Modelação Ecológica

AULA 18

19th November 2019
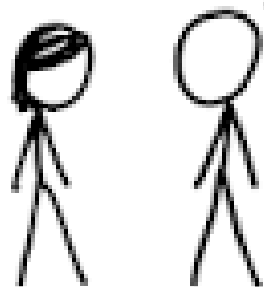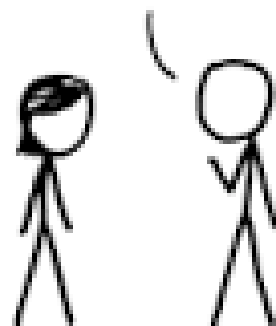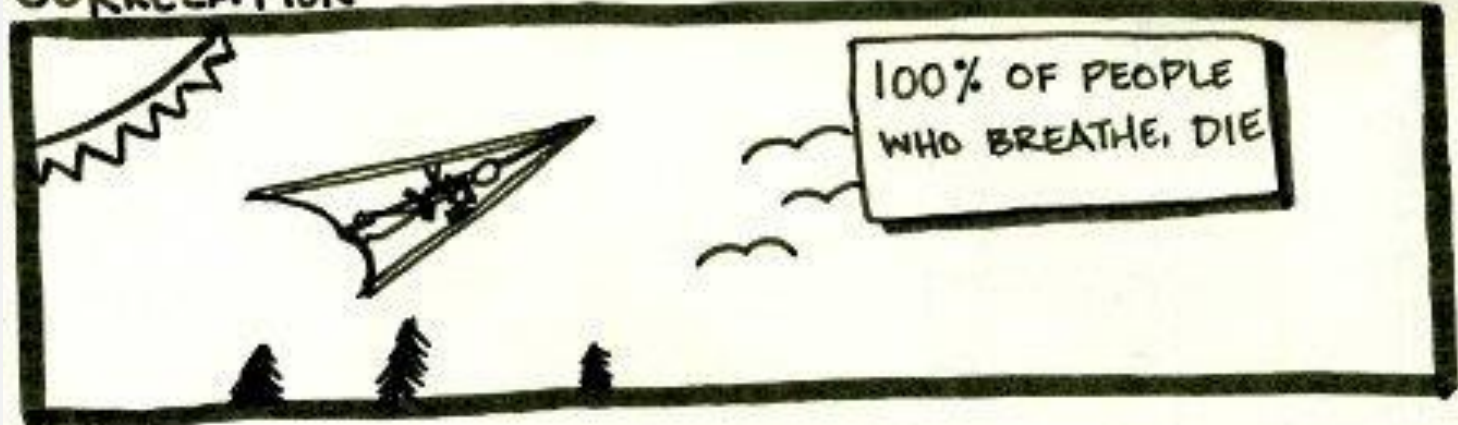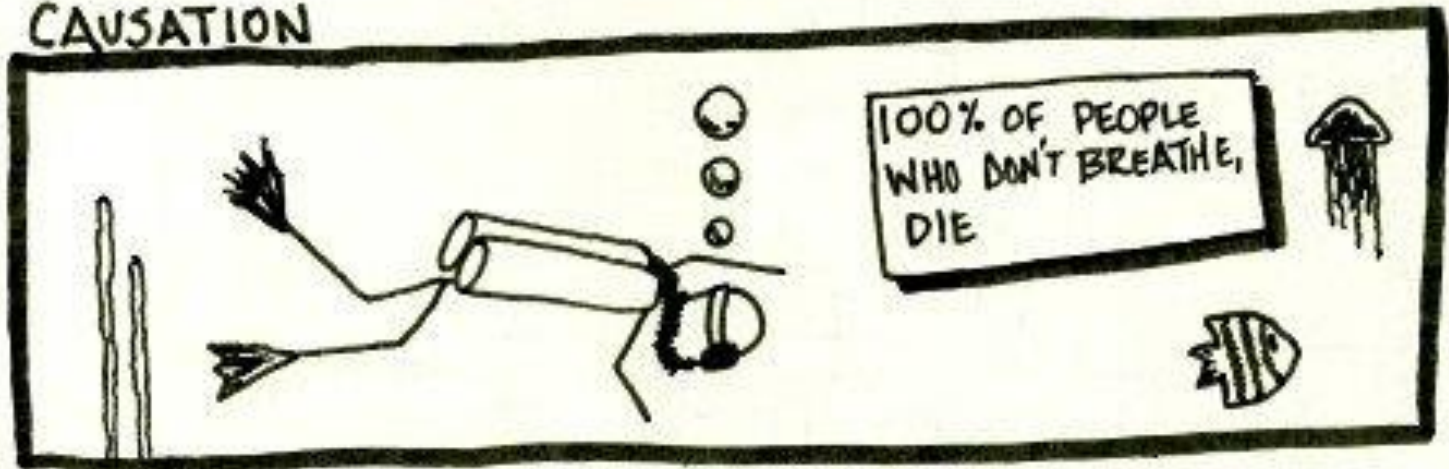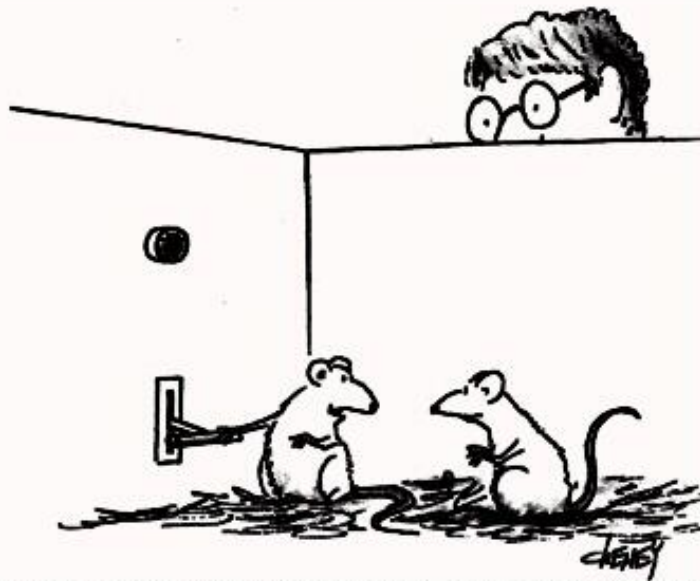
It's a rather interesting phenomenon. Every time I press this lever, that post-graduate student breathes a sigh of relief.

"Do you think all these film crews brought on global warming or did global warming bring on all these film crews?"

CORRELATION DOES NOT IMPLY CAUSATION.

AL-QAEDA
IRAQ

"WE'VE ESTABLISHED A CLEAR LINK."

ICE CREAM

causation

correlation

causation
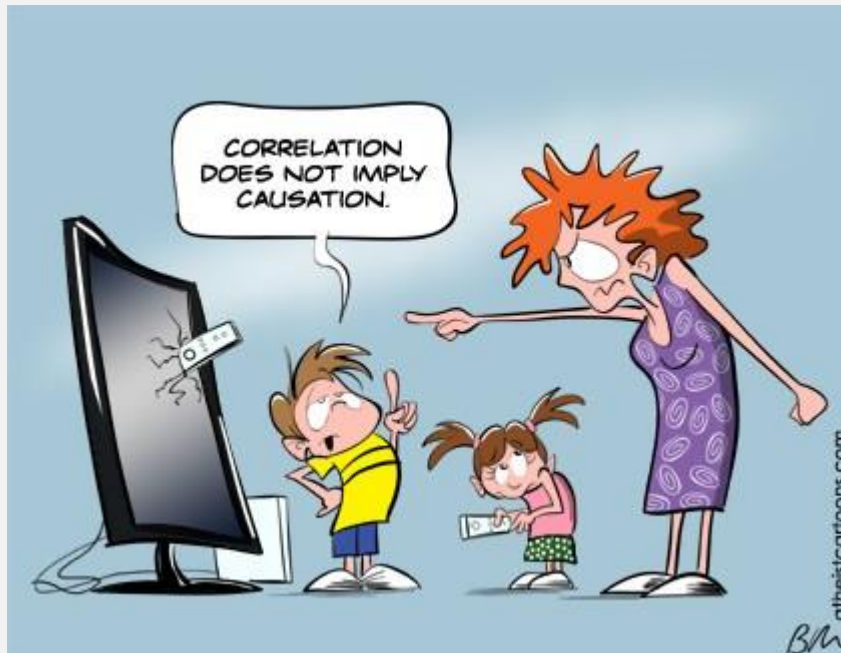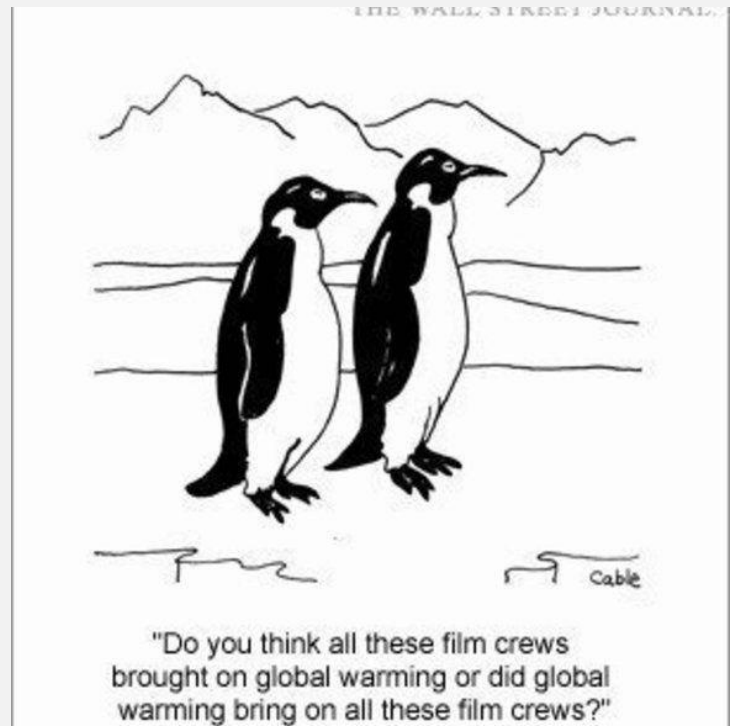
DRY, HOT AND SUNNY
SUMMER WEATHER

SUNBURN

ESDM: A tool for creating and exploring ensembles of predictions from species distribution and abundance models

Samuel M. Woodman ✉, Karin A. Forney, Elizabeth A. Becker, Monica L. DeAngelis, Elliott L. Hazen, Daniel M. Palacios, Jessica V. Redfern

Links to: theoretical work, multi-model inference, model averaging!

Dear MARMAM community,

On behalf of my co-authors, I am pleased to announce the following publication in Methods in Ecology and Evolution:

Woodman, S.M., Forney, K.A., Becker, E.A., DeAngelis, M.L., Hazen, E.L., Palacios, D.M., Redfern, J.V. (2019). eSDM: A tool for creating and exploring ensembles of predictions from species distribution and abundance models. Methods in Ecology and Evolution. 2019;00:1-11. doi:10.1111/2041-210X.13283

The paper is open access and is available at https://doi.org/10.1111/2041-210X.13283

The abstract is below, while more information about the eSDM R package and accompanying GUI can be found at https://github.com/smwoodman/eSDM (the package is also on CRAN).

Dealing with correlation
Random Effects, Mixed Models &
Generalized Estimating Equations

Mixed Models

Until now, all the independent variables were fixed effects. Today we introduce another type of variables, **random effects**.

The term **random effect** is widely used in the statistical literature, and unfortunately, it can mean a lot of different things.

Here we begin by introducing it as a variable that we are not interested in its effect per se, but one which we know might influence the response variable.

The use of **random effects** will allow us to cope with a different number of circumstances. As an example, when we collect the response variables within sampling units, the sampling unit will be typically a **random effect**.

Examples:

1. We are estimating the effect of different diets on baby dogs. Each litter is assigned to a diet. We collect the weight at birth and the weight after 1 month of the dogs. Litter is a **random effect** – we know different litters might have different weight gains irrespective of diet, due to e.g. genetic effects, but what we want is to evaluate the impact of the diet.
2. One is looking at estimating the effect of drinking on the performance on a given test. To do so each person is given x (x=1,2,3,4) glasses of beer in each of 4 days, selected at random in a given 15 day period. Person is a **random effect** – we want to account the variability of each person, but we do not care that some people react better or worse to drinking, our main interest is on the overall effect of drinking on the task.

By far the most common and easily understood type of random effect are the blocks in experiments or observational studies that are replicated across sites or times.

Random effects also encompass variation among individuals (when multiple responses are measured per individual, such as survival of multiple offspring or sex ratios of multiple broods), genotypes, species and regions or time periods.

When we have a model that includes both **fixed effects** and **random effects** we call it a **mixed model**

	package	function
Mixed models can be implemented in R via:	`nlme`	`lme`
	`lme4`	`lmer`

Just as for (simpler) linear models, **mixed models** can be extended in two ways:

1. Generalized linear mixed models – GLMMs (implemented e.g. via `lme4`, `glmm`)
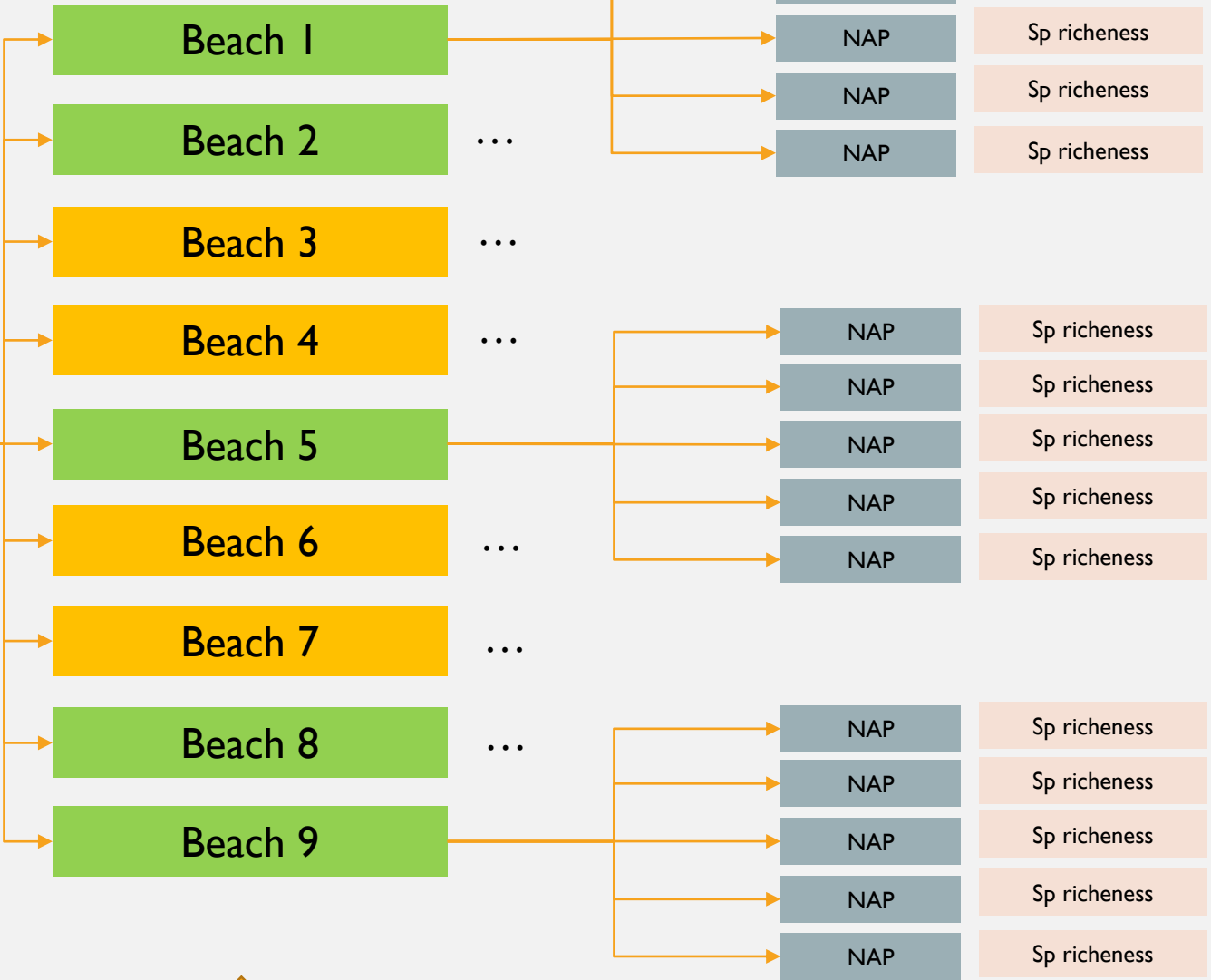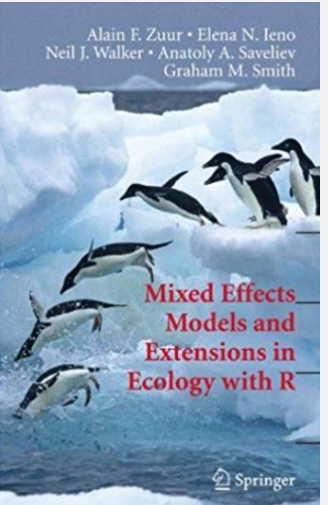2. Generalized additive mixed models – GAMMs (implemented e.g. via `mgcv`, `gamm`)

A note: fitting these is no longer a trivial matter (as fitting GLMs and GAMs perhaps was), and we often find ourselves facing strange error messages with convergence problems, or variances estimated as 0, etc. There are different ways to implement these models and which might be the best is beyond what we can deal with in "Modelação Ecológica". Some of these topics are still current research topics, so we have reached the point where no one really has definitive answers on what is the best way to proceed. Conceptually we know what to do, in practice, it might be hard to do it! But at least, you are now aware these options exist!

RIKZ.txt – species richness, as a function of station height and a measure of exposure (for more details on this data see chapter 5 in Zuur et al. 2009)
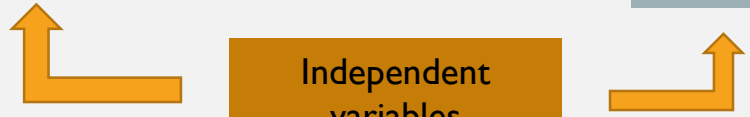
Mixed Effects Models and Extensions in Ecology with R

Alain F. Zuur · Elena N. Ieno
Neil J. Walker · Anatoly A. Saveliev
Graham M. Smith

Springer

NAP = height of sampling station compared to mean tidal level

Sampling unit: a site within a beach

Exposure=0   Exposure=1

Beach 1
Beach 2 …
Beach 3 …
Beach 4 …
Beach 5
Beach 6 …
Beach 7 …
Beach 8 …
Beach 9

RIKZ

NAP   Sp richeness
NAP   Sp richeness
NAP   Sp richeness
NAP   Sp richeness
NAP   Sp richeness

NAP   Sp richeness
NAP   Sp richeness
NAP   Sp richeness
NAP   Sp richeness
NAP   Sp richeness

NAP   Sp richeness
NAP   Sp richeness
NAP   Sp richeness
NAP   Sp richeness
NAP   Sp richeness

Independent variables

How do we explain the variability in species richness (R) as a function of NAP and Exposure

$$R_{ij} = a + b \times NAP_{ij} + c \times Exposure_i + e_{ij}$$

$i=1,2,\ldots,9$ beaches
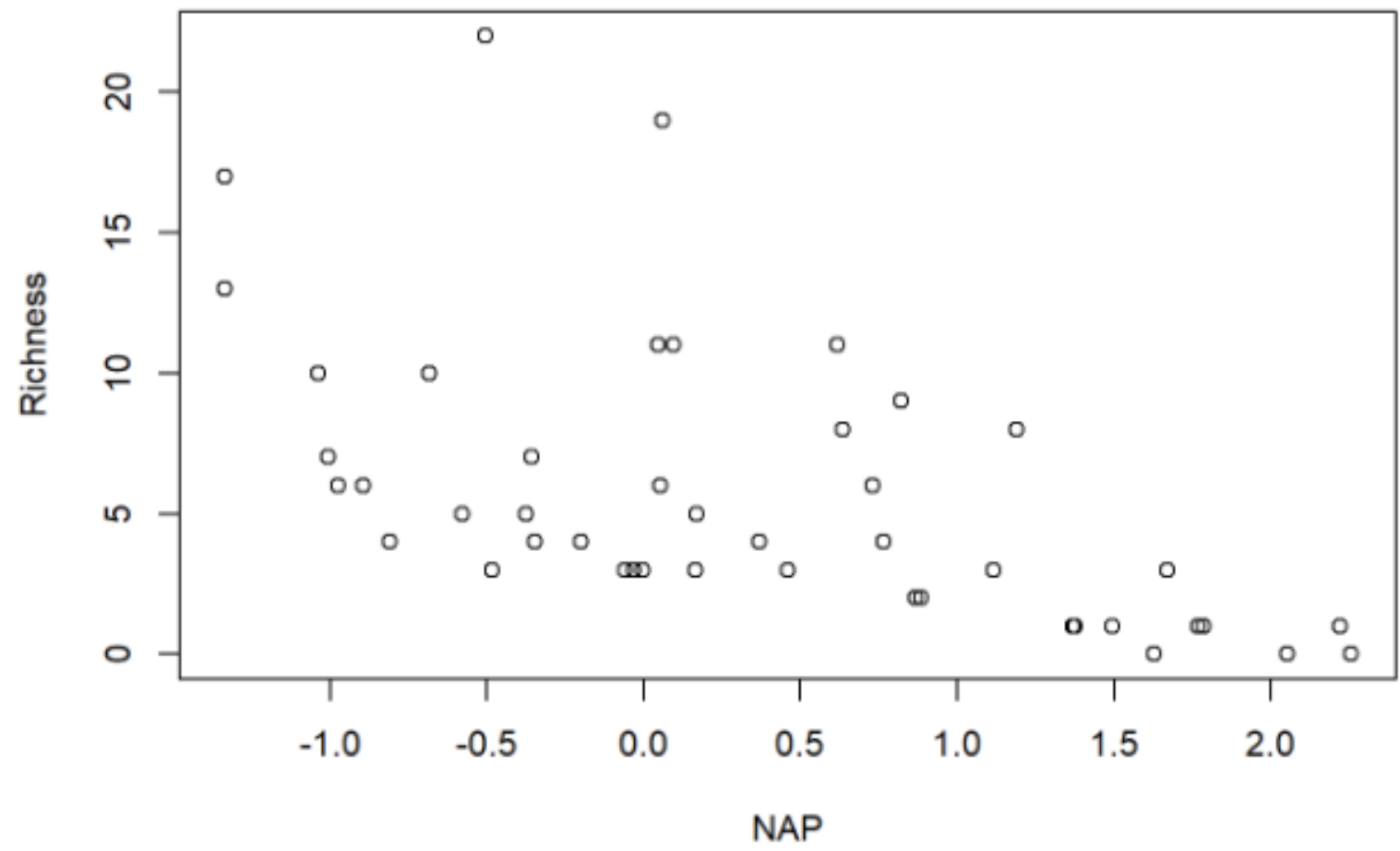$j=1,2,\ldots,5$ replicates

```
str(RIKZ)
```

```
## '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 ...
```

```
head(RIKZ)
```

```
##   Sample Richness Exposure    NAP Beach
## 1      1       11       10  0.045     1
## 2      2       10       10 -1.036     1
## 3      3       13       10 -1.336     1
## 4      4       11       10  0.616     1
## 5      5       10       10 -0.684     1
## 6      6        8        8  1.190     2
```
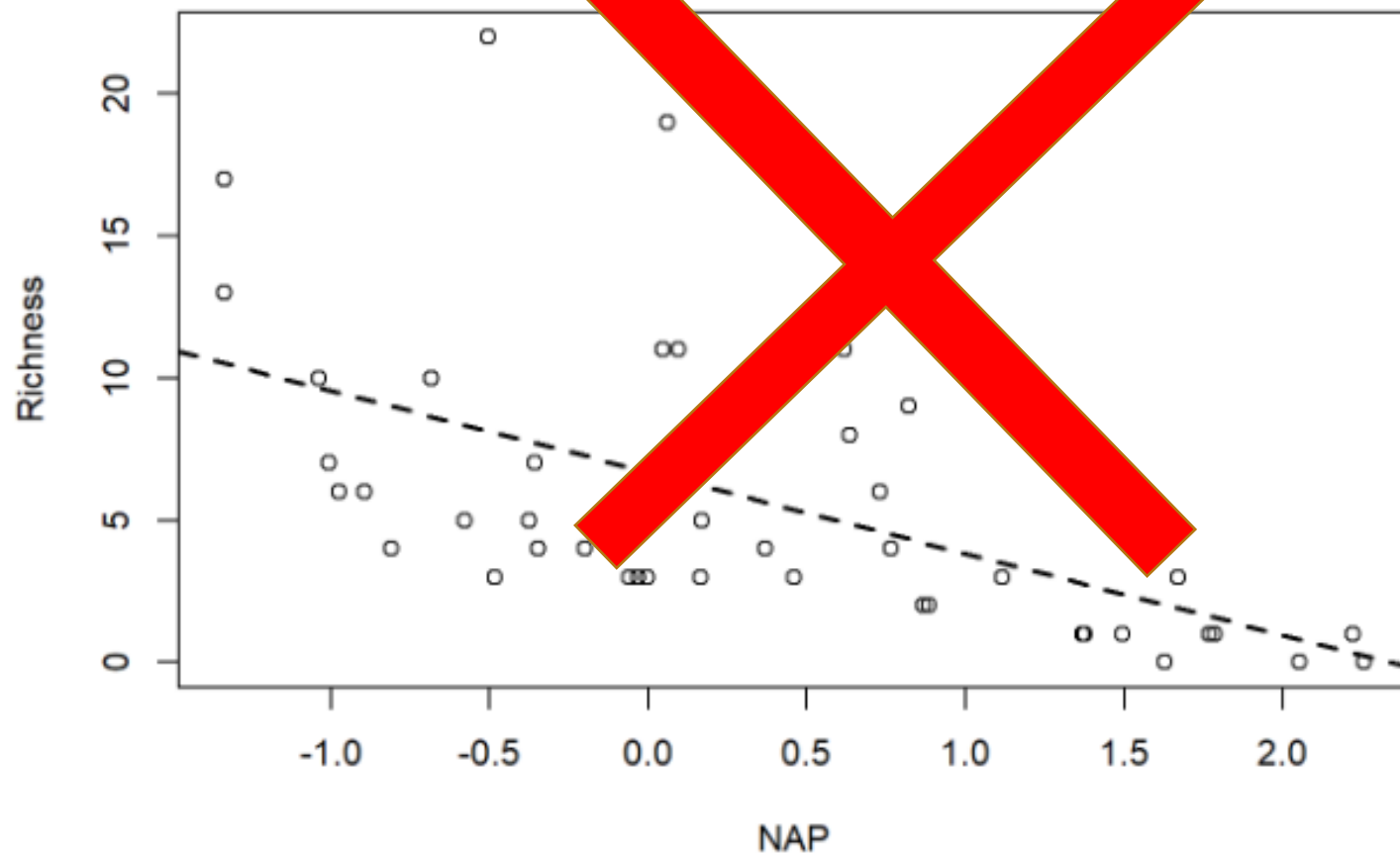
```
plot(Richness~NAP,data=RIKZ)
```

Ignoring the hierarchical nature of the data – assuming independent observations

```
plot(Richness~NAP,data=RIKZ)
abline(lm(Richness~NAP,data=RIKZ),lwd=2,lty=2)
```

$Richness_r = a + b*NAP_r + e_r$, r=1,2,...,45

Let's consider a two-stage approach:

1. We consider the relation between R and NAP, for each beach
2. We model the estimated coefficients per beach as a function of exposure

Stage 1 - Model the relationship within each beach

```
plot(Richness~NAP,pch=as.character(Beach),data=RIKZ)
as=numeric(9);bs=numeric(9);exp=numeric(9)
for(i in 1:9){
  m=lm(Richness~NAP,data=RIKZ[RIKZ$Beach==i,])
  cs=coef(m)
  as[i]=cs[1]
  bs[i]=cs[2]
  exp[i]=RIKZ$Exposure[i*5]
  abline(cs,lty=exp[i]+1,col=i)
}
legend("topright",lty=c(1,2),legend=c("Exposure 0","Exposure 1"),inset=0.05)
```
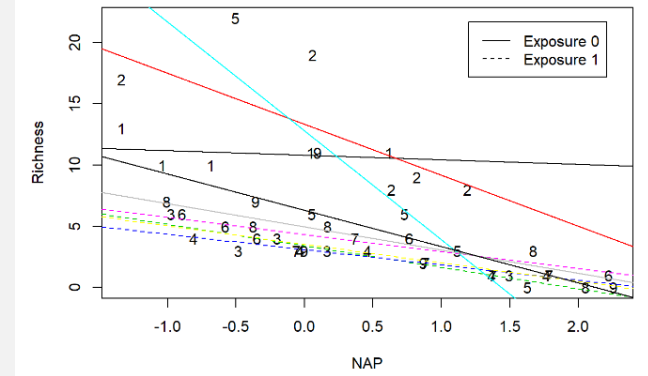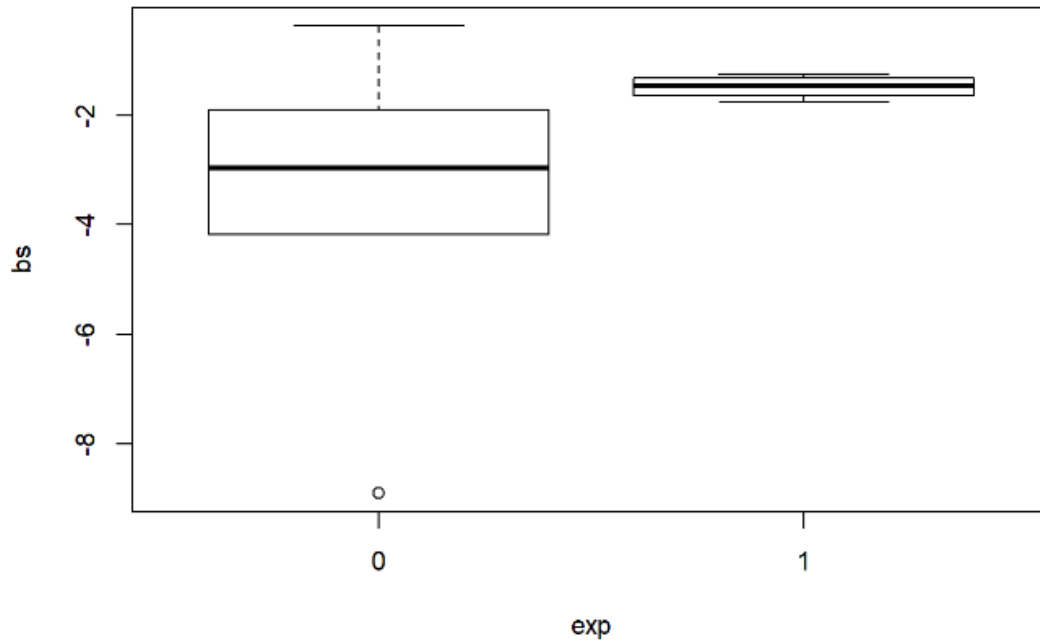
For each beach i: $Richness_{ij} = a_i + b_i*NAP_{ij} + e_{ij}$
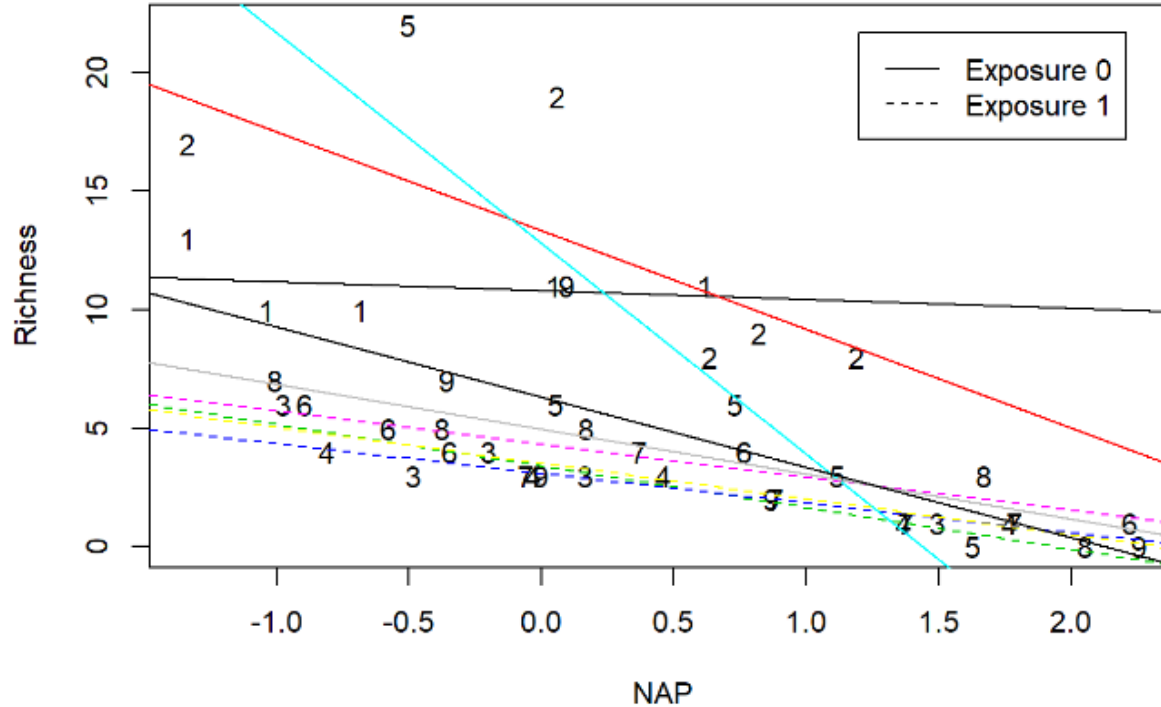
Stage II - Model the parameters as a function of exposure
(here consider the slope only)

$$b_i = d + c*Exposure_i + e_i$$

The mixed model way – a random intercept model

$$Richness_{ij} = \beta_0 + \beta_{1i} \times Beach_i + \beta_2 \times NAP_{ij} + e_{ij} \implies Richness_{ij} = \beta_0 + \beta_{1i} + \beta_2 \times NAP_{ij} + e_{ij}$$

A fixed effect, one parameter per beach

A random effect, a value that varies by beach, with mean 0 and some variance

Using `lme` from `nlme`

```
library(nlme)
RIKZ$fbeach=as.factor(RIKZ$Beach)
lme1=lme(Richness~NAP,random=~1|fbeach,data=RIKZ)
summary(lme1)
```

Using `lmer` from `lme4`

```
library(lme4)
lme2=lmer(Richness~NAP+(1|fbeach),data=RIKZ)
summary(lme2)
```

The mixed model way – a random intercept model

```
## Linear mixed-effects model fit by REML
##   Data: RIKZ
##        AIC      BIC    logLik
##   247.4802 254.525 -119.7401
##
## Random effects:
##  Formula: ~1 | fbeach
##         (Intercept) Residual
## StdDev:   2.944065   3.05977
##
## Fixed effects: Richness ~ NAP
##                  Value Std.Error DF   t-value p-value
## (Intercept)   6.581893 1.0957618 35  6.006682       0
## NAP          -2.568400 0.4947246 35 -5.191574       0
##  Correlation:
##      (Intr)
## NAP -0.157
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med        Q3       Max
## -1.4227495 -0.4848006 -0.1576462 0.2518966 3.9793918
##
## Number of Observations: 45
## Number of Groups: 9
```

lme by nlme

The mixed model way – a random intercept model

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Richness ~ NAP + (1 | fbeach)
##     Data: RIKZ
##
## REML criterion at convergence: 239.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4228 -0.4848 -0.1576  0.2519  3.9794
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  fbeach    (Intercept) 8.668    2.944
##  Residual              9.362    3.060
## Number of obs: 45, groups:  fbeach, 9
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   6.5819     1.0958   6.007
## NAP          -2.5684     0.4947  -5.192
##
## Correlation of Fixed Effects:
##     (Intr)
## NAP -0.157
```

lmer by lme4

No p-values in lmer…!

A highly controversial topic…:

https://stats.stackexchange.com/questions/118416/getting-p-value-with-mixed-effect-with-lme4-package

Now we have two levels of predictions. Those at the global level, and those at the beach level

$$\text{Richness}_{ij} = \beta_0 + \beta_{1i} + \beta_2 \times \text{NAP}_{ij} + e_{ij}$$

Make predictions from lme
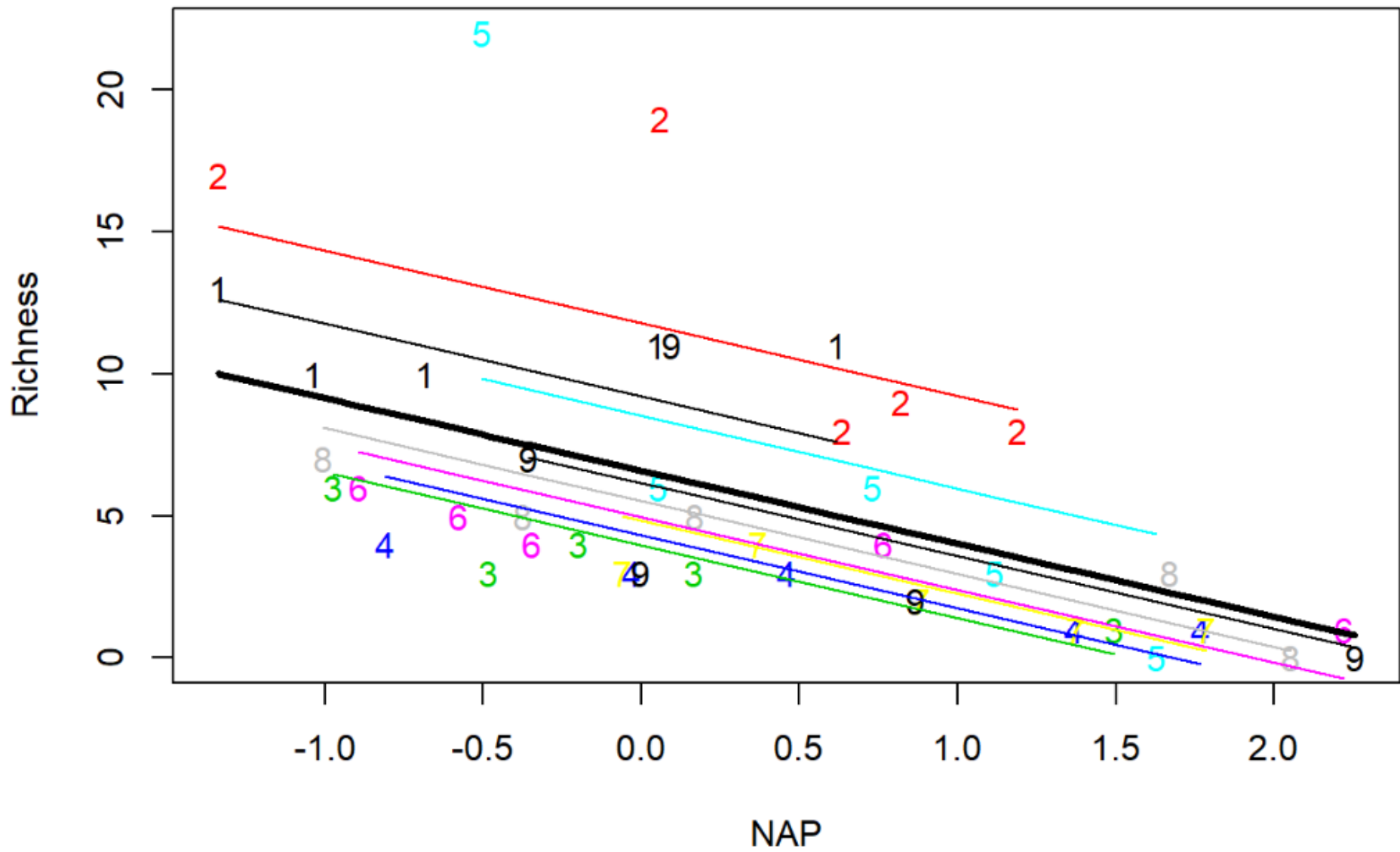
```
Level0=fitted(lme1,level=0)            →   population or global level
Level1=fitted(lme1,level=1)            →   beach level
```

Look at the model plotted

```
I=order(RIKZ$NAP)
NAPs=sort(RIKZ$NAP)
plot(Richness~NAP,pch=as.character(Beach),col=Beach,data=RIKZ)
lines(NAPs,Level0[I],lwd=3)
for(j in 1:9){
  #index for beach
  bi=RIKZ$Beach==j
  xs=RIKZ$NAP[bi]
  ys=Level1[bi]
  Oxs=order(xs)
  lines(sort(xs),ys[Oxs],col=j)
}
```

The mixed model way
– a random intercept and random slope model

$$Richness_{ij} = \beta_0 + \beta_{1i} \times Beach_i + \beta_2 \times NAP_{ij} + \beta_{3i} \times Beach_i \times NAP_{ij} + e_{ij}$$

A fixed effect, one parameter per beach

A fixed interaction, one parameter per beach

$$Richness_{ij} = \beta_0 + \beta_{1i} + \beta_{2i} \times NAP_{ij} + e_{ij}$$

Two random effects, one for the slope and one for the intercept, both varying by beach, with mean 0 and some variance (2 variances, one for slopes and one for intercepts)

Using `lme` from `nlme`

```
library(nlme)
RIKZ$fbeach=as.factor(RIKZ$Beach)
lme3=lme(Richness~NAP,random=~NAP|fbeach,data=RIKZ)
summary(lme3)
```

Using `lmer` from `lme4`

```
library(lme4)
lme5=lmer(Richness~NAP+(NAP|fbeach),data=RIKZ)
summary(lme5)
```

```
## Linear mixed-effects model fit by REML
##   Data: RIKZ
##        AIC      BIC    logLik
##    244.3839 254.9511 -116.1919
##
## Random effects:
##  Formula: ~NAP | fbeach
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev   Corr
## (Intercept) 3.549064 (Intr)
## NAP         1.714963 -0.99
## Residual    2.702820
##
## Fixed effects: Richness ~ NAP
##                 Value Std.Error DF   t-value p-value
## (Intercept)  6.588706  1.264761 35  5.209448   0e+00
## NAP         -2.830028  0.722940 35 -3.914610   4e-04
##  Correlation:
##     (Intr)
## NAP -0.819
##
## Standardized Within-Group Residuals:
##        Min        Q1        Med        Q3        Max
## -1.8213326 -0.3411043 -0.1674617  0.1921216  3.0397049
##
## Number of Observations: 45
## Number of Groups: 9
```

lme by nlme

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Richness ~ NAP + (NAP | fbeach)
##    Data: RIKZ
##
## REML criterion at convergence: 232.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8216 -0.3415 -0.1673  0.1931  3.0412
##
## Random effects:
##  Groups    Name        Variance Std.Dev. Corr
##  fbeach    (Intercept) 12.630   3.554
##            NAP          2.942   1.715   -0.99
##  Residual               7.303   2.702
## Number of obs: 45, groups:  fbeach, 9
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)    6.588       1.266   5.203
## NAP           -2.830       0.723  -3.914
##
## Correlation of Fixed Effects:
##     (Intr)
## NAP -0.820
## convergence code: 0
## Model failed to converge with max|grad| = 0.00721061 (tol = 0.002, component 1)
```

lmer by lme4
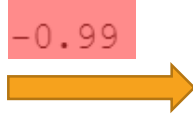
Note lower variability remains unexplained
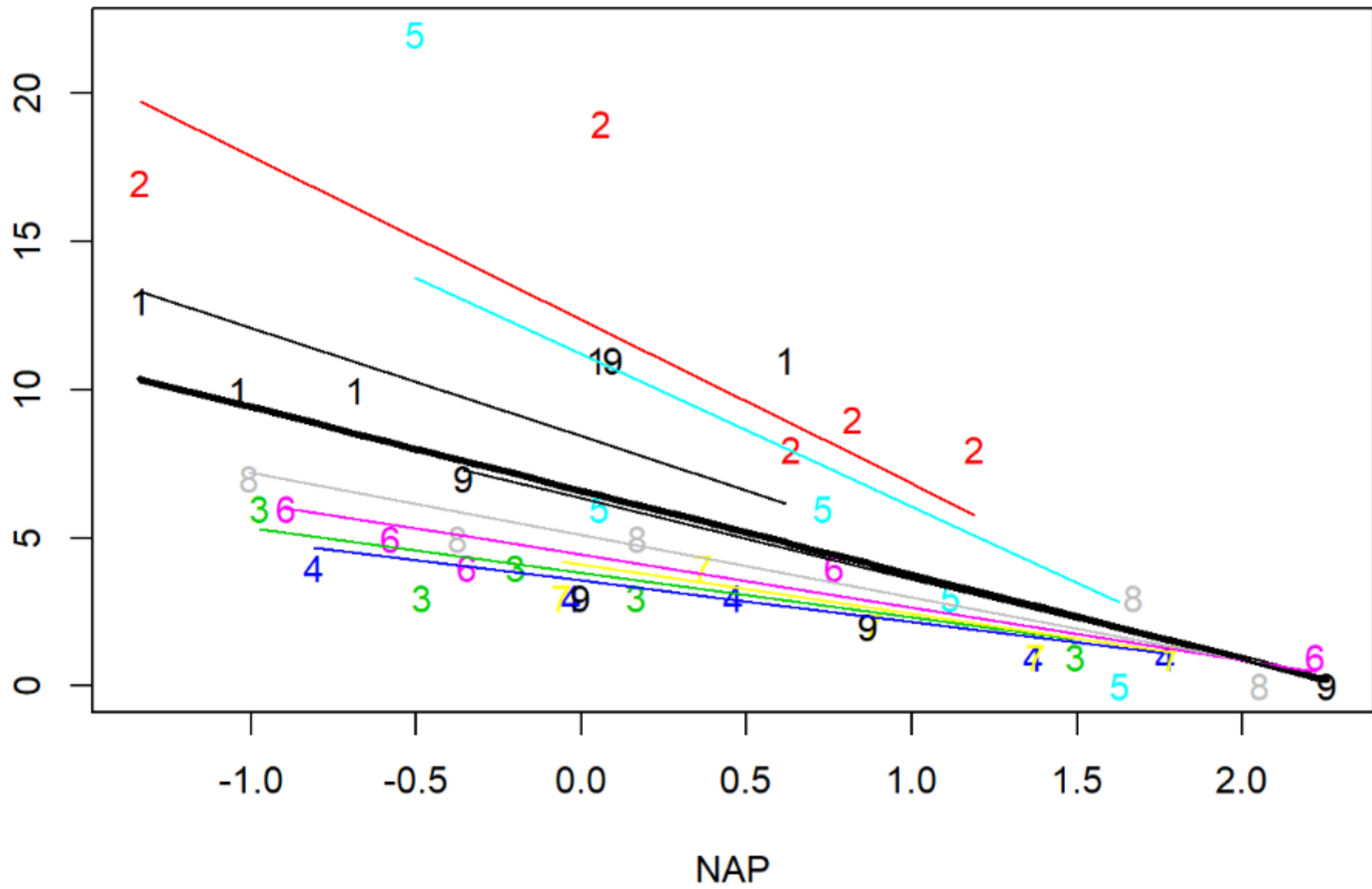compared to random-intercept-only model

Make predictions from lme

```r
Level0.3=fitted(lme3,level=0)
Level1.3=fitted(lme3,level=1)
```

and ploting

```r
I=order(RIKZ$NAP)
NAPs=sort(RIKZ$NAP)
plot(Richness~NAP,pch=as.character(Beach),col=Beach,data=RIKZ)
lines(NAPs,Level0.3[I],lwd=3)
for(j in 1:9){
  #index for beach
  bi=RIKZ$Beach==j
  xs=RIKZ$NAP[bi]
  ys=Level1.3[bi]
  Oxs=order(xs)
  lines(sort(xs),ys[Oxs],col=j)
}
```

A random effects model (a different random mean per beach, NAP not relevant)

```
lme6=lme(Richness~1,random=~1|fbeach,data=RIKZ)
```

```
summary(lme6)
```

```
## Linear mixed-effects model fit by REML
##  Data: RIKZ
##       AIC      BIC    logLik
##   267.1142 272.4668 -130.5571
##
## Random effects:
##  Formula: ~1 | fbeach
##         (Intercept) Residual
## StdDev:    3.237112 3.938415
##
## Fixed effects: Richness ~ 1
##                 Value Std.Error DF  t-value p-value
## (Intercept) 5.688889  1.228419 36 4.631066       0
##
## Standardized Within-Group Residuals:
##         Min         Q1        Med         Q3        Max
## -1.77968689 -0.50704111 -0.09795286  0.25468670  3.80631705
##
## Number of Observations: 45
## Number of Groups: 9
```

Comparing the 3 models

```
AIC(lme1,lme3,lme6)
```

```
## Warning in AIC.default(lme1, lme3, lme6): models are not all fitted to the
## same number of observations
```

```
##      df      AIC
## lme1  4 247.4802
## lme3  6 244.3839
## lme6  3 267.1142
```

Model selection in a mixed model context (a possible top-down approach)

1. Start from a full model with all relevant fixed effects

2. Fit different random components

3. Select the most parsimonious one (e.g. via AIC)

4. Conditional on that random effect structure, select the relevant (fixed) effects

Your task: find the best model for the RIKZ data set

A talk delivered at

See slides "A17 MarquesSPE2019.pptx"

Just as in GLMs, non-normal responses and link functions get added to models with random effects

Cell PRESS

Generalized linear mixed models: a practical guide for ecology and evolution

Benjamin M. Bolker[1], Mollie E. Brooks[1], Connie J. Clark[1], Shane W. Geange[2], John R. Poulsen[1], M. Henry H. Stevens[3] and Jada-Simone S. White[1]

"… Despite the availability of accurate techniques for estimating GLMM parameters in simple cases, complex GLMMs are challenging to fit and statistical inference such as hypothesis testing remains difficult…"

"…GLMMs are surprisingly challenging to use even for statisticians. Although several software packages can handle GLMMs (Table 1), few ecologists and evolutionary biologists are aware of the range of options or of the possible pitfalls.…"

https://biologyforfun.wordpress.com/2014/03/12/generalized-linear-mixed-models-in-ecology-and-in-r/

If you are looking at this slide,
I have pressed click one time too many
☺
!