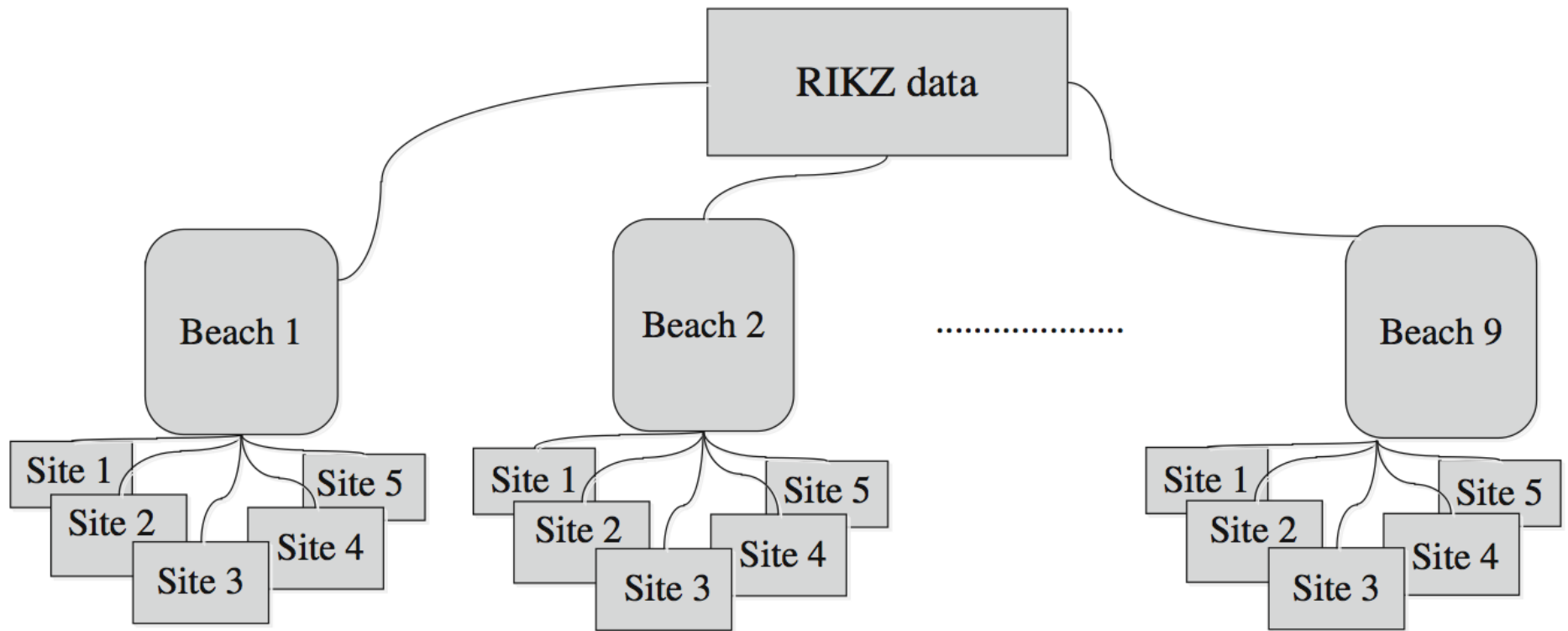


Multi-level Models

Chapter 5, Zuur et al.



response variable (R) is species richness (number of species)

Predictor variables are:

NAP: height of a sampling station compared to mean tidal level (site)

Exposure: index composed of wave action, length of surf slope, etc. (beach)

Could fit a standard Gaussian regression (i indexes beach, j indexes sample):

$$R_{ij} = \alpha + \beta_1 \times NAP_{ij} + \beta_2 \times Exposure_i + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

This treats the count-valued response variable as Gaussian and ignores similarity of measures within beach.

Consider a two-stage modeling procedure:

Could model each beach separately (just focusing on NAP for now)...

$$R_{ij} = \alpha + \beta_i \times NAP_{ij} + \varepsilon_{ij} \quad j = 1, \dots, 5$$

$$\begin{pmatrix} R_{i1} \\ R_{i2} \\ R_{i3} \\ R_{i4} \\ R_{i5} \end{pmatrix} = \begin{pmatrix} 1 & NAP_{i1} \\ 1 & NAP_{i1} \\ 1 & NAP_{i1} \\ 1 & NAP_{i1} \\ 1 & NAP_{i1} \end{pmatrix} \times \begin{pmatrix} \alpha \\ \beta_i \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \end{pmatrix} \Leftrightarrow \mathbf{R}_i = \mathbf{Z}_i \times \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$$

```

> library(AED); data(RIKZ)
> Beta <- vector(length = 9)
> for (i in 1:9){
  Mi <- summary(lm(Richness ~ NAP,
                   subset = (Beach==i), data=RIKZ))
  Beta[i] <- Mi$coefficients[2, 1]}

```

estimated betas for the different beaches:

-0.37 -4.17 -1.75 -1.24 -8.90 -1.38 -1.51 -1.89 -2.96

and then ... model the betahat's in a second stage:

$$\hat{\beta}_i = \eta + \tau \times Exposure_i + b_i \quad i = 1, \dots, 9$$

$$\begin{pmatrix} -0.37 \\ -4.17 \\ -1.75 \\ -1.24 \\ -8.90 \\ -1.38 \\ -1.51 \\ -1.89 \\ -2.96 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \times \begin{pmatrix} \eta \\ \tau \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \\ b_9 \end{pmatrix} \Leftrightarrow \hat{\beta}_i = \mathbf{K}_i \times \boldsymbol{\gamma} + \mathbf{b}_i \quad i = 1, \dots, 9$$

$$\mathbf{R}_i = \mathbf{Z}_i \times \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$$

$$\hat{\boldsymbol{\beta}}_i = \mathbf{K}_i \times \boldsymbol{\gamma} + \mathbf{b}_i$$

The “Linear Mixed Effects Model” puts the two stages together:

$$\mathbf{R}_i = \mathbf{X}_i \times \boldsymbol{\beta} + \mathbf{Z}_i \times \mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

has the 5 richness values for beach i

fixed effects
(all beaches)

random effects
(beach specific)

has the 5 NAP values for beach i

The complete mixed effects model:

$$\mathbf{Y}_i = \mathbf{X}_i \times \boldsymbol{\beta} + \mathbf{Z}_i \times \mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

$$\mathbf{b}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{D})$$

$$\boldsymbol{\varepsilon}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$$


$$\mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_N \text{ independent}$$

now consider some specific example...

The random intercept model

could use lm...

lm with beach-specific intercept: factor with nine levels:

$$R_{ij} = \alpha + \beta_1 \times Beach_i + \beta_2 \times NAP_{ij} + \varepsilon_{ij}$$


The random intercept model:

for any one beach (i)

$$\mathbf{R}_i = \mathbf{X}_i \times \boldsymbol{\beta} + \mathbf{Z}_i \times \mathbf{b}_i$$

$$\begin{pmatrix} R_{i1} \\ R_{i2} \\ R_{i3} \\ R_{i4} \\ R_{i5} \end{pmatrix} = \begin{pmatrix} 1 & NAP_{i1} \\ 1 & NAP_{i2} \\ 1 & NAP_{i3} \\ 1 & NAP_{i4} \\ 1 & NAP_{i5} \end{pmatrix} \times \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \times b_i + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \end{pmatrix}$$

$N(0, d^2)$ points to b_i
 $N(0, \sigma^2)$ points to ε_{i1}

```
> library(nlme)
> RIKZ$fBeach <- factor(RIKZ$Beach)
> Mlme1 <- lme(Richness ~ NAP, random = ~1 | fBeach,
               data = RIKZ)
> summary(Mlme1)
```

Linear mixed-effects model fit by REML

AIC	BIC	logLik
247.48	254.52	-119.74

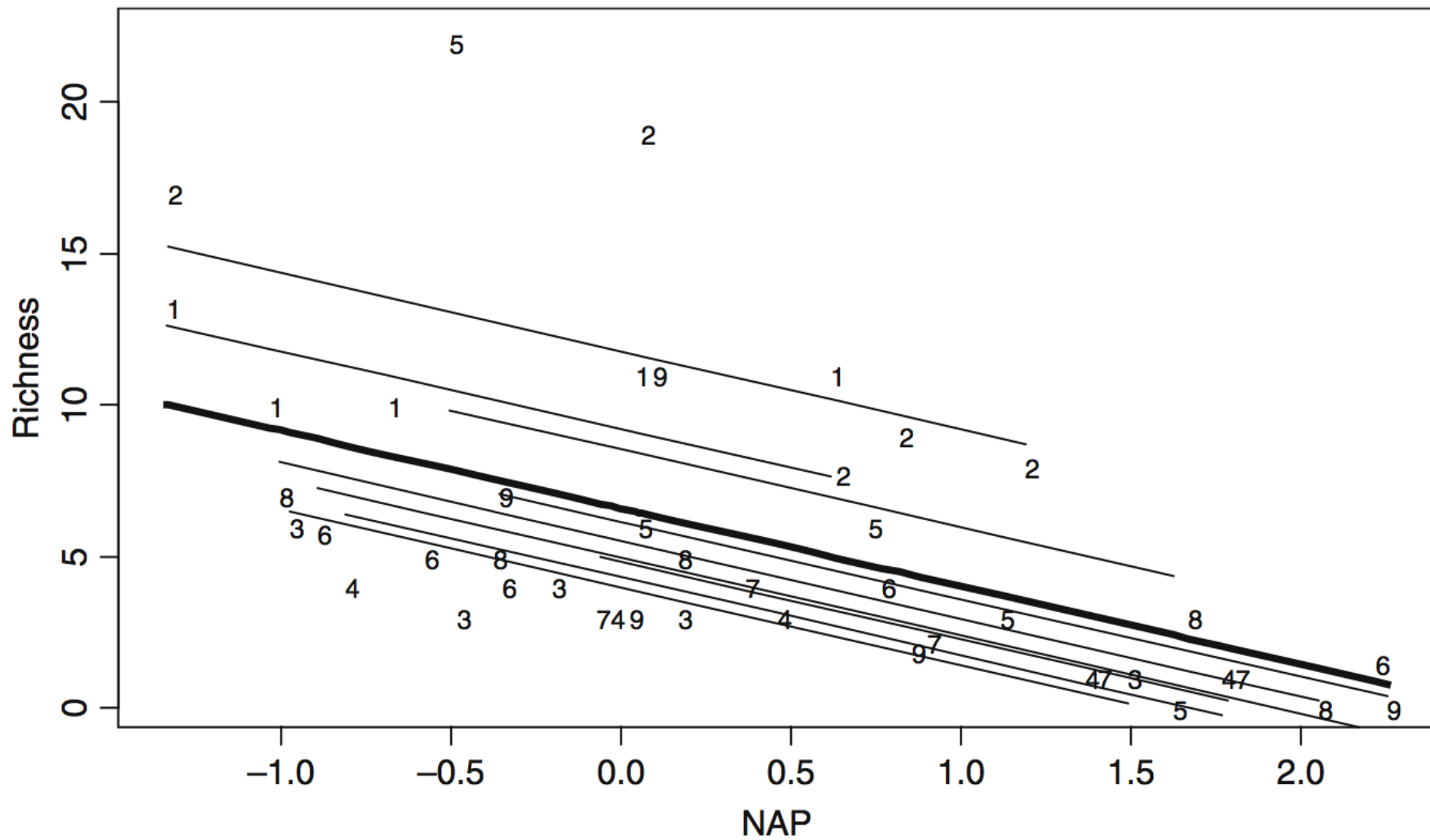
Random effects:

Formula: $\sim 1 \mid \text{fBeach}$

	(Intercept)	Residual
StdDev:	2.944	3.059

Fixed effects: Richness \sim NAP

	Value	Std.Error	DF	t-value	p-value
(Intercept)	6.58	1.09	35	6.00	<0.001
NAP	-2.56	0.49	35	-5.19	<0.001



```
> F0 <- fitted(Mlme1, level = 0)
> F1 <- fitted(Mlme1, level = 1)
> I <- order(RIKZ$NAP); NAPs <- sort(RIKZ$NAP)
> plot(NAPs, F0[I], lwd = 4, type = "l",
       ylim = c(0, 22), ylab = "Richness", xlab = "NAP")

> for (i in 1:9) {
  x1 <- RIKZ$NAP[RIKZ$Beach == i]
  y1 <- F1[RIKZ$Beach == i]
  K <- order(x1)
  lines(sort(x1), y1[K])
}
> text(RIKZ$NAP, RIKZ$Richness, RIKZ$Beach, cex = 0.9)
```

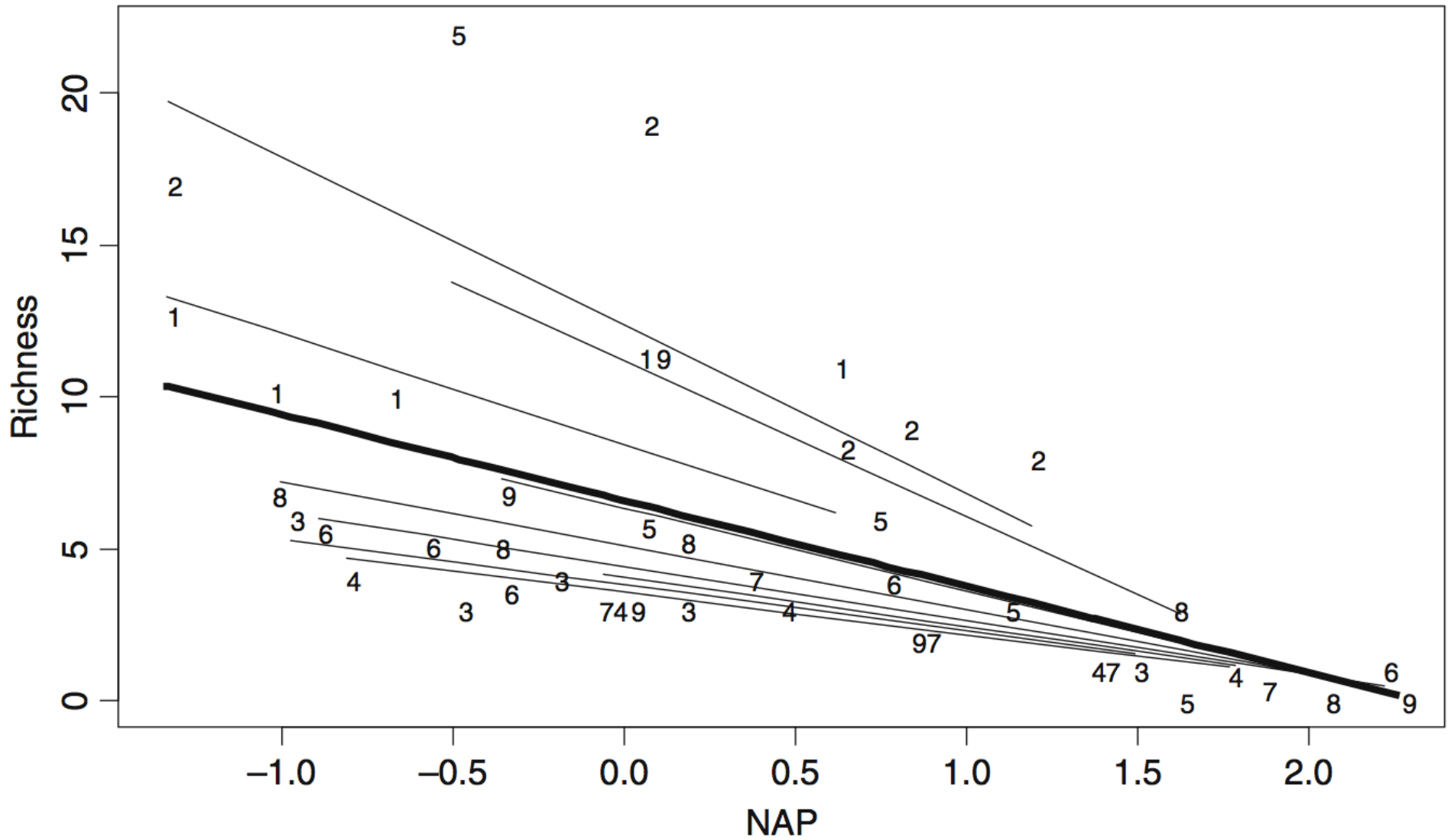
Can show that the covariance between two sites on the same beach is d^2 in this model

Covariance between two sites on different beaches is zero

Can achieve the same objective using the gls function:

```
> M.gls <- gls(Richness ~ NAP, method = "REML",  
              correlation = corCompSymm(form = ~ 1 | fBeach),  
              data = RIKZ)
```

The random intercept & slope model:



```
> Mlme2 <- lme(Richness ~ NAP,
               random = ~1 + NAP | fBeach, data = RIKZ)
> summary(Mlme2)
```

Linear mixed-effects model fit by REML

AIC	BIC	logLik
244.38	254.95	-116.19

Random effects:

Formula: ~1 + NAP | fBeach

	StdDev	Corr
(Intercept)	3.549	(Intr)
NAP	1.714	-0.99
Residual	2.702	

Fixed effects: Richness ~ NAP

	Value	Std.Error	DF	t-value	p-value
(Intercept)	6.58	1.26	35	5.20	<0.001
NAP	-2.83	0.72	35	-3.91	<0.001

$$\begin{pmatrix} R_{i1} \\ R_{i2} \\ R_{i3} \\ R_{i4} \\ R_{i5} \end{pmatrix} = \begin{pmatrix} 1 & NAP_{i1} \\ 1 & NAP_{i2} \\ 1 & NAP_{i3} \\ 1 & NAP_{i4} \\ 1 & NAP_{i5} \end{pmatrix} \times \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} 1 & NAP_{i1} \\ 1 & NAP_{i2} \\ 1 & NAP_{i3} \\ 1 & NAP_{i4} \\ 1 & NAP_{i5} \end{pmatrix} \times \begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \end{pmatrix}$$

Pure random effects model

```
> Mlme3 <- lme(Richness ~ 1, random = ~1 | fBeach,  
              data = RIKZ)
```

Linear mixed-effects model fit by REML

AIC	BIC	logLik
267.11	272.46	-130.55

Random effects:

	(Intercept)	Residual
StdDev:	3.23	3.93

Fixed effects:

	Value	Std.Error	DF	t-value	p-value
(Intercept)	5.68	1.22	36	4.63	<0.001

A Possible Mixed Effect Modeling Strategy:

- begin with a fixed component that contains as many variables as possible
- with this fixed component, find the optimal random component
- find the optimal fixed component

```
> B1 <- gls(Richness ~ 1 + NAP * fExp,  
            method = "REML", data = RIKZ)  
> B2 <- lme(Richness ~1 + NAP * fExp, data = RIKZ,  
            random = ~1 | fBeach, method = "REML")  
> B3 <- lme(Richness ~ 1 + NAP * fExp, data = RIKZ,  
            random = ~1 + NAP | fBeach, method="REML")
```

```
> AIC(B1,B2,B3)  
df  AIC  
B1 5 243.0875  
B2 6 236.4925  
B3 8 237.1331
```

> summary(B2)

Linear mixed-effects model fit by REML

Data: RIKZ

AIC	BIC	logLik
236.4925	246.7739	-112.2462

Random effects:

Formula: ~1 | fBeach

(Intercept) Residual

StdDev: 1.818578 2.942872

Fixed effects: Richness ~ 1 + NAP * fExp

	Value	Std.Error	DF	t-value	p-value
(Intercept)	8.861084	1.020845	34	8.680147	0.0000
NAP	-3.463651	0.627858	34	-5.516613	0.0000
fExp11	-5.255617	1.545229	7	-3.401189	0.0114
NAP:fExp11	2.000464	0.946126	34	2.114374	0.0419

Correlation:

	(Intr)	NAP	fExp11
NAP		-0.181	
fExp11		-0.661	0.120
NAP:fExp11		0.120	-0.664