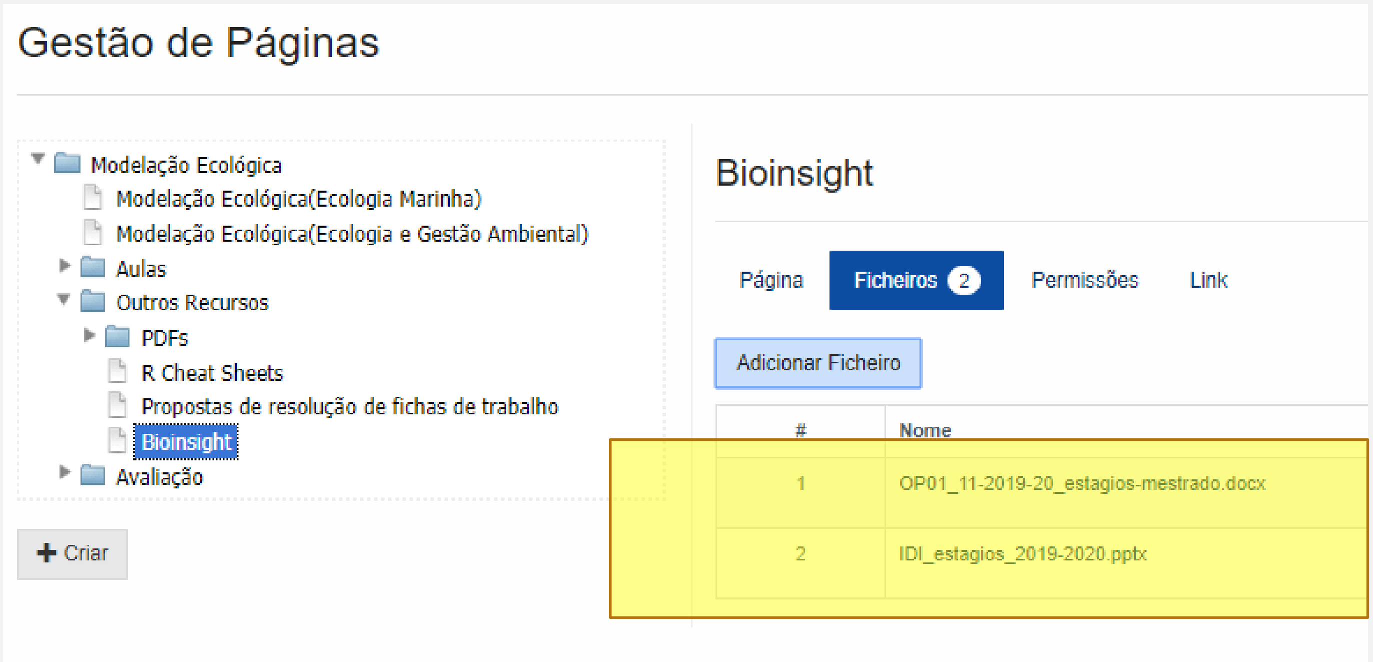
Modelação Ecológica

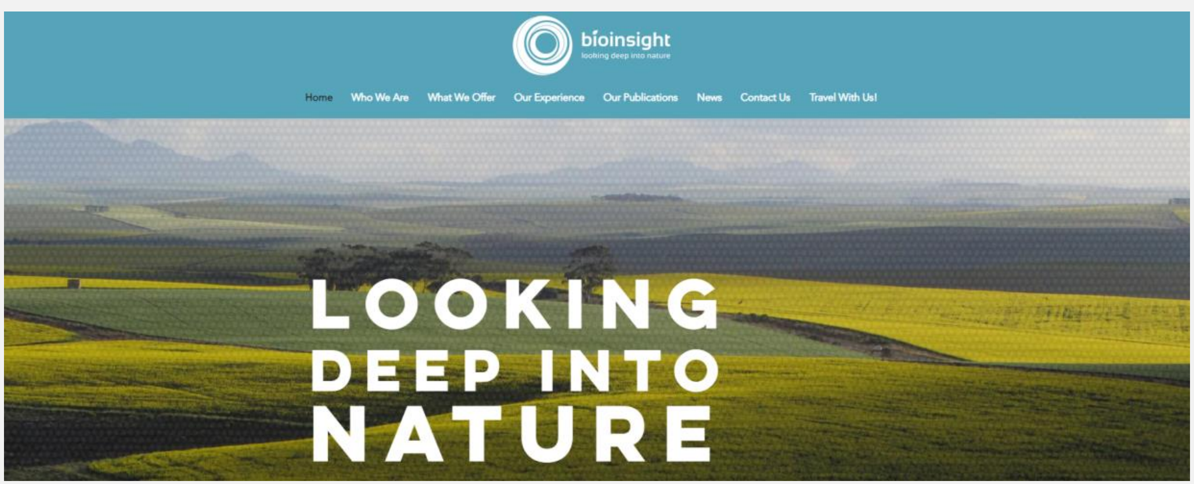
AULA 12

23 October 2019 – 14:00-16:30 – room 2.3.16

Tiago A. Marques

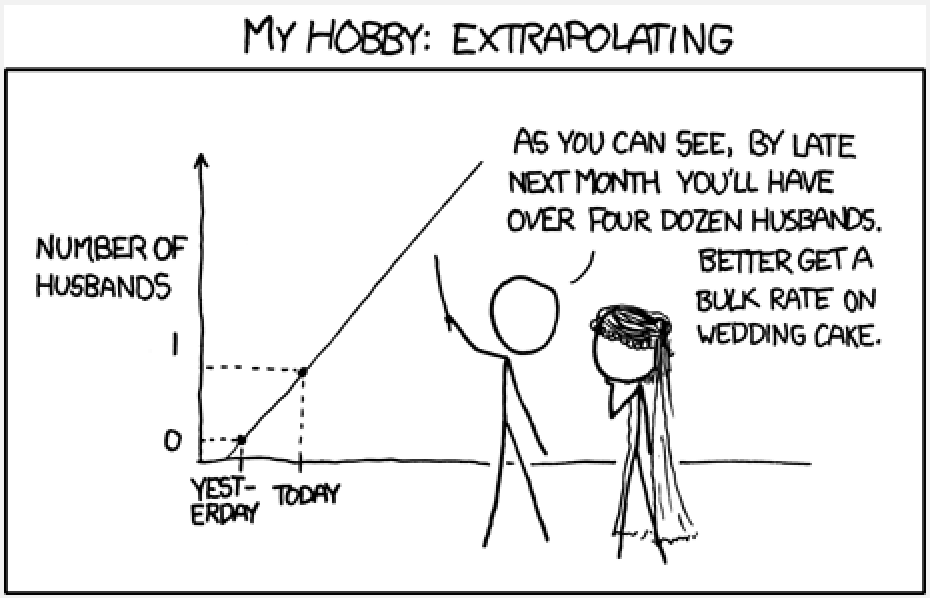
Visita da Bioinsight – ontem (22 / 10 / 2019)





http://www.bioinsight.pt/

EXTRAPOLATION!



If she loves you more each and every day,
by linear regression she hated you before you met.

Calendário Escolar 2019/2020

Início do ano letivo: 09 de setembro de 2019

1.º Semestre			
Inscrições e escolha de horários	03-09-2019	a	06-09-2019
Inscrições (1.º ano/1.ª vez)	09-09-2019	a	13-09-2019
Período de aulas (c/ 1.º ano)	**16-09-2019**	**a**	**20-12-2019**
Férias de Natal	23-12-2019	a	03-01-2020
Pausa letiva	04-01-2020	a	08-01-2020
Exames de Época Normal	**09-01-2020**	**a**	**21-01-2020**
Pausa letiva	22-01-2020	a	27-01-2020
Exames de Recurso	**28-01-2020**	**a**	**08-02-2020**
Pausa entre semestres	09-02-2020	a	16-02-2020

2.º Semestre			
Início	**17 de fevereiro de 2020**		
Período de aulas	**17-02-2020**	**a**	**29-05-2020**
Férias de Carnaval	24-02-2020	a	26-02-2020
Férias da Páscoa	08-04-2020	a	14-04-2020
Pausa letiva	30-05-2020	a	02-06-2020
Exames de Época Normal	**03-06-2020**	**a**	**20-06-2020**
Pausa letiva	21-06-2020	a	22-06-2020
Exames de Recurso	**23-06-2020**	**a**	**04-07-2020**

Exames de Época Especial	**14-07-2020**	**a**	**21-07-2020**
Férias de Verão	**27-07-2020**	**a**	**01-09-2020**

Época especial de conclusão	**Até 30 de setembro de 2020**

Obs.: A receção aos novos alunos terá lugar no dia 16 de setembro de 2019. A 30 de outubro celebrar-se-á o *Dia da Investigação* e não haverá aulas. No dia 22 de abril de 2020 (*Dia de Ciências*) não há aulas das 14h00 às 17h00.

…A 30 de outubro celebrar-se-á o *Dia da Investigação* e não haverá aulas…

https://ciencias.ulisboa.pt/pt/calend%C3%A1rio-escolar

CIÊNCIAS RESEARCH DAY

A melhor Ciência faz-se em CIÊNCIAS!

An opportunity for faculty researchers to share their work with the *in* and *out* community. **Spark your curiosity and your imagination. SAVE THE DATE!**

PROGRAMME

09:00-09:15 Welcome words (L. Carriço, Dean)

09:15-09:30 Facts and figures about research@CIÊNCIAS (M Santos-Reis, Vice Dean for Research)

09:30-10:15 **SESSION I – Top Notch Science**

10:15-10:45 Coffee-break

10:45-12:00 **SESSION I – Top Notch Science (cont)**

12:00 – 14:00 *Bring a sandwich, look at the posters and have a speed date*

14:00-15:00 **SESSION II – Recognising Excellence**

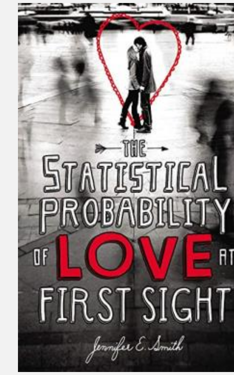
15:00-16:00 **SESSION III – Networking and Science for Society**

16:00-16:30 Coffee Break

16:30-17:00 **SESSION IV – Challenging Ideas for Ciências Creative Minds Contest**

17:00-17:15 Closing remarks and Awards (Pedro Almeida, Vice Dean for Communication and Image)

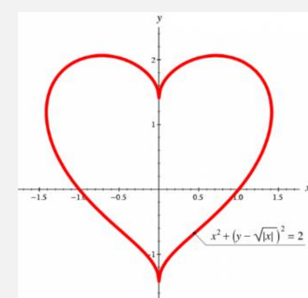
SPEED DATING A STATISTICIAN



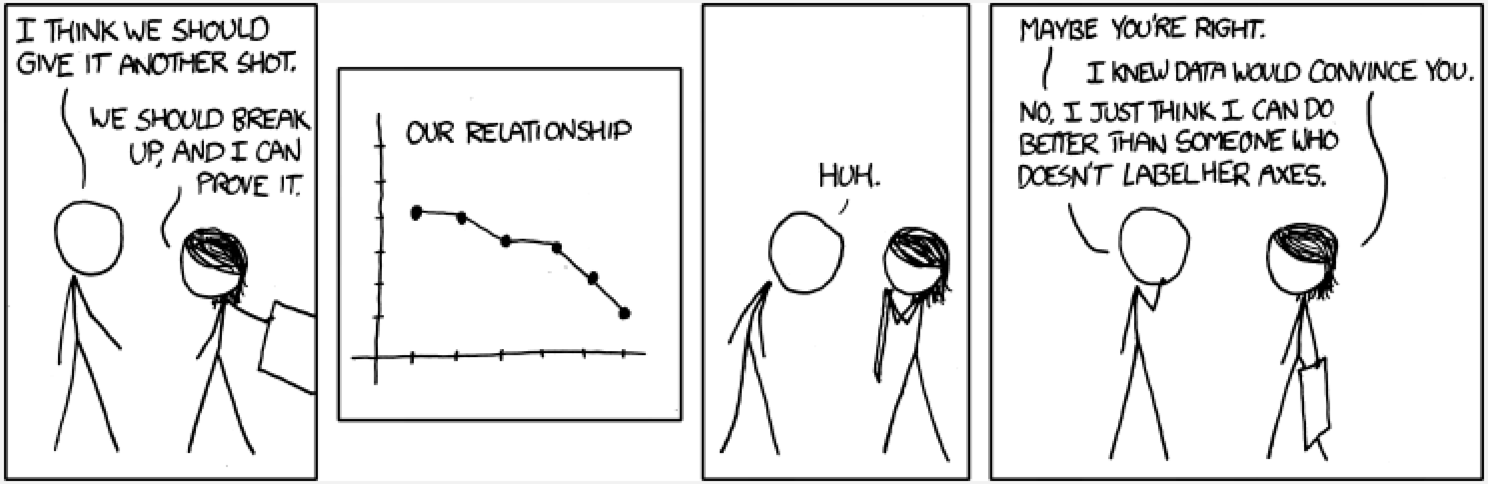
In a time where you can be "Married at first sight" or wonder about "Who wants to marry a farmer?", the concept of speed dating has gained new meanings.

Here we raise the bar and propose a date with a statistician. Does not sound that exciting? What if we propose to solve your statistical problem in 5 minutes? Surely that must mean we are headed to love at first sight! Come and see the romance behind statistics by "Speed dating a statistician".

It will significantly improve your day ☺ !



YOU CAN DO BETTER!



What does it all mean?

The misunderstanding of "significance" is part of a wider problem with the way statistical language is used – and a change is long overdue, says **Neil Sheldon**

The call to abandon – or, at the very least, rethink – the use of "significance" in statistics is increasingly familiar. This call has, in fact, two strands to it – strands that are distinct but related.

An old argument, one that can be found in R. A. Fisher's early writings on the subject, makes the point that statistical significance is not the same thing as practical significance. In more recent years, the practical significance of a result has been quantified by the use of effect sizes. Very roughly, the effect size provides an indication of the practical importance of a finding.

A newer argument is that the very concept of statistical significance is flawed in that it fails to deliver what it appears to offer. Statistical significance is a measure of the likelihood of the data given the null hypothesis, but what we actually want to know is the likelihood of the hypothesis given the data.

Neil Sheldon was a teacher for more than 40 years. He is a Chartered Statistician and a former vice-president of the Royal Statistical Society.

focuses on the humps as if they can be distributed, in a sort of latter-day *Just So* story, among camels. Clearly none of these is a good way of conveying the information. Yet it would not be surprising to come across statements like this in the media – or even in some academic writing.

For example, a newspaper website reported in 2015 that "[t]he average woman in Britain, on reaching 45, has had 1.9 children" (bit.ly/2QJELTb). This style of reporting is, of course, very common. Indeed, talking about mythical beasts – in this case, the average woman – is perhaps the most common way of presenting statistics in the popular press. The Office for National Statistics (ONS) does a little better, but when it reports that "average household size has remained at 2.4 people" (bit.ly/2QGLHAi) it is tempting to read this as if it is about the average household – a mythical beast – rather than the average *size* of households.

What does it all mean?

The misunderstanding of "significance" is part of a wider problem with the way statistical language is used – and a change is long overdue, says **Neil Sheldon**

Gestão de Páginas

- Modelação Ecológica
  - Modelação Ecológica(Ecologia Marinha)
  - Modelação Ecológica(Ecologia e Gestão Ambiental)
  - Aulas
  - Outros Recursos
    - PDFs
    - R Cheat Sheets
    - Propostas de resolução de fichas de trabalho
    - Bioinsight
  - Avaliação

+ Criar

PDFs

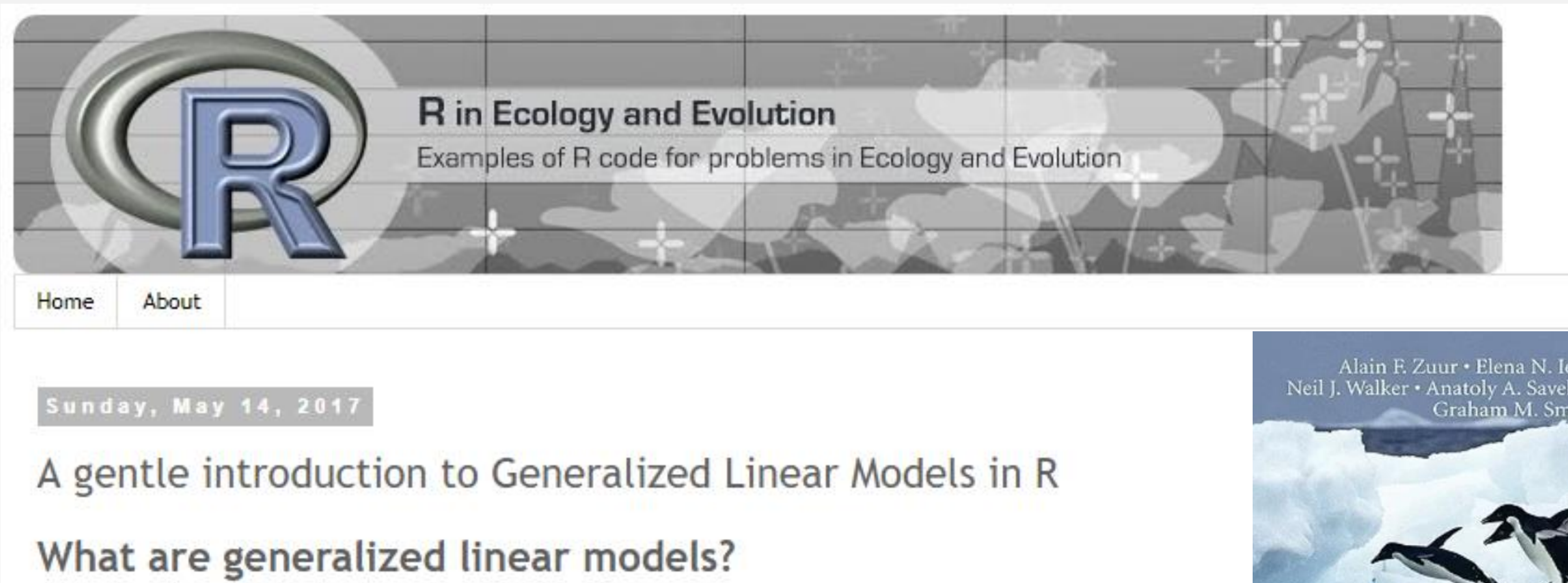
Página | Ficheiros 12 | Permissões | Link

Adicionar Ficheiro

#	Nome
1	Significance: What does it all mean? *Sheldon-2019-Significance.pdf*
2	Publication bias: What are the challenges and can they be overcome? *jpn-37-149.pdf*
3	The R Book.pdf
4	Mixed Effects Models And Extensions In Ecology With R *Zuur_Mixed-effects-models-and-extensions-in-ecology-with-R.pdf*

Ordem alterada, a partir de agora na pasta PDFs as coisas estão organizadas do mais recente para o mais antigo

Generalized Linear Models (continued!)


R in Ecology and Evolution

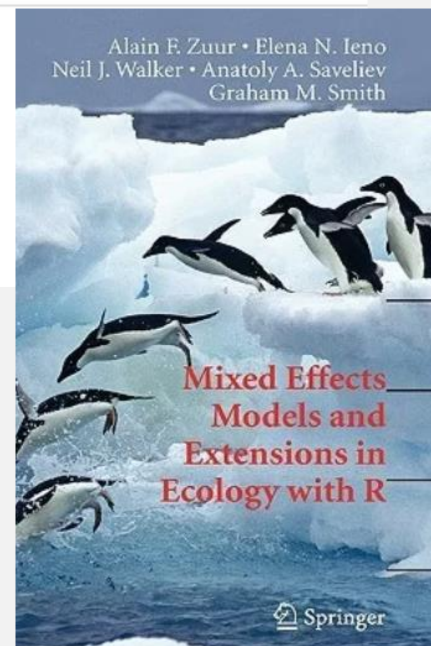
Examples of R code for problems in Ecology and Evolution

Home | About

Sunday, May 14, 2017

A gentle introduction to Generalized Linear Models in R

What are generalized linear models?

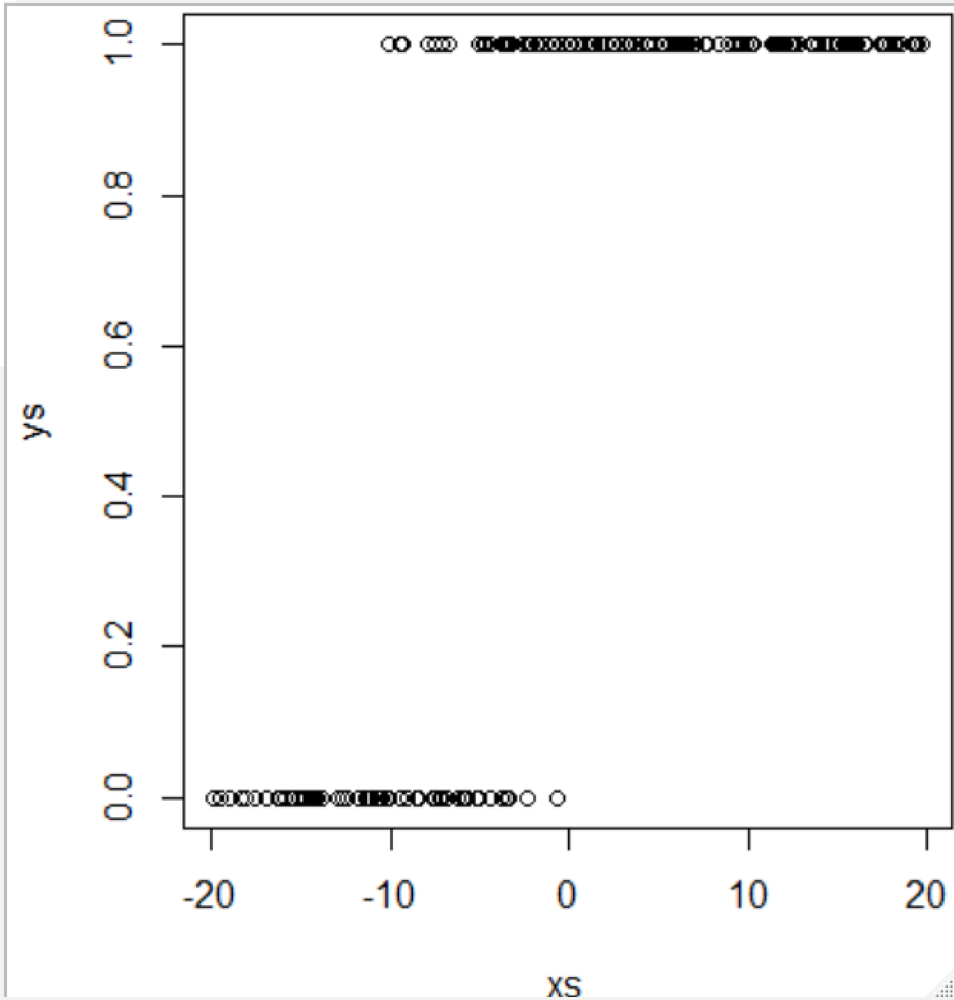


http://r-eco-evo.blogspot.com/2017/05/generalized-linear-models.html

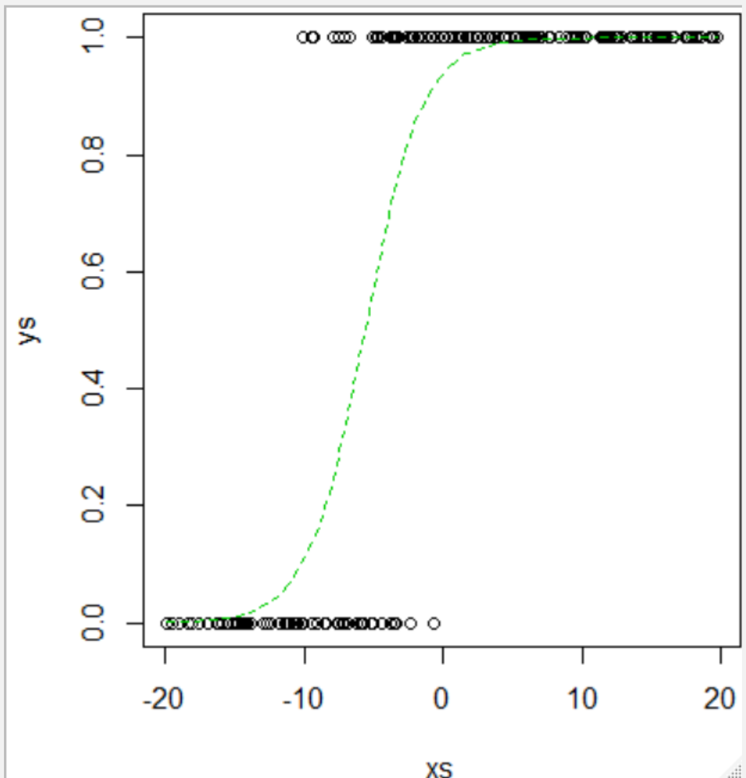
http://spatialecology.weebly.com/r-code--data/category/glm

Logistic regression: a simulated example (aka Vasco's data)

```r
#
set.seed(123)
n=200
#a covariate
xs=runif(n,-20,20)
#get the mean value
ilogit=function(x){
  il=exp(x)/(1+exp(x))
return(il)
  }
Ey=ilogit(2+0.4*xs)
#generate data
ys=rbinom(n,size = rep(1,n),prob = Ey)
#plot data
plot(xs,ys)
```



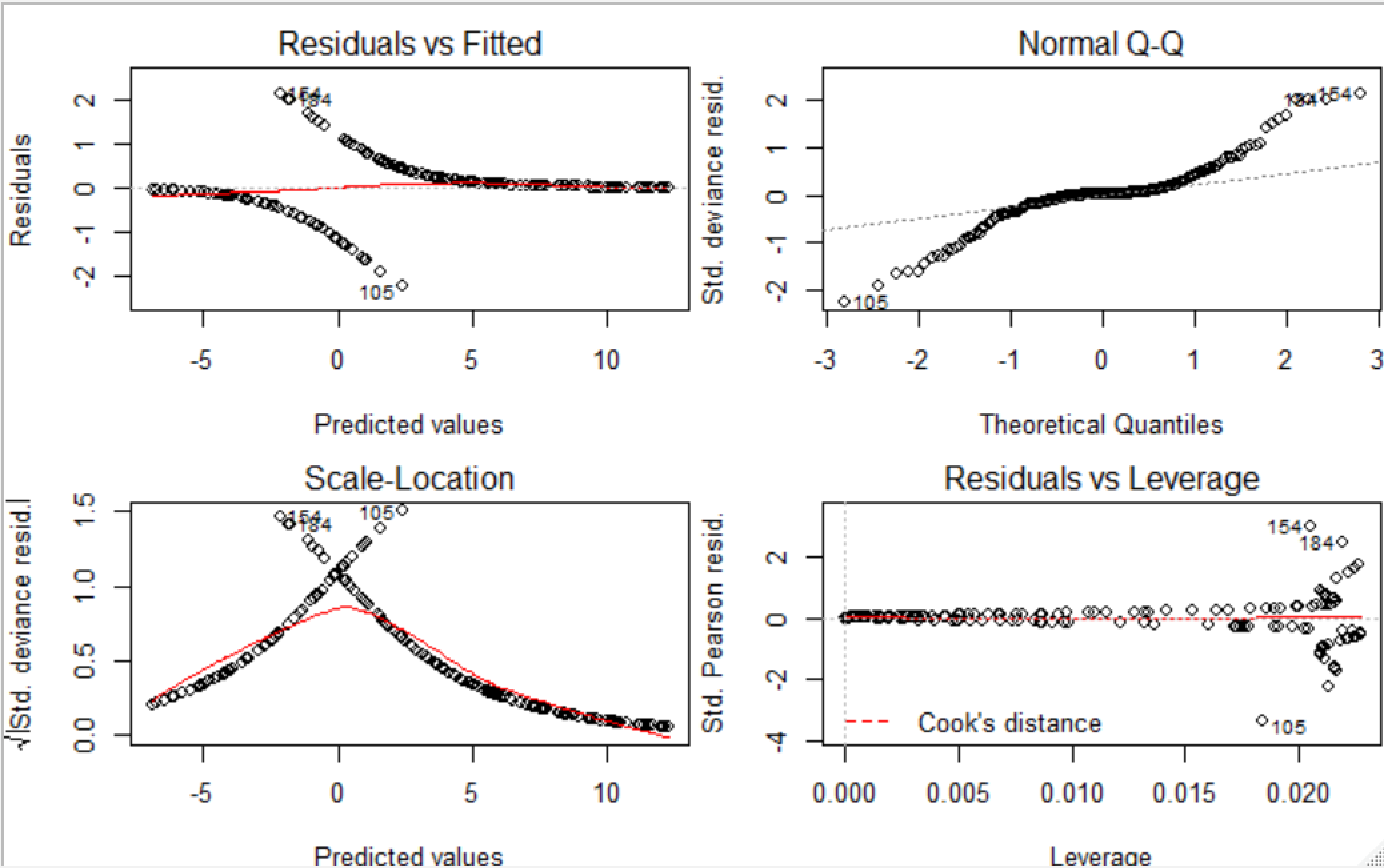
```
#run a glm
glmLR1=glm(ys~xs,family=binomial(link="logit"))
summary(glmLR1)

xs4pred=seq(min(xs),max(xs),length=100)
predglmLR1=predict(glmLR1,newdata = data.frame(xs=xs4pred),
type="response")

par(mfrow=c(1,1),mar=c(4,4,0.2,0.2))
plot(xs,ys)
lines(xs4pred,predglmLR1,col=3,lty=2)

summary(glmLR1)
```




```
> summary(glmLR1)

Call:
glm(formula = ys ~ xs, family = binomial(link = "logit"))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.22875  -0.16918   0.01288   0.14838   2.13624

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.72703    0.53664   5.082 3.74e-07 ***
xs           0.48324    0.08136   5.940 2.85e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 257.719  on 199  degrees of freedom
Residual deviance:  77.546  on 198  degrees of freedom
AIC: 81.546

Number of Fisher Scoring iterations: 7
```

```
#check model fit
par(mfrow=c(2,2))
plot(glmLR1)
```

The diagnostic plots look horrible… and yet, true model was used: next to impossible to do model diagnostics for logistic regression!



When covariates are strongly correlated

Imagine the following reality, which is actually a highly likely reality:

1. You have multiple environmental covariates, correlated amongst themselves
2. You have one response variable, that depends on some variables but not others

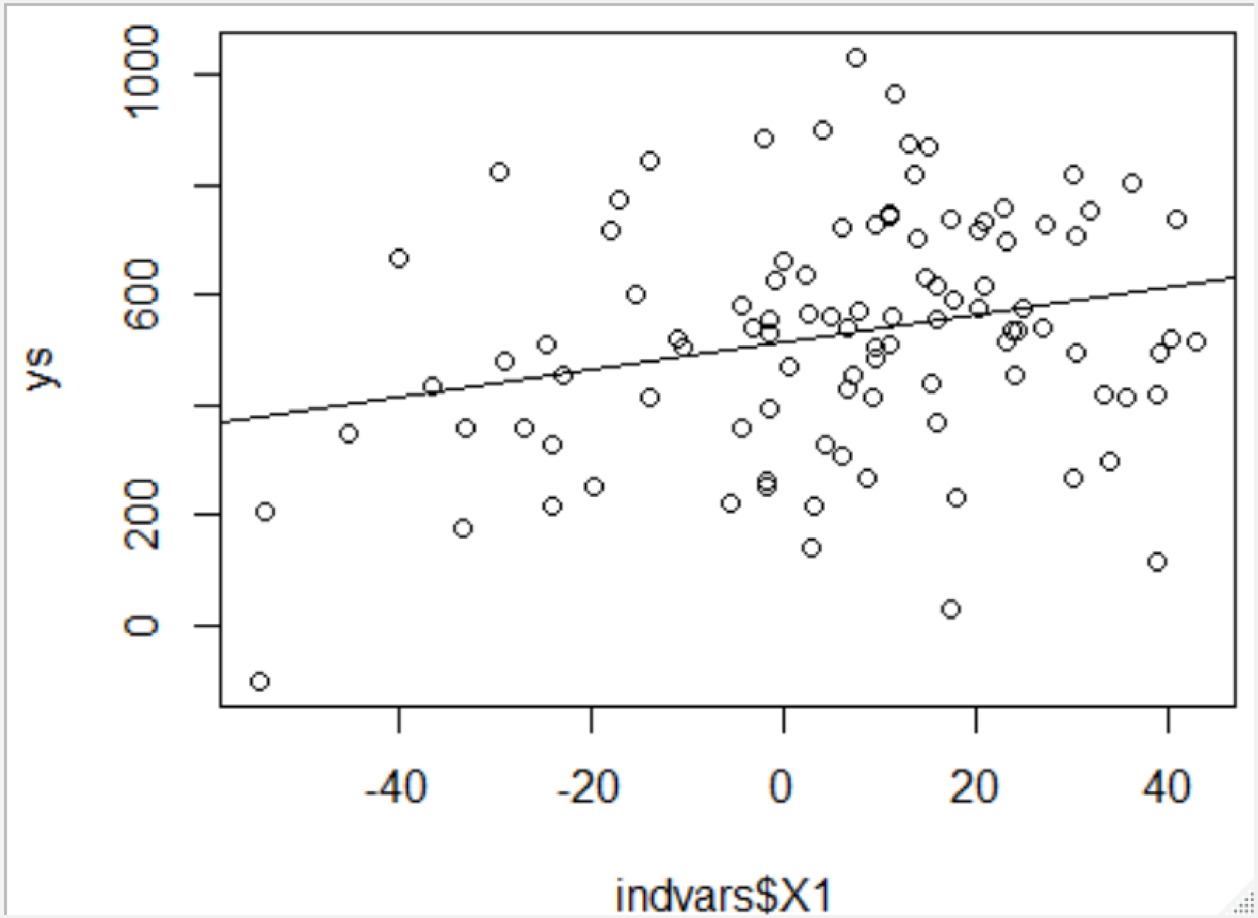
```
library(MASS)
set.seed(1234)
n=100
means <- c(2,4,6,8,10,12)
ncovs=(36-6)/2
covs<- rnorm(ncovs,mean=10,sd=2)
varcovars=matrix(NA,6,6)
varcovars[lower.tri(varcovars)]=covs
varcovars=t(varcovars)
varcovars[lower.tri(varcovars)]=covs
diag(varcovars)=means
varcovars=t(varcovars) %*% varcovars
indvars <- mvrnorm(n = n, mu=means, Sigma=varcovars)
```

This generates covariates with strong dependence – don't ask me about this, it took me hours to get it. Just assume this was your data!

```
> round(cor(indvars),2)
     X1   X2   X3   X4   X5   X6
X1 1.00 0.91 0.83 0.84 0.95 0.88
X2 0.91 1.00 0.91 0.95 0.96 0.97
X3 0.83 0.91 1.00 0.98 0.92 0.98
X4 0.84 0.95 0.98 1.00 0.93 0.98
X5 0.95 0.96 0.92 0.93 1.00 0.96
X6 0.88 0.97 0.98 0.98 0.96 1.00
```

```
ys <-510+4*indvars$X1+rnorm(n,mean=0,sd=200)
par(mfrow=c(1,1),mar=c(4,4,0.5,0.5))
plot(ys~indvars$X1)
lmX1 <- lm(ys~indvars$X1)
abline(lmX1)
summary(lmX1)
```

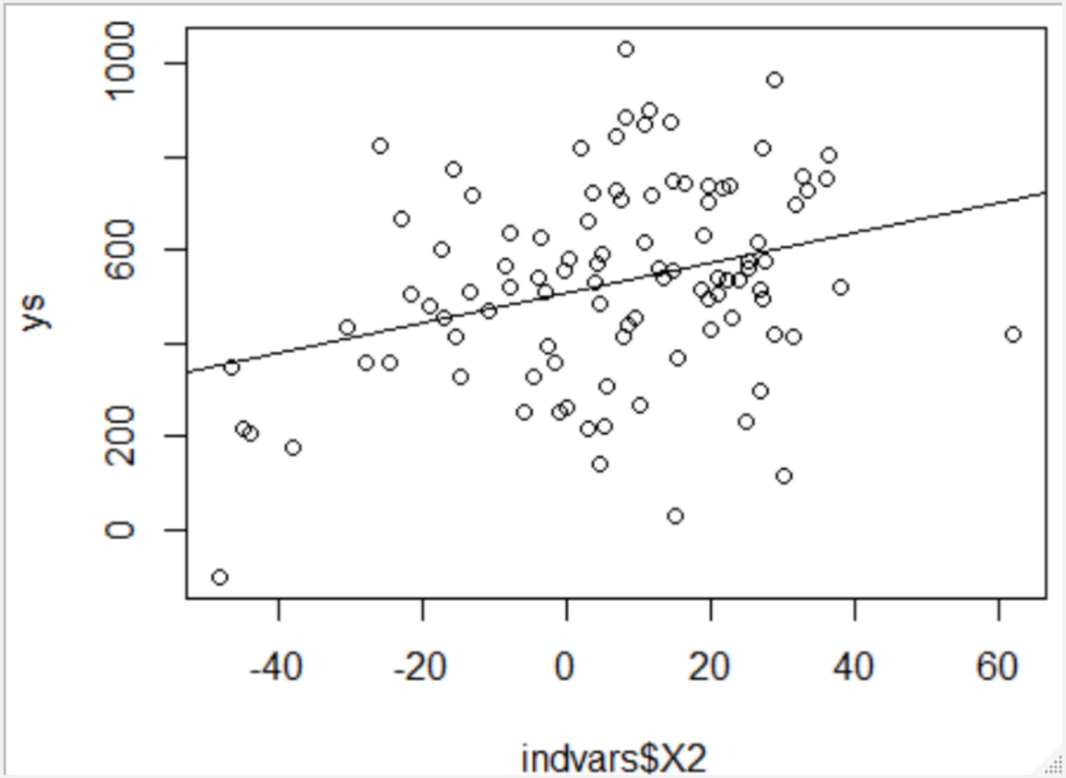
In reality, the dependent variable is explained by X1 alone!



```
> summary(lmX1)

Call:
lm(formula = ys ~ indvars$X1)

Residuals:
   Min     1Q Median     3Q    Max 
-528.5 -121.7   10.2  147.9  498.7 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 513.0964    21.5620  23.796   <2e-16 ***
indvars$X1    2.5040     0.9571   2.616   0.0103 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 208.1 on 98 degrees of freedom
Multiple R-squared:  0.06528,	Adjusted R-squared:  0.05574 
F-statistic: 6.844 on 1 and 98 DF,  p-value: 0.0103
```

Now fit a model with just X2

```
plot(ys~indvars$X2)
lmX2 <- lm(ys~indvars$X2)
abline(lmX2)
summary(lmX2)
```



```
> summary(lmX2)

Call:
lm(formula = ys ~ indvars$X2)

Residuals:
    Min      1Q  Median      3Q     Max 
-527.68 -130.77    1.38  144.74  497.36 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 507.5932    21.4188  23.698  < 2e-16 ***
indvars$X2    3.2117     0.9985   3.217  0.00176 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 204.7 on 98 degrees of freedom
Multiple R-squared:  0.0955,	Adjusted R-squared:  0.08627 
F-statistic: 10.35 on 1 and 98 DF,  p-value: 0.001758
```

Now fit a model with both X1 and X2

```
> lmx1x2 <- lm(ys~indvars$x1+indvars$x2)
> summary(lmx1x2)

Call:
lm(formula = ys ~ indvars$x1 + indvars$x2)

Residuals:
   Min     1Q Median     3Q    Max 
-522.8 -120.5    1.8  141.9  497.4 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  507.246     21.483  23.612   <2e-16 ***
indvars$x1    -1.600      2.325  -0.688   0.4930    
indvars$x2     4.763      2.466   1.931   0.0564 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 205.2 on 97 degrees of freedom
Multiple R-squared:  0.09989,	Adjusted R-squared:  0.08133 
F-statistic: 5.382 on 2 and 97 DF,  p-value: 0.006071
```

- X1 seems irrelevant
- X2 seems potentially
- relevant

- X1 seems irrelevant
- X3 seems potentially relevant

```
> summary(lmX3)

Call:
lm(formula = ys ~ indvars$X3)

Residuals:
    Min      1Q  Median      3Q     Max 
-528.99 -136.87    0.87  144.17  473.00 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 497.2065    21.9163  22.687  < 2e-16 ***
indvars$X3    3.5515     0.9808   3.621 0.000467 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 202.1 on 98 degrees of freedom
Multiple R-squared:  0.118,	Adjusted R-squared:  0.109 
F-statistic: 13.11 on 1 and 98 DF,  p-value: 0.0004667
```

```
> summary(lmX1X3)

Call:
lm(formula = ys ~ indvars$X1 + indvars$X3)

Residuals:
    Min      1Q  Median      3Q     Max 
-525.00 -131.39    0.54  149.75  467.68 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 495.6560    22.1678  22.359   <2e-16 ***
indvars$X1   -0.9352     1.6733  -0.559    0.578    
indvars$X3    4.3705     1.7652   2.476    0.015 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 202.8 on 97 degrees of freedom
Multiple R-squared:  0.1208,	Adjusted R-squared:  0.1027 
F-statistic: 6.666 on 2 and 97 DF,  p-value: 0.001938
```

- X1 seems irrelevant
- X4 seems potentially relevant

```
> summary(lmX4)

Call:
lm(formula = ys ~ indvars$X4)

Residuals:
    Min      1Q  Median      3Q     Max 
-522.86 -138.12    5.71  146.73  485.03 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 493.8915    22.5298  21.922  < 2e-16 ***
indvars$X4    3.1886     0.9159   3.481 0.000747 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 203 on 98 degrees of freedom
Multiple R-squared:  0.1101,	Adjusted R-squared:  0.101 
F-statistic: 12.12 on 1 and 98 DF,  p-value: 0.0007471
```

```
> summary(lmX1X4)

Call:
lm(formula = ys ~ indvars$X1 + indvars$X4)

Residuals:
    Min      1Q  Median      3Q     Max 
-518.61 -131.69    4.46  150.79  482.77 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 491.6768    23.1619  21.228   <2e-16 ***
indvars$X1   -0.7701     1.7280  -0.446   0.6568    
indvars$X4    3.8229     1.6948   2.256   0.0263 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 203.9 on 97 degrees of freedom
Multiple R-squared:  0.1119,	Adjusted R-squared:  0.09356 
F-statistic: 6.109 on 2 and 97 DF,  p-value: 0.003171
```

- X1 seems irrelevant
- X5 seems potentially relevant

```
> summary(lmX5)

Call:
lm(formula = ys ~ indvars$X5)

Residuals:
    Min      1Q  Median      3Q     Max 
-526.45 -123.38    6.28  141.59  485.74 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 488.5660    24.0643  20.303  < 2e-16 ***
indvars$X5    3.0305     0.9686   3.129  0.00231 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 205.2 on 98 degrees of freedom
Multiple R-squared:  0.09081,	Adjusted R-squared:  0.08154 
F-statistic: 9.789 on 1 and 98 DF,  p-value: 0.002313
```

```
> summary(lmX1X5)

Call:
lm(formula = ys ~ indvars$X1 + indvars$X5)

Residuals:
    Min      1Q  Median      3Q     Max 
-517.50 -121.88    0.69  144.25  473.86 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  467.729     31.522  14.838   <2e-16 ***
indvars$X1    -3.091      3.021  -1.023   0.3088    
indvars$X5     6.044      3.100   1.949   0.0541 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 205.2 on 97 degrees of freedom
Multiple R-squared:  0.1005,	Adjusted R-squared:  0.08197 
F-statistic:  5.42 on 2 and 97 DF,  p-value: 0.005869
```

- X1 seems irrelevant
- X6 seems potentially relevant

```
> summary(lmX6)

Call:
lm(formula = ys ~ indvars$X6)

Residuals:
    Min      1Q  Median      3Q     Max 
-523.22 -133.43   11.37  140.37  482.69 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 483.2604    23.8719  20.244  < 2e-16 ***
indvars$X6    2.9523     0.8345   3.538 0.000619 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 202.7 on 98 degrees of freedom
Multiple R-squared:  0.1132,	Adjusted R-squared:  0.1042 
F-statistic: 12.52 on 1 and 98 DF,  p-value: 0.0006186
```

```
> summary(lmX1X6)

Call:
lm(formula = ys ~ indvars$X1 + indvars$X6)

Residuals:
   Min     1Q Median     3Q    Max 
-513.7 -128.9    5.5  131.7  475.9 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  472.213     26.714  17.677   <2e-16 ***
indvars$X1    -1.832      1.982  -0.924   0.3577    
indvars$X6     4.399      1.774   2.479   0.0149 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 202.8 on 97 degrees of freedom
Multiple R-squared:  0.121,	Adjusted R-squared:  0.1029 
F-statistic: 6.676 on 2 and 97 DF,  p-value: 0.001922
```

```
> AIC(lmX1,lmX2,lmX3,lmX4,lmX5,lmX6,lmX1x2,lmX1x3,lmX1x4,lmX1x5,lmX1x6)
        df      AIC
lmX1     3 1355.347
lmX2     3 1352.061
lmX3     3 1349.540
lmX4     3 1350.439
lmX5     3 1352.577
lmX6     3 1350.079
lmX1x2   4 1353.574
lmX1x3   4 1351.219
lmX1x4   4 1352.234
lmX1x5   4 1353.504
lmX1x6   4 1351.202
```

True model

Best model

Model lmX3 is the best,
and it allows to predict Y is a reasonable way,
BUT
we would be misled in thinking that X2 drives Y, when it is X1 that drives Y

A key difference between the use of a model, for two different objectives:

1. Prediction
2. Explanation

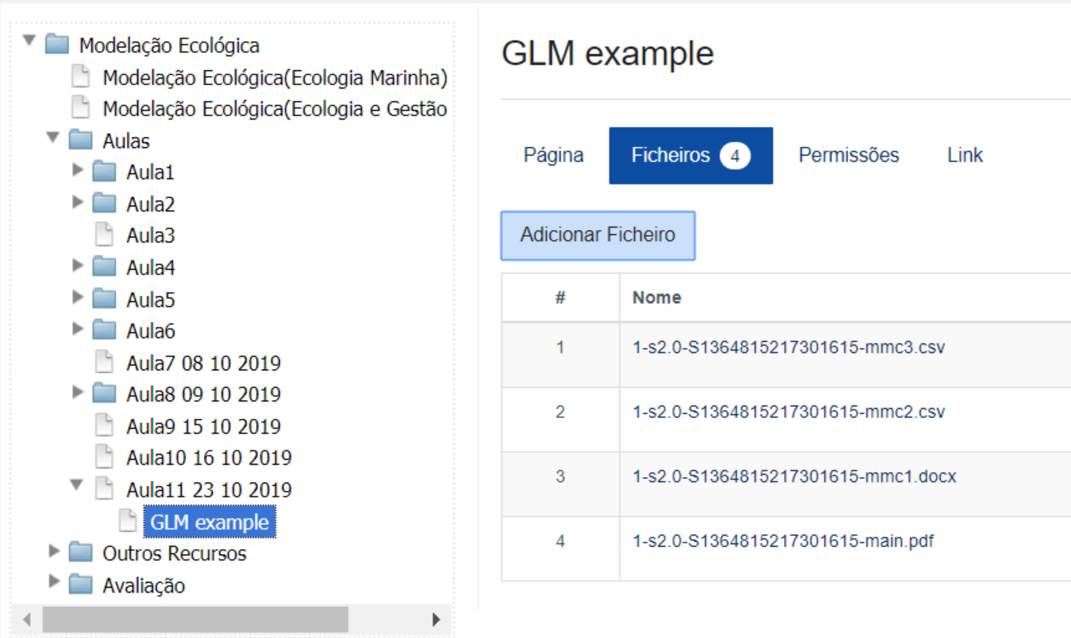
Explanation is what we are typically interested in "Ecological Modelling". But prediction is "good enough" for say "Machine Learning"!

Hands-on
GLM example(s)

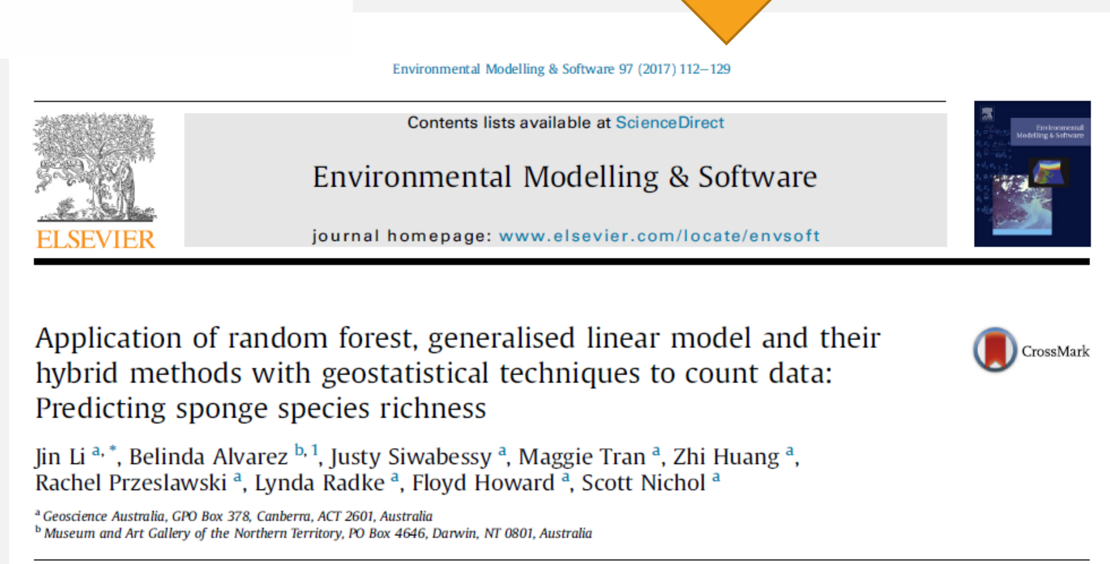
A count regression (try Poisson, but try Neg Bin too!)

A logistic regression (try logit link, but try other link functions too!)

Using the data in file "1-s2.0-S1364815217301615-mmc2.csv" (FENIX folder "Count data GLM") explain the variation in the response variable "sponge species richness" (species.richness) as a function of the other variables in said file.



This data set is used in the paper below, feel free to explore the paper for details ☺, brief variable description in next slides



Environmental Modelling & Software 97 (2017) 112–129

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

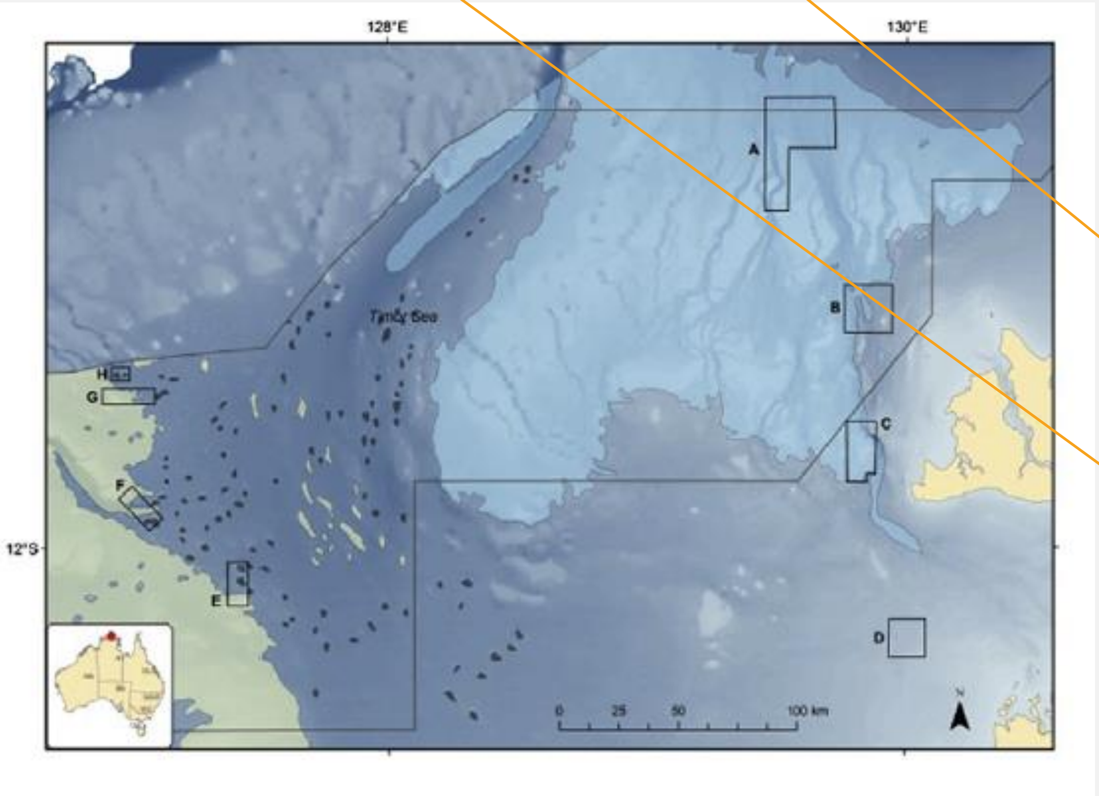
ELSEVIER

Application of random forest, generalised linear model and their hybrid methods with geostatistical techniques to count data: Predicting sponge species richness

Jin Li [a,*], Belinda Alvarez [b,1], Justy Siwabessy [a], Maggie Tran [a], Zhi Huang [a], Rachel Przeslawski [a], Lynda Radke [a], Floyd Howard [a], Scott Nichol [a]

[a] Geoscience Australia, GPO Box 378, Canberra, ACT 2601, Australia
[b] Museum and Art Gallery of the Northern Territory, PO Box 4646, Darwin, NT 0801, Australia

collection (Schlacher et al., 2007). There were 85 samples collected, and of which eight samples were excluded due to the uncertainty about transect length. In total, 77 samples were selected and used in this study. SSR is count data based on the presence/absence data, ranging from 1 to 39, with a mean of 10.48 and a standard deviation of 10.53. The point locations of samples are the mid-point of each transect.



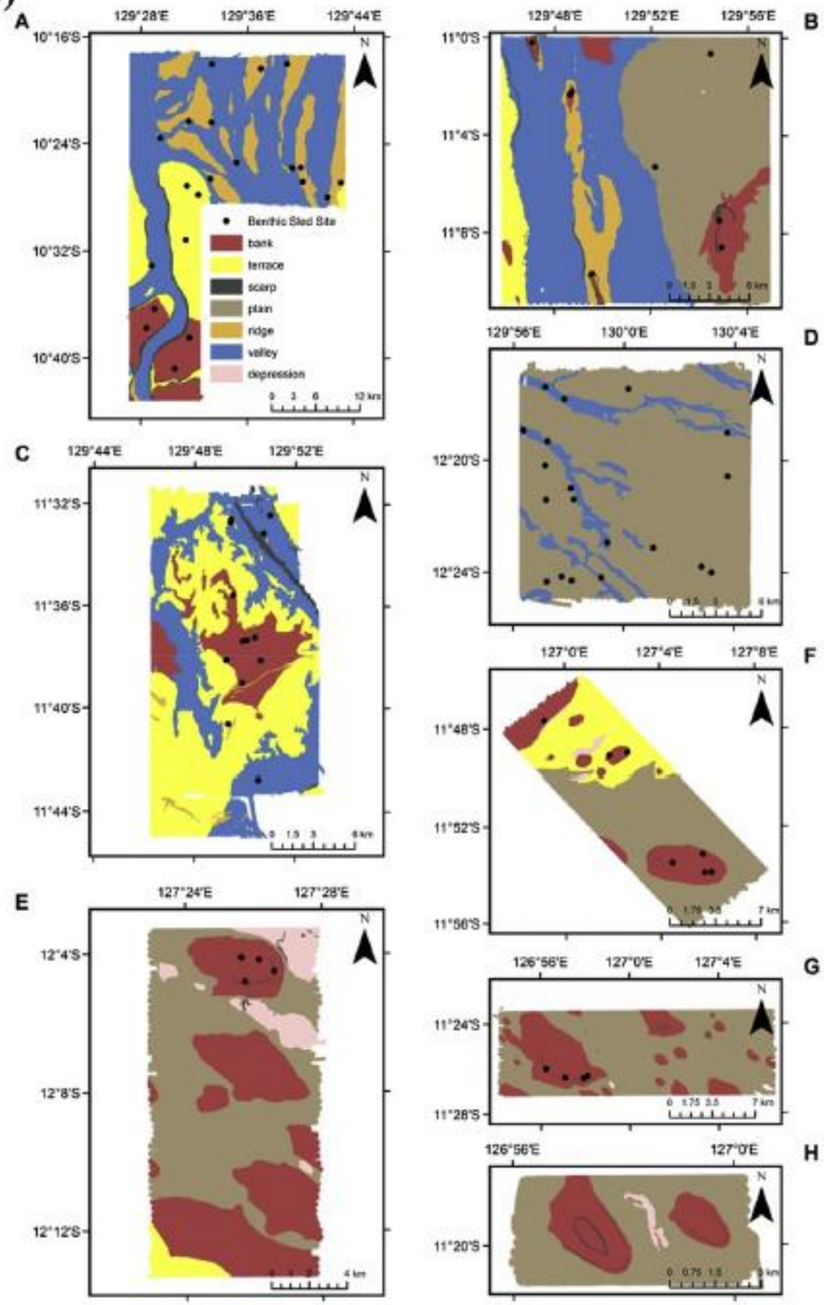

```
> with(data,range(species.richness))
[1]  1 39
> #the range of the response variable
> with(data,mean(species.richness))
[1] 10.48052
>
> #the range of the response variable
> with(data,sd(species.richness))
[1] 10.52517
```

2.3. *Predictive variables*

Following a preliminary analysis based on data availability and the relationships with seabed hardness as discussed above and in previous studies, 80 predictive variables were available for this study. They are:

1) Two location variables: latitude (lat) and longitude (long),
2) Three sediment variables: mud, sand and gravel,
3) Bathymetry (bathy),
4) Twenty-seven backscatter (bs) variables (bs10 to bs36): a diffused reflection of acoustic energy due to scattering process back to the direction from which it's been generated, measured as the ratio of the acoustic energy sent to a seabed to that returned from the seabed, normalised to incidence angles between 10° and 36°,
5) Seventeen derived variables from bs25 based on object and windows (30 m, 50 m and 70 m) approach:
   a. bs_o,
   b. homogeneity (bs_homo_o, bs_homo3, bs_homo5, bs_homo7),
   c. entropy (bs_entro_o, bs_entro3, bs_entro5, bs_entro7),
   d. Local Moran I (bs_lmi_o, bs_lmi3, bs_lmi5, bs_lmi7),
   e. Variance (bs_var_o, bs_var3, bs_var5, bs_var7).
6) Twenty-nine derived variables from bathy using object and windows (30 m, 50 m and 70 m) approach:
   a. bathy_o,
   b. lmi_o, lmi3, lmi5, lmi7,
   c. Topographic position index (tpi_o, tpi3, tpi5, tpi7),
   d. Seabed slope (slope_o, slope3, slope5, slope7),
   e. Planar curvature (plan_cur_o, plan_cur3, plan_cur5, plan_cur7),
   f. Profile curvature (prof_cur_o, prof_cur3, prof_cur5, prof_cur7),
   g. Topographic relief (relief_o, relief3, relief5, relief7),
   h. Seabed rugosity (rugosity_o, rugosity3, rugosity5, rugosity7).
7) Distance to coast (dist.coast)

(b)



PLOS | ONE

RESEARCH ARTICLE

Why sampling ratio matters: Logistic regression and studies of habitat use

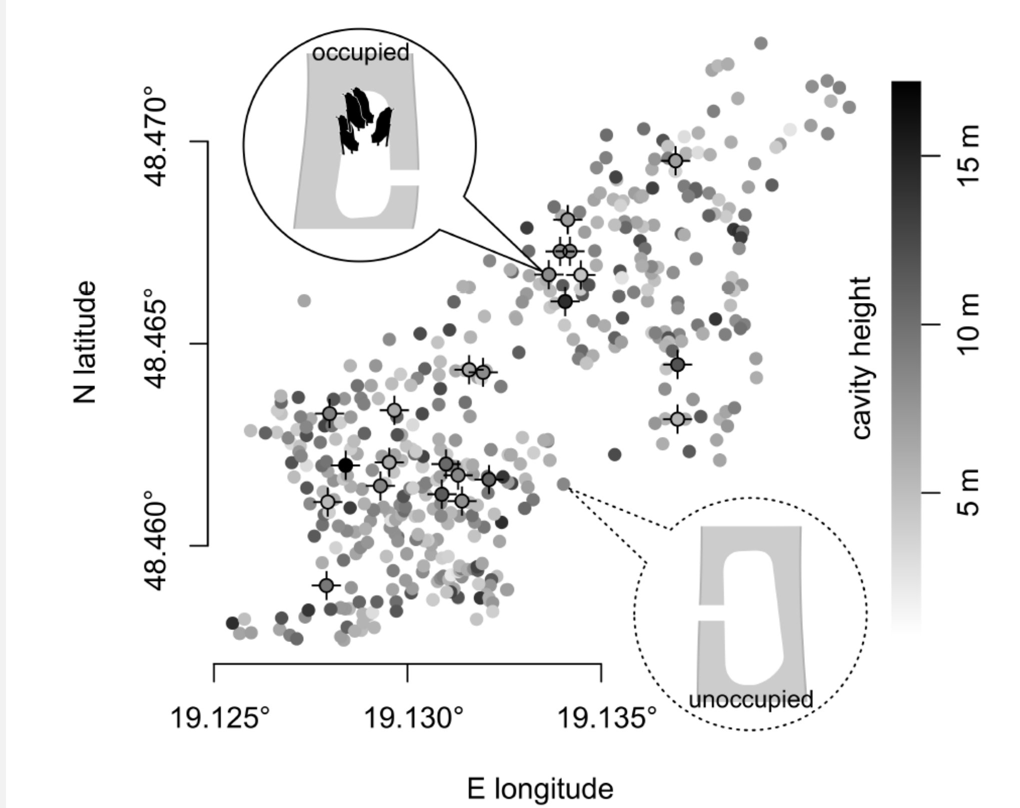
Ladislav Nad'o[1] *, **Peter Kaňuch**[1,2]

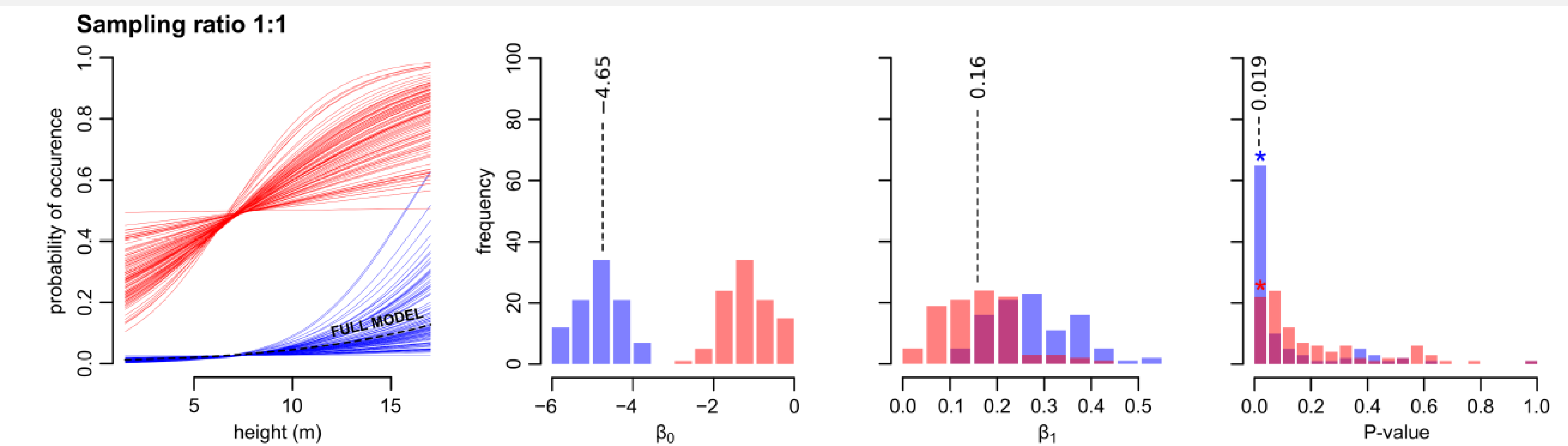
1 Institute of Forest Ecology, Slovak Academy of Sciences, Zvolen, Slovakia, **2** Institute of Biology and Ecology, Faculty of Science, P. J. Šafárik University in Košice, Košice, Slovakia

* ladislav.nado@gmail.com

Data in file "journal.pone.0200742.s002.csv" inside folder "Presence Absence GLM"

S1 Table. Dataset containing GPS coordinates and heights of 932 cavities (those used by bats are marked by 1).
(CSV)


occupied
unoccupied
N latitude
48.470°
48.465°
48.460°
19.125°
19.130°
19.135°
E longitude
cavity height
15 m
10 m
5 m


Sampling ratio 1:1
probability of occurence
FULL MODEL
height (m)
frequency
−4.65
β_0
0.16
β_1
0.019
P-value