Homework 3: Linear Mixed Models

Your_Name_Here

Turn in via blackboard next Friday by the end of the day. please title the document **hw3_yourlastname_yourfirstname**. Ideally it will be a PDF file generated from knitting an RMarkdown document.

In this lab you will be working with the RIKZ data (Zuur et al. 2007). They are data on the macrofauna species richness from five sites at each of nine beaches. The authors were interested in how species richness is affected by exposure and NAP (height of sampling station relative to mean tidal level). However, they likely have to account for the potential correlation of data from sites collected at the same beach.

1. Read in the RIKZ data from blackboard using `read.table` since it's a text file and not a csv. Assign it to an object and examine the structure of the object. How many rows and columns are there?

```
df <- read.table(file = "Data/RIKZ.txt", header = TRUE, stringsAsFactors = FALSE)
str(df)
```

```
## 'data.frame':    45 obs. of  5 variables:
##  $ Sample  : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Richness: int  11 10 13 11 10 8 9 8 19 17 ...
##  $ Exposure: int  10 10 10 10 10 8 8 8 8 8 ...
##  $ NAP     : num  0.045 -1.036 -1.336 0.616 -0.684 ...
##  $ Beach   : int  1 1 1 1 1 2 2 2 2 2 ...
```

```
ncol(df)
```

```
## [1] 5
```

```
nrow(df)
```

```
## [1] 45
```

45 rows and 5 columns

2. Use a linear regression to examine the effects of NAP on species richness.

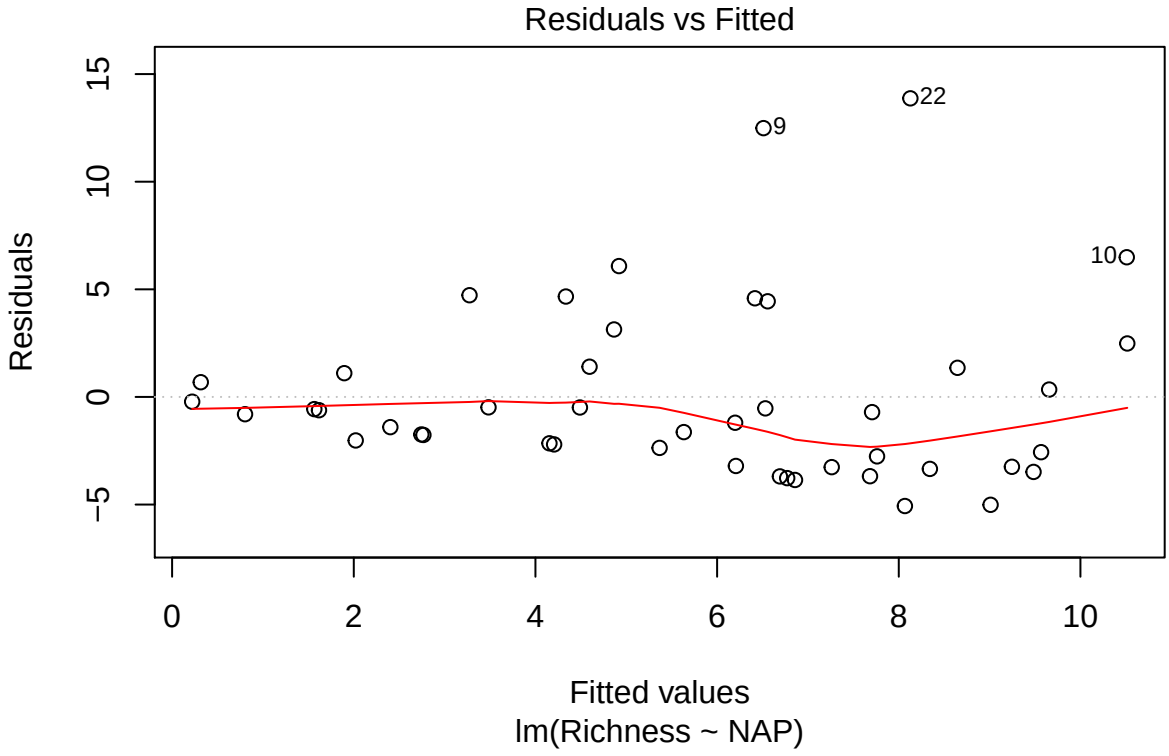
```
lm1 <- lm(Richness ~ NAP, data = df)
summary(lm1)
```

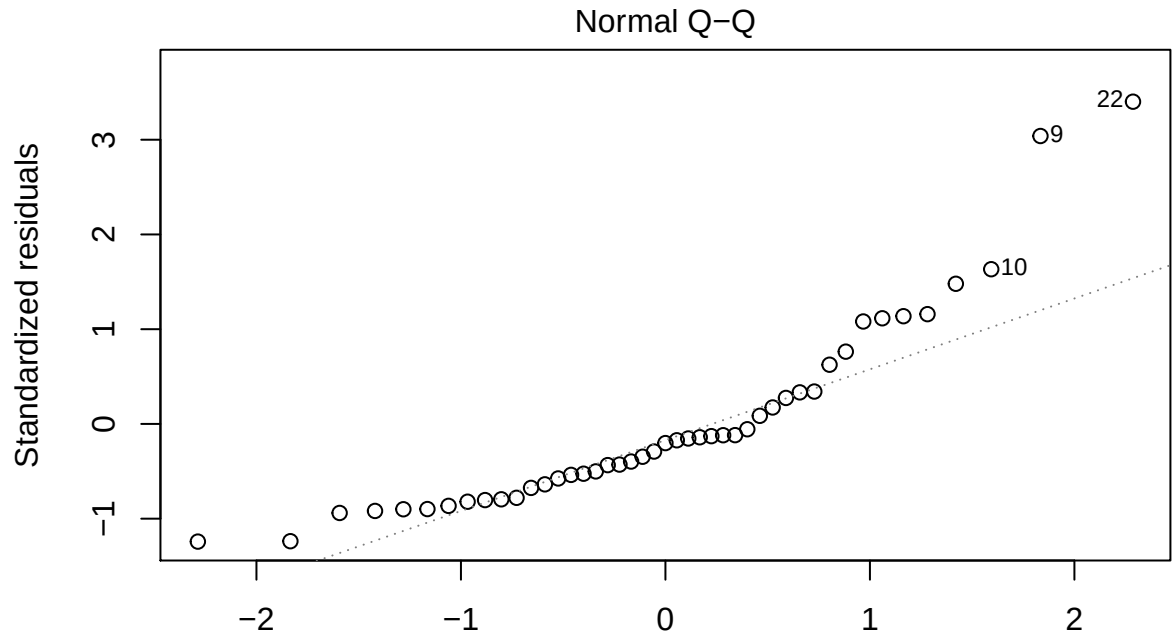
```
##
## Call:
## lm(formula = Richness ~ NAP, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0675 -2.7607 -0.8029  1.3534 13.8723
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.6857     0.6578  10.164 5.25e-13 ***
## NAP          -2.8669     0.6307  -4.545 4.42e-05 ***
```

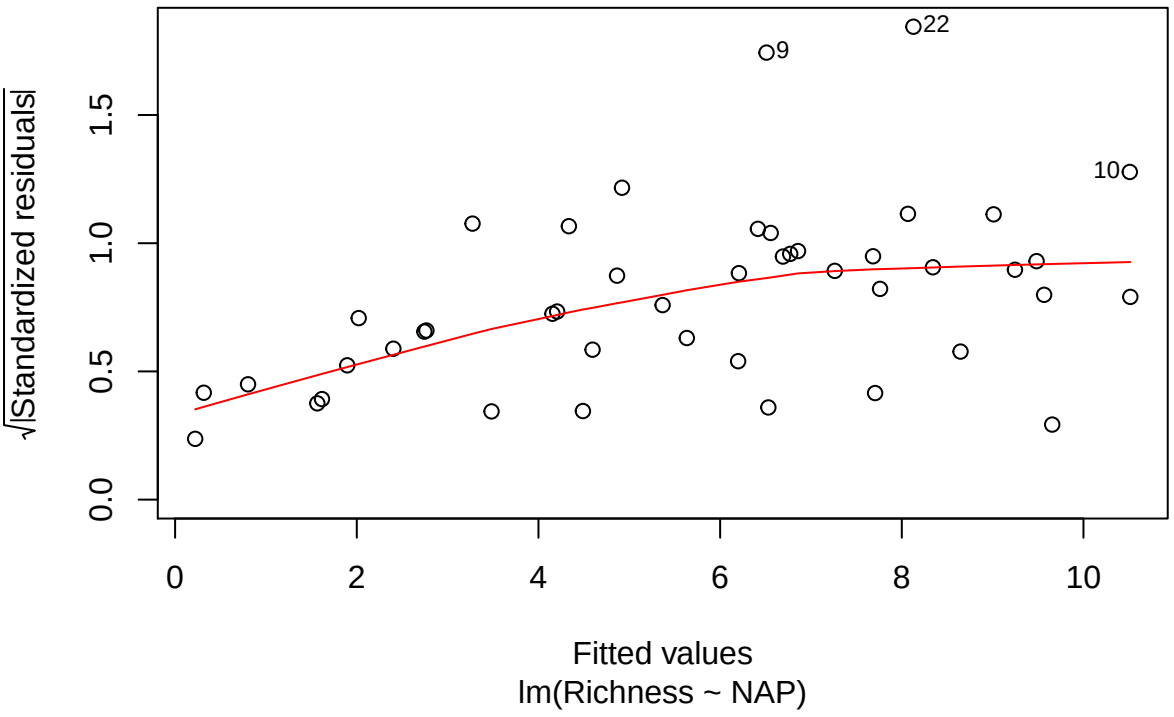
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.16 on 43 degrees of freedom
## Multiple R-squared:  0.3245, Adjusted R-squared:  0.3088
## F-statistic: 20.66 on 1 and 43 DF,  p-value: 4.418e-05
```

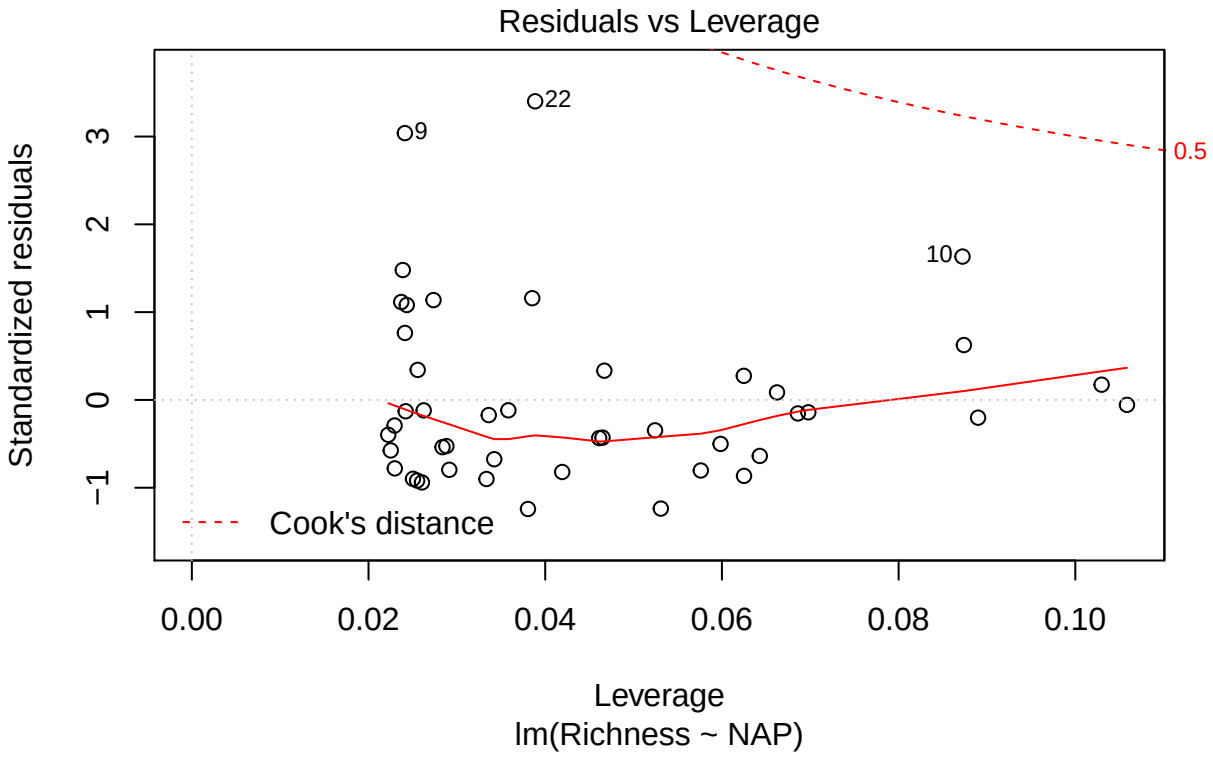
3. Check if the model assumptions fit the data. Describe the results.

```
plot(lm1)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Richness ~ NAP)
22
9
10
−2
−1
0
1
2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
Fitted values
lm(Richness ~ NAP)
22
9
10
0
2
4
6
8
10
0.0
0.5
1.0
1.5



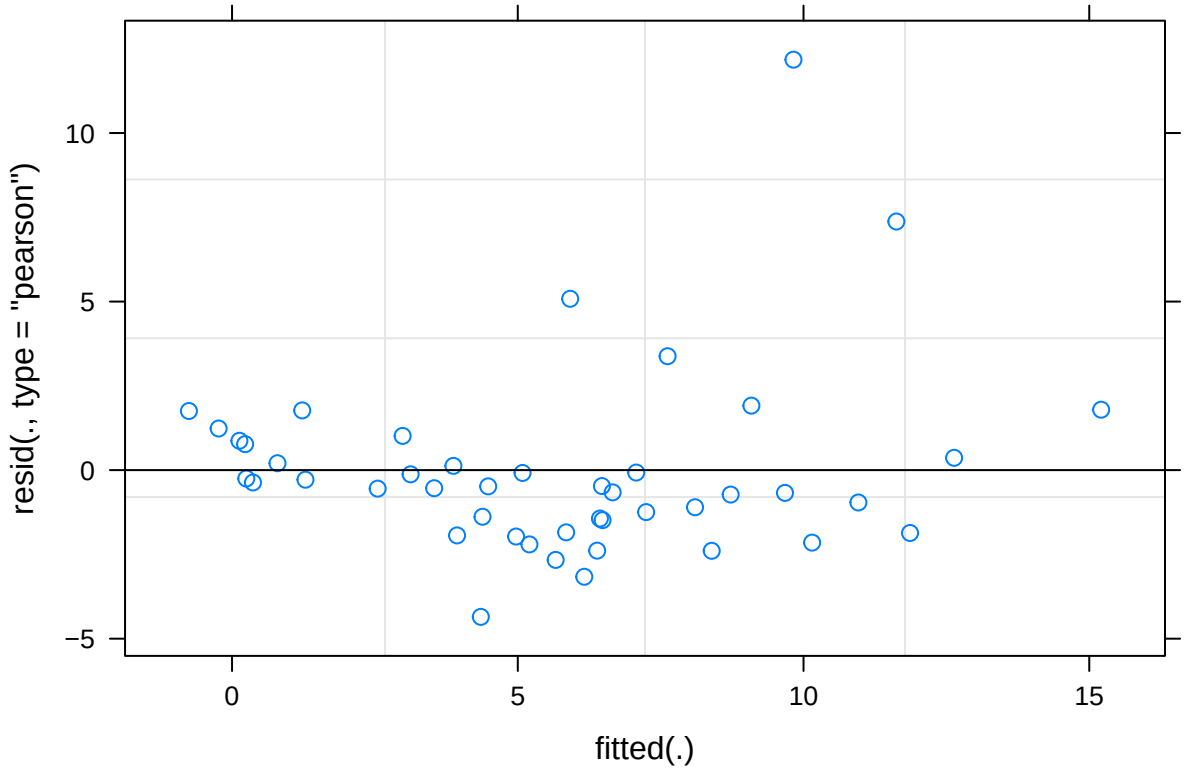
Heterogeneity of residuals with increasing variance with larger values. Also some potential patterning in the residuals. Significant violations of normality based on the QQplot. I would not use this model with these data.

4. Analyze the same patterns, this time using a linear mixed effects model with a random intercept for beach. Check the model assumptions including homogenity by group (beach).

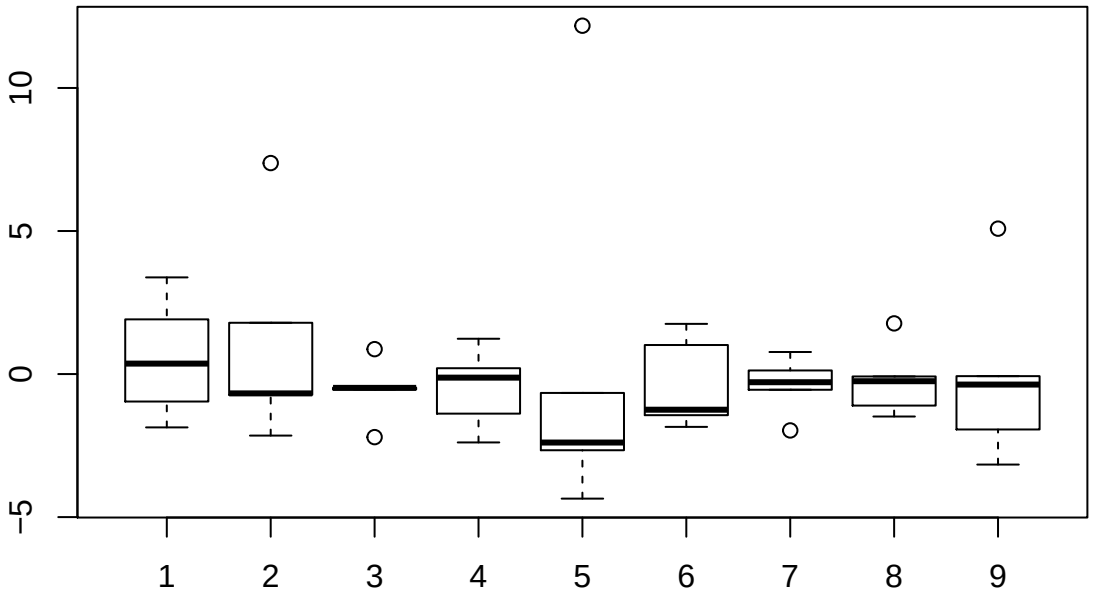
```
library(lme4)
```

```
## Loading required package: Matrix
```

```
lme1 <- lmer(Richness ~ 1 + NAP + (1 | Beach), data = df)
# par(mfrow = c(1,2))
plot(lme1)
```



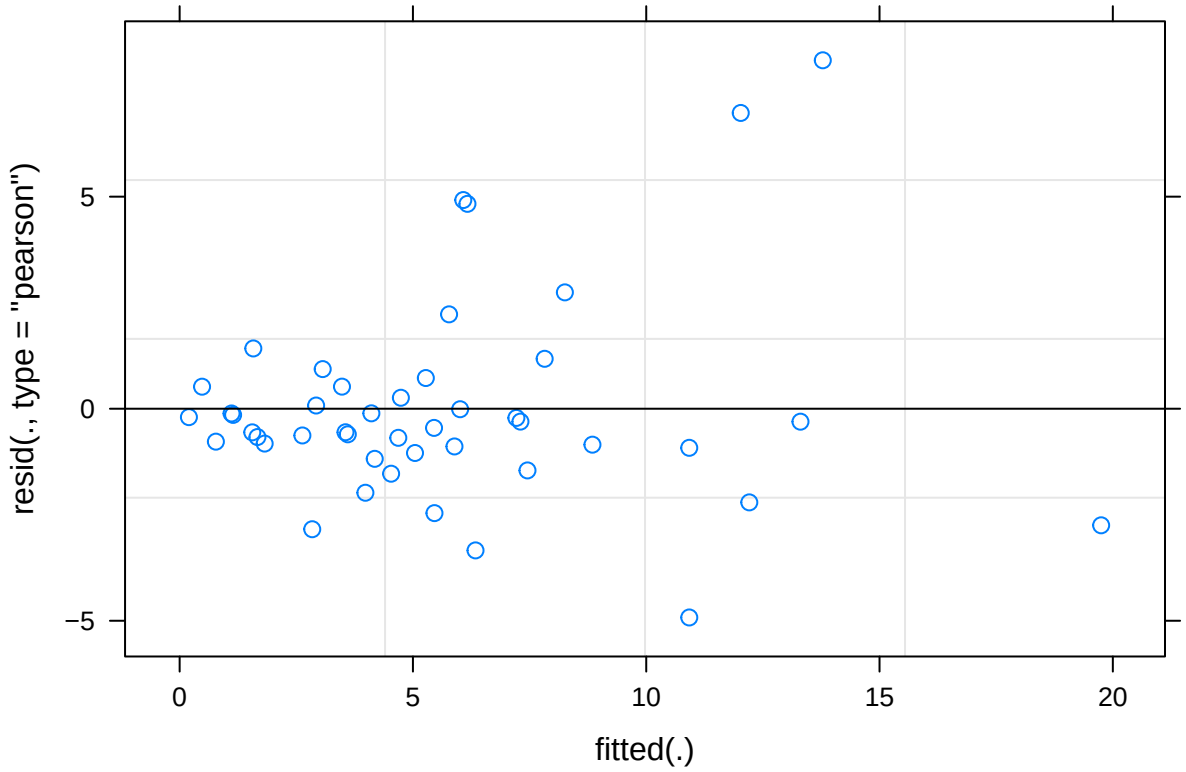

```
boxplot(resid(lme1) ~ df$Beach)
```



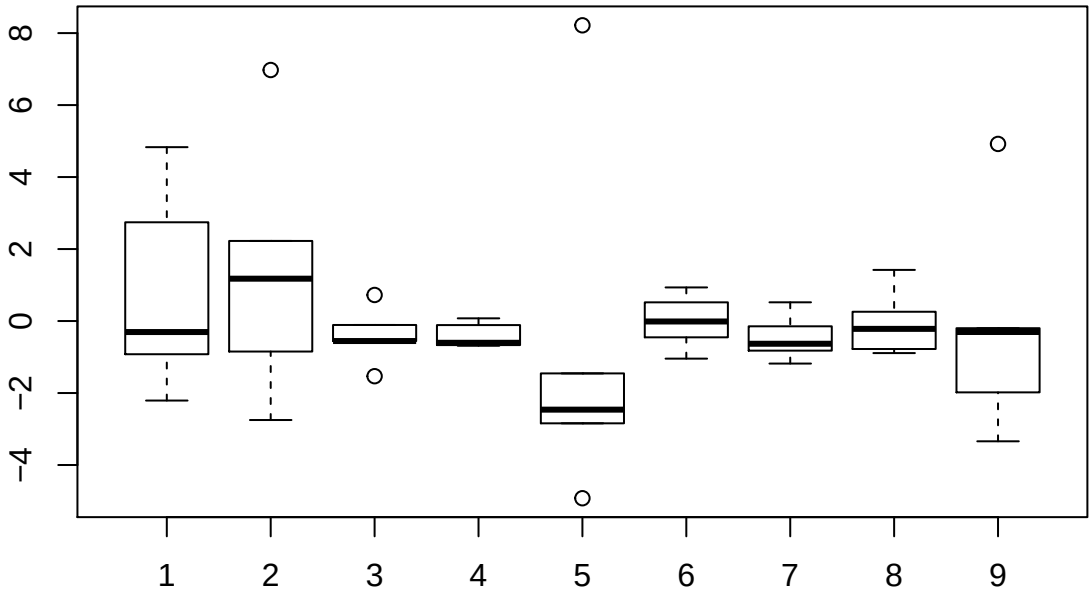

```
# par(mfrow = c(1,1))
```

5. Run a third model using a random intercept and random effect of beach on NAP (random slope). Check the model assumptions including homogenity by group (beach).

```
lme2 <- lmer(Richness ~ 1 + NAP + (1 + NAP | Beach), data = df)
plot(lme2)
```




```
boxplot(resid(lme2) ~ df$Beach)
```



6. Which model would you use and why?

I would use the random intercepts and slopes model of these options but there still might be some heterogeneity of variance in the residuals and I would explore other models for these data.

7. Describe the results of this model.

```
summary(lme2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Richness ~ 1 + NAP + (1 + NAP | Beach)
##    Data: df
##
## REML criterion at convergence: 232.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8213 -0.3411 -0.1675  0.1921  3.0397
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Beach    (Intercept) 12.596   3.549
##           NAP          2.941   1.715    -0.99
##  Residual              7.305   2.703
## Number of obs: 45, groups:  Beach, 9
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   6.5887     1.2648   5.209
## NAP          -2.8300     0.7229  -3.915
##
## Correlation of Fixed Effects:
##     (Intr)
## NAP -0.819
```

Richness decreases with increasing NAP but this affect varies modestly among beaches based on the random slope SD. There is large variation in richness among beaches (SD = 3.5) not explained by the deterministic part of the model (NAP).