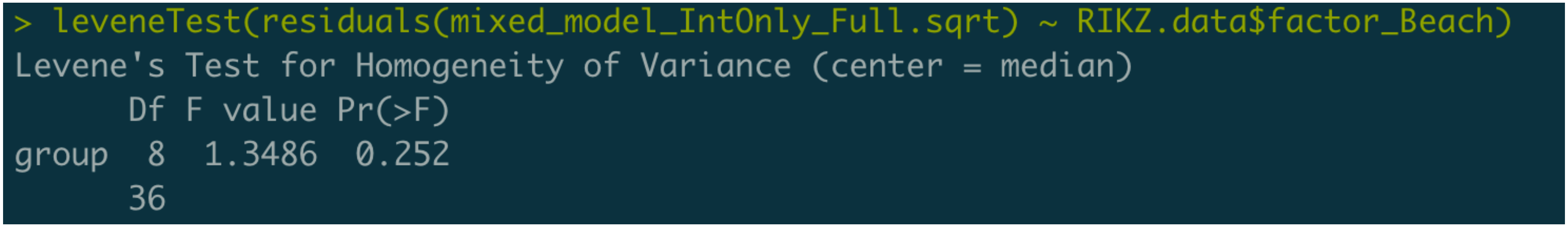
```
Random effects:
 Groups       Name        Variance Std.Dev.
 factor_Beach (Intercept) 2.208    1.486
 Residual                 8.210    2.865
Number of obs: 45, groups:  factor_Beach, 9

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  14.1320     2.0618  9.7520   6.854 5.02e-05 ***
NAP          -5.5175     1.3936 40.7436  -3.959 0.000295 ***
Exposure     -5.2625     1.3583  9.9914  -3.874 0.003092 **
NAP:Exposure  2.0252     0.9155 40.2485   2.212 0.032688 *
```

Assumptions: Normality (after square root transformation of Richness)



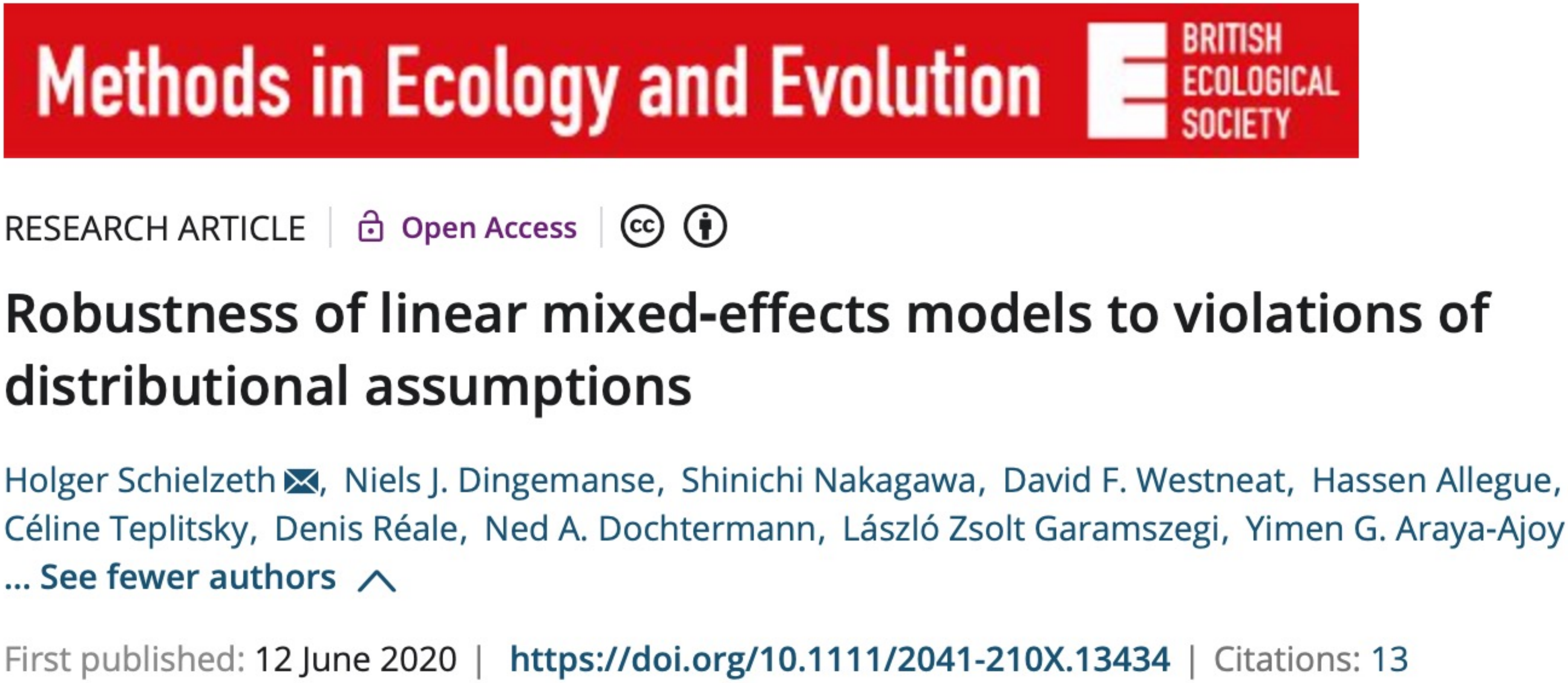
Assumptions: residual homoscedasticity (residuals against predicted values)





```
> leveneTest(residuals(mixed_model_IntOnly_Full.sqrt) ~ RIKZ.data$factor_Beach)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  8  1.3486  0.252
      36
```

Assumptions: Recently shown that mixed-effects models are robust against normality and heteroscedastic assumptions



Methods in Ecology and Evolution

BRITISH
ECOLOGICAL
SOCIETY

RESEARCH ARTICLE | Open Access

Robustness of linear mixed-effects models to violations of distributional assumptions

Holger Schielzeth, Niels J. Dingemanse, Shinichi Nakagawa, David F. Westneat, Hassen Allegue, Céline Teplitsky, Denis Réale, Ned A. Dochtermann, László Zsolt Garamszegi, Yimen G. Araya-Ajoy

... See fewer authors

First published: 12 June 2020 | https://doi.org/10.1111/2041-210X.13434 | Citations: 13